Bounding Treatment Effects Under Interference in Geographic Natural Experiments: An Application to All-Mail Voting in Colorado*

Luke Keele[†] Rocío Titiunik[‡]

April 7, 2014

Preliminary Draft

Abstract

Political scientists often seek to understand whether modes of voting influence turnout rates. Two states and many localities have either adopted or experimented with all-mail elections, where voters no longer travel in-person to central locations to vote. Instead voters are mailed ballots and voting often starts as much as two weeks before election day. We analyze a natural experiment from the 2010 Colorado primary on all-mail voting. In that election, counties in Colorado had the option of an all-mail election or could retain traditional in-person voting on election day. We found that the town of Basalt, in the southwestern part of the state, was split in half by two counties that chose different modes of voting. We exploit this natural experiment to understand whether turnout levels were altered by all-mail elections. We adopt three different designs in our analysis: a geographic regression discontinuity design, differences-in-differences, and adjustment via a penalized match on geographic distance. However, social interactions often give rise to spillover effects in which the exposure of one individual to the treatment may affect outcomes of untreated units. In our application, treated and control voters lived in very close proximity and spillovers are probable. We develop a method of bounds to investigate whether our inference is a function of treatment spillovers. Using a model of treatment spillovers based on geographic proximity, we adjust unit level outcomes under different possible patterns of treatment spillovers. We then re-estimate the vote by mail treatment effect using these adjusted outcomes to create bounds on the estimate. These bounds allow us to observe whether an inference based on the assumption of no interference is plausible.

*Authors are in alphabetical order. We thank participants at the Causality in Political Networks Conference (University of Chicago, 2013) for valuable comments and discussion.

[†]Associate Professor, Department of Political Science, 211 Pond Lab, Penn State University, University Park, PA 16802 Phone: 814-863-1592, Email: ljk20@psu.edu

[‡]Assistant Professor, Department of Political Science, University of Michigan, 5700 Haven Hall, 505 South State St., Ann Arbor, MI 48109 Phone: 734-615-9114, Email: titiunik@umich.edu

1 Introduction

In the recent decades, the U.S. has witnessed an increase in methods of voting that differ from traditional in-person voting on election day. These reforms, commonly referred to as "convenience voting," include in-person early voting (voters may cast a vote in person before election day), no-excuse absentee voting (voters may apply for an absentee ballot without providing a reason for doing so), and all-mail voting (voting by mail is mandatory) – see Gronke et al. (2008) for a review. These policies are often implemented with the goal of reducing the costs of voting, which is in turn expected to increase voter participation.

Of the different convenience voting policies that have been adopted, all-mail voting is the most drastic, since in-person precinct voting is eliminated and voters can *only* vote via mail. Under all-mail voting, there are no longer polling places, and citizens receive a ballot on the mail several weeks in advance of election day and then return it by mail to the election administration office. For this reason, the effects of all-mail voting on turnout could be different from the effects of other types of convenience voting: although all reforms share the goal of making voting more convenient by giving citizens the opportunity to cast a vote during longer periods of time and by a more accessible method, all-mail voting is the only convenience method that *eliminates* precinct-place voting. Thus, while most convenience voting policies are offered in addition to more traditional polling-place voting and therefore increase the number of ways in which a ballot can be cast, all-mail voting replaces precinct-place voting and therefore reduces the number of available voting alternatives.

Given the far-reaching nature of all-mail voting reforms, scholars have been interested in studying whether all-mail voting affects voter turnout. In principle, these effects could either be negative or positive. On the one hand, all-mail voting makes the act of voting more convenient: instead of having to go to a polling place on election day, voters receive a ballot in the mail several weeks in advance of election day, and they can return the ballot by mail or drop it off at a specified location any time before election day. This reduces transportation costs and increases flexibility; this reduction in the costs of voting should lead to an increase in turnout. On the other hand, all-mail voting could decrease turnout for at least two reasons. First, the elimination of alternative voting methods may itself lead to a decrease in turnout. Even when voting by mail seems to be the most convenient alternative for the average voter, some voters may find polling-place voting more convenient (e.g., those with transitory addresses), thus the complete elimination of this alternative may lead some voters to abstain from voting. Second, the move to all-mail elections reduces the social aspect of voting by preventing citizens from gathering at the polling place and voting together and it also removes from voters and campaigns the possibility of using election day as a focal point, potentially discouraging some voters from casting a vote.

We examine this question using a geographic natural experiment in Colorado, where in 2010 counties were given the choice to require that votes be cast by mail during the primary election. Counties that adopted all-mail elections removed other alternative methods of voting, while counties that did not offered traditional polling-place voting on election day and also allowed by-mail no-excuse absentee voting. Given that voter administration is conducted by county governments, counties may chose the mode of voting that will maximize turnout within that county. Such selection will complicate statistical inferences. To that end, we compare voters in the town of Basalt which is split by the border between Eagle and Pitkin counties. Pitkin county retained in-person voting, while Eagle changed to an all-mail election. Thus it we make the case that voters in the town of Basalt are split in a haphazard fashion that can aid our ability to draw inferences about the effects of all-mail elections.

However, our inference is complicated by the fact that, by construction, our treated and control groups are spatially very close to one another, a phenomenon that could lead some treated voters to interact with control voters (engage in political discussion, read newspaper articles focused on election day) and change their voting decision as a result of that interaction. This phenomenon of interference between voters would undermine the interpretation of our estimates as the effect of all-mail voting.

The problem of interference may in principle affect any study of voting. But since we base our inferences on a design that uses the variation in treatment induced by a geographic boundary and compares voters that are spatially close on either side of this boundary, the possibility of interference is more likely than in other, non-geographic studies. We develop a method to evaluate the robustness of estimated effects to spatial interference between units. Our approach assumes that interference is spatially based and voters are more likely to interact with other voters who live near them. Adopting a particular model of interference, we then calculate bounds for the estimated effect of all-mail voting on turnout. We start by assuming the worst case scenario where voters in the treated area change their voting decision if as few as one control voter lives near them, and analyze how the estimated effect changes in this case. We then vary the particular form of interference that is allowed in order to obtain less conservative bounds. This allows us to pose a model of interference, and subject to that model, we may ask did our inference change under interference?

This article is organized as follows: Section 2 provides an overview of other studies of all-mail voting. Section 3 describes the natural experiment that is the focus of our analysis. Section 4.3 outlines our notation, describes the our causal estimands and identifiability conditions. In this Section, we also develop the method of bounds for understanding the presence of interference. In Section 5.2, we analyze the data from the Colorado natural experiment on all-mail voting. Section 6 includes discussion and concluding remarks.

2 All-Mail Voting

The literature on the effects of all-mail elections on turnout has been mostly focused on the state of Oregon, where polling-place voting was gradually eliminated during the 1990s, and since 1998 all statewide primary and general elections are conducted by mail only (see Gerber et al. (2013) and Gronke and Miller (2012) for recent comprehensive reviews). Soon after the change from polling-place to all-mail voting, several authors compared turnout before and after the reform and found that the swtich to all-mail voting had increased turnout.

For example, Southwell and Burchett (2000) find that all-mail voting increased turnout by 10%, and Karp and Banducci (2000) find that the increase in turnout is concentrated local elections. Positive findings in Oregon were also reported by Berinsky et al. (2001) and Richey (2008). However, a recent study by Gronke and Miller (2012) that includes several Oregon election cycles after the all-mail reform, finds that the large positive effects found by Southwell and Burchett (2000) cannot be replicated when a longer time period is used, and argues that the effects seen right after the reform were due to a novelty effect and not to a long-term effect of all-mail voting on turnout. (But Gronke and Miller (2012) do find positive effects of all-mail voting on primary and special elections.)

Most studies of the studies that focused in Oregon relied on a comparison of turnout before and after the reform was implemented. In contrast, research on other states where all-mail voting has not been adopted statewide relies on a comparison of nonexperimental treated and control groups. Kousser and Mullin (2007) and Bergman and Yates (2011) study turnout in California, where county election officials can assign voters to all-mail voting precincts in low-population areas. Using this variation as a natural experiment, both studies find that all-mail voting decreases turnout.[1] And in a recent paper, Gerber et al. (2013) use the large-scale move from polling-place to all-mail elections in the state of Washington where, unlike Oregon, counties did not switch to all-mail elections simultaneously, creating variation in when all-mail voting was adopted that the authors exploit to compare turnout in all-mail counties to turnout in polling-place counties during the same election. Using this natural experiment, the authors find that all-mail voting increases turnout by two to four percentage points.

[1] Bergman and Yates (2011) perform an individual level analysis and find that all-mail voting decreases turnout by 13%. Kousser and Mullin (2007) compare all-mail precincts to polling-place precincts and find that overall turnout decreases by 3%, although they find positive effects for special local elections.

3 All-Mail Voting in 2010 Colorado Primary

In recent years, the state of Colorado has implemented several reforms aimed at reducing the costs of voting and increasing voter turnout. After a series of difficulties experienced in several elections during the early and mid 2000s, Colorado's General Assembly created in 2008 the Election Reform Commission, whose task was to provide legislative recommendations regarding the state's voting and elections systems. Colorado has adopted several measures that have expanded the possibilities to vote by mail. Starting in 2008, voters could choose to be placed on a permanent vote-by-mail list. For the 2010 primary, the Secretary of State allowed each county to choose whether to hold either an all-mail election, use voter centers or hold traditional in-person voting at precincts. Figure 1 contains a map which shows the counties that selected each mode of election. While urban areas generally selected all-mail elections, many rural counties elected to use in-person voting.

Given that counties were able to select mode of voting, this creates a natural comparison for understanding how election mode affects turnout. Many studies of election mode are forced to make cross-state comparisons. Keele and Minozzi (2012) show that this approach generally fails given that it can be very difficult to isolate cross-state differences in election mode from other factors that affect turnout. However, given that counties are responsible for election administration and were able to select their mode of election in the 2010 primary, comparisons across counties might be biased by unobserved confounding. To overcome this selection process, we examined the state for a possible geographic discontinuity by county border. That is we look for some location where a town or city is split by a county border, where one county uses all-mail voting while the other county uses in-person voting. While many towns and cities in Colorado are split by county borders, we found that one town, Basalt, in the southwestern part of the state was exact split in this fashion.

Figure 1 also highlights the location of the town of Basalt. According to the 2010 census, Basalt has a total population of 3857. The population is largely white though about 20%

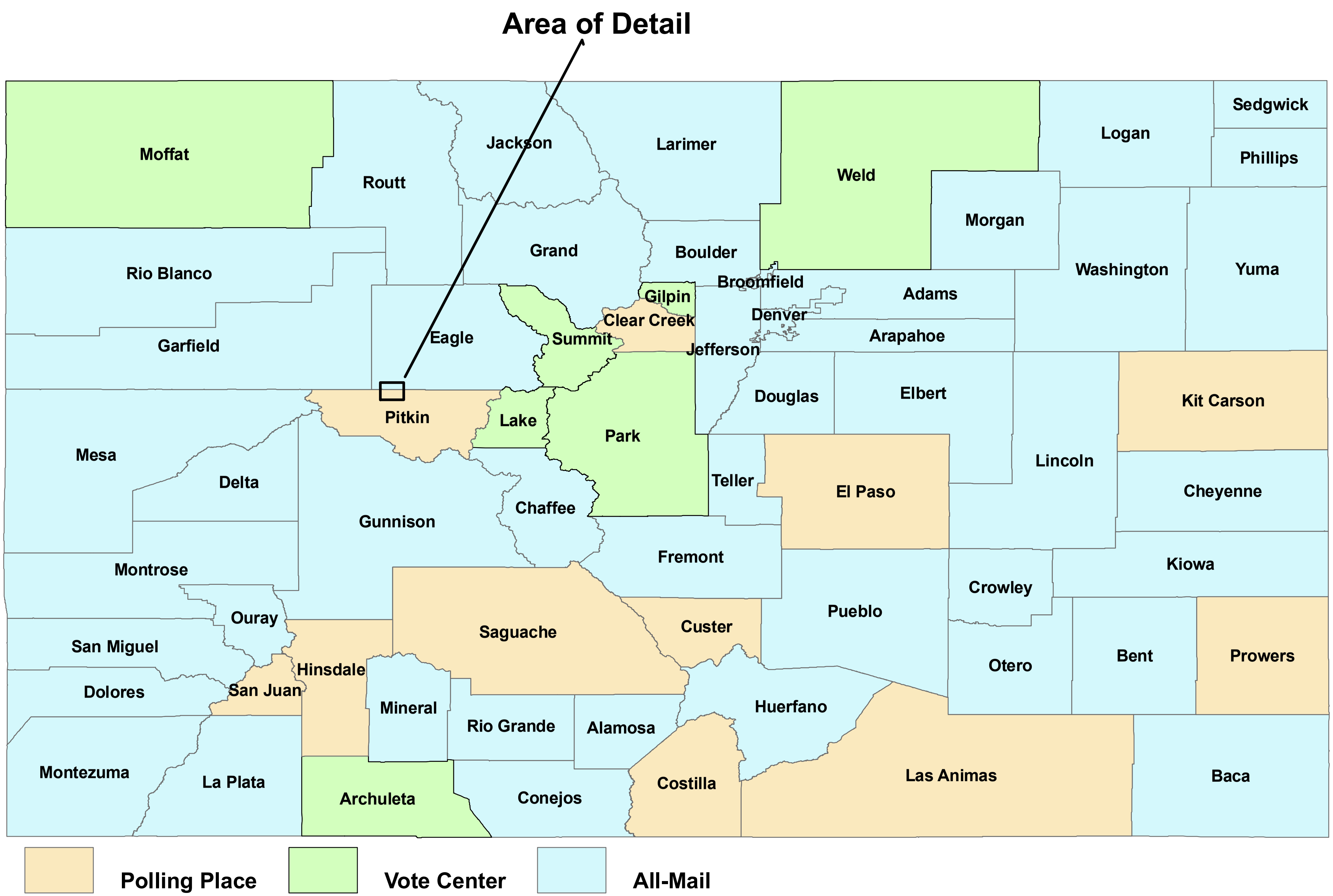

Figure 1: County map of Colorado with model of voting and location of Basalt highlighted.

of the population identifies as Hispanic. The town is close to the resort city of Aspen, and using property sale records, we found that the median house price in 2010 was over $600,000. Figure 2 contains a map which shows the town in greater detail. The central part of the town is spilt by the county border which defines mode of election. This part of Basalt contains the main shopping district, residential areas and schools. While property taxes in Colorado have a county component, we found that property taxes are based on five different tax zones with school district contributing the most to overall property tax burden. While the county border splits the town, the entire area is within the same school district. Moreover, all residents of Basalt attend the same set of public schools which are located within the central part of the town.

We focus our analysis on the town of Basalt, since it forms a natural experiment. In a natural experiment, some units obtain treatment and other are denied treatment in a haphazard manner. We argue that the county border divides town residents into vote by mail and in person voting districts in a haphazard manner. The difficulty, of course, is that haphazard assignment to treatment if often a far cry from a randomized experiment where randomization is a known fact. However, we argue that haphazard treatment assignment is preferable to situations where units are allowed to fully self-select treatment assignment. Moreover, natural experiments like the regression discontinuity design are predicated on reducing heterogeneity in the study population by making a more focused comparison. That is, clearly the case here as we focus on a small but homogenous part of the state of Colorado in hopes of comparing comparable voters.

Primary elections often hold little interest for voters since primary races are often uncompetitive. The 2010 Colorado primary, however, had three high profile elections on the ballot. In the Republican gubernatorial primary, a Tea Party insurgent beat Scott McInnis a 6 term U.S. representative after it came to light that McInnis plagiarized a water study he was paid to conduct. In the Democratic U.S. Senate primary, the candidate endorsed by President Obama narrowly beat a more liberal candidate endorsed by Bill Clinton. In the

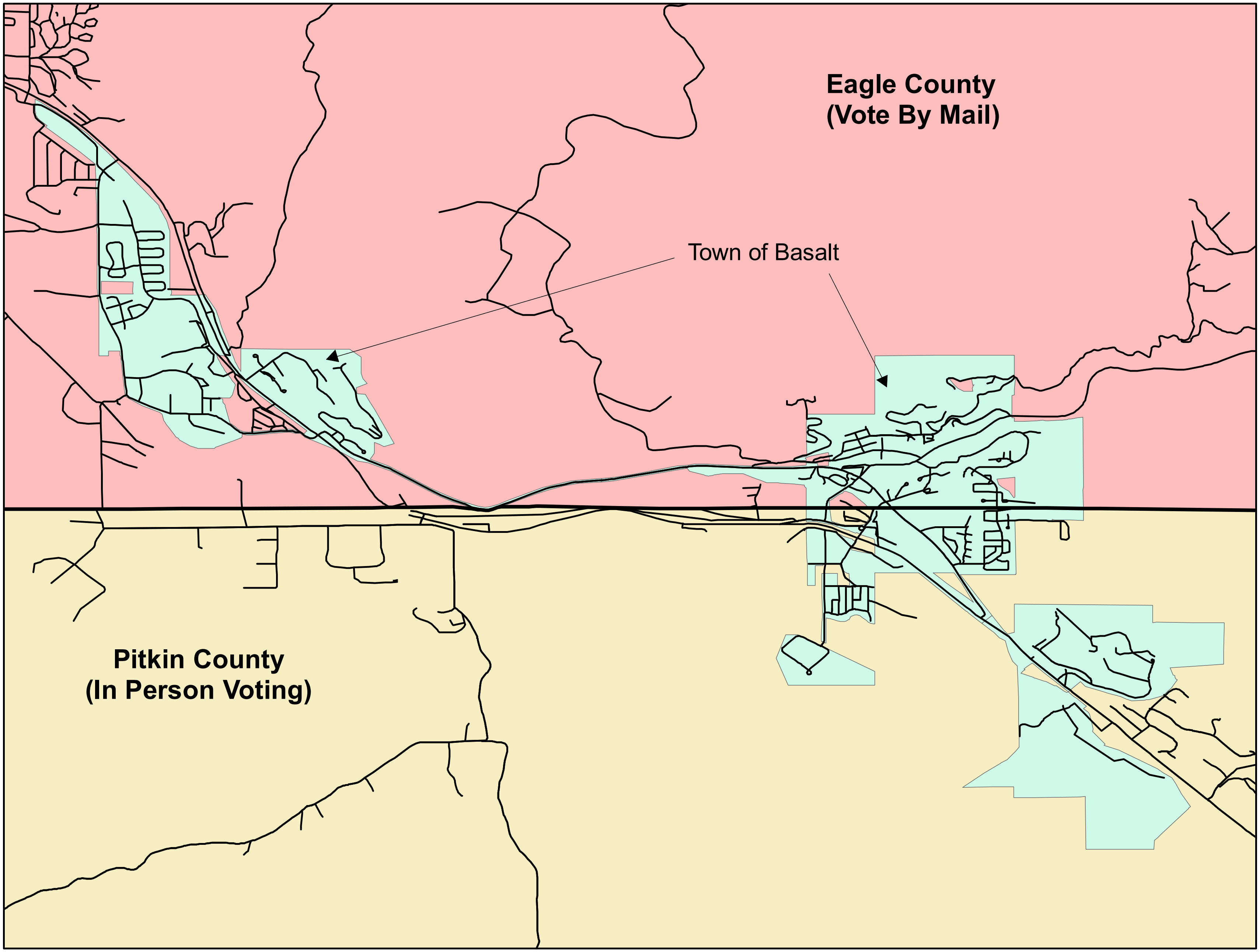

Figure 2: The county discontinuity: Basalt split by county boundary

Republican U.S. primary, Ken Buck, a Tea Party candidate, beat Jane Norton the candidate endorsed by the Colorado Republican party establishment. Results from the primary received national coverage and were featured on the front page of the *New York Times*. We now outline the methods we use for analysis.

4 Methodology

We adopt the potential outcomes framework for causal inference (see Holland 1986; Rubin 2005). Within this framework, we first describe a mode of statistical analysis that assumes voters do not interfere with one another. We then generalize the framework to allow for generalized patterns of interference.

4.1 Framework for Causal Inference

We conduct our analysis with individual level voter data. For each voter, there are two potential outcomes, $Y_i(1)$ and $Y_i(0)$. $Y_i(1)$ is the potential turnout status for voter i when the voter is assigned to an all-mail mode of elections, while $Y_i(0)$ is the potential outcome for voter i when the voter is assigned to in-person voting on election day. We use the binary variable $D_i \in [0, 1]$ to denote treatment status for voter i. $D_i = 1$ if unit i resides in Eagle County and is assigned to an all-mail election, and 0 if the voter resides in Pitkin county and may vote in-person on election day for $i = 1, 2, \dots, n$. We collect all treatment indicators in the n-vector $\mathbf{D}$, and let $Y_i(\mathbf{D})$ be the potential outcome of unit i. In general, if we let the treatment status of every unit affect the potential outcome of every other unit, every i will have one distinct potential outcome for every value that the treatment vector $\mathbf{D}$ might take, which is 2^n. This form of interference, while being very flexible, is not practically useful — the set of possibilities is simply too large.

As such, we make the following assumption about interference among subjects.

Assumption 1 (No interference)**.** *The potential outcome of each unit depends only on the treatment received by that unit and not on the treatment assigned to any other unit: For all*

$\dot{\mathbf{D}} \neq \ddot{\mathbf{D}}$, $Y_i(\dot{\mathbf{D}}) = Y_i(\ddot{\mathbf{D}})$ *if* $\dot{D}_i = \ddot{D}_i$, *for* $i = 1, 2, \ldots, n$.

Under this assumption, we can write $Y_i(\mathbf{D}) = Y_i(D_i)$, since i's potential outcome only depends on the treatment received by i. Under this assumption, we define the unit level vote by mail treatment effect for the ith voter as

$$\tau_i = Y_i(1) - Y_i(0)$$

Given the close geographic proximity between treated and control voters, Assumption 1 may be implausible. Assumption 1 would be violated if a control voter were to urge a treated voter to vote on election day due to the fact that he or she had just returned from voting in-person. While interference can occur in any experiment or observational study, in settings where units are geographically close, interference becomes more likely. We first outline a mode of analysis assuming Assumption 1 holds. We then outline a method of bounds to understand whether interference may affect our inference.

4.2 Estimating the Vote by Mail Treatment Effect

To estimate the vote by mail treatment effect, we utilize three different identification strategies. An identification strategy is a set of assumptions that warrant inferences about unobservable counterfactual quantities based on observable quantities. The first identification strategy exploits geographic proximity to the boundary between treated and control areas. Specifically, we use the Geographic Regression Discontinuity (GRD) design, where 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 (Keele and Titiunik 2013). In our application, we assume that the county border that divides the town of Basalt does so in an as-if random fashion.

Under the GRD 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 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. We define a score that uniquely represents unit 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. We call the set that collects all boundary points $\mathcal{B}$, and denote a single boundary point $(b_1, b_2) = \mathbf{b}$, with $\mathbf{b} \in \mathcal{B}$. The geographic location of individual i is given by two coordinates such as latitude and longitude, $(S_{i1}, S_{i2}) = \mathbf{S}_i$. We use $\mathbf{s}_t \in A_t$ to refer to locations in the treatment area and $\mathbf{s}_c \in A_c$ to refer to locations in the control area. Thus, $\mathcal{A}^t$ and $\mathcal{A}^c$ are the sets that collect the scores of units with $T = 1$ and $T = 0$, respectively. Assignment of T_i is now a deterministic function of this score, which has a discontinuity at the known boundary $\mathcal{B}$. Identification in the GRD design holds under the following assumption.

Assumption 2 (Continuity in two-dimensional score)**.** *The conditional regression functions are continuous in* $\mathbf{S}$ *at all points* $\mathbf{b}$ *on the boundary:*

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

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

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

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*, under Assumption 2 the GRD design will identify the treatment effect at *each* of the boundary points. Since the length of the border is relatively short, we adopt the following simplifying assumption.

Assumption 3 (Spatially Constant Average Potential Outcomes)**.** *Average Potential outcomes are constant with respect to location along the discontinuity border:*

$$E\left\{Y_{i0}|\mathbf{S}_i = \mathbf{b}^p\right\} = E\left\{Y_{i0}|\mathbf{S}_i = \mathbf{b}^q\right\}$$

$$E\left\{Y_{i1}|\mathbf{S}_i = \mathbf{b}^p\right\} = E\left\{Y_{i1}|\mathbf{S}_i = \mathbf{b}^q\right\}$$

for all $\mathbf{b}^p \in \mathcal{B}, \mathbf{b}^q \in \mathcal{B}$*.*

Under this assumption, we have constant treatment effects at all boundary points, $\tau(\mathbf{b}^p) = \tau(\mathbf{b}^q)$ for all $\mathbf{b}^p \in \mathcal{B}, \mathbf{b}^q \in \mathcal{B}$. Estimation of the vote by mail treatment effect under Assumption 3 is straightforward. We simply estimate proportions within $\mathcal{A}^t$ and $\mathcal{A}^c$ and take the difference within proportions. To condition on geographic proximity, we defined bands of 50, 100, and 300 meters on either side of the Eagle county limit and calculated the average treatment effect for voters within each band via liner regression models.

Under the GRD design, we only condition on geography under the assumption that spatially proximate voters are as-if randomly assigned. Such an assumption may be unreasonable if voters sort around the county boundary that separates all-mail voters from in-person voters. As an alternate identification strategy, we use differences-in-differences (DID), which exploits longitudinal variation. First, we define some additional notation. Let $t \in \{0, 1\}$ indicate time, where $t = 0$ before the treatment is administered and $t = 1$ after. Under the DID identification strategy, we re-write the outcome as $Y_i(t) = \tau_{DID} D_i \times t + \delta_t + \eta_i + \nu_{it}$, where δ_t is time-specific, η_i is individual-specific, and ν_{it} represents unobservable characteristics. The DID estimand is

$$\tau_{DID} = \{E[Y|D = 1, t = 1] - E[Y|D = 0, t = 1]\} - \{E[Y|D = 1, t = 0] - E[Y|D = 0, t = 0]\}.$$

Estimation of the vote by mail treatment effect is simple under the DID identification strategy. Let μ_{dt} be the conditional sample moment for group d in time t. The DID estimate is $\hat{\tau}_{DID} = (\mu_{11} - \mu_{01}) - (\mu_{10} - \mu_{00})$. This quantity can also be estimated with least squares under the following linear model: $Y_{it} = \beta_0 + \beta_1 t + \beta_2 D_i + \beta_3 t \times D_i + \epsilon_{it}$. Abadie (2005) shows that $plim\ \beta_3 = (\mu_{11} - \mu_{01}) - (\mu_{10} - \mu_{00})$.

Identification, here, requires the expected potential outcomes for treated and control units to follow parallel paths in the absence of treatment. Formally, the assumption can be written as $E[Y_0(t = 1) - Y_0(t = 0)|D = 1] = E[Y_0(t = 1) - Y_0(t = 0)|D = 0]$. Under DID we assume that all differences with respect to differences in turnout are time invariant.

The third and final identification strategy that we use combines the GRD design with a matching estimator. Keele et al. (Forthcoming) develop a matching estimator for use with geographic discontinuities. Their matching estimator allows one to compare treated and control units that are spatially proximate while balancing observed covariates. Our data contains a small number of covariates on which to match. As such we minimize distance between treated and control units while matching on sex, race, age and voting status in the 2008 general and primary election and the 2006 general and primary election. We elected to match sex, race, and age without constraints, while we enforced an exact match on voter history. We also exactly matched on whether a voter resided in central Basalt. Next, we describe the function we optimize to form the matches.

Let j_t index the subjects in the treated area, A_T, and similarly let j_c index the subjects in A_C. Define d_{j_t,j_c} as the geographic distance between treated unit j_t and control j_c. To enforce specific forms of covariate balance, define $\mathscr{e} \in \mathcal{E}$ as the index of the covariates for which it is needed to match exactly, and $\mathscr{b}_{\mathscr{e}} \in \mathcal{B}_{\mathscr{e}}$ as the categories that covariate $\mathscr{e}$ takes, so that $x_{j_t;\mathscr{e}}$ is the value of nominal covariate $\mathscr{e}$ for treated unit j_t with $x_{j_t;\mathscr{e}} \in \mathcal{B}_{\mathscr{e}}$. Finally, let $\mathscr{m} \in \mathcal{M}$ be the index of the covariates for which it is desired to balance their means, so that $x_{j_t;\mathscr{m}}$ is the value of covariate $\mathscr{m}$ for treated unit j_t, and $x_{j_c;\mathscr{m}}$ is the value of covariate $\mathscr{m}$ for control j_c.

To solve our problem optimally, we introduce binary decision variables

$$a_{j_t,j_c} = \begin{cases} 1 & \text{if treated unit } j_t \text{ is matched to control unit } j_c, \\ 0 & \text{otherwise,} \end{cases}$$

and, for a given scalar λ, we minimize

$$\sum_{j_t \in A_T} \sum_{j_c \in A_C} d_{j_t,j_c} a_{j_t,j_c} - \lambda \sum_{j_t \in A_T} \sum_{j_c \in A_C} a_{j_t,j_c} \tag{1}$$

subject to pair matching and covariate balancing constraints. Under this penalized match, if geographic distance can be minimized it will be, and if it cannot be minimized in every case, it will be minimized as often as possible. In particular, the pair matching constraints require each treated and control subject to be matched at most once,

$$\sum_{j_c \in A_C} a_{j_t,j_c} \leq 1, \quad \forall j_t \in A_T, \tag{2}$$

$$\sum_{j_t \in A_T} a_{j_t,j_c} \leq 1, \quad \forall j_c \in A_C. \tag{3}$$

This implies that we match without replacement, which we do to simplify inference. The covariate balancing constraints are defined as follows

$$\sum_{j_t \in A_T} \sum_{j_c \in A_C} \left| \mathbb{1}_{\{x_{j_t;e} = \ell_e\}} x_{j_t;e} - \mathbb{1}_{\{x_{j_c;e} = \ell_e\}} x_{j_c;e} \right| a_{j_t,j_c} = 0, \quad \forall e \in \mathcal{E}, \tag{4}$$

$$\left| \sum_{j_t \in A_T} \sum_{j_c \in A_C} a_{j_t,j_c} x_{j_t;m} - \sum_{j_t \in A_T} \sum_{j_c \in A_C} a_{j_t,j_c} x_{j_c;m} \right| \leq \varepsilon_m \sum_{j_t \in A_T} \sum_{j_c \in A_C} a_{j_t,j_c}, \quad \forall m \in \mathcal{M}, \tag{5}$$

where $\mathbb{1}$ is the indicator function.

These constraints enforce exact matching and mean balance, respectively. See Zubizarreta (2012) for a discussion of these and other covariate balance constraints in the context of a more general mixed integer program. We also incorporated optimal subset matching into the integer programming framework in the objective function (1) via the λ parameter. Subject to the pair matching constraints (2) and (3) and the covariate balancing constraints (4)–(5), this form of penalized optimization addresses the lack of common support problem in the distribution of observed covariates of the treated and control groups. Including this penalty allows us to keep the largest number of matched pairs for which distance is minimized and the balance constraints are satisfied.

With the matched analysis, we can also perform a well-known form of sensitivity analysis for matching estimators developed by Rosenbaum (2002). In a sensitivity analysis, we quantify the exact degree to which the identification assumption must be violated in order for our inference to be changed. In the sensitivity analysis, we manipulate the Γ parameter which measures the degree of departure from random assignment of treatment. Two subjects with the same observed characteristics may differ in the odds of receiving the treatment by at most a factor of Γ. In a randomized experiment, randomization of the treatment ensures that $\Gamma = 1$, that is the odds of treatment are the same across treated and control. In an observational study, Γ may depart from one. For example, if Γ is two for two subjects, one treated and one control, that are identical on matched covariates, then one subject is twice as likely as the other to receive the treatment because they differ in terms of an unobserved covariate (Rosenbaum 2005). While the true value of Γ is unknown, we can try several values of Γ and see how the conclusions of the study change. Specifically, we calculate an upper bound on the p-value for a range of Γ values. If the upper bound on the p-value exceeds the conventional 0.05 threshold, then we conclude that a hidden confounder of that magnitude would explain the observed association. If the study conclusions hold for higher Γ values, the estimate is fairly robust to the presence of a hidden confounder.

4.3 Estimating Effects Under Interference

The three identification strategies outlined above all assume that there is no interference between treated and control units. Given that the design relies on a set of the data where voters are in close geographic proximity, we might expect interference across units. We next outline what we argue is the most plausible form of interference given the empirical application.

As we outlined in Section 1, turnout often declines in all-mail elections especially when political parties fail to mobilize voters. Moreover, voters in all-mail elections do not experience the social aspects of election day, which can become a focal point of conversations and behavior about voting. For example, in counties where voting occurs at the polling place, campaigns are able to focus their strategies around election day and newspapers mention election day. All this may create voter enthusiasm in the control area. In the treatment area, where voting is all-mail, such a focal point is lacking, as there is no time and place where most people get together to cast their votes. But it is conceivable that treated voters that live very near the control area experience some of enthusiasm and attention to politics that occurs in the control area. If this enthusiasm spills over, it may cause some treated citizens to vote when they otherwise would not have voted. In other words, our model of interference assumes that some of people who voted in the treatment group did so not because of the greater convenience introduced by all-mail voting, but rather because they absorbed the enthusiasm and mobilization efforts of those in the control group who had access to traditional polling-place voting.

The specific model of interference that we adopt assumes that voters that reside in the control area where voting is conducted at polling places may transmit election day enthusiasm to treated voters. Some treated units may come into contact with control voters on election data due to spatial proximity. If so, these treated units may vote due to interference from control voters. Under this interference process, we must consider how to perform statistical inference.

One approach to statistical estimation under interference has been to decompose the overall treatment effect into direct and indirect effects, where the indirect effects arise due to interference between units. Some propose hierarchical models in which interference occurs within but not between groups (e.g. Hudgens and Halloran 2008; Tchetgen Tchetgen and VanderWeele 2012; VanderWeele et al. 2013). Aronow and Samii (2012) generalize this framework to allow for the estimation of average causal effects under general forms of interference. Their approach requires specifying a treatment exposure model for interference, and combines that model with the known randomization distribution of treatment that arises from the design of the experiment. Bowers et al. (2013) use a Fisherian inference that also requires specifying a model of interference, which is then used to adjust observed outcomes and test hypothesis generated by the model.

We use a method of bounds in the spirit of Manski (1990) to identify the possible role of interference. Since our focus is on a natural experiment rather than an experimental design, we do not rely on the distribution of treatment being known, as other approaches do. We first assume the model of interference from above. Under that model, treatment effects estimated under SUTVA should be too small, since some treated voters voted due to interference and not the vote by mail treatment. To obtain a bound on treatment effects under possible patterns of interference, we adjust observed outcomes consistent with our model of interference and then re-estimate the treatment effect to form a bound.

To calculate the bounds, we alter treated outcomes based on their spatial proximity to control voters. Among the treated voters, we calculate what we call the interference set which we denote $\mathcal{I}$. The interference set is the set of treated voters that are close enough to control units to be interfered with. To form $\mathcal{I}$, we identify the l subset of treated voters that have j control voters within distance k. For treated voters in $\mathcal{I}$, we assume that their observed outcomes, Y_i, are a function of interference, and we compute $\tilde{Y}$, the interference free outcome. We replace the observed outcomes with $\tilde{Y}$ for those voters in $\mathcal{I}$ and recompute the estimated treatment effect. This estimate forms a bound on the treatment effect under our model interference. If this estimate is close to the estimate under no interference, this implies that interference does not alter our inference. We formalize this process of calculating bounds under the following algorithm.

Algorithm 4.1: INTERFERENCE BOUND(τ_{int}, k, j, l)

for $i \leftarrow 1$ **to** m treated units

do
- Locate all controls within k distance of treated i
- c $\leftarrow$ number of control units that are k distance from treated i
- Place treated unit i in $\mathcal{I}$ **if** $c > j$
- Replace Y with $\tilde{Y}$ $\forall$ units in $\mathcal{I}$
- Recompute treatment effect to form bound τ_{int}

return (τ_{int})

Our bounding approach assumes that treated voters who are surrounded by control voters will be subject to spillovers, while those treated voters who are further away will not be subject to spillovers. Our approach allows the analyst to vary both proximity and density of control voters that is considered necessary for interference. This allows for the calculation of bounds under a wide variety of possible forms of interference. The bounds allow us to ask whether the observed results are robust under the no interference assumption.

To form $\tilde{Y}$, we change Y_i from 1 having voted to 0 not voted for all treated voters that are in $\mathcal{I}$. This change in outcomes forms our model for the interference free outcome. The size of $\mathcal{I}$ and the bound will vary depending on the values we choose for k and j. For example, we might calculate one set of bounds with k set to 250 meters and j to 1. Under this set of parameters, we change treated outcomes using our model for the interference free outcome under the assumption that any treated voter that is within 250 meters of a single control voter had an outcome due to interference. In our empirical application, we considered four different scenarios. First, we assumed that those treated voters whose residence was within 100 meters of at least one control residence voted in the election due to spillover effects from the polling-place control area and not because of all-mail voting. We therefore changed the outcomes from 'voted' to 'not voted' of those treated voters that had at least one control observation within 100 meters. Next, the interference set is formed from all treated voters that are within 100 meters of at least 3 control voters. Finally, the interference set is formed from all treated voters that are within 100 meters of at least 5 control voters. We repeated the bounds analysis with the same j values but set k to 250 meters. The bounds will be more conservative as we make k larger and j smaller. Here, our most conservative set of bounds set k to 250 and j to 1.

5 Analysis of the 2010 Colorado Primary

In this section we analyze the data from the 2010 Colorado primary. We first focus on estimating the vote by mail treatment effect within the town of Basalt. In this section we

Table 1: Estimates of the vote-by-mail treatment based on a geographic discontinuity. Estimate is based on difference in proportions for varying distances around the county border.

	Basalt Area	300m from Boundary	100m from Boundary	50m from Boundary
Point Estimate	0.9	0.4	0.1	-0.5
s.e.	1.8	2.3	3.2	3.6
t	0.50	0.18	0.02	-0.16
p	0.63	0.86	0.98	0.87
N	1625	1104	654	544

use three different identification strategies in hopes of finding agreement across different assumptions. We then explore the possibility that the estimated treatment effects were contaminated by interference.

5.1 Estimates of the Vote by Mail Treatment Effect

We first estimate the vote by mail treatment effect conditioning on geography. For these estimates we assume that those that live near the Eagle county boundary are more comparable than those who live farther away. In Table 1 we report the estimated vote by mail treatment effect for the larger Basalt area and then for subsets of this larger sample based on geographic proximity. For these analyses, we restrict the estimates to those that live within 300, 100, or 50 meters from the Eagle county boundary. The Basalt area sample includes those that live in central Basalt township, but excludes residents that fall within the Basalt city limit but are farther north of the Eagle county boundary. Based on the estimates that condition on geography alone, we find little evidence that all-mail voting increased turnout. The largest estimate is 0.9, which implies that turnout was nearly one percentage point higher among those voters with an all-mail election. However, the 95% confidence interval for this estimate cover zero. As we condition on geography the estimates shrink in magnitude until for those who live with 50 meters of the county limit, the estimate implies that vote by mail depressed turnout by half a percentage point. None of these estimates are statistically significant.

Table 2: Differences-in-Differences estimates of vote-by-mail treatment effect for three different geographic areas.

	Eagle County	Extended Basalt Township	Central Basalt Township
Point Estimate	-8.3	-7.6	-4.6
s.e.	0.00	0.00	0.00
t	0.00	0.00	0.00
p	0.00	0.00	0.00
N	6927	3913	1625

The estimates in Table 1 assume that conditioning on geographic location is all that is needed to render voters comparable. While voters were quite similar in terms of demographic characteristics, we find that voters in Eagle County vote at higher rates than do voters in Pitkin County. On average, turnout in Eagle County was five percentage points higher in the four elections prior to the 2010 primary. The method of difference-in-differences allows us to condition on these baseline differences in turnout. We use the 2008 primary as the pre-treatment election, and estimated DID via least squares with robust standard errors based on a sandwich estimator.

For two of the DID estimates, we condition on geography through stratification. The first analysis includes all voters that live in the immediate area around the Eagle-Pitkin County boundary. The next analysis is restricted to voters that live within the Basalt city limits. The final set of estimates is limited only to voters that live within the Basalt town center. The final DID estimate is based on the same sample as the full sample analysis in Table 1.

Table 2 contains the three DID estimates. While these estimates vary in magnitude, all three estimates imply that the all-mail election depressed turnout. In the full sample, turnout was 8.3 percentage points lower among voters who were required to vote by mail. The magnitude of the treatment effect is smaller for those within central Basalt. In every case, however, the estimates are statistically significant. Thus once we account for the fact that turnout is higher in the treated area, we find that the all-mail election in 2010 appears

Table 3: Covariate balance before and after matching around the geographic discontinuity. Std. Diff.= absolute standardized difference. Distances are in kilometers. Means for turnout are proportion of registered voters voting in that election.

	Unmatched			Matched		
	Mean Treated	Mean Control	Std. Diff.	Mean Treated	Mean Control	Std. Diff.
Nonwhite	0.07	0.08	0.03	0.09	0.09	0.02
Age	48.4	45.6	0.19	45.7	46.1	0.02
Female	0.49	0.50	0.01	0.51	0.49	0.03
2008 General Election Turnout	0.71	0.60	0.24	0.62	0.62	0.00
2008 Primary Election Turnout	0.07	0.02	0.27	0.02	0.02	0.00
2006 General Election Turnout	0.48	0.36	0.24	0.38	0.38	0.00
2006 Primary Election Turnout	0.06	0.01	0.22	0.01	0.01	0.00
Median Geographic Distance		0.89			0.59	

to have depressed turnout. Next, we use matching to condition on covariates more generally.

Before presenting results after matching on covariates, we examine whether observed covariates are balanced in the unmatched sample and how that balance improves after we match. Table 3 contains sample means and the absolute standardized differences in means (difference in means divided by the pooled standard deviation between groups before matching) for three demographic characteristics and for turnout in the last four elections. While demographic characteristics are generally balanced in the unmatched sample, turnout is always higher in the treated county. Through matching we are able to remove all imbalances in prior turnout. Unlike the DID estimates, which only conditions on the fixed difference for a single past election, the matching estimator removes all observable differences in past turnout.

Table 4 contains the results from the matched sample. We find that among the subset of the sample that was nearly identical in terms of observed covariates, turnout was lower by five percentage points for those voters who participated in the all-mail election. The 95% confidence interval for this estimate is bound away from zero. These results are consistent with the DID estimates which condition on only a single past election. The fifth row of Table

4 contains the value of Γ at which the p-value exceeds 0.05. Here, we see that if an unobserved covariate caused two identically matched voters to differ in their odds of treatment by as little 1.3, that would explain the estimated effect. As such, the sensitivity analysis indicates an unobserved confounder could easily explain the association we observe.

Table 4: Outcomes for the 2008 primary election by treatment status among matched pairs

	Control, $T_i = 0$	Treated, $T_i = 1$
Didn't Vote, $Y_i = 0$	452	481
Voted, $Y_i = 1$	89	60
Difference in proportions	-5.0	
p^a	0.002	
Γ^b	1.30	

[a] The one-sided p-value from McNemar's test is 0.254.

[b] The Γ value at which the upper bound on a one-sided p-value exceeds the 0.05 threshold.

5.2 Bounds Under Interference

Thus far we have assumed that voters did not interfere with one another as a function of treatment status. Next, we relax that assumption to estimate bounds on the vote by mail treatment effect under six forms of possible interference. Here, we change outcomes among treated voters based on proximity to control voters. After we change those outcomes, we then re-estimate the vote by mail treatment effect. Given that conditioning on past turnout appears to be critical, we apply DID to the data once we have altered outcomes based on possible interference.

Table 5 contains the bounds on the vote by mail treatment under interference. Under the first scenario, we assume that treated voters who lived within 100 meters of at least one control unit were affected by that control unit and changed their behavior accordingly. This is a conservative model since a rather small amount of geographic proximity is all that is required for interference. The DID estimate under this scenario implies that the vote by

Table 5: Bounds on vote by mail treatment effect estimates under six different interference

100 meters			
Number of control voters	1	3	5
Estimate[a]	-5.6	-4.7	-4.7
p-value	0.006	0.03	0.03
250 meters			
Number of control voters	1	3	5
Estimate[a]	-9.6	-5.7	-5.5
p-value	0.000	0.005	0.007

[a] Estimate is calculated via differences-in-differences.

mail treatment decreased turnout by 5.6 percentage points. This estimate is quite similar to the estimates based on DID and matching in Tables 2 and 4.

The next two bounds analysis change treated outcomes when there at least 3 or 5 control units within 100 meters of each treated unit. In both cases, we find that the DID estimate is -4.6 percentage points. As such, if we assume that interference is confined to a 100 meter neighborhood around treated units, we find that our inference is essentially unchanged. Next, we allow the zone of interference to be 250 meters. The first bound changes treated outcomes for treated voters that are within 250 meters of at least one control voter. Under this form of interference, we find that the point estimate for the vote by mail treatment effect nearly doubles in magnitude. Here, the DID estimate is -9.6. However, once we require that a treated voter must be within 250 meters of 3 or 5 control voters the DID estimate is -5.7 and -5.5 percentage points respectively. These estimates are nearly identical to those from both the matching design and the DID estimates.

The bounds analysis implies that only under the most extreme form of interference would our inference be altered. That is, only when we assume that every treated voter within 250 meters of a control had his or her outcomes changed by interference, do we find that the magnitude of the effect increases. Under every other scenario, the estimates are remarkably similar to those that assume that interference is not present.

6 Discussion

In this article, we have demonstrated how one can account for interference in a natural experiment. Often policymakers cannot use classical randomized experiments to understand the effects of voting regulations such as method of registration or all-mail elections. Natural experiments such as we examine here provide important opportunities for researchers to draw causal inferences about all-mail elections. However, natural experiments are also more likely to be subject to interference since researchers are unable to control experimental units. We developed a method of bounds that allows us to estimate treatment effects under different patterns of interference based on geographic proximity.

In our analysis of the 2010 Colorado primary, we find the vote by mail elections appear to suppress turnout. Using our methods of bounds, we found that unless we assume a fairly strong patten of interference, the treatment effects estimated under a no-interference assumption are plausible. Our result is consistent with our studies that have found that all-mail elections, while making voting easier, can actually decrease turnout. Since we focus on a primary election, we are already focusing on a subset of voters that vote in most elections. Moreover, in the 2010 Colorado primary, voters had three salient races on the ballot. Our finding suggests that the social aspects of voting that occur on election day should not be discounted.

References

Abadie, A. (2005), "Semiparametric Difference-in-Difference Estimators," *Review of Economic Studies*, 75, 1–19.

Aronow, P. M. and Samii, C. (2012), "Estimating average causal effects under general interference," Working paper.

Bergman, E. and Yates, P. A. (2011), "Changing election methods: How does mandated vote-by-mail affect individual registrants?" *Election Law Journal*, 10, 115–127.

Berinsky, A. J., Burns, N., and Traugott, M. W. (2001), "Who votes by mail?: A dynamic model of the individual-level consequences of voting-by-mail systems," *Public Opinion Quarterly*, 65, 178–197.

Bowers, J., Fredrickson, M. M., and Panagopoulos, C. (2013), "Reasoning about Interference Between Units: A General Framework," *Political Analysis*, 21, 97–124.

Gerber, A. S., Huber, G. A., and Hill, S. J. (2013), "Identifying the Effect of All-Mail Elections on Turnout: Staggered Reform in the Evergreen State," *Political Science Research and Methods*, 1, 91–116.

Gronke, P., Galanes-Rosenbaum, E., Miller, P. A., and Toffey, D. (2008), "Convenience voting," *Annu. Rev. Polit. Sci.*, 11, 437–455.

Gronke, P. and Miller, P. (2012), "Voting by Mail and Turnout in Oregon Revisiting Southwell and Burchett," *American Politics Research*, 40, 976–997.

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–960.

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

Karp, J. A. and Banducci, S. A. (2000), "Going postal: How all-mail elections influence turnout," *Political Behavior*, 22, 223–239.

Keele, L., Titiunik, R., and Zubizarreta, J. (Forthcoming), "Enhancing Geographic Discontinuities Through Matching," *Journal of the Royal Statistical Society: Series A*.

Keele, L. J. and Minozzi, W. (2012), "How Much is Minnesota Like Wisconsin? Assumptions and Counterfactuals in Causal Inference with Observational Data," *Political Analysis*, 21, 193–216.

Keele, L. J. and Titiunik, R. (2013), "Geographic Boundaries as Regression Discontinuities," Unpublished Manuscript.

Kousser, T. and Mullin, M. (2007), "Does Voting by Mail Increase Participation? Using Matching to Analyze a Natural Experiment," *Political Analysis*, 15, 428–445.

Manski, C. F. (1990), "Nonparametric Bounds on Treatment Effects," *The American Economic Review Papers and Proceedings*, 80, 319–323.

Richey, S. (2008), "Voting by Mail: Turnout and Institutional Reform in Oregon*," *Social Science Quarterly*, 89, 902–915.

Rosenbaum, P. R. (2002), *Observational Studies*, New York, NY: Springer, 2nd ed.

— (2005), "Observational Study," in *Encyclopedia of Statistics in Behavioral Science*, eds. Everitt, B. S. and Howell, D. C., John Wiley and Sons, vol. 3, pp. 1451 – 1462.

Rubin, D. B. (2005), "Causal inference using potential outcomes," *Journal of the American Statistical Association*, 100.

Southwell, P. L. and Burchett, J. I. (2000), "The effect of all-mail elections on voter turnout," *American Politics Research*, 28, 72–79.

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

VanderWeele, T. J., Hong, G., Jones, S. M., and Brown, J. L. (2013), "Mediation and Spillover Effects in Group Randomized Trials: A Case Study of the 4Rs Educational Intervention," *Journal of the American Statistical Association*, 108, 469–481.

Zubizarreta, J. R. (2012), "Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure After Surgery," *Journal of the American Statistical Association*, Forthcoming.