How long lines affect minority precincts

Stephen Pettigrew
pettigrew@fas.harvard.edu
March 3, 2015

Abstract

In recent elections long lines to vote seem to have become the new normal in some areas. Anecdotal evidence suggests that long waiting times disproportionately affect minority neighborhoods. In this paper, I evaluate the size of the racial gap in line length and show that it persists even when accounting for the factors which are out of the control of local election officials. I find that mostly minority neighborhoods experience average wait times which are about eight minutes longer than predominantly white neighborhoods. I use a novel dataset to show that this relationship is not a result of differential arrival rates based on race. In fact, I find that arrival rates should be increasing wait times in white neighborhoods. I then provide initial evidence which suggests that waiting in a long line diminishes a voter's probability of participating in future elections. These results highlight the importance of election administration in questions of racial politics, participation, and electoral outcomes.

1 Introduction

In the November 2012 general election, one in every ten voters waited in line for more than a half hour to cast a ballot. About 3.5 million voters waited in excess of an hour, with some standing in line for as many as three or four hours. Long lines at the polls became such a hot topic in the media that President Obama acknowledged in his victory speech that the issue was one that needs to be fixed. Despite the growing media attention to the problem of length lines at precincts, very little political science work has investigated the consequences of long waiting times.

The goal of this paper is to estimate the size of racial differences in precinct lines, and to explore one downstream effect of long lines. I show that white precincts have average waits that are several minutes shorter than minority precincts, even when accounting for many of the factors which may confound the relationship. Next, I show that an unsatisfactory voting experience in the form of a lengthy wait decreases an individual's probability of voting in subsequent elections. Taken together, these results expose long Election Day lines as a consequential problem that is more likely to impact minority voters.

I begin the paper with an outline of the major factors which contribute to long lines on Election Day. I categorize these factors into those which a local election administrator can control and those which they cannot. I then hypothesize that waiting in a long line to vote will discourage voters from participating in future elections.

In my analysis I demonstrate that there are tremendous differences between racial groups in the average waiting times in three recent elections. I use regression to show that the racial effects are robust to potential confounding by the factors out of the control of election officials. In particular, I find that a voter in a nearly all white neighborhood will wait in line that is approximately 8 minutes shorter than a voter in a predominantly minority neighborhood. The magnitude of this finding is striking, given that the national average wait time in 2012 was 13 minutes.

I then use two data sources, including a novel dataset from Maryland, to demonstrate that racial differences in lines cannot be explained by different arrival rates at precincts. In fact, the conclusion to be drawn from this evidence is that lines in white precincts ought to be *longer* than those in minority precincts. This contradiction suggests that election administrators are over-compensating for the possibility of long lines in white neighborhoods, thereby creating a racial gap in the opposite direction.

I conclude with some initial evidence that long wait times may impact the composition of the electorate in subsequent elections. Using data from the city of Boston, I show that individuals who voted at a precinct with a long line in the 2012 presidential election were less likely to turn out in several elections during the following year.

2 What explains why lines develop at precincts?

In one of the earliest texts about queueing theory, Thomas Saaty (1961, p.3) defines a queue as a system of "arriving items that wait to be served at the facility which provides the service they seek." In the context of election administration, we can think of the precinct location as the facility to which voters arrive to cast their ballots. The precinct system generally has two (or sometimes three) stages that a voter must go through in order to cast a ballot. First, voters check in and have their registration record verified by a precinct worker.[1] In some states where the law requires, photo identification cards must be presented at this stage in the voting process. After checking in, the voter then fills out their ballot. After that, some states have a check-out stage wherein the voter must scan their paper ballot into a machine or go through some other form of post-voting check.[2]

[1] In states which allow Election Day registration, new voters go through the added stage of registering to vote.

[2] In Massachusetts, for example, voters must have their name re-verified in the poll book before leaving the precinct.

A queues can form before any of the stages in this system, and a bottleneck at any stage will make it more likely that queues will form at prior stages. There are several factors which can contribute to the likelihood of bottlenecking at a voting precinct, and thus are predictive of how long voters wait in line to vote. In this section I divide several of these factors into those that are manipulable by a election administrator and those that are beyond their control.[3]

2.1 Factors which election administrators can control

The principal way in which election administrators can minimize lines is by efficiently allocating resources to the precincts in their jurisdiction. There are three major sets of resources that an election official must allocate: poll workers, check-in books or machines, and vote casting stations or machines. Election officials are limited in the number of resources that they have available for use on Election day. When there are a finite number of people able to work at the polls on Election Day, or when a county cannot afford to invest in additional voting machines, election officials must decide how to optimally allocate these resources to their hundreds, or even thousands, of precincts.

A major reason that resource allocation is so critical is that that, according to the queueing theory literature, a system with enough capacity to keep up with new arrivals will experience backlogs when one service station is removed from the system. Even worse, queues have strong non-linear properties. When the removal of a voting machine or check-in station generates a queue, that line will tend to grow exponentially rather than linearly and will not disappear until the arrival rate of new voters into the queue diminishes(Gross and Harris, 1985). For a hypothetical precinct with a slight under-allocation of resources and at which voters arrive at a constant rate all day, the amount of time that a voter waits in line

[3]For most of the country elections are administered at the county level, with the exception of a few New England states where elections administration decisions are made at the town level.

will be an exponential function of the time of their arrival at the precinct. The line will not decrease in size until new arrivals are prevented from entering the system.

Another factor which election officials can control is the level of training provided to the people working at the polls. If a precinct does not have at least one worker sufficiently trained in vote machine troubleshooting, a broken down machine could create a situation no different from a scenario where the machine was not allocated to the precinct in the first place. Poll workers must also deal with procedural problems which arise throughout the day, and misinterpretation or misapplication of election law and rules may create problems which slow down the voting process for everybody at a precinct location.

The other main factor which election administrators might be able to control is the vote technology used in their jurisdiction.[4] Although other technologies have been used in the past, nearly every jurisdiction used one of two ballot casting technologies in 2012. About 59% of jurisdictions used a paper ballot optical scan system, where the voter fills out a paper ballot by hand and then has the vote counted by an optical scanning machine. Roughly 22% of jurisdictions used a Direct Recording Electronic voting system (DRE), where a voter fills out their ballot electronically at a computer kiosk. The remaining jurisdictions use either hand counted paper ballots, punch card systems, or a mixture of systems.

Prior research (Edelstein, 2006) has found that jurisdictions which use DREs tend to have longer lines than than jurisdictions which use optical scanning machines. In the November 2012 election, the average wait in jurisdictions with a DRE was 17 minutes and 34 seconds (se: $<$ 1 second), while in optical scan jurisdictions it was 11 minutes and 23 seconds (se: $<$ 1 second). This difference is partly a result of the flexibility afforded by optical scanning machines. Because the scanning process only takes a few seconds, the limiting factor in an optical scan system is how many stations are available for voters to

[4] This is not universally controlled by local election administrators. This decision is made at the state level in most states.

privately fill out their ballots. These stations only require a table and privacy dividers, so it is not especially costly to have a surplus of voting stations. DRE systems, on the other hand, are significantly less flexible. The machines are more expensive, so it is not cost effective to have more machines available at a precinct than is necessary. DREs are also more sophisticated technologically than paper ballot, so there is a much higher risk of machine failure. If there is not a poll worker at the precinct with the expertise to fix the problem, a bottleneck could occur.

2.2 Factor which election administrators cannot control

In contrast to the factors listed above, there are several other things which are out of the control of election officials which may systematically affect wait times. One of these is the length of the ballot and how long it takes voters to fill it out. Ballots that take more than five minutes to fill out require, on average, double the number of voting machines in each precinct to keep waiting times to a reasonable level throughout the day (Edelstein and Edelstein, 2010). Within a state, the length of the ballot is roughly uniform, since most referendums are included on the ballot statewide and these items tend to be the longest on the ballot in terms of word count. Allen and Bernshteyn (2006) use simulations to show that increasing a ballot from an average of 3 minutes to fill out to 3.5 minutes can add half an hour to the average wait time at a precinct.

State laws regulating the administration of elections also constrain election administrators. One such regulation, voter identification card requirements, have been found to increase the average length of lines (Kimball, 2013). Voter ID laws slow down the speed with which poll workers can check in voters, particularly when problems arise regarding a voter not having the proper identification.

Another factor which influences wait times is the number of registered voters assigned to a precinct. Given two precincts with equal resource allocation, the one with a higher number of registered voters is more likely to form long lines. Most election officials are not able to evenly distribute voters across precincts by regularly redrawing precinct boundaries. However, the number of registered voters at a precinct is a known quantity which can be accounted for when election officials decide how to allocate machines or poll workers.

Closely related to the number of registrants is voter turnout. Unlike the number of registrants, the turnout is not knowable prior to an election. Election officials must use statistical models, or more commonly, heuristics to predict the number of voters who will seek service at a precinct. As I will show later, under-allocation of resources is the most likely explanation for the development of long lines, and systematic errors in the prediction of turnout is a likely culprit for this miscalculation.

The final factor is the arrival rate of voters. If voters were to arrive at a constant rate through the day, then allocation of resources is straightfoward. If arrival rates fluctuate through the day, then it is necessary to allocate resources based on the peak arrival rate during the day, rather than the average (Allen and Bernshteyn, 2006). Arrival rates are not uniform, however. Edelstein (2006) emphasizes that although a precinct may have voters arrive in a Poisson process[5] with an average arrival rate of 50 voters per hour, it is not unreasonable to have 125 show up after work at 5:00pm, while only 25 show up at 2:00pm. As I will show in a later in this paper, there is good empirical evidence which shows that the biggest spikes in arrivals occur very early in the morning, before many people go to work and after 5:00pm, when many people are done with their workday. It is not uncommon for a precinct to have arrival rates during these busy periods which are an order of magnitude bigger than the rate at other points during the day. Although the arrival rates are not known before an election, if election officials naively assume a constant arrival rate through a day based on expected turnout, they are almost guaranteed to under-allocate resources.

[5] A common assumption in these models.

3 Long lines and turnout

Long lines could potentially impact who votes in the current election as well as future elections. Potential voters may anticipate a long line and never show up as a result. These individuals are called "balkers" in the queueing literature (Gross and Harris, 1985). Others may show up to vote, but leave the line before voting–"renegers" in the literature. To date, there is very little work about balkers and renegers, and virtually none about the impacts on future elections.

The Final Report for the 2008 SPAE indicates that 11% of non-voters in 2008 failed to vote because of lengthy waits at the polls (Alvarez et al., 2009). This figure rose to 14.5% in 2012, based on that year's SPAE (Stewart, 2013). The survey does not provide any indication of which of these individuals balked and which reneged. It is also difficult to assume that, absent long lines, every one of them would have voted, although given that there were about 16 million such non-voters in 2012, it is not unreasonable to assume that at least some of them would have participated. Another study suggests that as many as 49,000 people in Central Florida alone balked at voting in 2012 because of long lines (Allen, 2013).

The best data available about the number of people who reneged was collected by Spencer and Markovits (2010). During the 2008 presidential primary election in California, the researched conducted a field study at 30 precincts in three San Francisco-area counties. Observers at each precinct recorded arrival rates, waiting times, and vote casting times for everybody that showed up at the polling station. To date, this is the only data which measures the rate at which people get into line but leave before casting a ballot. They find that about two percent of people who showed up to vote eventually renege. They also show that this rate increases as the line grows. The probability of reneging is four times longer when there are five people in line compared to when there is no line. If the people who balk or reneg are systematically different in their political preferences from those who vote, then the the distribution of votes will be skewed as a result.

The other way in which lines can affect the composition of the electorate is by diminishing the probability that an individual votes in future elections. If a voter waits in line for six hours to cast their ballot, as some in Florida did in 2008[6], then they will be turned off from the electoral process and might be less likely to vote in 2010 or 2012. In terms of the basic turnout calculus model proposed by Riker and Ordeshook (1968), waiting in a long line represents an additional cost to voting, which the voter may factor into their turnout decision in the next election. Later in the paper I will test the hypothesis that waiting in a long line in election i will make a voter less likely to participate in future elections $i + j$.

From a normative perspective, such alterations to the electorate are particularly problematic if they are more likely to occur for specific subsets of voters. If it is the case, for example, that every voter in the country has waits the exact same amount of time to vote, say two hours, then each person bears the same nominal cost of voting.[7] If however half of voters have to wait two hours to vote and the other half wait only 10 minutes, then some voters are being asked to bear a higher cost for the same service than other voters. It becomes even more consequential if these two groups of voters are different in terms of the political preferences, since this could cause a measurable change in the election outcome.[8] Such an effect is unlikely to be large enough to change the result of a presidential election. In local, state, or even congressional races, however, one precinct with an unusually long line may alter the outcome of an election decided by just a few dozen or hundred votes.

[6] "Photo of my 9 hour wait to vote in Florida." www.dailykos.com/story/2012/11/03/1155089/-Photo-of-My-9-Hour-Wait-to-Vote-in-Florida. Accessed Oct. 13, 2014.

[7] Certainly it is likely that a two hour wait will have a different relative cost for each voter, depending on other factors like whether they receive an annual salary or hourly wages at their job.

[8] A controversial report by the House Judiciary Committee Democratic Staff (2005) argued that this exact thing happened in Ohio in 2004. The report suggested that the misallocation of resources had turned away likely-Kerry voters from the polls and cost him enough votes to lose the entire state. A followup study (Highton, 2006) shows that the relationship between machine allocation and turnout was not strong enough to have swung the election result. but that it may have cost Kerry a few thousand votes.

4 How strongly is raced tied to line length?

Before testing the turnout hypothesis, I consider which factors have the strongest predictive influence on waiting times. Although I do not have a wealth of data regarding resource allocation to individual precincts, I am able to account for many of the other factors. After accounting for those factors, any remaining variance in average wait times must be attributable to either random error or to factors that I was not able to account for. I will show that this unexplained variance–largely attributable to the allocation of resources–is strongly correlated with race and other demographic characteristics of neighborhoods.

The ideal setup of this analysis would be to have the exact waiting times for every voter, in addition to demographic traits, neighborhood characteristics, and other precinct-level covariates. Although it is possible to collect data for some of these variables, the key one–individual wait times–is not available for most voters and most precincts. To get around this problem, I rely in this section on survey data from 2006, 2008, and 2012, available in the Cooperative Congressional Election Study (CCES) cumulative file (Ansolabehere and Pettigrew, 2014). I removed respondents who said that they did not vote in that year's November general election and those whose official vote history file indicates that they did not vote.

Wait time for each respondent in the survey is measured by asking, "Approximately how long did you wait in line to vote?" Respondents are then presented with five possible responses: "not at all", "less than 10 minutes", "10 to 30 minutes", "31 minutes to an hour", and "more than an hour." Those who waited longer than an hour were prompted to specify their wait time in an open-ended followup question. The responses from this question were recoded to be on a time scale, in minutes. Respondents who fall into the first four categories were recoded to be the midpoint of their response category (i.e. 0, 5, 20, and 45 minutes).[9]

[9] Alternatively, I could have used an ordered logit or probit model to model the response categories directly. Or instead of imputing mid-points, I could have drawn a value from a uniform or some other distribution

Respondents who waited longer than one hour had their wait time coded as their response to the open ended question.[10] Because the dependent variable is mostly continuous in nature, I estimate the models below using OLS.[11] Any respondents who failed to answer the question were dropped from the analysis, leaving 65,775 observations to analyze.

Figure 1: Average waiting time by race and election year

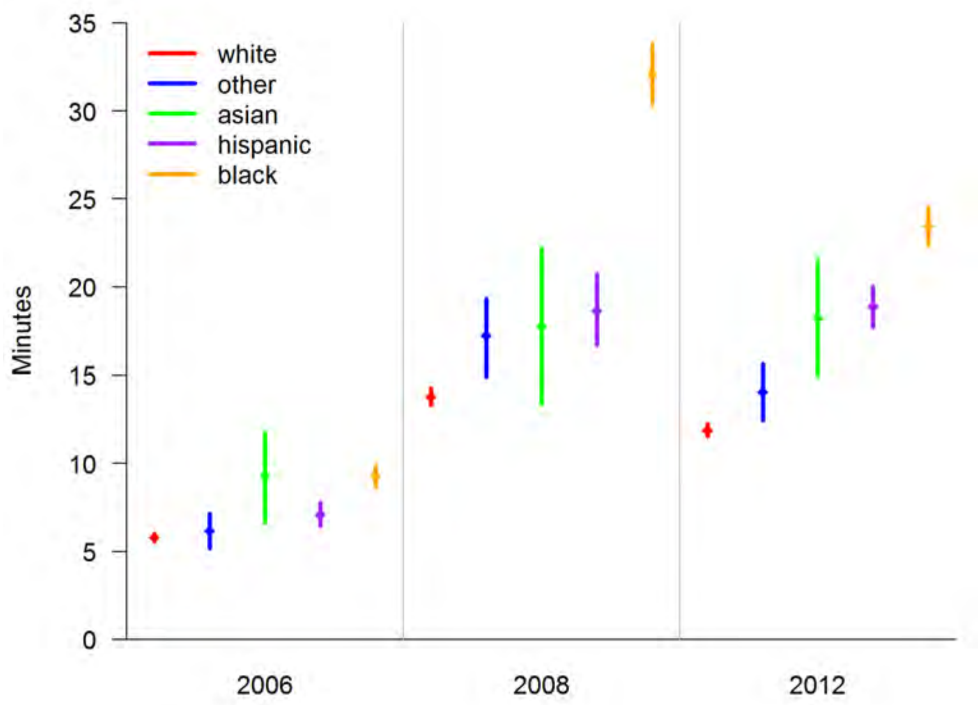


Figure 1 shows the unconditional mean and 95% confidence intervals of the wait times experienced by different racial groups in the 2006, 2008, and 2012 elections. The first thing to notice is that the shortest average wait in each election was experienced by white voters. In each year, there was a statistically significant difference in the average wait for white

for that observation's category. Neither of these options impacts the substantive conclusions of the analysis.

[10]In the handful of instances in which these respondents did not answer the open ended question, their wait time was coded to be the mean of the wait times for all other 1+ hour respondents in their survey.

[11]Because the dependent variable is truncated at zero, a more complicated likelihood function, such as one based on the half-normal or log-normal functional form may perhaps be more appropriate. I chose OLS because of its ease of interpretation and because it yields results that are substantively similar to the more complicated models.

voters versus black voters and white voters versus Hispanic voters. The difference between white and black is most pronounced in 2008, when white voters waited an average of 13 minutes and 47 seconds (se: 28 seconds) and black voters waited 32 minutes and 7 seconds (1 min, 41 s). In the same year, the average wait for Hispanic voters was 18 minutes and 38 seconds (se: 2 min, 4 s).

The other major trend to notice from Figure 1 is that across all racial groups, the average wait in the 2006 midterm election was lower than the average wait during either of the two presidential elections. The likely explanation for this difference is that turnout in midterm elections is typically much lower than in presidential elections. The number of voting machines in most counties or towns will tend to be fixed, meaning that the number of voters per machine will be much higher in presidential years, thereby increasing the probability of longer lines. Given the magnitude of this intercept shift, I use year fixed effects in all models which pool data across years.

4.1 Accounting for potential confounding variables

While the racial differences shown in Figure 1 are suggestive, just looking at the bivariate relationship does not rule out the possibility of other things confounding the relationship. To better assess the robustness of this correlation, I will account for as many of the factors described in section 2 and then estimate the race gap. If race and line length remain strongly correlated after controlling for these observable factors, then we can conclude one of two things. It could be that race is having a direct effect on line length in a manner that we may characterize as overt racial discrimination. Alternatively, and perhaps more plausibly, we could conclude that an unobserved variable is confounding the relationship between race and line length. I show below that the racial effect is robust to controlling for these other variables, and I take this as evidence of systematic misallocation of resources, since that is the biggest factor for which I am not able to account.

In the OLS models I present below, I use several strategies to account for the factors which contribute to longer lines. State fixed effects eliminate variance in wait times that results from state-specific factors like voter identification laws and ballot length. In a separate specification, I use county fixed effects[12] instead of state ones. This specification carries the benefits of the state fixed effects, but has the added benefit of controlling for vote technology.[13] Also, because resource allocation decisions are made at the electoral jurisdiction level, county fixed effects will control for the raw number of poll workers and voting machines available. They also account for the level of training received by poll workers because training practices will be mostly standard within an electoral jurisdiction.

One variable for which I am not able to directly account is turnout. The smallest geographic unit at which turnout figures are readily available is the county, so a model with county fixed effects would perform no worse than a model with state fixed effects and county-specific turnout numbers. Ideally, I would include a control variable for turnout which is measured at the sub-county level. Precinct-level data is available in many states, although it is not possible to match respondents in the CCES to specific precincts. The smallest geographic unit at which respondents are identified is zip codes. Aggregating the precinct data to zip codes, however, is nearly impossible.[14]

Neglecting turnout may seem to create problems in drawing conclusions about race, particularly when studying the 2008 and 2012 elections in which turnout among African-

[12] I use county fixed effects as a proxy here for electoral jurisdiction fixed effects. The only difference is that jurisdiction fixed effects would be different in a few northeastern states where election administration is done at the town level. Because I lack data to be able to definitively pinpoint CCES respondents to towns, I use county fixed effects as instead. Dropping these states from the analysis does not have an impact on the results.

[13] Only 87 of the 2856 counties in the sample use a mixture of technologies within their jurisdiction

[14] Creating zip code-level turnout measures requires precinct shapefiles or precinct addresses. Shapefiles are available for some states on the Harvard Election Data Archive (Ansolabehere, Palmer, and Lee, 2014), but the shapefiles cannot be matched to election returns, except in a couple states. Precinct addresses are available through the Voting Information Project (Voting Information Project, 2014), but that data does not include a common precinct identifier which would allow for geolocated addresses to be matched to the Harvard Election Data Archive.

Americans spiked tremendously. There is good reason, however, to think that the bias created by this omitted variable should actually be working against finding an effect of race. If we assume that county election officials anticipated increased turnout among African-Americans, then they should have allocated more machines to these precincts than they usually would. This should would tend to reduce the wait times for predominantly black precincts, and it would perhaps increase lines in mostly white precincts if machines were being redirected away from them. This means if any election officials strategically reallocated resources toward black precincts, my estimates of the racial gap in wait times are smaller than that they would have been otherwise.

Table 1 shows the regression results where respondent wait time is the outcome variable. Models 1 and 2 include state fixed effects, while models 3 and 4 use county fixed effects.[15] In addition to election year dummy variables, I control for neighborhood demographic variables and individual respondent traits. The neighborhood demographics come from the 2008 and 2012 ACS surveys and are measured at the zip code level, because this was the most fine grained geographic unit for which there was data in the CCES. The most important of these neighborhood demographics is the percentage of white people in the zip code. Such a specification returns the most easily interpreted results, compared to using the percent black, hispanic, asian, and other races all in the same model. Other zip code level variables I have included are the percentage of homeowners and median income, which proxy for the economic affluence of the neighborhood. I also control for population density and percentage of senior citizens. I also include respondent-level covariates for race, age, income bracket, party identification, and whether or not the respondent voted early or on Election Day.

In stark contrast to Figure 1, model 4 shows that there is not a statistically significant

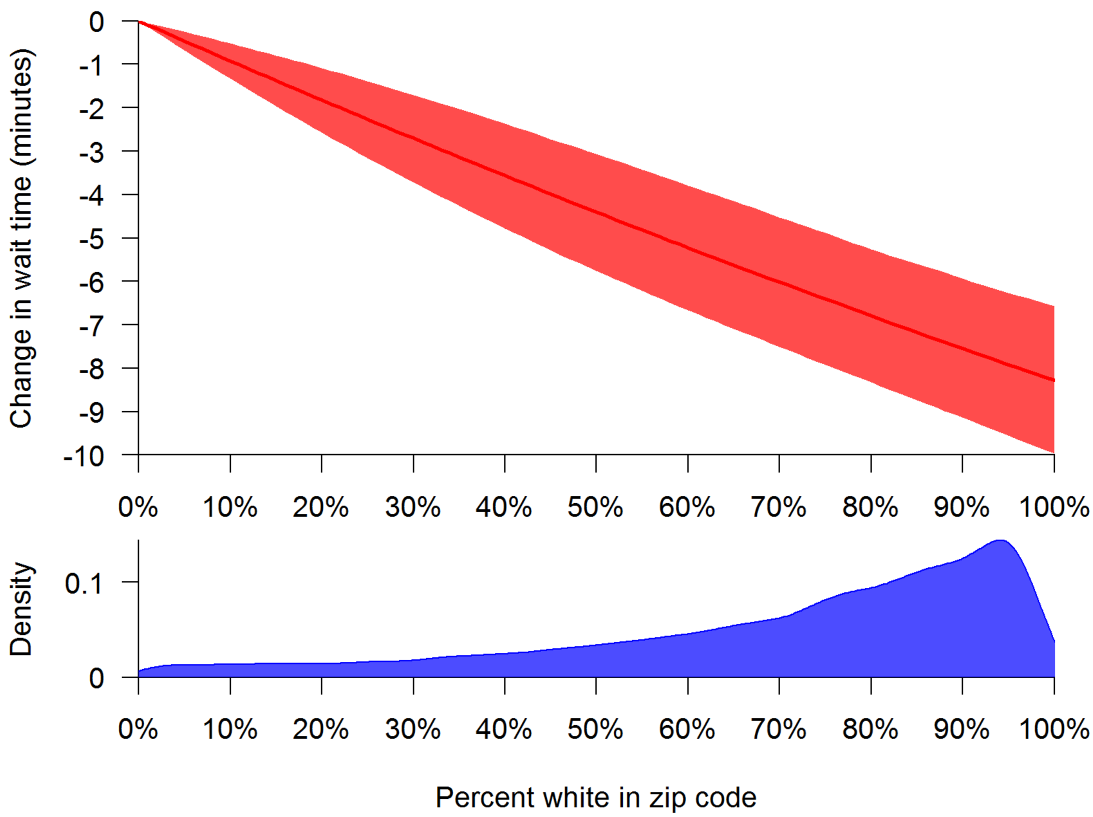
[15]The county fixed effects models do not include an intercept term because the *lfe* package in R, which provides the fastest estimates of models with a large number of fixed effects, saturates the fixed effects and does not calculate a model intercept.

Table 1: Predictors of waiting times

	(1)	(2)	(3)	(4)
WhitePct	-19.435^{***}	-25.713^{***}	-14.613^{***}	-9.323^{***}
	(1.794)	(2.063)	(2.048)	(2.340)
WhitePct^2	1.237	8.674^{***}	5.534^{**}	1.068
	(1.560)	(1.801)	(1.838)	(2.117)
HomeownerPct		-3.997^{***}		1.074
		(1.021)		(1.155)
Population density		0.0001^{***}		0.0001^{***}
		(0.00002)		(0.00003)
Pct. over 65		-0.986		-1.334
		(2.147)		(2.512)
Median income		2.936^{***}		3.034^{***}
		(0.408)		(0.406)
Race - Black		1.244^{**}		0.183
		(0.433)		(0.431)
Race - Hispanic		1.485		2.181
		(1.156)		(1.128)
Race - Asian		0.985		1.030
		(0.536)		(0.532)
Race - Other		-0.027		-0.012
		(0.039)		(0.039)
Age		0.125^{***}		0.040^{***}
		(0.007)		(0.009)
Age^2		-0.0004		-0.001
		(0.0004)		(0.0004)
Income - 50k-100k		1.144^{***}		0.794^{**}
		(0.248)		(0.247)
Income - 100k+		1.593^{***}		1.216^{***}
		(0.297)		(0.295)
Party - I		0.124		0.187
		(0.356)		(0.353)
Party - R		-0.890^{***}		-0.866^{***}
		(0.221)		(0.221)
Early voter		6.277^{***}		7.450^{***}
		(0.318)		(0.321)
2008	9.195^{***}	7.957^{***}	9.754^{***}	8.255^{***}
	(0.250)	(0.272)	(0.246)	(0.270)
2012	6.905^{***}	6.205^{***}	7.248^{***}	6.299^{***}
	(0.221)	(0.264)	(0.218)	(0.262)
Intercept	16.558^{***}	15.575^{***}		
	(0.894)	(1.454)		
Fixed effects	State	State	County	County
Observations	65,554	55,471	65,554	55,471
R^2	0.115	0.135	0.217	0.230
Adjusted R^2	0.114	0.134	0.183	0.191

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

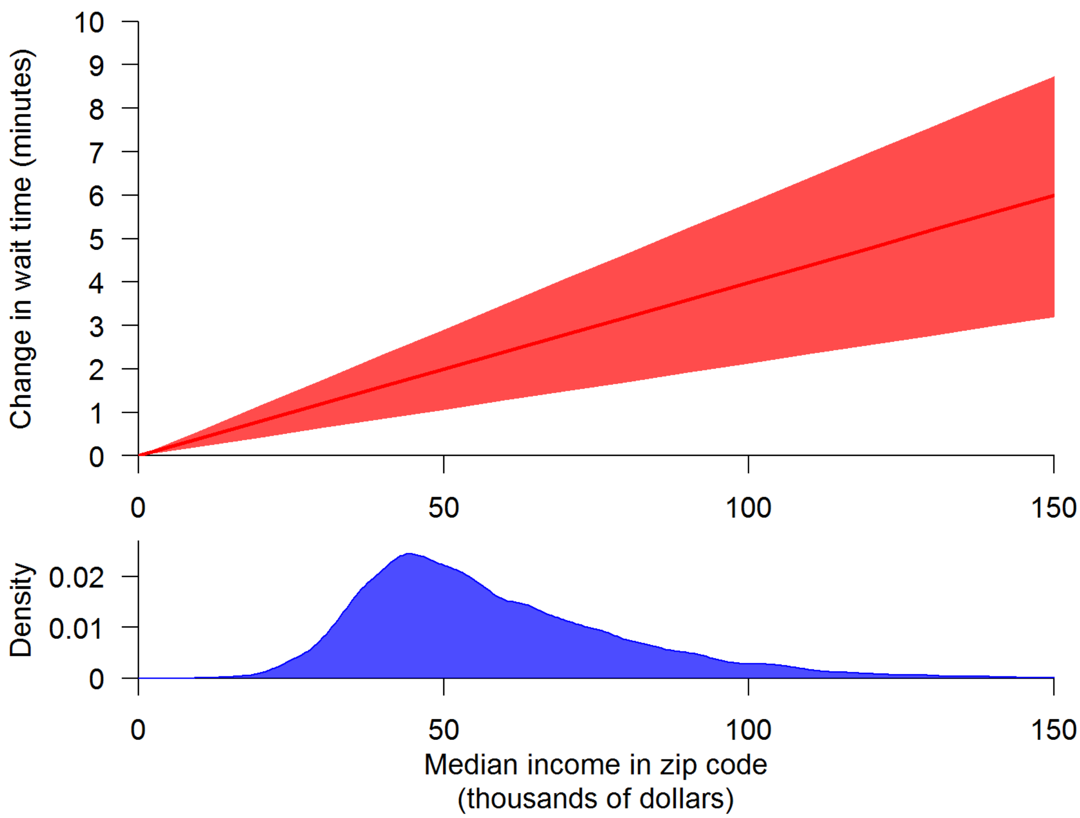
Figure 2: Change in average wait by white percentage in neighborhood



effect of the individual voter's race on how long they waited in line. There is, however, a sizable impact of neighborhood racial composition. Figure 2 shows the marginal effect of different levels of racial composition on expected wait time (from model 4), as well as a density plot showing the number of people who live in each neighborhood. As the figure shows, the average wait times in a completely white neighborhood are about five minutes shorter than a neighborhood that is only 50% white and about eight minutes shorter than a neighborhood of almost all minorities. These may seem like only a short amount of time, but consider that the average wait nation-wide in 2012 was only about 13 minutes and 30 seconds (se: < 1 second).

The fact that the model estimates a large positive marginal effect for neighborhood racial demographics, but a null effect for individual race also is important. Finding an effect for individual race, but not neighborhood demographics would have suggested that individual black or Hispanic voters were somehow being forced to wait longer than white voters at the same precinct. This story makes little sense, given that long lines affect precincts as a whole, rather than individuals. What the model implies is that a white voter who lives in a predominantly black neighborhood is likely to wait longer than a black voter in a predominantly white neighborhood. This finding lends credibility to the idea that lines result from inefficient resource allocation to precincts, rather than some sort of overt discrimination by poll workers against individual voters.

Figure 3: Change in average wait by median income in neighborhood



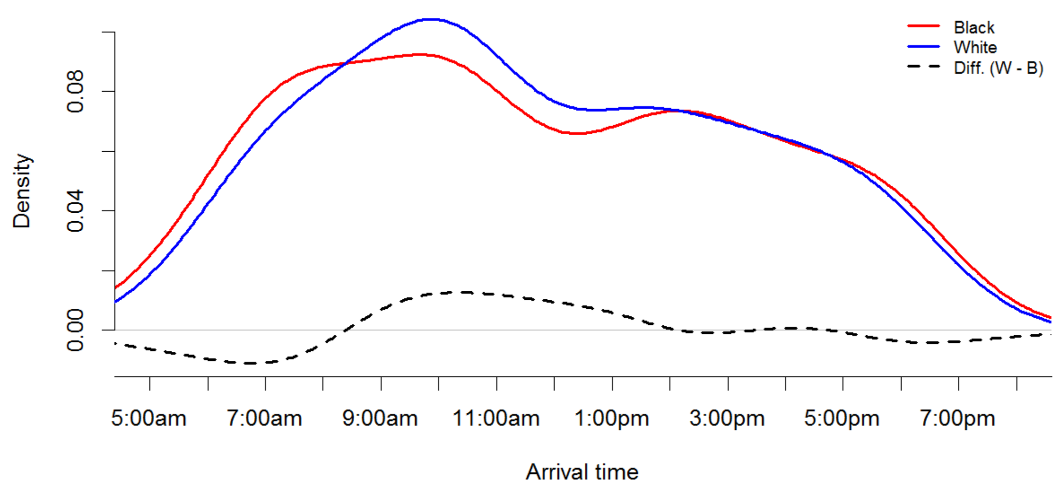
Another important takeaway from Table 1 relates to the income variables. Both the median income of the respondent's zip code and the respondent's own family income have positive and statistically significant effects on wait times. Figure 3 shows the marginal effect of neighborhood median income on wait. Neighborhoods with a median income of $50,000 (roughly the 50^{th} percentile) wait about two minutes less than neighborhoods with a median of $100,000 ($95^{th}$ percentile). The coefficients for individual income also point in the same direction. Individuals with a family income between $50,000 and $100,000 experience lines that are about thirty seconds longer than people in the $100,000+ bracket.

These findings are likely a consequence of turnout effects. Higher income individuals have a higher propensity to vote (Campbell et al., 1960), resulting in a higher probability of long lines at precincts. The finding, however, contradicts previous work which found that high income neighborhoods were allocated more machines and had more efficient poll workers. Spencer and Markovits (2010) find that a $10,000 increase in median household income translated into eleven fewer seconds in the check-in process for each voter. Such a difference could have a multiplier effect which would result in longer waits of 5 to 45 minutes in low income areas.

5 Arrival rates of voters

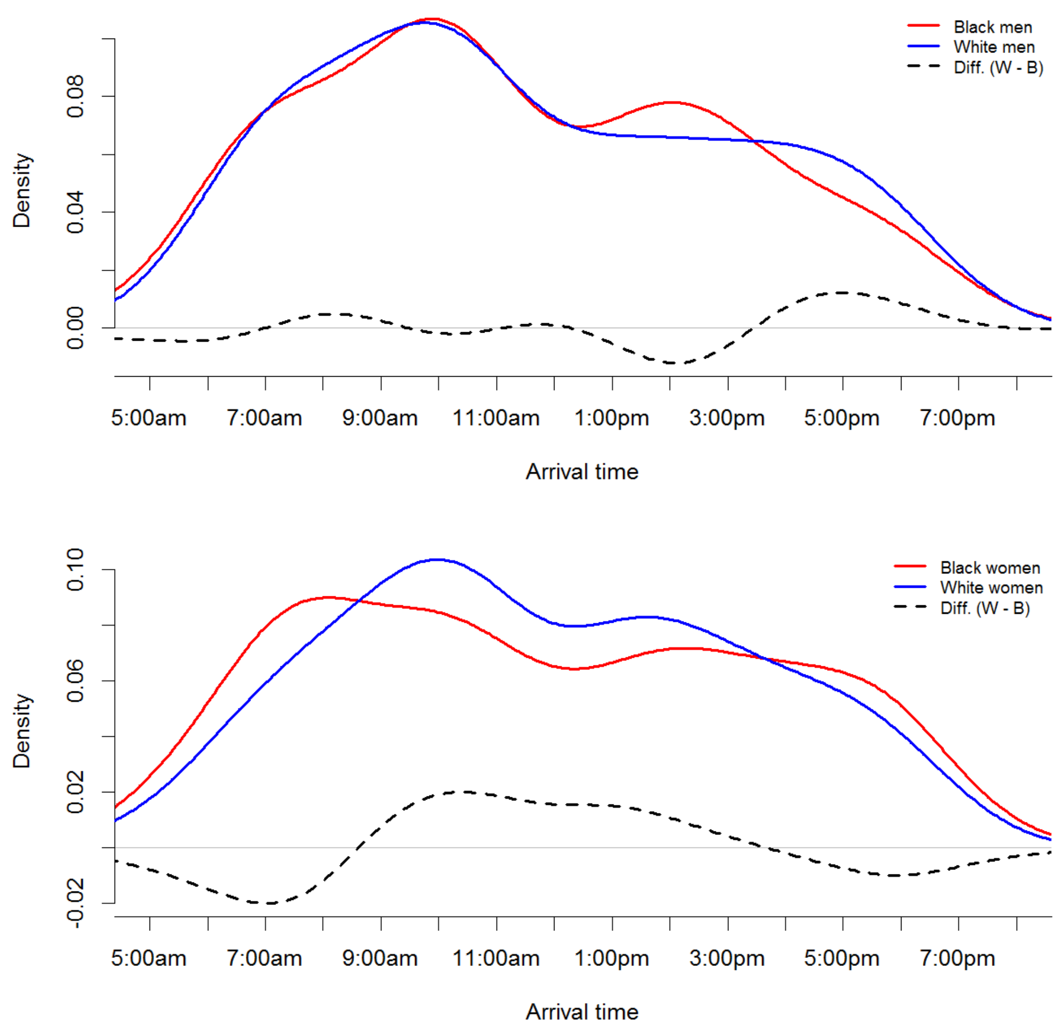
Aside from precinct-level resource allocation data, the other major factor which I do not account for in Table 1 is the arrival rate of voters. The CCES does not include an item asking when voters arrived at their precincts to vote, so I am not able to incorporate temporal trends in lines, like the that the longest lines tend to occur when precincts open in the morning. Arrival rates are important because they can have a big effect on lines. Consider two precincts which have an equal number of voting machines, poll workers, and people who show up to vote on Election Day. Imagine that voters in the first precinct arrive at a constant rate all day, but that 95% of voters in the second arrive after they leave work at 5:00pm. Despite the fact that the ratio of poll workers and voting machines to voters is identical in the two precincts, the voters in the second precinct are significantly more likely to wait in line to vote. If the variability of arrival rates are systematically correlated with other measurable traits, then wait times will also be correlated with those traits when election officials do not account for arrival rates when allocating resources.

Figure 4: Arrival times based on respondent race (SPAE data)



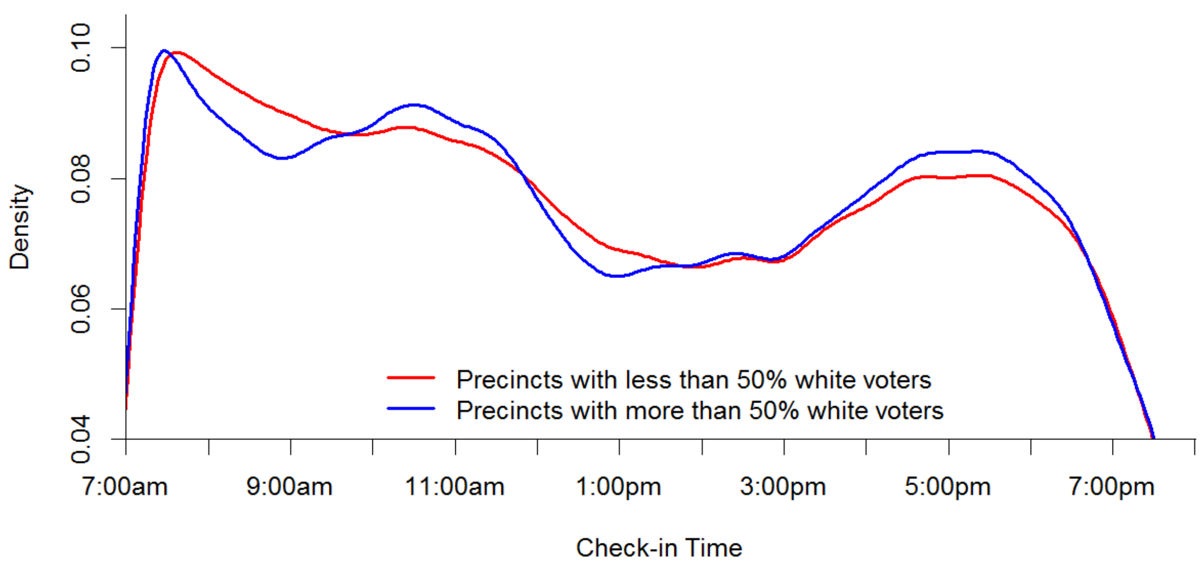
I assess whether arrival rates differ based on race, I use two sources of data which include arrival times. The first is the Survey of the Performance of American Elections from 2009 and 2012 (Stewart, 2010, 2013), which provides the time that 9,496 respondents arrived to vote (in hour-long intervals). The second is the timestamp of the exact second at which each of the over 2 million voters in Maryland checked-in in 2012. Figure 4 shows the distribution of arrival times among white and black respondents in the SPAE. The blue line in the graph displays the density of arrival times among white voters; the red one shows the density among black voters. The biggest thing that stands out is that the the peak arrival rate among white voters is higher than the peak rate among black voters. This evidence implies that, given equal resource distribution among equally sized precincts, a precinct with more white voters should experience lines that are *longer* than a precinct with more black voters. This clearly runs counter to the findings in the previous section and implies that election officials might be overcorrecting for different arrival rates by allocating so many resources to white precincts that it causes their wait times to drop below those of black precincts.

Figure 5: Arrival times based on respondent race and gender (SPAE data)



One possible explanation for this difference in arrivals is that white voters have less constrained schedules during the late morning. Figure 5 provides more insight into this possibility by dividing the sample based on both race and gender. The top graph in the figure shows that the arrival rates among white and black men are very similar in the morning, although in the afternoon white men seem to arrive at a more uniform rate, whereas black men cluster their arrivals around 2:00pm. The bottom graph shows the arrival rates of women and demonstrates that the racial gap in arrival rates at 10:00am are mostly driven by women. Black women have an arrival rate that is nearly constant throughout the day, whereas white women have a mid-morning spike in arrivals.

Figure 6: Check-in times based on precinct demographics



So far, this evidence about arrival times has been based on the race of individual voters. Data from Maryland in 2012 allows me to evaluate how arrival rates vary based on precinct-level demographics. Although I do not know the race of the voter associated with each timestamped check-in time, I matched precincts to census block groups to get a measure of the racial composition of the precincts. Figure 6 compares the check-in times of voters in precincts that are more than half white and those in precincts that are less than half white. The peak of each curve occurs during the first hour in which the polls were open, 7:00am. After that point, the precincts with a larger concentration of minority voters seem to have a check-in rate that is less variable than whiter precincts. The mid-morning and evening bumps in voter check-ins are larger in the more white precincts. Likewise, the troughs around 8:30am and 12:30pm are much bigger in white precincts.

This evidence from Maryland is similiar to that from the SPAE. The important conclusion to be drawn here is that the racial gap evident in waiting times cannot be explained away by differential arrival rates. If it were the case that election officials were neglecting arrival rates and allocating resources based only on the total number of expected voters, then we would see longer lines in white precincts. The evidence in this section suggests, however, that election officials are allocating so many resources to white precincts that they go beyond just accounting for the maximum arrival rate.

6 Effect on turnout

Given that race is a strong predictor of how long an individual waits in line to vote, a logical followup question is whether lines have downstream effects on vote behavior. To answer this, I consider some preliminary evidence which suggests that waiting in a long line today may diminish your probability of voting in future elections. In this section I use the closing times of the 255 precincts in the city of Boston as a proxy measure of how long the lines were at each location. Precincts must remain open until everybody who was in line at the designated closing time has an opportunity to cast a ballot. Later closing times, therefore, signify that there was a longer line at the end of the day. I merge this data with the Boston voter file, which includes the turnout history of every registered voter in each of the five general, primary, and special elections that occurred in Boston between November 2012 and

November 2013.[16]

I use logit models to estimate the impact of precinct closing time on the probability of voting in each of these five elections, conditional on several covariates. The voter file contains very little demographic information about each voter, however it does provide addresses which I geolocated to determine several characteristics of the voter's neighborhood. In addition to controlling for age, I controlled for the percent black, Hispanic, or other race, percent of residents over 65 years old, median age, average household size, and proportion of home owners in each voter's Census block group. I also included a control for the ratio of check-in booths to registered voters in each precinct and whether the individual voted in the November 2010 election.

The top five bars in Figure 7 show the estimated coefficients for the 2012 closing time covariate in each of the five city-wide elections in Boston in 2013.[17] Although certainly not conclusive, the figure provides some initial evidence in support of the hypothesis that long lines diminish a voter's probably of turnout out in future elections. A statistically significant ($p < .05$) negative effect was found in three of the five 2013 elections.[18] The figure also includes the coefficient from a placebo test, where I evaluated the effect of 2012 line length on the probability of turning out in November 2011. The model passes this placebo check because the coefficient is statistically indistinguishable from zero.

Figure 8 provides a sense of how large these effect sizes are in terms of the probability of turning out. The graph shows the effect of the closing time in 2012 on the probability of turning out in the June 25, 2013 special election.[19] Voters in a precinct which closed on time (8:00pm) likely experienced shorter waits than those in precincts which closed laterand

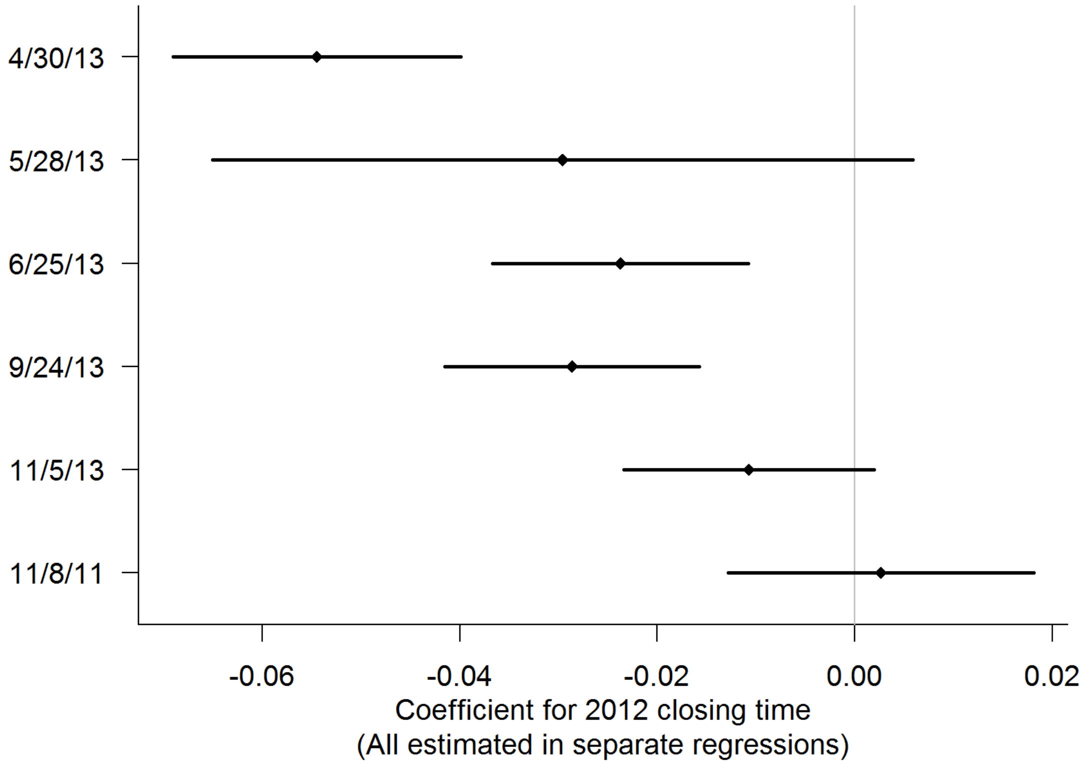
[16] These elections are the April 30, 2013 special state primary, the May 28, 2013 special state election and primary, the June 25, 2013 special state election, the September 24, 2013 preliminary municipal election, and the November 5, 2013 municipal election.

[17] The full results from these models can be found in Table 2 and 3 in the appendix.

[18] The coefficient in a fourth model, November 5, 2013, is statistically significant at the .10 level.

[19] I chose this election as an example because its effect size was in the middle of the range of the five elections considered. The marginal effect sizes presented here are similar to those from the other election.

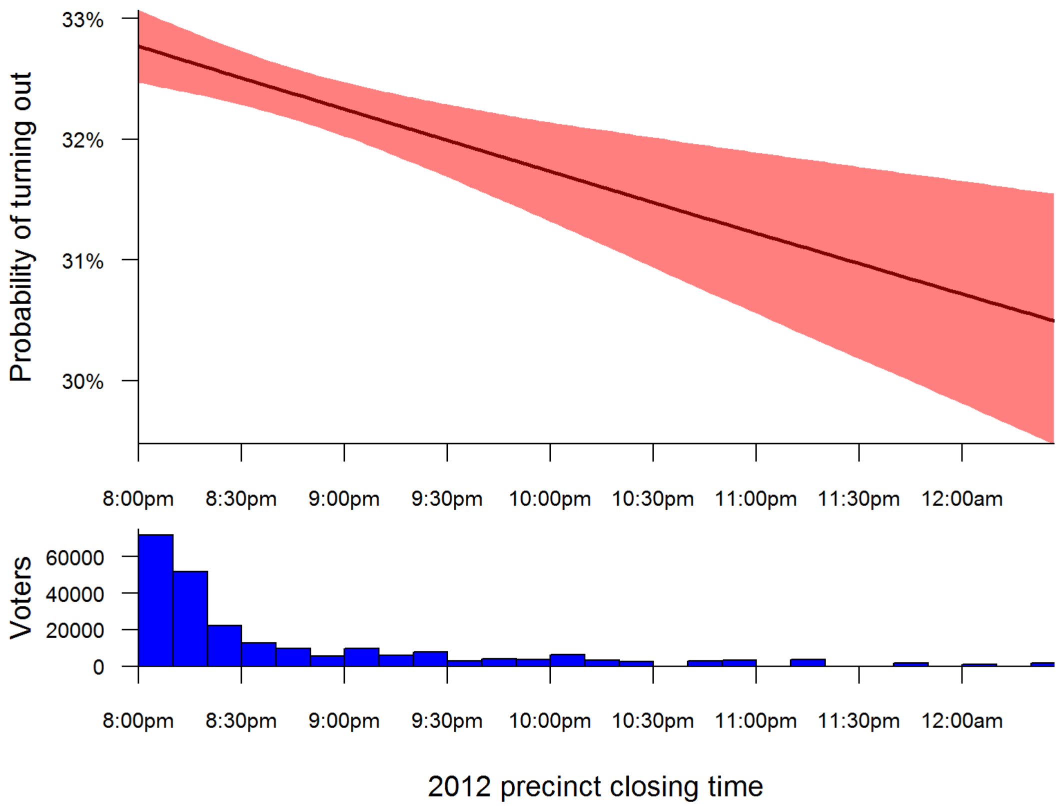
Figure 7: Estimated effect of 2012 precinct closing time on turnout in subsequent elections



were also the most likely to vote in the June 25 special election. Individuals from precincts that closed at 9:00pm were half a percentage point less likely to vote than those in precincts that closed on time at 8:00pm. The 38,000 voters in a precinct that closed later than 9:30 had their probability of turning out in the special election decrease by more than a full percentage point.

The size of this marginal effect may seem trivial, although there are a few reasons to believe that these effect sizes are attenuated. First, closing time of the precinct is nothing more than a proxy measure of individual wait times. The tremendous amount of measurement error in the explanatory variable of interest will pull the size of the coefficients toward zero, making a significant result difficult to identify. Also the off-year elections in 2013 had

Figure 8: Marginal effect of closing time in Boston (June 25 election)



a very low turnout. If a turnout effect exists, it would probably be larger in a midterm or presidential election when voters have a reasonable expectation that lines might be as long as they were in November 2012. The elections considered here provide a tough case for finding an effect, so the statistically significant results suggest that more work should be done on this question.

In future versions of this paper I plan to expand this analysis by looking at several other data sources. First I plan to use the CCES panel study in 2012 and 2014 to look at how 2012 wait times affected turnout at the individual level. The survey will also provide me with other covariates which I can use to explore the possibility of heterogeneous effects. I will also use the Maryland timestamped data as another test of the hypothesis using individual level data. I can use algorithms described in the queueing theory literature to back out approximate wait times from the timestamp data (Larson, 1990). By merging the wait times in 2012 with voter history file in 2014, I can test the hypothesis using the full population of voters in a medium-sized state. I also will replicate the Boston analysis using the 2012 poll closing times from Florida, Maryland, and South Carolina.

7 Conclusion

The findings of this paper have been twofold. First, I find that significant differences in precinct wait times based on race are not explained away by observable factors for which we can account statistically. Controlling for these factors, I find that a voter in an almost entirely white precinct will experience a line that is, on average, eight minutes shorter than a voter in an almost entirely minority neighborhood. This difference is enormous when you consider that that nation-wide average wait to vote in 2012 was 13 minutes and 30 seconds. I also show that African-Americans arrival times at precincts tend to be more diffusely spread out throughout the day, compared to those of whites who tend to arrive in clusters. This difference should create longer lines in white precincts, not minority ones. Because we observe the exact opposite, we have evidence suggesting that the racial gap is, in large part, attributable to differential resource allocation to precincts.

The second contribution takeaway from the paper is related to the downstream consequences of long lines at precincts. I show that long lines at precincts in Boston in 2012 correlated with a diminished probability of voting in subsequent elections. While the marginal effect was only a couple percentage points, the evidence highlights the need for future research on the topic.

Taken together, these two findings have serious implications on the importance that electoral administration has on our democracy. The results suggest that not only do minority neighborhoods bear a larger cost of voting than white neighborhoods, the impact of that cost persists months and possibly years later. The possibility of the turnout effect impacting election margins is not zero, especially given that racial minorities tend to have political preferences that are much different from white voters. This highlights the need for further exploration of this topic and opens a range of possibilities for future research.

References

Allen, Theodore, and Mikhail Bernshteyn. 2006. "Mitigating Voter Waiting Times." *Chance* 19.

Allen, Theodore T. 2013. "Delving into the reasons for long lines can bring solutions.".

Alvarez, Michael R., Stephen Ansolabehere, Adam Berinsky, Gabriel Lenz, Charles Stewart III, and Thad Hall. 2009. *2008 Survey of the Performance of American Elections - Final Report*.

Ansolabehere, Stephen, and Stephen Pettigrew. 2014. "Cumulative CCES Common Content (2006-2012)." http://dx.doi.org/10.7910/DVN/26451.

Ansolabehere, Stephen, Maxwell Palmer, and Amanda Lee. 2014. "Precinct-Level Election Data." http://hdl.handle.net/1902.1/21919.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York: Wiley.

Edelstein, William A. 2006. "New Voting Systems for New York–Long Lines and High Cost." *New Yorkers for Verified Voting* 10.

Edelstein, William A., and Arthur D. Edelstein. 2010. "Queuing and Elections: Long Lines, DREs and Paper Ballots.".

Gross, Donald, and Carl M. Harris. 1985. *Fundamentals of Queueing Theory, Second Edition*. New York: John Wiley & Sons.

Highton, Benjamin. 2006. "Long Lines, Voting Machine Availability and Turnout: The Case of Franklin County, Ohio in the 2004 Presidential Election." *PS: Political Science* .

House Judiciary Committee Democratic Staff. 2005. "Preserving Democracy: What Went Wrong in Ohio." U.S. House of Representatives.

Kimball, David C. 2013. "Why are Voting Lines Longer for Urban Voters?" *Conference presentation at the Southwestern Social Science Association annual conference* .

Larson, Richard C. 1990. "The Queue Inference Engine: Deducing Queue Statistics from Transactional Data." *Management Science* 36.

Riker, William H., and Peter C. Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Political Science Review* 62: 25–42.

Saaty, T.L. 1961. *Elements of queueing theory: with applications*. McGraw-Hill.

Spencer, Douglas M., and Zachary S. Markovits. 2010. "Long Lines at Polling Stations? Observations from an Election Day Field Study." *Election Law Journal* 9.

Stewart, Charles. 2010. “November 2009 Survey of the Performance of American Elections.” http://hdl.handle.net/1902.1/20948.

Stewart, Charles. 2013. “2012 Survey of the Performance of American Elections.” http://hdl.handle.net/1902.1/21624.

Voting Information Project. 2014. “Voter Information Tool.” https://votinginfoproject.org/.

Appendix

Table 2: Impact of 2012 line length on future turnout

	Apr. 30, 2013	May 28, 2013	June 25, 2013
Lateness of closing (hours)	-0.054^{***}	-0.030	-0.024^{***}
	(0.007)	(0.018)	(0.007)
Voted in Nov. 2010	1.629^{***}	1.528^{***}	1.585^{***}
	(0.016)	(0.042)	(0.014)
Unregistered in Nov. 2010	0.420^{***}	0.302^{***}	0.614^{***}
	(0.021)	(0.058)	(0.017)
Voter age	0.030^{***}	0.025^{***}	0.027^{***}
	(0.0003)	(0.001)	(0.0003)
Registrants per check-in booth	-0.0003^{***}	-0.001^{***}	-0.0001^{**}
	(0.00003)	(0.0001)	(0.00003)
Turnout change 04/08	0.583^{***}	6.764^{***}	-0.839^{***}
	(0.220)	(0.517)	(0.195)
Black pct.	-0.585^{***}	1.048^{***}	-0.151^{***}
	(0.030)	(0.064)	(0.027)
Hispanic pct.	-1.267^{***}	-11.619^{***}	0.158^{***}
	(0.055)	(0.205)	(0.048)
Other race pct.	-0.898^{***}	0.236^{*}	-0.399^{***}
	(0.066)	(0.130)	(0.060)
Senior citizen pct.	-1.223^{***}	-5.118^{***}	-0.930^{***}
	(0.154)	(0.386)	(0.139)
Median age	0.020^{***}	0.042^{***}	0.016^{***}
	(0.002)	(0.005)	(0.002)
Household size	0.284^{***}	1.167^{***}	-0.056^{***}
	(0.020)	(0.046)	(0.018)
Renters pct.	-0.148^{***}	1.207^{***}	-0.252^{***}
	(0.051)	(0.111)	(0.046)
Constant	-4.071^{***}	-8.572^{***}	-2.905^{***}
	(0.107)	(0.251)	(0.096)
Observations	232,228	232,228	232,228
Log Likelihood	−111,130.800	−30,644.640	−130,191.900
Akaike Inf. Crit.	222,289.600	61,317.290	260,411.900

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Impact of 2012 line length on future turnout

	Sept. 24, 2013	Nov. 5, 2013	Nov. 8, 2011 (Placebo)
Lateness of closing (hours)	-0.029^{***}	-0.011^{*}	0.003
	(0.007)	(0.006)	(0.008)
Voted in Nov. 2010	1.365^{***}	1.416^{***}	1.872^{***}
	(0.012)	(0.012)	(0.018)
Unregistered in Nov. 2010	0.027^{*}	0.081^{***}	-0.843^{***}
	(0.016)	(0.014)	(0.033)
Voter age	0.027^{***}	0.028^{***}	0.031^{***}
	(0.0003)	(0.0003)	(0.0003)
Registrants per check-in booth	-0.0003^{***}	-0.0003^{***}	-0.0005^{***}
	(0.00003)	(0.00003)	(0.00004)
Turnout change 04/08	0.376^{*}	-0.278	1.059^{***}
	(0.195)	(0.191)	(0.237)
Black pct.	-0.466^{***}	-0.753^{***}	-0.503^{***}
	(0.027)	(0.027)	(0.032)
Hispanic pct.	0.230^{***}	0.513^{***}	-1.045^{***}
	(0.047)	(0.046)	(0.057)
Other race pct.	-0.411^{***}	-0.887^{***}	0.118^{*}
	(0.059)	(0.058)	(0.070)
Senior citizen pct.	-0.634^{***}	-1.222^{***}	-0.216
	(0.139)	(0.138)	(0.165)
Median age	0.027^{***}	0.035^{***}	0.014^{***}
	(0.002)	(0.002)	(0.002)
Household size	0.419^{***}	0.363^{***}	0.303^{***}
	(0.018)	(0.018)	(0.022)
Renters pct.	-0.409^{***}	-0.240^{***}	-0.055
	(0.046)	(0.046)	(0.054)
Constant	-3.550^{***}	-3.190^{***}	-4.511^{***}
	(0.097)	(0.095)	(0.116)
Observations	232,228	232,228	232,228
Log Likelihood	−132,240.200	−133,131.800	−97,744.990
Akaike Inf. Crit.	264,508.400	266,291.500	195,518.000

Note: *p<0.1; **p<0.05; ***p<0.01