Improving Access to Voting: Matching Voters to Polling Stations Optimally

Guillermo Durán[1,2,3], Mauro Giormenti[2], Mario Guajardo[4], Pablo M. Pinto *[5], Pablo A. Rey[3] and Nicolás E. Stier-Moses[1,6]

[1]CONICET, Argentina
[2]Instituto de Cálculo and Departamento de Matemática, FCEyN, Universidad de Buenos Aires, Argentina
[3]Departamento de Ingeniería Industrial, FCFM, Universidad de Chile, Chile
[4]Department of Business and Management Science, NHH Norwegian School of Economics, Bergen, Norway
[5]Department of Political Science, University of Houston, USA
[6]Universidad Torcuato Di Tella School of Business, Buenos Aires, Argentina

April 17, 2017

Abstract

Research on political representation has traditionally focused on the design of electoral systems. Yet, there is evidence that voting costs result in lower turnout and undermine voters' confidence in the electoral system. Election administrators can selectively manipulate participation costs for different individuals and groups, leading to biased electoral outcomes. Quantifying the costs of voting and designing fair, transparent and efficient rules for voter assignment to polling stations are important for theoretical and practical reasons. We rely on analytical models to quantify the differential costs of participation faced by different voters. To estimate the parameters of these models we use real world data on the 2013 midterm elections in Argentina. The matchings produced by our model cut average voting time by more than 27%, underscoring the inefficiencies of the current method of alphabetical assignment. Our strategy generates better estimates of the role of geographical and temporal conditions on electoral outcomes.

Keywords: Elections, Voting Costs, Linear Programming, Queuing Theory, Geocoding

*Corresponding author. E-mail: ppinto@central.uh.edu. The University of Houston, 3551 Cullen Blvd, Houston, TX 77204-3011.

Acknowledgements

We would like to sincerely thank the Tribunal Superior de Justicia (Ciudad de Buenos Aires) for providing the information to conduct this work, Carlos Prieto for encouraging us to study this problem, and Francisco Cantu, Jeronimo Cortina, Justin Kirkland and Kenneth Rivkin for his comments and suggestions. This study was partly funded by UBACyT grant no. 20020130100808BA (Argentina), ANPCyT PICT grant no. 2012-1324 (Argentina), CONICET-CONICYT International Cooperation Grant Argentina-Chile, and FONDECyT grant no. 1140787 (Chile) as well as by the Complex Engineering Systems Institute ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816).

1 Introduction

Scholarly work on political representation mainly focuses on problems associated with electoral system and ballot design. However, the location and administration of polling stations have sizable effects on electoral outcomes (Alvarez et al. 2008, p. 248). Hurdles to participation increase the opportunity costs of voting, which result in lower turnout (Dyck and Gimpel 2005) and undermine confidence in the electoral system (Claassen et al. 2008). There is now robust evidence that distance to polling stations and waiting in line to vote impose real costs on voters and discourage political participation (Dyck and Gimpel 2005, Stewart III and Ansolabehere 2013). Yet voting costs are not evenly distributed across the electorate. In the US, for instance, racial minorities tend to experience longer voting times than white voters (Stewart III 2012), and urban voters travel less but wait longer to vote (Stewart III and Ansolabehere 2013). Hence varying voting costs results in disenfranchisement, affecting political representation. As Lijphart (1997, p. 2) eloquently puts it: "low voter turnout means unequal and socioeconomically biased turnout."

These insights have piqued academic interest in issues associated with electoral administration. Decisions by authorities in charge of planning and administering elections affect the distance and wait time faced by different groups of voters (Haspel and Knotts 2005, p. 560). These choices could be subjected to opportunistic political manipulation by those authorities, facilitating access to the ballot of their supporters and making it more difficult for their opponents to cast a vote (Nagler 1991, Brady and McNulty 2011). Given existing evidence that the costs of voting vary significantly across individuals of different socioeconomic status, matching voters to polling stations has become a politically salient and controversial problem (Stewart III 2012, Herron and Smith 2016).

Providing electoral administrators with transparent rules for voter assignment and resource allocation has important implications for the functioning of the electoral process, the fundamental pillar of democratic governance.[1] The usual recommendation to reduce voting costs is to better allocate resources across polling stations. However, electoral authorities lack a clear benchmark on how to allocate those resources in an efficient, objective and fair way. Our paper aims at filling that void: Employing tools of mathematical programming we develop a modeling strategy for optimal voter assignment to polling stations, reducing voting costs.[2]

[1] "The use of automated districting procedures is at least able to ensure that systematic distortions of the electoral outcome are avoided, and, denying political parties the opportunity to manipulate districts, they can be very useful to provide fair district maps." (Ricca et al. 2013, p. 250).

[2] Our approach is similar to Allen and Bernshteyn (2006), Orford et al. (2011), Yang et al. (2013), Herron and Smith (2016). This modeling strategy can also be applied to other problems in political science including districting, political mobilization, conflict and resource allocation.

Our modeling strategy allows us to quantify the costs to voting under different voter assignment scenarios. First, we develop a benchmark model to match voters to polling stations that satisfies the fundamental principles of efficiency, objectivity and fairness. We use this benchmark to quantify the costs of participation for voters in Argentina, where voting in national elections is mandatory. The current electoral system in Argentina, which simply matches voters to stations in alphabetical order, does not take into consideration voting costs resulting from travel distance to polling stations and wait times. Using real data from voter assignment to polling stations in one district in the city of Buenos Aires, we document large differences in voting costs resulting from the current electoral system.

We find that minor changes in voter assignment to polling stations resulting from our model lead to sizable changes in the expected costs of voting: on average the matching that results from the benchmark model reduces average voting time by more than 27% in comparison to the current official assignment. This is a very important result, given that even small changes in transit and wait times have a significant impact on participation rates (McNulty et al. 2009) and on perceptions about the legitimacy of the electoral process (Spencer and Markovits 2010, Gerber et al. 2013). We further develop alternative models that vary the capacity of polling places, relocation of polling stations, or both; we compare the results of the current and benchmark matchings, and assess the limitations and obstacles to their implementation.

The rest of this paper is divided into six sections. Section 2 places our contribution in the extant literature; Section 3 describes the current official matching process of voters to polling stations used in Argentinean elections and the matching problem variants that will be solved; Section 4 introduces the proposed voter-matching models, beginning with the benchmark assignment version, which we use to document the extra participation costs experienced by voters under the current system; Section 5 discusses the methodologies for obtaining and processing the data on queues and waiting times at the polling places and the voters' geographical location data; Section 6 implements the models and analyzes the results; and finally, Section 7 presents our closing comments and conclusions.

2 Related literature

Voting is one of the central features of democratic governance. In theory the individual decision to vote is affected by expected benefits and costs of turning out (Downs 1957, Riker and Ordeshook 1968, Aldrich 1993). Since the probability of casting a decisive vote is negligible, any small cost of participation would lead to abstention. The choice to participate in an election will thus vary across voters according to individual characteristics, including interest and motivation. Yet we do observe that many individuals vote systematically, suggesting that political participation is not solely based on self interest (Palfrey and Rosenthal 1983, 1985, Aldrich 1993, Green et al. 1994). Hence, even in highly salient elections large numbers of potential voters abstain, suggesting civic-mindedness is far from universal.[3] Hurdles to participation also play an important role in other major perspectives on the individual and group level determinants of voting (Gomez et al. 2007).[4] In all of these perspectives distance to polling stations and wait time increase the opportunity costs of voting resulting in lower turnout (Dyck and Gimpel 2005, p. 532).

There is now robust evidence that distance to polling stations depress turnout in elections in the U.S. and around the world (Gimpel and Schuknecht 2003, Haspel and Knotts 2005, Dyck and Gimpel 2005, Highton 2006, McNulty et al. 2009, Spencer and Markovits 2010, Brady and McNulty 2011, Herron and Smith 2016). There is also a strong relationship between proximity to polling stations and the propensity to vote in municipal elections in Denmark (Bhatti 2012). A closely related problem that has received attention in recent years is that of queues at polling places and the consequent voter waiting times. Allen and Bernshteyn (2006) provide empirical evidence that people are deterred from voting due to waiting times. Long lines can deter voters, increasing attrition through balking or reneging (Spencer and Markovits 2010). Stewart III and Ansolabehere (2013) analyze polling place queues and their negative effects in the United States.[5] Herron and Smith (2016) refer to several problems recently experienced by voters in USA and Canada, some of them who were unable to vote despite having queued for hours. Wait times also affect the perception that own vote counts, undermining confidence in the electoral system (Claassen et al. 2008).

The problem of defining and modifying the boundaries of an electoral district has been tackled by a number of scholars in the field of election administration. Research in this area focuses on a wide range of issues including voter registration (Rosenstone and Wolfinger 1978, Nagler 1991, Herron and Smith 2016), location of polling stations and allocation of voting machines (Knack 1995, Highton 2006, Wang et al. 2014), the agencies in charge of monitoring the electoral process (Herron and Smith 2016), and voting technologies (Edelstein and Edelstein 2010). Voting location –school, church, fire station, etc.– and the immediate surroundings of a polling place have been shown to have an impact on voting decisions and choices (Wheeler et al. 2008, Rutchick 2010). Interest in election administration is also a response to concerns about

[3]The rational voter hypothesis has been criticized because it fails to explain the levels of voting; however: the rational hypothesis provides important insights into explaining participation at the margin (Aldrich 1993, Blais et al. 2000, Bhatti 2012).

[4]These perspectives place emphasis on socio-economic status (Almond and Verba 2015, Verba and Nie 1987), or group level conditions affecting voter mobilization (Rosenstone and Wolfinger 1978, Rosenstone et al. 1993).

[5]The 2008 Survey of the Performance of American Elections (SPAE) reports that 11% of non-voters were discouraged by long wait times. The number for 2012 was 14.5% (Pettigrew 2013). During the 2008 presidential election 2% of individuals who arrived at polling stations reneged due to long wait times (Spencer and Markovits 2010).

gerrymandering or manipulation of the boundaries of an electoral district to favor a particular political party. Cain et al. (2008), for example, analyze the existence of gerrymandering in California, while Chambers and Miller (2013) introduce measures that can incorporate geographic features to assess the compactness of districts.

On a related problem associated with the administration of voting resources, scholars have used simulations to balance the assignments on voting machines across polling stations driven by equity and efficiency considerations (Yang et al. 2009, 2014, Wang et al. 2014). Yang et al. (2015) apply a robust optimization approach to the voting machine allocation problem aimed at reducing the number of voters experiencing extremely long waiting times. The numerical results reported in these studies are derived from data for the 2008 election in Franklin County, Ohio, USA. Despite these contributions, a recent review of the empirical literature on voting in the U.S. by Ansolabehere and Shaw (2016) concludes that we have limited knowledge on how resource allocation affects voting costs and the quality of the voting experience. Moreover, while political districting and voting machine assignment problems in the aforementioned literature are important, the problem of matching voters to polling stations given predefined districts and resource allocations has received much less attention. This is an important oversight, especially when considering democracies in developing countries, which have limited access to resources and technology. In these settings elections usually take place on a single day, voters are assigned beforehand to a single location, and voting and tallying votes is conducted manually.

The previous discussion underscores the importance of designing voter matching mechanisms that fulfill two basic criteria: first, the matching mechanism should be globally efficient, reducing the costs of participating in the electoral process; and, second, the matching should follow objective criteria, such as geography and wait times, which are less susceptible to political manipulation. Analytical models can help designing optimal mechanisms that fulfill these criteria.[6] Our contribution is three-fold: First, we use well established analytical frameworks, specifically from mathematical programming and queuing theory, to address the problem of efficiently assigning voters to polling stations.[7] This strategy allows us to capture the decision problem according to objective measures of efficiency. Lastly, we carry out an empirical analysis using real data for the 2013 national election in Argentina, in which we are able to document unnecessary hurdles to participation faced by Argentine voters.

Naturally, the problem of minimizing voting costs, including location of polling places and allocation of

[6]The use of optimization to issues of election administration has been investigated since the 1970's (Garfinkel and Nemhauser 1970), and is receiving stronger attention in recent years (Tasnádi 2011, Ricca et al. 2013, Kalcsics 2015)

[7]See also Lenzi (2013) who develops a similar strategy and obtains preliminary results for Pergamino, an electoral district in Argentina.

resources to polling places (poll stations, voting machines, capacities), interacts with political districting, ballot design and other attributes of the electoral system. We take the electoral system, districting and resources as given, to model the voting costs resulting from the matching of voters to polling places. We augment the model by incorporating the allocation of resources as decisions to minimize total voting times.

3 Measuring voting costs

The approach we propose employs mathematical programming to match voters to polling stations. Our method relies on a matching problem whose formulation and resolution strategies are already well established (Nemhauser and Wolsey 1988). We use this model to quantify the costs of participation using real data on voters in one district in Argentina during the 2013 national midterm elections. Voting in Argentina's national elections is mandatory, but a sizable portion of the electorate does not turn out to vote. All eligible voters must cast their ballots at polling stations, one or more of which are set up for the purpose in a single polling place. Typically, these polling places are located within schools, and will generally be referred to as such hereafter. The country is partitioned into electoral divisions (or districts) and voters are assigned to vote at polling stations in schools situated within the division they live in. Voters cast votes for state-wide party lists. The City of Buenos Aires elects 21 representatives to the National Congress; congressional seats are allocated proportionally to party lists following the D'Hondt formula. We take the electoral system as a given, focusing our analysis on the method of voter assignment to polling stations.

Under the current National Elections Code (National Code), the matching of voters to specific polling stations, and therefore schools, is performed according to an essentially random process. Although electoral divisions usually are relatively small in area, the assignments are often not geographically efficient, resulting in higher costs of participation for the average voter. More specifically, it is possible to define an alternative mechanism where the matching of two voters to schools could be interchanged in such a way that the distances to their polling places would be shortened. Also, changes to the current distribution of polling stations and their (legally defined) voter capacity could reduce the time voters spend queuing at polling stations once they arrive at their assigned school. The purpose of this study is to develop an efficient and fair voter assignment strategy, which minimize distances to polling stations and waiting times, and quantify the costs of voting under the current matching mechanism in relation to that benchmark.

We begin our analysis by measuring the time each voter requires to walk from their residence to their assigned school and back. Next, we investigate the dynamics of the voting process at the polling station

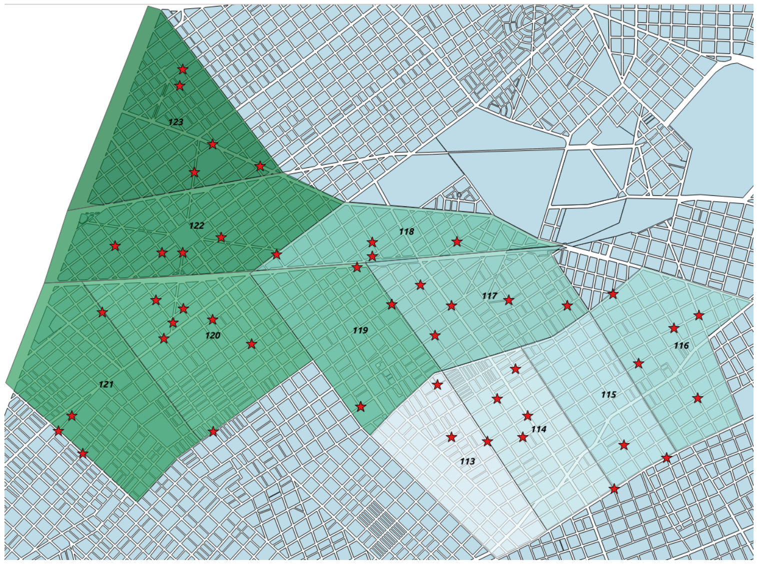

Figure 1: Partition of the 11th Comuna of Buenos Aires into electoral divisions. Red stars indicate schools designated as polling places.

and compute queue waiting times. We obtained voter data for the 11th *Comuna* (administrative district) of the City of Buenos Aires.[8] District 11 is representative of an average urban district in Argentina. It is partitioned into 11 electoral divisions numbered 113 to 123, configured as shown on the map in Figure 1.

Each voter in the district is assigned to a polling station (*mesas electorales*) within a school (polling station). The locations of schools designated as polling stations are indicated by red stars on Figure 1. The information gathered also included the schools' and voters' addresses. Under the current system each polling station has an average capacity of 350 electors. Table 1 reports the turnout rates for each of the polling stations within the 11 electoral divisions in the district. The table shows that there is ample variation in turnout rates per polling station for each electoral division. Although voting is mandatory, as seen in the table not everybody votes in practice.

Having collected this real-world data enabled us to rigorously analyze voting costs resulting the current official matching defined by the election authorities, in comparison to the wait times and distance from matchings generated by our models. We start by characterizing the voter matching problem.

[8] *Comunas* are geographical and administrative districts of the City of Buenos Aires, identified ordinally (1st to 15th) and ranging in population from 150,000 to 230,000 inhabitants. Each one is subdivided into several electoral divisions.

Electoral Division	Mean	Std. Dev.	Median	Min	Max
113	0.79	0.03	0.79	0.74	0.86
114	0.78	0.03	0.78	0.69	0.83
115	0.78	0.02	0.78	0.73	0.82
116	0.78	0.03	0.79	0.70	0.84
117	0.79	0.02	0.79	0.73	0.84
118	0.79	0.03	0.79	0.71	0.88
120	0.79	0.03	0.79	0.73	0.85
121	0.80	0.04	0.80	0.61	0.89
122	0.77	0.05	0.78	0.53	0.84
123	0.79	0.05	0.80	0.49	0.84
119	0.78	0.09	0.80	0.14	0.84

Table 1: Turnout per polling station

3.1 The polling place voter matching problem

The National Elections Directorate, an agency of the Argentinean Ministry of the Interior and Transport, is the entity in charge of scheduling, organizing and carrying out all activities related to national elections. Among other tasks it is responsible for matching voters to polling stations, and therefore to the schools, that is, the polling places where the stations are located.[9]

The official matching process currently in force (hereafter described simply as "current system") is set out in detail in the most recent National Elections Code and related regulations, but the following sums up the main points required for purposes of the present analysis.

- *Geographic subdivisions.* As noted above, Argentina is partitioned for electoral purposes into units known as *circuitos electorales* or electoral divisions. The Code currently requires that every adult citizen must vote within the division they officially reside in.
- *Polling stations.* All voters resident in a given electoral division are assigned across *mesas* or polling stations distributed among multiple schools within the division. The capacity of each station is 350 voters over the course of the election day.
- *Assignment of voters to stations.* An appropriate number of polling stations must be set up within each school. All persons on the official voters list are sorted alphabetically by surname and then assigned to the stations in that order.

Note that the number of polling stations in each school is known in advance. If no geographical data

[9] The Directorate's authority to administer elections emanates from Decree No. 682/2010.

is taken into account other than the partition into electoral divisions, the assignment just described will be equivalent to any other random assignment of voters to stations.

3.2 Basic problem

The first step to identifying improvements to the current system using the available geographical data is to formulate the basic voter matching problem as described below. Given a specific electoral division, we know:

- The list of voters and their addresses.
- The list of schools available for use as polling places, their addresses and the number of polling stations set up in each one.

We further assume that the assignments for all persons on the voter registry to a polling station in the previous election (i.e., the last place they voted) are also known. A model for voter matching is efficient if it assigns each voter to a school in such a way as to minimize the total distance walked by all voters in the electoral division; for this aspect of the problem, the station assignments within a school are irrelevant. To ensure that no voters will be negatively affected by the implementation of this improved assignment, we incorporate a "fairness restriction" in the form of a constraint imposing that no voter may be assigned to a school further from their residence than that assigned in the previous election. This is in line with previous literature acknowledging that in many applications it is not sufficient to reduce the waiting times for some voters at the expense of longer waiting times of others (Yang et al. 2013, 2014).

3.3 Model extensions

To enrich the analysis and explore other possible improvements to the current matching system, we consider the following variants and extensions:

1. The problem is broadened to include a concept of total voting time, equal to walking time plus waiting (queuing to vote) time.
2. The polling stations are interchangeable, that is, they may be redistributed to other schools provided the total number of stations is maintained.
3. The maximum number of voters per station is an optimization variable rather than just an input. All stations within a school are assumed to have the same capacity.

In the latter two cases, the "capacity" of a school is an output variable assigned by the mathematical problem. In the third case, we highlight that waiting time depends on the number of voters per station.

4 Optimization models

We now set up the model for the basic problem described in Section 3.2 above and then generalize this simple formulation to those described in Section 3.3. As it will be explained further in Section 5.3, all models rely on voters' residence data grouped by the closest intersection rather than their exact physical location. This is done for privacy considerations as well as to reduce the complexity when solving the problem.

4.1 Basic model

Sets and parameters

Let I be the set of intersections at which voters are grouped and from which the voters are assumed to walk to their assigned polling places. Let S be the set of schools designated as polling places. The following parameters are also defined:

Parameter	Description
$vot_i \in \mathbb{Z}_+, i \in I$	Number of voters from intersection i.
$dist_{is} \in \mathbb{R}_+, i \in I, s \in S$	Distance between intersection i and school s (in km.).
$cap_s \in \mathbb{Z}_+, s \in S$	Maximum capacity of school s.
$current_{is} \in \mathbb{Z}_+, i \in I, s \in S$	Number of voters from intersection i *currently* assigned to school s .

Decision variables

A single set of non-negative integer variables for counting the voters from each intersection that are assigned to the different schools is defined as follows:

$x_{is} \in \mathbb{Z}_+, i \in I, s \in S$ Number of voters from intersection i assigned to school s.

Objective function

The objective function minimizes the total distance walked by all voters:

$$\min \sum_{i \in I} \sum_{s \in S} 2 \cdot dist_{is} \cdot x_{is} \,.$$

Note that the distance traveled by voters to vote is actually twice the distance from their residence to their assigned polling station. This double factor does not affect the optimization of the basic model but plays an important role once the model is extended to account for total voting time.

Constraints

1. All voters are assigned to a school:

$$\sum_{s \in S} x_{is} = vot_i \qquad \forall i \in I\,.$$

2. No school's voter capacity is exceeded:

$$\sum_{i \in I} x_{is} \leq cap_s \qquad \forall s \in S\,.$$

3. No voter is assigned to a school further from their residence than the current assignment (the fairness restriction):

$$\sum_{j \in S : dist_{ij} > dist_{is}} (current_{ij} - x_{ij}) \geq 0 \qquad \forall i \in I, s \in S\,.$$

4.2 General model with interchangeable polling stations and optimizable polling station capacity

We now generalize the basic model by extending it in various dimensions. First, the objective function is expanded to incorporate total voting time and thus includes both travel time to and from the school and waiting time while queuing to vote. Second, we relax constraints limiting the redistribution of polling stations between schools and the maximum number of voters per station.

The notation for the general model is similar to that of the basic formulation. For ease of reference we repeat the definitions of the common sets and parameters below as well as setting out the new ones needed for expressing the model extensions.

Sets and parameters

Set	Description
I	Set of intersections at which voters are grouped.
S	Set of schools designated as polling places.
K	Set of possible polling station capacities (number of voters).
M	Set of possible number of polling stations per school.

Parameter	Description
$vot_i \in \mathbb{Z}_+, i \in I$	Number of voters from intersection i.
$dist_{is} \in \mathbb{R}_+, i \in I, s \in S$	Distance between intersection i and school s (in km.).
$current_{is} \in \mathbb{Z}_+, i \in I, s \in S$	Number of voters from intersection i *currently* assigned to school s .
$stations_s \in \mathbb{Z}_+, s \in S$	Current number of polling stations at school s.
$wait_k \in \mathbb{R}_+, k \in K$	Mean waiting time at polling station with a capacity of k voters (in minutes).

Decision variables

In addition to the single set of decision variables in the basic model, the general model incorporates a set for modeling decisions regarding the number of polling stations and their voter capacity and a second set for the resulting total capacity for each school. These variables are as follows:

Variables	Description
$x_{is} \in \mathbb{Z}_+, i \in I, s \in S$	Number of voters from intersection i assigned to school s.
$cap_s \in \mathbb{Z}_+, s \in S$	Number of voters assigned to school s.
$statcap_{skl} \in \{0,1\}, s \in S, k \in K, l \in M$	Binary variable indicating whether or not there are l polling stations at school s with k or $k-1$ voters.

Recall that the polling stations at each school must by law have the same number of assigned voters. However, if the total number of voters is not a perfect multiple of the number of stations, the best that can be done is to assign numbers of voters such that any two stations differ by no more than 1 voter. Thus, if the stations have a capacity of k voters, each one will be assigned k or $k-1$ voters. This is captured by the $statcap_{skl}$ variables.

Objective function

The objective function of the general model minimizes the total voting time, defined above as the sum of the walking time to and from the school and the waiting time queuing to vote at the polling station. Thus,

$$\min \sum_{i \in I} \sum_{s \in S} 30 \cdot dist_{is} \cdot x_{is} + \sum_{s \in S} \sum_{k \in K} \sum_{l \in M} k \cdot l \cdot wait_k \cdot statcap_{skl} \,.$$

The first term is now walking time instead of distance as such, calculated by assuming a walking speed of 4 km/h (i.e., 15 minutes per km). This translates into multiplying the distance to and from the school by a factor of 30. The second term is the queue waiting time.

Constraints

1. All voters are assigned to a school:

$$\sum_{s \in S} x_{is} = vot_i \qquad \forall i \in I \,.$$

2. The school capacity is exactly the number of voters assigned to it:

$$\sum_{i \in I} x_{is} = cap_s \qquad \forall s \in S \,.$$

3. No voter is assigned to a school further from their residence than the current assignment system (the fairness restriction):

$$\sum_{j \in S : dist_{ij} > dist_{is}} (current_{ij} - x_{ij}) \geq 0 \qquad \forall i \in I, s \in S \,.$$

4. The capacity of a school depends on the number of polling stations and their capacity:

$$1 + \sum_{k \in K} \sum_{l \in M} l \cdot (k-1) \cdot statcap_{skl} \leq cap_s \leq \sum_{k \in K} \sum_{l \in M} l \cdot k \cdot statcap_{skl} \quad \forall s \in S \,.$$

5. A certain number of polling stations and their respective voter capacities must be defined for each

school:

$$\sum_{k \in K} \sum_{l \in M} statcap_{skl} = 1 \qquad \forall s \in S \, .$$

6. The number of polling stations after all relocations must be the same as the number in the current assignment (i.e., before redistribution):

$$\sum_{s \in S} \sum_{k \in K} \sum_{l \in M} l \cdot statcap_{skl} = \sum_{s \in S} stations_s .$$

5 Estimation of model parameters

In this section we describe the methodology for estimating waiting times using a queuing theory model (see, e.g., Kleinrock 1975), and how the geographical data from the voters list was processed to determine voters' distances from schools designated as polling places.

5.1 Modeling of voter queues

To implement the general model incorporating voter waiting time we must capture certain details of the voting process that leads to the formation of voter queues outside polling stations. In particular, we must determine the relationship between average queue waiting time and polling station capacity, which is a function of the voter arrival and voting rates. The interarrival time is the time that elapses between two consecutive voter arrivals at a polling station queue while service time is the time a voter takes to cast a ballot once it is their turn to leave the queue and enter the voting booth.

To estimate these parameters we collected data during the two rounds of the Argentine midterm elections of 2013: the mandatory primaries (PASO in its Spanish acronym) held on August 11, and the general election that took place on October 27. These elections are well documented in Singer (2014). On both elections we recorded the voting process dynamics at several polling stations of a school in the district. The information we registered included the exact times voters arrived, entered a voting booth, and then left the booth allowing the next voter in the queue to enter. We also noted the gender and (approximate) age of each voter, although neither characteristic turned out to be a statistically significant predictor of voting behavior differences.

Not surprisingly, voter arrival rates varied throughout the day. To control for that phenomenon we grouped the data into time slots and analyzed each group separately. Analyses of these data suggest that arrivals approximately follow a Poisson process with a different rate for each group. The service times, on the other hand, approximated a log-normal distribution that did not vary throughout the day. For some time slots we were not able to collect enough data to estimate the average arrival rate; in such cases the rates were extrapolated from those that were accurately estimated.

5.2 Estimation of waiting times and queue simulation

Having inferred the characteristics of the voter arrival and service times, we conducted a series of simulations to estimate the voter queue waiting times. Simulation techniques have proved to be useful in previous electoral studies, such as Yang et al. (2009, 2014), Herron and Smith (2016). We built and implemented a simple queue simulator in the Ruby programming language. The voter arrival rates for various periods during the election day were inputted to the simulator, which then generated voter arrivals following a non-homogeneous Poisson process with a constant rate for each period and simulated the service times using a log-normal distribution fitted to the data.

These results were validated by comparing the average simulated waiting times to the real data, that is, the results observed in the field. We then used the simulator to estimate the average voter waiting times as a function of the number of voters assigned to each polling station. Thus, we performed simulations in which the arrival rates were varied in proportion to the station capacities. The formula used to compute the arrival rate λ_k for a given time slot and station with k assigned voters was $\lambda_k = k\lambda_{350}/350$. For each unit of station capacity between 60 and 500 voters we ran 100 simulation iterations and took the average waiting time for each run. The output was highly correlated to an exponential functional form and therefore was fitted to one, as shown in Figure 2. The resulting function was

$$wait_k = \exp(-2.5897 + 0.0133k) + 1.17\,.$$

5.3 Coding of voter residence data

The geographical data used to define voters' residences for purposes of calculating their walking distances was derived from the voters list for the 11th administrative district of Buenos Aires as noted earlier. The list contained three data fields for each voter: address, assigned school and polling station number. Unfortunately, this information was not entirely well ordered or even wholly accurate. The main defect was the inconsistencies in street name spellings, but the use of a single field for the street name and the house,

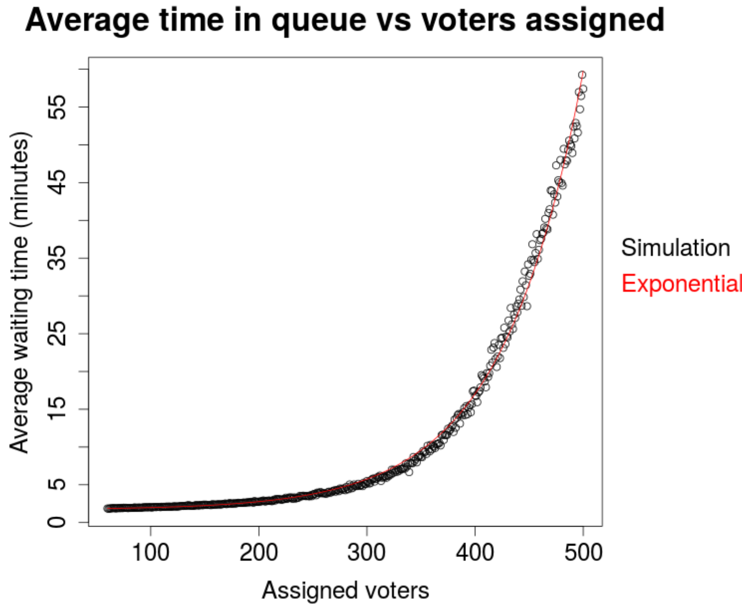

Figure 2: Each circle represents the average voter waiting time generated by 100 simulations executed for each possible number of assigned voters. The curve is the exponential function that best fits the circled points.

building and/or apartment number also complicated the automatic processing of the data.[10] To rectify this situation we implemented some rudimentary speech recognition and grammar corrector heuristics that did spare us much manual work, but considerable manual revision was still required to fix cases the heuristics could not handle. If the model were applied at a much larger scale, these problems would present a more serious obstacle.

Once the data were finally clean and well ordered, we then used Google's geocoding API (Google 2016) to map each voter's address to its latitude and longitude. These coordinates were then loaded into a geographic information system (GIS) along with layers that included the streets and intersections of Buenos Aires. The tool used for this process was Quantum GIS (QGIS Development Team 1999).

Since our solution does not require extreme precision and privacy was a concern, we grouped voters at the intersection nearest to their exact addresses. The data actually fed to the model were thus the number of voters "located" at each intersection, the latter acting as a sufficiently accurate estimate of voter residence location. Using the GIS, we then calculated the matrix of distances from every intersection to every school. As well as maintaining voter anonymity, this approach significantly reduced the size of the problem.

[10]For example, the street properly spelled as Calle Joaquín V. González was found on the list under numerous different variations such as Joaquin V Gonzalez, Av Joaquin V Gonzalez, J V Gonzalez, J B Gonzalez, J Gonzalez, V Gonzalez,

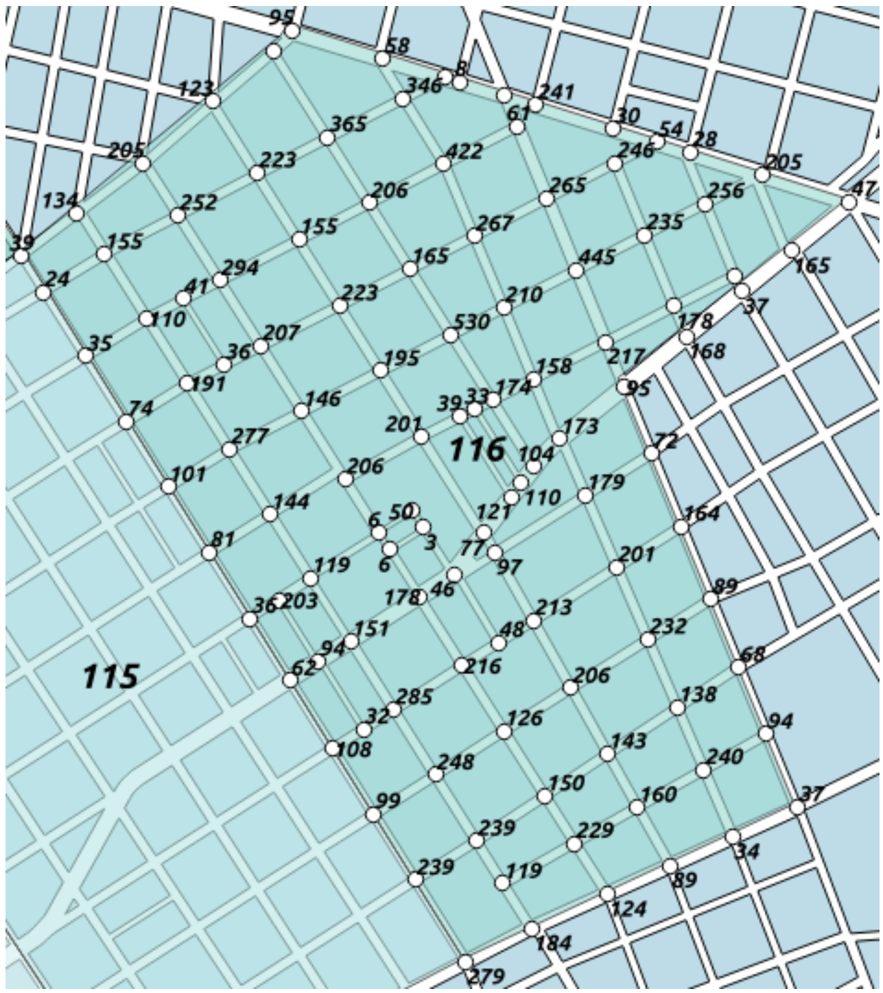

Figure 3: Number of voters assigned to each intersection in electoral division 116.

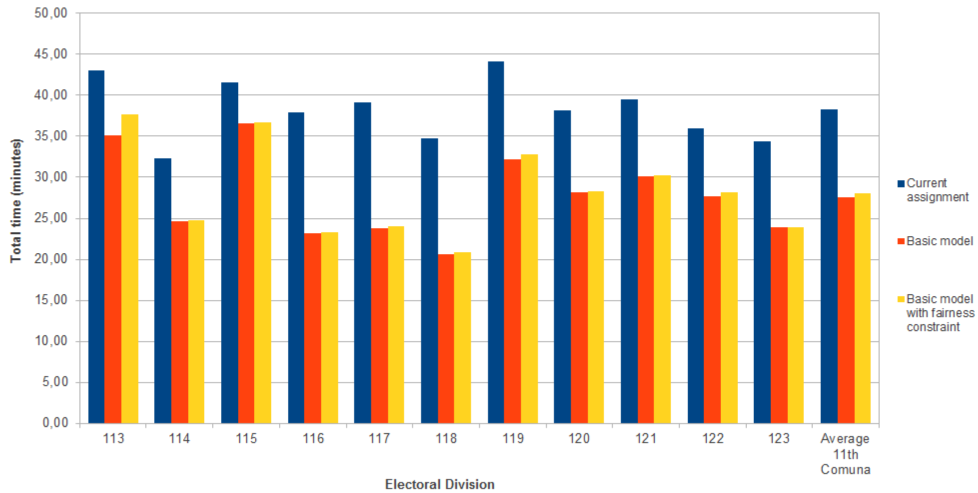

Figure 4: Total voting time per person under the current official assignments and the basic model assignments (with and without the fairness restriction).

Finally, any intersection that had no grouped voters was discarded as irrelevant. An example of the numbers of voters grouped at each intersection in a single electoral division is shown in Figure 3.

6 Computational results

The voter matching models were implemented using Pyomo, an open-source optimization modeling language. The GNU Linear Programming Kit (GLPK), also freeware, was used as the solver. The optimization problems were fed with the real-world voters list data described above. Below, we describe the results of that computational study.

The expected total voting time under the current official assignments and the assignments generated by the basic model with and without the optional fairness restriction (Constraint 3) are shown in Figure 4. Both basic models produced significant improvements for every electoral division. The total time savings per voter ranged from 5 minutes (Division 115) to more than 15 minutes (Division 117), the average for the entire 11th *Comuna* being 10 minutes and 40 seconds. Since the average time under the current official assignments is 38 minutes and 19 seconds, the savings amount to more than 27%.

Confronting these results with Figure 1 suggests there may be a correlation between the variation in the improvements among the electoral divisions and the geographical distributions of the schools within them.

Gonzalez, and Joaquin V Gonzales.

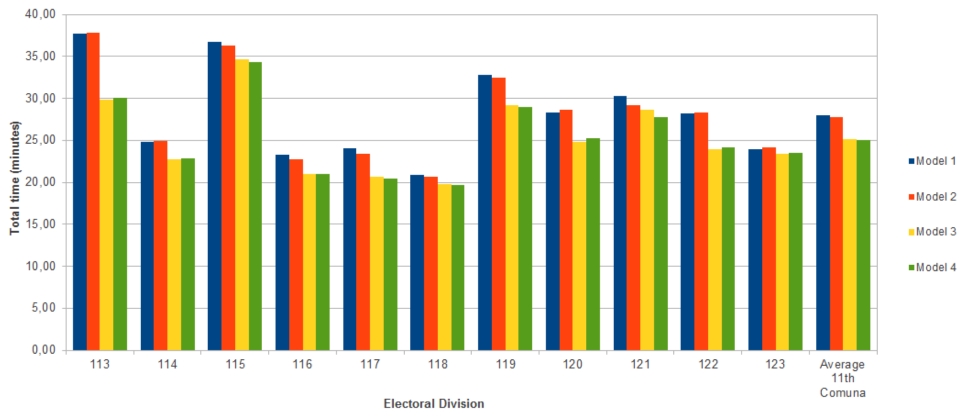

Figure 5: Total voting time per person under assignments generated by the basic model (Model 1) and the general model variants (Models 2, 3 and 4).

The smallest improvements are found in divisions where most of the schools are close to each other (such as 113, 114 and especially 115) while the largest savings occurred in divisions where the school distribution is relatively evenly spaced (for example, 116, 117 and 118). This is what we would intuitively expect given that the model attempts to assign voters to schools in such a way that their walking distances are as short as possible. If the designated schools are all close together, it will make little difference which ones voters are assigned to.

The results of the basic model with the fairness restriction (eliminating model assignments that increase any voter's total voting time beyond that for the current official assignments) are also shown in Figure 4. As can be seen, with the sole exception of Division 113 the negative impact of the added condition was negligible. We therefore opted to keep the constraint in all model runs.

Turning now to the general model, we implemented three variants denoted Models 2, 3 and 4 in what follows. Model 2 allowed capacity to be varied from school to school; Model 3 allowed polling stations to be relocated from one school to another; and Model 4, specified in Section 4.2, allowed both modifications. The results of the three variants together with those of the basic model, now denoted Model 1, are shown in Figure 5.

What stands out in these results is that Model 2's performance was almost the same as that of Model 1 while Models 3 and 4 were also all but identical. Since the difference within both pairs was that school capacities are allowed to vary, the fact that in neither case did performance change with this factor suggests that flexibility in school capacity affords no advantage, at least as long as the mean number of voters per

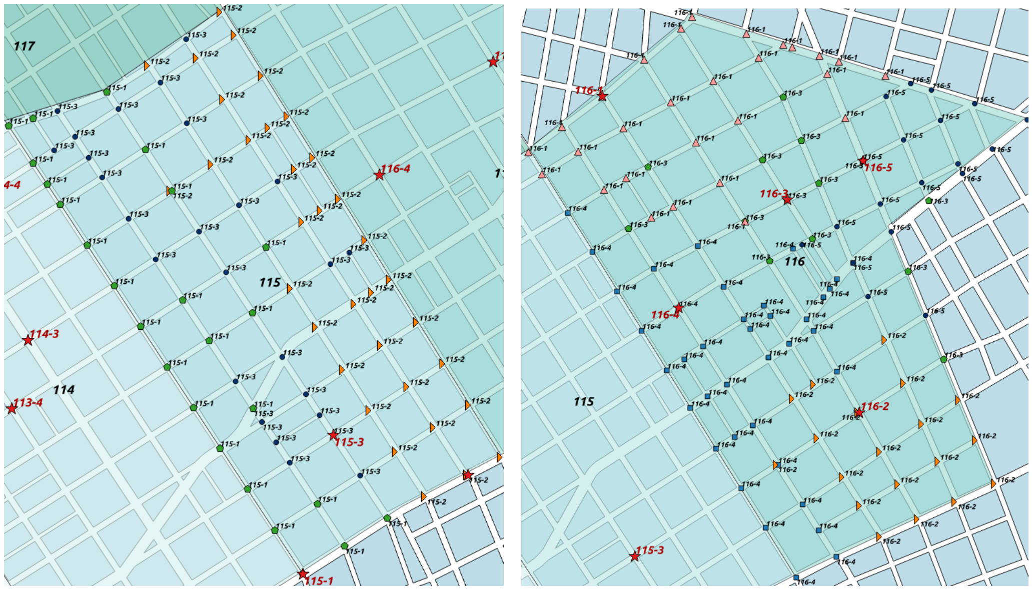

Figure 6: Model 1 (basic model): distribution of voter assignments by intersection in electoral divisions 115 and 116. Intersections are numbered to match the school their grouped voters are assigned to.

polling stations remains near 350.

We highlight the substantial improvement between the first pair (Models 1 and 2) and the second pair (Models 3 and 4), an average reduction of 3 minutes. The option provided by the latter two formulations to relocate polling stations meant that they could be moved between schools wherever this would improve voting times. How effective was this flexibility can be appreciated by comparing the geographic distributions of the intersection (i.e., voter) assignments generated by Model 4 with those of Model 1. The respective distributions for the two models are shown in Figures 6 and 7 for electoral divisions 115 (on the left) and 116 (on the right). The intersections are tagged with a number whose final digit matches the final digit of the school (a red star) it was assigned to; for greater readability, tags of intersections assigned to the same school have the same shape and color.

In the Model 1 case (Fig. 6), we see that Division 115 voters assigned to Schools 1 and 3 are concentrated roughly in two parallel vertical bands while those assigned to School 2 are scattered throughout the division. This pattern reflects the bunching of the 3 schools at the “bottom” of the division map.

These improvements suggest that the basic model’s performance is enhanced by allowing polling stations to be relocated from one school to another. It should be noted, however, that we did not set any limit to the number of stations a school may have. This implies that the Model 3 and 4 results are in fact the upper bound of what could be attained.

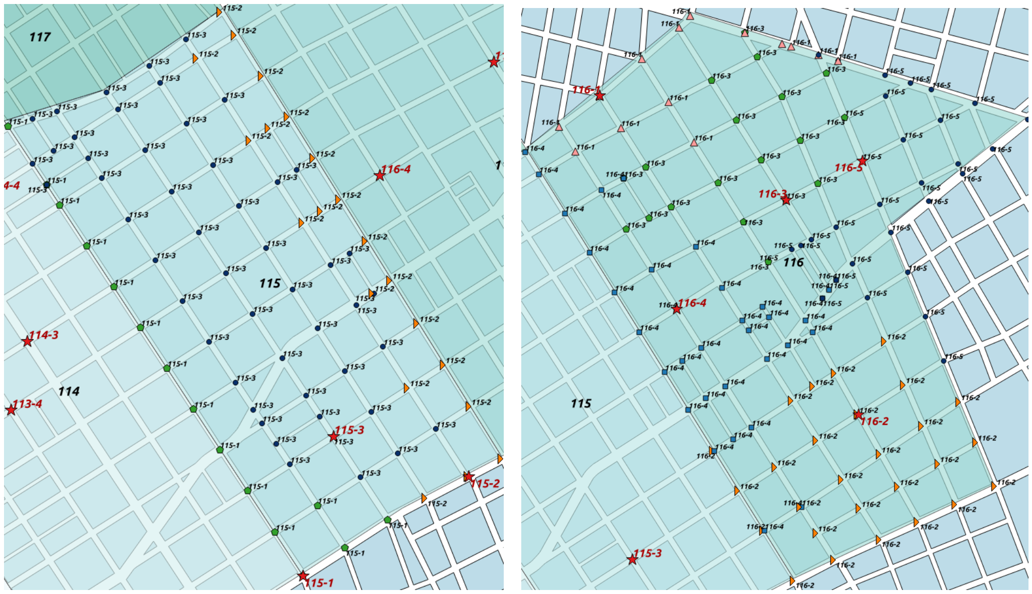

Figure 7: Model 4 (with station relocation option): distribution of voter assignments by intersection in electoral divisions 115 and 116. Intersections are numbered to match the school their grouped voters are assigned to.

7 Conclusion

A recent body of literature has produced persuasive evidence that voting costs are not evenly distributed across the electorate, affect political representation and undermine confidence in democratic governance. For these reasons scholars have turned their attention to understanding issues associated with the administration of elections. Our paper relies on analytical models to quantify the costs of voting and assess the fairness, transparency and efficiency of voter matching processes. We illustrate our contribution using voter matching data from an electoral district in the 2013 midterm elections in Argentina. Our models, along with the ensuing computational experiments, provide ample evidence that the voter matching resulting from the current system is inefficient. Indeed, the implementation of our benchmark model would cut the average voting time under the current official matching by more than 27%. These models fulfill basic criteria of efficiency, fairness and objectivity: they reduce voting time, ensure that no individual voter is made worse off under the optimized matching, and are based on objective criteria associated with minimizing distance travelled and wait time.

An analysis of the results has also revealed that relaxing the constraint that imposes an equal number of voters assigned to each polling station, as stipulated under the current election regulations, would be of little benefit since the resulting reductions in voting times were almost the same as those obtained with the unrelaxed version. By contrast, a redistribution of stations among the various polling places within an electoral division may contribute to significant voting time improvements. Computing how large these reductions might be for a voting division and measuring their impact on turnout and voter satisfaction would necessarily require the collection of additional data. If the benefits prove to be considerable, changes could be implemented promptly since current regulations allow for such changes, unlike modifying the number of rotations assigned to each voting station. It should be noted that the district used for this study is a densely populated area in Buenos Aires. If the proposed models were applied to districts in rural or lower-density urban areas where the divisions are significantly larger, we would expect the voting time reductions achieved by the models' more-efficient assignments would be even larger. Comparisons of the different electoral divisions in the district suggest that significant improvements could also be obtained if polling places were distributed within the divisions more uniformly. To attain an improvement in this sense would require a political decision to redesign electoral subdivision and would possibly be more difficult to implement. In any case, district redesign is an optimization problem in its own right that we leave for a future investigation.

In addition, we remark that although the current official method of matching voters to polling stations in alphabetical order is simple and intuitive, these virtues by themselves are of limited value. The implementation of any of the models presented here is eminently feasible and would generate results that are more efficient than the current allocation. The improved matching we have proposed could be easily implemented under Argentina's current electoral rules, as it requires a minor administrative decision: dropping the requirement that the voters in each electoral division are assigned to voting stations in alphabetical order. In any event, our modeling strategy can be used in many different settings, and could help generate better estimates of the role of geographical and temporal conditions on the electoral experience of voters.

Finally, we underscore the methodological contribution of our paper. Assigning voters to polling stations is analogous to a matching problem whose formulation and resolution are well studied within the *Operations Research* community. The modeling strategy presented in the paper can be extended to analyze other fundamental problems in politics, including allocation of scarce resources, redistricting, political mobilization, and conflict. We believe this is a promising area of research for political scientists and methodologists.

References

J. H. Aldrich, 1993. Rational choice and turnout. *American Journal of Political Science*, 37(1):246–278.

T. Allen and M. Bernshteyn, 2006. Mitigating voter waiting times. *Chance*, 19(4):25–34.

G. A. Almond and S. Verba, 2015. *The civic culture: Political attitudes and democracy in five nations*. Princeton University Press.

R. M. Alvarez, S. Ansolabehere, A. Berinsky, G. Lenz, C. Stewart III, and T. Hall, 2008. Survey of the performance of American elections: Final report.

S. Ansolabehere and D. Shaw, 2016. Assessing (and fixing?) election day lines: Evidence from a survey of local election officials. *Electoral Studies*, 41:1–11.

Y. Bhatti, 2012. Distance and voting: Evidence from Danish municipalities. *Scandinavian Political Studies*, 35(2):141–158.

A. Blais, R. Young, and M. Lapp, 2000. The calculus of voting: An empirical test. *European Journal of Political Research*, 37(2):181–201.

H. E. Brady and J. E. McNulty, 2011. Turning out to vote: The costs of finding and getting to the polling place. *American Political Science Review*, 105(01):115–134.

B. Cain, I. Uhui, and K. MacDonald, 2008. Sorting or self-sorting? competition and redistricting in California. In F. Douzet, T. Kousser, and K. Miller, editors, *The New Political Geography of California*, pages 245–265. Berkeley Public Policy Press.

C. P. Chambers and A. D. Miller, 2013. Measuring legislative boundaries. *Mathematical Social Sciences*, 66 (3):268–275.

R. L. Claassen, D. B. Magleby, J. Q. Monson, and K. D. Patterson, 2008. 'At your service': Voter evaluations of poll worker performance. *American Politics Research*, 36(4):612–634.

A. D. Downs, 1957. *An economic theory of democracy*. Harper and Row, New York, NY.

J. J. Dyck and J. G. Gimpel, 2005. Distance, turnout, and the convenience of voting. *Social Science Quarterly*, 86(3):531–548.

W. Edelstein and A. Edelstein, 2010. Queuing and elections: Long lines, DREs and paper ballots. In *Proceedings of the 2010 Electronic Voting Technology Workshop*. Workshop on Trustworthy Elections (EVT/WOTE '10). URL `http://www.usenix.org/events/evtwote10/tech/full_papers/Edelstein.pdf`.

R. S. Garfinkel and G. L. Nemhauser, 1970. Optimal political districting by implicit enumeration techniques. *Management Science*, 16(8):B–495.

A. S. Gerber, G. A. Huber, D. Doherty, and C. M. Dowling, 2013. Is there a secret ballot? Ballot secrecy perceptions and their implications for voting behaviour. *British Journal of Political Science*, 43(01):77–102.

J. G. Gimpel and J. E. Schuknecht, 2003. Political participation and the accessibility of the ballot box. *Political Geography*, 22(5):471–488.

B. T. Gomez, T. G. Hansford, and G. A. Krause, 2007. The Republicans should pray for rain: Weather, turnout, and voting in us presidential elections. *Journal of Politics*, 69(3):649–663.

Google. *Google's geocoding API documentation*, 2016. URL `https://developers.google.com/maps/documentation/geocoding`. Last accessed Nov 2016.

D. P. Green, I. Shapiro, and I. Shapiro, 1994. *Pathologies of rational choice theory: A critique of applications in political science*. Cambridge Univ Press.

M. Haspel and H. G. Knotts, 2005. Location, location, location: Precinct placement and the costs of voting. *Journal of Politics*, 67(2):560–573.

M. C. Herron and D. A. Smith, 2016. Precinct resources and voter wait times. *Electoral Studies*, 42:249–263.

B. Highton, 2006. Long lines, voting machine availability, and turnout: The case of Franklin County, Ohio in the 2004 presidential election. *PS: Political Science & Politics*, 39(01):65–68.

J. Kalcsics, 2015. Districting problems. In G. Laporte, S. Nickel, and F. Saldanha da Gama, editors, *Location Science*, pages 595–622. Springer.

L. Kleinrock, 1975. *Queueing Systems*, volume I: Theory. Wiley Interscience.

S. Knack, 1995. Does 'motor voter' work? evidence from state-level data. *The Journal of Politics*, 57(03): 796–811.

J. Lenzi, 2013. Asignación eficiente de votantes a escuelas mediante programación matemática: El caso de la ciudad de Pergamino. Tesis de licenciatura, Universidad Nacional de Córdoba, Argentina. (in Spanish).

A. Lijphart, 1997. Unequal participation: Democracy's unresolved dilemma. *American Political Science Review*, 91(1):1–14. Presidential Address, American Political Science Association.

J. E. McNulty, C. M. Dowling, and M. H. Ariotti, 2009. Driving saints to sin: How increasing the difficulty of voting dissuades even the most motivated voters. *Political Analysis*, 17(4):435–455.

J. Nagler, 1991. The effect of registration laws and education on us voter turnout. *The American Political Science Review*, pages 1393–1405.

National Code, 1983. National Code for Elections. Law Number 19.945 of the Argentine Republic.

G. L. Nemhauser and L. A. Wolsey, 1988. *Integer and Combinatorial Optimization. Interscience Series in Discrete Mathematics and Optimization*. John Wiley & Sons.

S. Orford, C. Railings, M. Thrasher, and G. Borisyuk, 2011. Changes in the probability of voter turnout when resiting polling stations: A case study in Brent, UK. *Environment and Planning C: Government and Policy*, 29(1):149–169.

T. R. Palfrey and H. Rosenthal, 1983. A strategic calculus of voting. *Public Choice*, 41(1):7–53.

T. R. Palfrey and H. Rosenthal, 1985. Voter participation and strategic uncertainty. *American Political Science Review*, 79(01):62–78.

S. Pettigrew, 2013. Time tax: Which groups wait in the longest lines on election day? Working paper.

QGIS Development Team, 1999. Quantum Geographic Information System QGIS. Open Source Geospatial Foundation Project. URL `http://www.qgis.org/`.

F. Ricca, A. Scozzari, and B. Simeone, 2013. Political districting: from classical models to recent approaches. *Annals of Operations Research*, 204(1):271–299.

W. H. Riker and P. C. Ordeshook, 1968. A theory of the calculus of voting. *American Political Science Review*, 62(01):25–42.

S. J. Rosenstone and R. E. Wolfinger, 1978. The effect of registration laws on voter turnout. *American Political Science Review*, 72(01):22–45.

S. J. H. Rosenstone, J. M. J. Rosenstone, and J. M. Hansen, 1993. *Mobilization, participation, and democracy in America*. Macmillan Publishing Company,.

A. M. Rutchick, 2010. Deus ex machina: The influence of polling place on voting behavior. *Political Psychology*, 31(2):209–225.

M. M. Singer, 2014. The 2013 congressional elections in Argentina. *Electoral Studies*, 35:370–373.

D. M. Spencer and Z. S. Markovits, 2010. Long lines at polling stations? Observations from an election day field study. *Election Law Journal*, 9(1):3–17.

C. Stewart III, 2012. Waiting to vote in 2012. *Journal of Law & Politics*, 28:439.

C. Stewart III and S. Ansolabehere, 2013. Waiting in line to vote. Working paper 114, Caltech/MIT Voting Technology Project.

A. Tasnádi, 2011. The political districting problem: A survey. *Society and Economy*, 33(3):543–554.

S. Verba and N. H. Nie, 1987. *Participation in America: Political democracy and social equality*. University of Chicago Press.

X. J. Wang, M. Yang, and M. J. Fry, 2014. Efficiency and equity tradeoffs in voting machine allocation problems. *Journal of the Operational Research Society*, 66(8):1363–1369.

C. Wheeler, J. Berger, and M. Meredith, 2008. Can where people vote influence how they vote? In A. Lee and D. Soman, editors, *NA - Advances in Consumer Research*, volume 35, pages 173–175. Association for Consumer Research, Duluth, MN.

M. Yang, M. Fry, and W. Kelton, 2009. Are all voting queues created equal? In *Proceedings of WSC '09 Winter Simulation Conference*, pages 3140–3149.

M. Yang, T. T. Allen, M. J. Fry, and W. D. Kelton, 2013. The call for equity: simulation optimization models to minimize the range of waiting times. *IIE Transactions*, 45(7):781–795.

M. Yang, M. J. Fry, W. D. Kelton, and T. T. Allen, 2014. Improving voting systems through service-operations management. *Production and Operations Management*, 23(7):1083–1097.

M. Yang, X. J. Wang, and N. Xu, 2015. A robust voting machine allocation model to reduce extreme waiting. *Omega*, 57:230–237.