Black Candidates and Black Turnout: A Study of Mayoral Elections in the New South

Luke Keele† Paru Shah‡ Ismail White§ Kristine Kay¶

August 1, 2014

Abstract

What effect does candidate race have on co-racial voter turnout? Recent studies suggest that the presence of a black candidate results in an increase in black turnout. We argue that much of these findings can be attributed to the different design choices of previous researchers and the inattention to strategic candidate and voter behavior. In this study, we examine mayoral elections in the state of Louisiana between 1988 and 2011, matching cities along an important dimension–size of the black population. We find that when we compare turnout levels in places with similar black populations, the effect of a black mayoral candidate are modest in general elections, but substantial in runoff elections. However, we find that the result in runoff elections is driven by a small number of elections with large effects.

*We thank Eleanor Schiff for research assistance. For comments and criticisms, we thank seminar participants at Penn State University and Stony Brook University. An earlier version of this paper was present at the Annual Meeting of the Midwest Political Science Association, Chicago, IL 2014.

†Associate Professor, Department of Political Science, 211 Pond Hall, Penn State University, University Park, PA, 19130 Email: ljk20@psu.edu, corresponding author.

‡Assistant Professor, Department of Political Science, PO Box 413, University of Wisconsin, Milwaukee, Milwaukee, WI 53201 Phone: 414-229-3979, Email: shahp@uwm.edu

§Assistant Professor, Department of Political Science, 2008 Derby Hall, Ohio State University, Columbus, OH 43210 Phone: 614-292-4478, Email: white.697@polisci.osu.edu

¶Graduate Student, Department of Political Science, Derby Hall, Ohio State University, Columbus, OH 43210
1 Introduction

On Tuesday, November 7, 2006 Cedric Glover was elected the first African American mayor of Shreveport, LA. Glover’s election came as a surprise to many who doubted he could gather enough support to win in a southern city where in years past blacks had attempted but failed to gain control of city hall. Glover’s victory was ultimately attributed to his campaign’s keen ability to exploit the racial politics of Shreveport through the mobilization of African American voters. The day after the election the local newspaper, *The Times*, surmised that Glover’s victory was likely due at least in part to, “Several high-profile get-out-the-vote initiatives targeting black voters in the last week” that “energized Glover’s core voter base.” Similarly, one local commentator noted that, “Cedric did a stupendous job of turnout...Black turnout is way up from the primary, much closer to proportion” (Mahfoufi November 8, 2006).

The election of Shreveport’s first African American mayor illustrates a common explanation for black electoral success: that black candidates succeed by mobilizing black voters. This belief, well known among political pundits, has existed for decades as a kind of anecdotally substantiated fact about black political behavior. The academic literature has advanced one primary reason why minority candidates might increase participation among co-racial voters: empowerment. Under empowerment theory, a co-racial or co-ethnic candidate in political office sends a signal to the minority group that they have a stake in the political process and can influence policy (Barreto, Segura and Woods 2004; Bobo and Gilliam 1990; Browning, Marshall and Tabb 1984; Gilliam 1996; Gilliam and Kaufman 1998; Leighley 2001). This feeling of empowerment leads minority voters to participate by voting. There is also another important reason why minority candidates may increase minority turnout. Minority candidates should view minority voters as a natural voting bloc that is easily mobilized. As such, minority candidates may use targeted mobilization efforts aimed specifically at minority voters (Toure and Hamilton 1992).
While current theory has been devoted to understanding why we might expect an increase in minority participation, we would argue that there are plausible counter-arguments as well. First, many minority candidates may not be visible enough such that co-racial voters may be unaware that a minority candidate is running. Even elections for the U.S. House may not be visible enough for minority voters to even realize the candidate shares his or her racial identity, especially when many such races are uncompetitive which obviates the need for any mobilization efforts by minority candidates. Finally, it may be the case that empowerment wears off after one or two election cycles and thus may no longer mobilize minority voters (Gilliam and Kaufman 1998). Moreover, recent work has found little evidence that minority candidates spur increased turnout among co-racial voters (Sekhon, Titiunik and Henderson 2010; Keele and White 2011). While other studies have found relatively modest but positive effects using similar methodst (Fraga 2014). These last two studies carefully account for the selection of voters into districts that have minority candidates. Accounting for such selection clearly reveals differences between the types of citizens who live in districts that have minority candidates on the ballot compared to those that do not, raising serious concerns about the inferences drawn from earlier studies.

Moreover, the question is critical to understanding minority political participation. The U.S. has a long history of exclusion from politics based on race. Given this long history of racial discrimination in the U.S., it is important to understand the mechanisms by which minority participation might be increased. One method by which minority participation might be increased is if minority candidates stand for elected office. If minority candidates generally increase turnout among co-racial voters, the recruitment of minority candidates can serve as way to increase minority political participation. If minority voters do not turnout for co-racial candidates, then other avenues for increasing minority political participation will need to be found. While the empirical question is relatively simple: is minority turnout higher when minority candidates run, the logistics of answering such a question is rather complex. Both strategic behavior by candidates and voters complicates our efforts to answer
To test whether or not the presence of a black mayoral candidate on the ballot increases turnout among African American voters, we examine a unique data set on mayoral elections in Louisiana spanning more than two decades (1988-2011). What we find is that while black turnout is indeed higher in places that feature black mayoral candidates than it is in places that do not, this effect appears to be largely driven by the percentage of black residents in the municipality. Once we account for the racial composition of the municipality, we find that African American candidates only increase turnout when there is a clear signal of viability. In Louisiana, this signal of viability comes in the form of runoff elections. As such, the unique electoral system in Louisiana appears to play a key role in stimulating turnout among African Americans when there is a co-racial candidate on the ballot.

2 Mayoral Elections in Louisiana

In our study, we focus on the link between co-racial candidates and turnout in mayoral elections. Mayoral elections have several distinctive characteristics which are useful for this empirical question. First, municipal boundaries are generally not subject to manipulation via redistricting which often results in uncompetitive legislative elections. Thus municipal elections for mayor are not tailor made for African Americans to win office the way that majority-minority legislative districts often are. Second, the original empowerment claims were actually made about African American mayors (Bobo and Gilliam [1990]). Finally, mayoral candidates tend to be more visible to voters in ways that even U.S. House members may not be as the mayor may receive attention in the local media. Mayors often have closer links to voters and can credit claim in highly visible ways. Using elections other than general legislative elections often reveals important dynamics about race and politics (Branton 2009).

We created a data set to test whether black mayoral candidates increase turnout among African American voters in Louisiana. The state of Louisiana maintains three different data sources that we combined for our analysis. First, the state of Louisiana maintains a candidate
database. This database contains information on candidates for all state and local elections. The information in the database includes candidate name, address, office, data, sex, and most importantly for our purposes, race. From the database, we extracted all mayoral candidates from 1988 to 2011. While information on race is typically reported, it is at times missing. We found that for all candidates in our time period, information on race was missing 3% of the time. While we could have used an imputation model for the missing race data, instead we use a bounds approach. We generated two alternative treatment measures to assess the effect of missingness. In one, we code all missing data on race to white, and in the second we code all missing data on race to African American. We can then generate estimates for all three treatment indicators. We found the results were not sensitive to which measure we used, so we conclude that missingness on treatment is ignorable.

Next, we merged electoral returns (also maintained by the state of Louisiana) to the candidate database. Once election returns were matched to each candidate in the database, we converted the raw votes into percentages and denoted the candidates rank in the election outcomes. This step was necessary due to the unique structure of Louisiana elections. Mayoral elections in Louisiana are structured as a runoff system. In what is considered to be the general election all candidates for an office are placed on a single ballot. This general election serves in many cases as both a primary and general election mechanism. If one candidate in the general election receives more than 50% of the vote, that candidate is the winner and no further elections are held. However, if no candidate manages to receive more than 50% of the vote, the top two vote getters in the general election then advance to a runoff election which is typically held between two weeks and a month after the general election. We ranked candidates in the general election since there are often three or more candidates. Ranking the candidates allows us to know which candidate was in third place for general elections that led to a runoff election, which becomes an important feature in our research design.

The final part of our data collection consisted of compiling turnout for mayoral elections. The state of Louisiana records precinct level turnout numbers by race and party. These data
are online from 1998 to the present. For earlier years, the state has paper records, which we converted to an electronic format. Next, we matched the precincts to municipalities. In the electoral returns data, results for mayoral elections are reported at the precinct level which allowed us to map which precincts fall within specific city limits. We aggregated the precinct level turnout data for each municipality, which results in a municipal level data set with indicators for whether at least one of the candidates in the mayoral election was black. We also added in Census data from 1990 on the population of each municipality and the percentage of African American residents. While a larger number of census covariates are generally available, these covariates are not collected for most of the smaller towns that make up the bulk of our data.

3 The Selection Problem

Most extant studies of whether minority candidates increase turnout among co-racial voters focus on legislative elections. By only examining mayoral elections, we sidestep the complications of gerrymandering and the legal requirements of race and redistricting that are present in legislative elections. However, strategic behavior by candidates and voters creates important challenges to understanding this empirical question. Next, we review how strategic behavior complicates our empirical analysis.

3.1 Candidate and Voter Strategy

First, in saying that candidates behave strategically, we mean that competitive candidates will tend to run in elections where they have a relative higher probability of winning. Previous work has found that serious candidates with the credentials and money necessary to win an election tend not to run when they are unlikely to win. As [Jacobson (1989)](pg. 775) concluded about strong challengers in Congressional elections, they “do not emerge randomly; their occurrence varies with the prospects of victory.” Given the relative absence of racial cross-over voting, especially in the South, black candidates should only emerge in towns with substantial black populations.
Second, we expect black voters to be discerning in that they will back just any black candidate, but will carefully assess candidate “viability.” A number of studies have shown that links between racial or ethnic identities and vote choice depend on candidate quality. Several experiments have shown that voters are willing to vote for a less preferred candidate if that candidate has a greater chance of winning based on opinion polls or previous victories 

Forsythe et al. (1993). Similarly, Jackman and Vavreck (2010) made a connection between racial attitudes and candidate viability, finding that negative racial attitudes among white voters was associated with more doubt about President Obama’s ability to win in 2008. However, they did not investigate the consequences for black voters. Manzano and Sanchez (2010) demonstrate that candidate quality moderates the correlation between ethnic attachment and vote choice. Williams (1990) analysis of the 1987 JCPS/Gallup poll revealed that while black voters were likely to support black candidates based on race, they were often perceived to be less experienced and less likely to win. Simply put, we expect vote choice to be best predicted using an expected utility model that incorporates both voter preferences and their perceptions of candidate electability (Abramowitz 1989). In other words, a candidate being black is likely to affect the voters’ preferences, but may not affect the likelihood of voting unless the candidate appears viable.

As such strategic behavior by candidates implies that raw comparisons of elections with and without a black candidate are likely to inflate the effect of the black candidate on black turnout because places with low black turnout for other reasons are also less likely to have a black candidate, and vice versa. Second, black voters must view a candidate as electable to turnout in his or her favor, which will also overstate the effects of the black candidate in any race where the black candidate is viable. Next, we formalize the empirical complications created by strategic behavior.
3.2 Formalizing the Problem

Here, we formalize how the selection problem is a threat to our ability to draw causal inference from the data using notation from the potential outcomes framework (Holland 1986; Rubin 1974, 1990). Let $D_i \in \{0, 1\}$ be an indicator of treatment that is 1 if in a mayoral election at least one of the candidates is African American and 0 otherwise and $Y_i$ records the turnout among African Americans expressed as a percentage for each municipality. We denote that for each municipality $i$, there exists a pair of potential outcomes: $Y_i(1)$ for the level of turnout if exposed to the treatment and $Y_i(0)$ if not exposed. In this framework, we define the causal effect of the treatment as the difference: $Y_i(1) - Y_i(0)$. The fundamental problem is that we cannot observe both $Y_i(1)$ and $Y_i(0)$. Instead we must estimate average effects of treatments over populations: $E[Y_i|D = 1] - E[Y_i|D = 0] \[1\]$. This expression is the average observed difference in turnout across places with at least one African American candidate and places without an African American candidate.

This naive comparison of averages may tell us something about the true causal effect of African American candidates. This quantity may have have both causal and non causal components. The following decomposition of this observed difference reveals the problem caused by strategic candidates. We can write the observed difference in expectations as the following quantity:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

The second term captures the average causal effect of African American candidates on those who had an African American candidate. This is the average difference in turnout levels between those with an African American candidate and what would have happened had there

---

\[1\] In the analyses that follow, we estimate average treatment on the treated: $ATT = E[Y_i(1) - Y_i(0)|D_i = 1]$. 

8
been no African American candidate. The second term, $E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$, which we label selection bias, captures baseline differences in turnout regardless of treatment status. If strategic candidates select to run in cities with higher than average co-racial turnout, this term will be positive and may be so large that the observed difference in turnout may be solely a function of selection bias.

Simple summaries of the data are sufficient to reveal the selection problem. Figure 1 displays box plots for the proportion of African American residents in a municipality for general elections across treatment status. The differences are striking. In races with at least one black candidate, the median percentage of African American residents is 60%. In races without any black candidates, the median percentage of African American residents is typically 20% in that town or city. While the medians are clearly different, the overall distributions are also quite distinct. The grey boxes represent the inter-quartile range for each distribution. Here, we find that the inter-quartile ranges do not even overlap. Figure 1 provides compelling evidence that naive comparisons based on estimated differences across cities and towns with black candidates and cities and towns without black candidates are confounded given strategic behavior by candidates. We now describe the two different identification strategies that underpin our statistical analysis.

4 Research Design

Next, we outline the research design we use to identify the effect of black candidates on co-racial turnout. The difficulty we face is that our data only reveal statistical associations. These observed associations contain some unknown mix of causal and non-causal (spurious) components. A causal effect is identified when we have removed all non-causal components from the observed association. For this task, we need an identification strategy. An identification strategy is a set of assumptions needed to give estimated associations a causal interpretation. Below we outline the identification strategies that form the basis of our empirical analysis.
4.1 Observational Study

The first identification strategy we use is “selection on observables” (Barnow, Cain and Goldberger 1980). Under this approach, we assume that treatment selection is based on observable covariates. Under this strategy, analysts collect all known confounders and use a statistical estimator to make treated and control groups comparable while the treatment effect is estimated. In short, we must assume that all the differences between treated and control municipalities are observable. As we describe below, we use matching to make treated units similar to control units, but there be unobserved covariates that bias this comparison. Critically, selection on observables is strong and nonrefutable, insofar as it cannot be verified with observed data (Manski 2007). We readily acknowledge that given the limited set of covariates available and the fact that we do not observe a clearly identifiable treatment assignment mechanism, there is little reason to suspect that selection on observables is a credible identification strategy. To strength our design, we conduct a sensitivity analysis to
probe the plausibility of our identification assumption and use an alternative identification strategy, which we describe below.

4.2 Sensitivity Analysis

Given that one of our identification strategies is selection on observables, we need to probe whether our results would be sensitive to bias from a hidden confounder. In a sensitivity analysis, we quantify the degree to which a key assumption must be violated in order for our inference to be reversed. While there are a number of different methods of sensitivity analysis, we use a method of sensitivity analysis for matching estimators based on randomization inference discussed in [Rosenbaum (2002) ch. 4]. We first apply randomization inference as our mode of statistical analysis. See [Keele, McConnaughy and White (2012)] for a basic introduction to randomization inference. After matching, we apply Wilcoxon’s signed rank test, which is an appropriate randomization test for paired data with a continuous outcome. Based on Wilcoxon’s signed rank test, we estimate a point estimate and an associated 95% confidence interval via the method of Hodges-Lehmman [Hodges and Lehmann (1963)].

Under randomization inference, we assume that within matched pairs, receipt of the treatment is effectively random conditional on the matches. Formally, in our analysis, there are $I$ matched pairs, $i = 1, \ldots, I$, with two subjects, $j = 1, 2$, one treated and one control or $2I$ total subjects. If the $j^{th}$ subject in pair $i$ receives the treatment, write $Z_{ij} = 1$, whereas if this subject receives the control, write $Z_{ij} = 0$, so $Z_{i1} + Z_{i2} = 1$, for $i = 1, \ldots, I$. In our study, each matched pair consists of one municipality with at least one African-American candidate and one municipality without any African-American candidates. Matching on observed covariates $x_{ij}$ may have made cities and towns more similar in their chances of being exposed to the treatment. However, we may have failed to match on an important unobserved binary covariate $u_{ij}$ such that $x_{ij} = x_{ij'} \ \forall \ i, j, j'$, but possibly $u_{ij} \neq u_{ij'}$. If so, unlike in a randomized experiment, the probability of being exposed to treatment may not be the same within matched pairs due to the fact that we failed to match on a relevant
covariate. To explore this possibility, we use a sensitivity analysis.

First, we define $\pi_{ij}$ as the probability that municipality $j$ in pair $i$ receives the treatment. For two matched cities in pair $i$, say $j$ and $j'$, because they have the same observed covariates $x_{ij} = x_{ij'}$, it may be true that $\pi_{ij} = \pi_{ij'}$. If this condition holds, our estimates are valid since we controlled for all relevant covariates. The worry is that we may have failed to match on a relevant covariate, which would imply that $\pi_{ij} \neq \pi_{ij'}$, and our estimates will be biased. Rosenbaum (2002, ch. 4) shows that if two matched units differ in the probability of being treated due to an unobserved covariate, $u_{ij} \neq u_{ij'}$, then these two cities may differ in their odds of being exposed to the candidate race treatment by at most a factor of $\Gamma \geq 1$ such that

$$\frac{1}{\Gamma} \leq \frac{\pi_{ij}}{1 - \pi_{ij'}} \leq \Gamma, \quad \forall i, j, j', \text{ with } x_{ij} = x_{ij'}.$$ 

How is this useful? This inequality is useful since it shows that we can place bounds on quantities like point estimates for different possible levels of confounding due to $u_{ij}$. For example, if $\Gamma = 1$, then $\pi_{ij} = \pi_{ij'}$, and our point estimate for the effect of a black candidate is identified. If $\Gamma > 1$, and there is some level of confounding, then the true value of the point estimate is unknown but is bounded by a known interval.

In a sensitivity analysis, we use several values of $\Gamma$ to compute bounds on the point estimate for the treatment effect. We then observe at which value of $\Gamma$ the lower bound on the point estimate crosses zero. If that value of $\Gamma$ is large, we can be confident that it would take a large bias from a hidden confounder to reverse the conclusions of the study. If that value of $\Gamma$ is small, then we have little confidence in the results since a relatively small amount of confounding could overturn our conclusions. Thus the sensitivity analysis indicates the magnitude of bias due to an unobserved covariate that would need to be present to alter the conclusions reached under the selection on observables assumption.

12
4.3 Discontinuity Study

Under a second identification strategy, we exploit a discontinuity created by Louisiana’s runoff system. Under the Louisiana system, runoff elections are triggered when no one candidate receives a large enough vote share in the general election. When this happens, the two candidates with the highest vote shares run in a second runoff election. This creates a discontinuity between the candidates that barely qualifies for the runoff and the candidate that barely fails to qualify for the runoff.

We use this discontinuity by comparing turnout in runoff elections where at least one of the candidates is African American to runoff elections where both candidates are white, but an African American finished third in the general election. In sum, we are comparing African American turnout in a runoff election where one candidate is African American to another runoff election where no candidates are African American but an African American came in third and as such was relatively close to qualifying for that runoff. Ideally, we would restrict our analysis to runoff elections where an African American not only finished third but also garnered a share of the vote that would nearly qualify him or her for the runoff. Unfortunately, we did not have enough cases of this type, so instead we look at any election where a runoff was triggered and an African American candidate was in third place and was within 10 percentage points of qualifying for the runoff. We designate these runoff elections as African American third (AAT) runoff elections. The key advantage of ATT runoffs is that the design increases the comparability of the elections. If an African American candidate places third in the general election and a runoff is triggered, we think this makes that ensuing runoff election more comparable than comparing runoff elections with and without African American candidates. If true, there will be an observable implication we can test: balance in observables should be better for ATT runoffs when compared to all runoff elections.

In this type of runoff election, we might expect even larger effects since in one runoff election the African American electorate will have had a co-racial candidate come close to the runoff but fail. That would probably be true if the third place candidates were barely
losers, but that is sometimes not the case in our analysis. In general, we think the strength of the discontinuity is that it will increase comparability across treated and control and thus perhaps aid balance in unobservables. To that end, we will inspect observables to see if there is greater balance for these ATT runoff elections. Ideally, observables would be balanced by the design, but we believe this discontinuity is worth using so long as balance is better than in the analysis that uses all runoff elections. Should observables still be unbalanced, we can adjust for observed covariates and apply the method of sensitivity analysis outlined above.

5 Analysis

As we outlined above, we have only two covariates to adjust for: the population of the municipality and the percentage of residents that are African American. As we also noted above, we suspect that the percentage of African American residents will be particularly important. In the analysis, we use census data on African American population from 1990 instead of from 2000 or 2010. The reason we use 1990 data is that it may be the case that African Americans move to cities with African American mayors. If so, the percentage of black population may be affected by the treatment. Since the majority of our elections occur after 1990, conditioning on Census data from 1990 should reduce the possibility of bias from conditioning on a post-treatment variable. We didn’t include New Orleans in any of our analyses, since we found it incomparable to any other city in the state in terms of both population and the portion of African American residents.

We adjusted for these covariates using an integer programming based match as implemented in the R package \texttt{mipmatch} [Zubizarreta 2012]. Matching based on integer programming achieves covariate balance directly by minimizing the total sum of distances while constraining the measures of imbalance to be less than or equal to certain tolerances. This form of matching also allows us to impose constraints for exact and near-exact matching, and near and near-fine balance for more than one nominal covariate.

Under this form of matching, the analyst sets a tolerance for the imbalance for each
covariate. For example, we might set the tolerance on the difference in means on a covariate to be less than 2 points across treated and control groups. The algorithm then attempts to achieve that level of balance. If this balance constraint is infeasible the algorithm stops and reports an error. The analyst must then select a less strict balance constraint. For municipal population, we set a constraint such that matched pairs had to have a difference in population of no more than 3000 residents. For the proportion of African American residents, we constrained the average difference to be less than .01 or less than one percent on a percentage scale. For these two continuous measures, we focused on ensuring that both central moments and the higher moments of the treated and control group distributions were similar. To check imbalances in higher moments, we use the Kolmogorov-Smirnov (KS) test. The KS test measures the distance between the empirical distribution functions of the treated and control groups, which allows for the detection of discrepancies in higher moments of the distributions.

We also believe that election year is an important variable to match on. That is, we would prefer to find matches from within the same yearly electoral cycle, which will hold constant larger national or state level trends (like presidential elections) that might also drive turnout in a particular year. See the appendix for an alternative match that we considered but ruled out due to these considerations. Ideally, we would impose an exact match on election year. The difficulty, particularly for runoff elections, is that an exact match on election year tends to make balance very poor for other covariates. Instead, we used an almost exact match on election year. Under almost exact matching, if an exact match is possible, one will be produced, if not, the match will be as close to exact as possible, within a specified tolerance (Rosenbaum [2010], ch. 9). Section ?? in the appendix contains tables that report the extent to which we were able to exact match on election year. For general elections, we obtained an exact match on election year in every matched pair but one. In runoff elections, we obtained an exact match on election year in very matched pair but three.

We used the integer programming match in conjunction with optimal subset matching
Optimal subset matching allows us to match as many treated subjects as possible, recognizing that some treated subjects may be too extreme to match. The algorithm thus discards any treated units that are too extreme to match but does so such that balance differences are minimized but the number of treated units is optimized. This technique is analogous to trimming treated units via a caliper but in an optimal fashion. Under this strategy, it does change the estimand. That is, because we have discarded some treated units, the estimand is a more local version of the treatment effect on the treated. We found it impossible to retain all the treated units and achieve acceptable levels of balance.

6 Results

6.1 Balance Results and Unadjusted Estimates

We begin with by reporting levels of imbalance before and after matching. We do this to not only demonstrate that our matching estimator removed imbalances but also to assess the validity of the design based on the discontinuity. Ideally, observables would be balanced in the design based on the discontinuity, but at minimum the discontinuity should improve on balance compared to the design based on all runoff elections.

Table 1 contains the difference in means and a KS test $p$-value for general elections, runoff elections and AAT runoff elections. A few items are worth noting. In general elections, African American candidates tend to run in cities that are 53% black as compared to places that are 21% black. In runoff elections, we tend to find African American candidates in cities that are nearly 60% black as compared to cities that are only 27% African American.

The balance test results clearly vindicate using the discontinuity created by runoff elections. First, we find that under the discontinuity that municipal populations are balanced by the design. More importantly we find better comparability in terms of the percentage of African Americans. While the percentage gap is still large, in the control group the percentage of African Americans is now 41%. That is, we are more likely to have control elections where the African American candidate is viable. WE think this result validates the use of
Next, we present unadjusted estimates from general elections, runoff elections, and AAT runoffs. The unadjusted estimates are simply comparisons between races with African American candidates and races without African American candidates. We report these results to compare to the adjusted estimates. If the estimates change, we may be able to make inferences about the amount of bias reduction that occurs given our adjustments. Table 2 contains the unadjusted estimates. In all three cases, we find African American turnout is higher when an African American candidate is in the mayoral race. For general elections, African American turnout is 7.5 percentage points higher and for runoff elections turnout is 9.1 percentage points higher. In the AAT runoffs, turnout is 5.9 percentage points higher. Thus, estimates based on a naive comparison point to unambiguous racial effects. However, we should note that for the ATT runoffs, the treatment effect is smaller rather than larger than the estimate for the all runoff elections. As we noted, the balance is better for AAT runoffs, which suggests that when the treated and control municipalities are more compa-
rable the treatment effect size decreases. We next seek to further increase comparability through matching.

Table 2: Unadjusted Estimates of The Effect of African American Candidates in Mayoral Elections in Louisiana, 1988-2011

<table>
<thead>
<tr>
<th></th>
<th>General Election</th>
<th>Runoff Election</th>
<th>Runoff Election Af-Am Candidate Third&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point Estimate</td>
<td>7.2</td>
<td>6.9</td>
<td>2.6</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>[5.11, ∞, ]</td>
<td>[2.91, ∞]</td>
<td>[5.93, ∞]</td>
</tr>
<tr>
<td>p-value</td>
<td>.001</td>
<td>.001</td>
<td>.30</td>
</tr>
<tr>
<td>N</td>
<td>1006</td>
<td>192</td>
<td>54</td>
</tr>
</tbody>
</table>

Note: <sup>a</sup>This means an African-American candidate finished third in the general election and a runoff election was triggered. Point estimates are the Hodges-Lehmann estimates from the Wilcoxon sum rank test.

In Table 3 we report the balance after we match. For general elections, we achieve good balance. The difference in the percentage of African American residents is a mere tenth of a point. For municipal population, the difference in means is reduced from over six thousand to less than three thousand. To achieve this level of balance, we had to discard a number of treated observations. For general elections, we started with 335 treated elections. To balance the covariates, we discarded 220 treated observations and found suitable matches for the remaining 196 treated elections. In sum, before matching we have an apples to oranges comparison, but after matching we are more confident that this is an apples to apples comparison. For runoffs elections, we also produce a set of successful matches. For the percentage of African American residents, the difference is one percentage point. For municipal population, the difference in means is less than 600 residents. Again, we discarded a number of treated units. In the entire sample, we had 70 runoff elections with at least one African American candidate. We found successful matches for 26 of these elections. For the AAT runoffs, the difference in percentages for African American residents is eleven points. While this difference is somewhat large, we doubt comparability is hurt much since in both treated and control locations, African American candidates should be viable since
the average percentage of African American residents is 55% in the treated group and 44%
in the control group. In the data, we started with 16 ATT runoffs and found matches for 8
of these elections. Of course, in all three cases we may have failed to match on an important
hidden confounder. We consider that possibility through a sensitivity analysis.

6.2 Point Estimates and Sensitivity Analysis

We now turn to the estimates from the matched analysis. In Table 4, we report point
estimates, 95% confidence intervals, one-sided \( p \)-values, and the value for \( \Gamma \) the sensitivity
analysis parameter for both the \( p \)-value and the point estimate. First, we discuss the results
from the general elections. Recall that in the unadjusted data the point estimate was 7.2
percentage points, and once we match-assuming there are no hidden confounders-the point
estimate is now 2.6 percentage points. The sensitivity analysis, however, reveals that these
estimates are quite sensitive to bias from a hidden confounder. If the odds of treatment (\( \Gamma \))
differ by as little as 19\%, we would conclude that the actual treatment effect is zero, and
if the odds of treatment differ by as much as 7\%, the \( p \)-value would exceed the usual 0.05
threshold. How can we tell if these are large or modest values of \( \Gamma \)?
To provide a baseline, we regressed the treatment status on municipal population and percentage of African American residents using a logit model. We then calculated the odds-ratios for the covariates in this model. We can compare these odds-ratios from this model to those from the sensitivity analysis. For example, increasing the percentage of African Americans by one percentage point increases the odds of an African American candidate by 9%. Thus if we failed to match on a covariate that alters the odds of treatment the same magnitude as one percentage point increase in African American population that could explain the estimate we observe in general elections. Our inference is clear, once we compare cities with similar racial distributions, the effect of an African American candidate is much more modest in general elections, and the effect could easily be explained by unobserved confounding.

For runoff elections, however, the point estimate actually increases to 10.9 percentage points and is statistically significant ($p = .007$). The point estimate is fairly robust to the presence of a hidden confounder. A hidden confounder would have to change the odds of treatment by 470%, to explain this point estimate. Using our benchmark model, increasing African American population by 15 points increases the odds of treatment by only 380%. Thus even if we failed to match on a cofounder of this magnitude, we would still conclude that an African American candidate increases turnout. The $p$-value is more sensitive with a $\Gamma$ value of 1.42.

Finally, in the AAT runoffs, the point estimate suggests that African American candidates increase turnout by 4.3 percentage points. This estimate is notably smaller than the estimated effect in regular runoff elections. If we think the discontinuity better balances unobservables, this would suggest the effect is overstated in runoff elections. In terms of the sensitivity analysis, however, the magnitudes are similar as the $\Gamma$ value for ATT runoffs is 4.3 for the point estimate and 4.7 for all runoff elections.

The results from our empirical analysis suggest that an African American candidate for mayor increases turnout a modest amount in a general election. Moreover, this estimate
Table 4: Adjusted Estimates of The Effect of African American Candidates in Mayoral Elections in Louisiana, 1988-2011

<table>
<thead>
<tr>
<th></th>
<th>General Election</th>
<th>Runoff Election</th>
<th>Runoff Election Af-Am Cand. Places Third$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point Estimate</strong></td>
<td>2.6</td>
<td>10.9</td>
<td>4.3</td>
</tr>
<tr>
<td><strong>95% Confidence Interval</strong></td>
<td>[0.53, ∞]</td>
<td>[2.6, ∞]</td>
<td>[−2.7, ∞]</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>.01</td>
<td>.007</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Γ point estimate</strong></td>
<td>1.25</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td><strong>Γ p-value</strong></td>
<td>1.07</td>
<td>1.42</td>
<td>1.02</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>392</td>
<td>52</td>
<td>16</td>
</tr>
</tbody>
</table>

Note: $^a$In this design, an African-American candidate finished third in the general election and a runoff election was triggered. We match on proportion of the population that was African American in 1990, the log of municipal population in 1990, and election year.

could be easily explained by bias from a hidden confounder. In runoff elections, however, it appears that African American candidates increase turnout among co-racial voters by at nearly 11 percentage points. The estimates from the runoff elections are also far more resistant to bias from hidden confounders. Next, we consider the possibility of heterogenous treatment effects.

### 6.3 Effect Heterogeneity

Thus far we have explored whether a black mayoral candidate increases turnout among co-racial voters. We believe that the evidence indicates that this is likely in runoff elections. We now ask whether the treatment effect we observe is due to an uncommon but dramatic response to a small number of black candidates. Here, we might imagine that for many black candidates their effect on turnout is negligible, but in some cases the African American candidate is unusually charismatic or well funded and able to mobilize effectively, and he or she produces an uncommonly large turnout effect. Such a pattern as this is often missed by methodology that focuses on typical effects. Moreover, should we find such a pattern, it will clearly limit the generalizability of the effects we estimated. This pattern of treatment
effects suggests that while some African American candidates can activate voters and increase turnout, most do not.

We begin by plotting the data and looking for results suggestive of this pattern using a quantile-quantile plot, which plots two empirical distributions against each other. In Figure 2 we plot matched pair treated and control outcomes in a quantile-quantile plot. For general elections, the matched pair differences fall along the solid line, which suggests a fairly constant effect. For runoff elections, the matched pair differences diverge consistently from the solid line, suggesting that the treatment effect is quite large in some cases.

![Figure 2: Quantitle-quantile plots of turnout in treated and control groups. Solid line represents $y = x$. Under a constant effects model, treated control differences should fall along solid line.](image)

How might we probe for unusual effects in a more formal way? [Rosenbaum (2007)] develops a formal test for such a pattern of effects, which we now describe. Typically, in an analysis of treatment effects, we only compare treated outcomes to control outcomes. To understand whether some treated outcomes are unusual, we must compare treated outcomes to both control outcomes and other treated outcomes. [Rosenbaum (2007)] suggests comparing
subjects not two at a time but in groups of size \( m \), for fixed \( m > 2 \). If we set \( m = 8 \), we would compare outcomes among a set of 8 observations that are a mix of treated and control units. Within the block of \( m \) units, we calculate the aligned response as \( R_j - \bar{R}_m \), where \( R_j \) is the response for subject \( j \) and \( \bar{R}_m \) is the average response within the \( m \) units in the group. The aligned response measures the discrepancy between each unit in the group of size \( m \) and the group average. Say subject \( j \) in block \( i \) had the largest aligned response among the \( m \) subjects in that group. For this unit, we say it had the largest or peak response within that group. A peak response would be a small piece of evidence that this subject had an unusual outcome compared to other units. Moreover, if the unit with the peak response in the group is a treated unit, this treated unit has an unusual response compared to not only control units but other treated units as well. We then record for this group whether the peak response was from a control or treated unit. We then perform this same calculation for all the possible groups of size \( m \) that can be formed from the \( n \) units. For all these possible groups, we record how often the peak response was from a treated unit. When a high percentage of peak responses are due to treatment, this is evidence that some treated responses are large relative to all other treated and control responses.

Thus far, we have a method for counting whether the percentage of peak responses are among the treated. However, if a treated subject had a peak response, this might happen by chance or it might happened because the treatment caused a dramatic increase in the response for that unit. To understand, the role chance plays in a treated unit displaying a peak response, we use permutations of the statistic that counts number of treated peak responses to calculate a \( p \)-value for evidence against the null hypothesis that the percentage of peak responses among the treated occurs by chance.

We begin with the results from runoff elections. [Rosenbaum (2007)] suggests setting \( m \) to 20 for larger sample sizes and to 4 or 8 for smaller sample sizes. For runoff elections, we did separate analyses with \( m \) set to 4 and then 8. For 26 pairs, we can form 1,562,275 possible groups of size 8. When you examine groups of eight cities, and you pick the city
with the largest increase in turnout, 93.5% of the time the city with the most impressive increase in turnout was a city with a black mayoral candidate. This effect is statistically significant with a $p$-value of 0.018. When there are big gains in turnout, they mostly occur in the treated group of cities. For 26 pairs, we can form 14950 possible groups of size 4. For runoff elections, when you look at 4 cities at a time 79.9% of the time the unit with a peak response was a treated unit, which was also statistically significant with $p$-value of 0.005. That is, peak responses, or unusually large effects are concentrated in the treatment group.

Now, we turn to the results from general elections. Since we have 196 matched pairs, we set $m = 20$, so that we consider groups of twenty municipalities. For 196 pairs, we can form $1.1 \times 10^{27}$ possible groups of size 20. We found that in all these possible groupings a treated unit had a peak response 29.9% of the time, and we are unable to reject the null with a $p$-value of 0.10. When you look at 20 matched cities and you pick the city with the largest increase in turnout, about 30% of the time that city is a city with a black candidate. If we set $m = 10$, we find that a treated unit had a peak response 23% of the time, with a $p$-value of 0.081. In general elections then, then peak responses are far less common among the treated group.

This pattern suggests that for general elections, the effect is relatively constant across matched pairs. That is, peak responses occur among treated units about a third of the time when we examine effects among larger sets of units. For runoff elections, however, we find that it is often the case that the treated unit displays a peak response among a larger subset of units. This suggests that some African American candidates are able to produce uncommonly large effects, while in many other cases the effects are relatively small. This again limits the generalizability of our findings. That is, we find that African American candidates can increase turnout among co-racial voters, but this effect is concentrated in runoff elections and among a small subset of candidates within those runoff elections.
7 Discussion and Conclusion

The goal of this study was to understand whether the race of a candidate plays a role in turnout among co-racial voters. While no single result from observational data can be considered definitive, the pattern in our data was fairly clear. In general elections, once we evaluate comparable mayoral races, there was, at best, weak evidence that black voters turnout to vote more for black candidates. Specifically, the effect is relatively modest once we adjust for confounders, and the sensitivity analysis suggests that it would take a small amount of bias from a hidden cofounder to explain the observed difference across treated and control groups. In runoff elections, we found much larger estimates that were robust to bias from hidden confounders. Moreover, we found the estimates were consistent across two different designs: one design based entirely on selection on observables and one design based on a discontinuity. However, we found evidence that this effect is driven in large part by a small number of candidates in runoffs, who produce uncommonly large turnout effects.

This pattern of effects, however, creates a paradox in terms for what we might learn more broadly about African American participation. It appears that runoff elections serve as a signal of viability for African American candidates. In these cases, African Americans are more likely to vote when the candidate is co-racial. In general elections, however, without the viability signal, African Americans may be somewhat more likely to vote but the effects are weak and subject to bias from hidden confounders. This suggests that the conditions for higher turnout among co-racial voters are confined to the unusual electoral rules of Louisiana. That is, outside of Louisiana, African American candidates won’t receive the viability signal provided by runoff elections which appears to spur stronger levels of co-racial turnout. Thus our null result appears to be much more generalizable than the more confirmatory result from runoff elections. That is, the viability signal provided by runoff elections is unlikely to be present in general elections in Louisiana or in many elections outside of Louisiana.

Our analysis also highlights the role of candidate strategy. African American candidates
run for mayor in places with a large base of co-racial support and are much less likely to run in other places at least in Louisiana, but we would assert that this is probably true throughout the American South. As we demonstrated, places with at least one African American candidate on average have a population that is around 60% black. In legislative districts, we might expect this to occur since districts are often drawn to be majority-minority. Municipal elections, however, are not subject to such outside manipulation, and yet the behavior of candidates appears to be quite similar.

These findings also underscore the problem of selection effects that have plagued many of the early observational studies attempting to assess the effects of co-racial candidates on co-racial turnout. Given that black candidates are clearly not randomly assigned to elections, simple comparisons of co-racial/ethnic candidates and black and/or Latino voter turnout are problematic and likely overstate the impact that co-racial/ethnic candidates have on minority turnout. The data and analytical approach presented above represent a good guideline for accurately assessing the effects of co-racial/ethnic candidates on minority turnout.

We might also ask what implications strategic candidate theory has for empowerment theory. On the face of it, there would appear to be some incompatibility, since empowerment theory focuses on how black candidates change attitudes and feeling among voters. However, we would argue that the two theories are compatible, suggesting mutually strategic candidates and voters. If black candidates are indeed not running in places where they believe they are less likely to win because of a perceived racial disadvantage, then to the extent that black voters living in these places are also aware of this disadvantage, they have no reason to turnout and vote, even when a black candidates does run. Black voters in these places are essentially disempowered, not by the lack of black candidates but by a chronic lack of black electoral power.

For blacks in low black population districts (and especially in places where blacks have few viable coalition partners) it is not until black population numbers reach a certain threshold
that they will perceive the group to have any electoral strength. Thus, black turnout in these communities is likely to remain low as voters doubt the prospects of electing candidates that they think will represent their interest. It is, of course, this very logic that drove the creation of majority-minority districts under the Voting Rights Act.

Our predictions are also consistent with research on candidate emergence which suggests that the under-representation of blacks in elected office is not an issue of potential candidates, but a shortage of elections black candidates feel they have the potential to win. For example, one analysis of why so few black House members do choose to run for Senate seats points to the role of racial demographics (Johnson, Oppenheimer and Selin 2012). Similarly, Shah’s (2014) analysis of black candidate emergence in local offices across Louisiana concludes black candidates are most likely to run when the demographics of the jurisdiction are in their favor. Thus, the issue does not seem to be a shortage of potential candidates, but a shortage of elections black candidates feel they have the potential to win, based on demographics. Any examination of the ability of black candidates to influence turnout should take into account the size of the black population as this is likely the primary determinant of why a black candidate runs and the degree to which black voters see that candidate as viable.
References


Appendices

A.1 Alternative Form of Matching

One might imagine an alternative way to implement our selection on observables identification strategy. In our designs, we chose to match treated units to control units in the same election year so long as the two units were similar on observable covariates. An alternative way to implement our identification strategy, would be to use within unit comparisons. Here, instead of matching units to other cities, we would use a within unit comparison. That is, we could compare turnout from a general election without any African American candidates for mayor to turnout from a different general election with an African American candidate for the same municipality.

We don’t implement this second match for one reason. Turnout is subject to large cross-temporal shocks. The most prominent example of this is the large fluctuation in turnout due to presidential election. The worry here is that we might attribute higher turnout from year to year fluctuations to the presence of a African American candidate for mayor. That is, we would not want to compare a treated election in a presidential year to a control election in almost any other year. However, even if we matched across two presidential years, turnout can fluctuate across presidential election years depending on competitiveness and mobilization. In general, we think that year-to-year fluctuations in turnout are too large to make the within-municipality design credible. We also would argue that this same year to year variation in turnout also make a differences-in-differences (DID) design implausible. (Keele and Minozzi 2012) provide on example of how year to year fluctuations in turnout violate the key identification assumption in a DID design.