Unifying the Demand and Supply-Side Theories of Minority Representation: A Logical Model

Yuki Atsusaka*

atsusaka@rice.edu

June 10, 2020

Abstract

I derive a *quantitatively predictive logical model* of minority candidate emergence. Drawing on previous research, I demonstrate that the probability of minority candidates running for office can be computed by a function of what I call the racial margin of victory in the last elections and the percentage of minority voters. To show the validity of the model, I visualize its quantitative predictions and test them with novel data sets of Louisiana mayoral elections from 1986 to 2016 and state legislative general elections in 2012 and 2014. The logical model can correctly predict 89% of minority candidate emergence and its in-sample and out-of-sample predictive performance is higher than those of multivariate regressions. The logical model unifies demand and supply-side theories of minority candidate emergence and provides new insight that the district racial composition is only half of the story of minority descriptive representation and that previous elections also matter.

Keywords: Quantitatively predictive logical models; minority representation; candidate emergence; electoral engineering

Word Count: 10923

*Yuki Atsusaka is a Ph.D. Candidate in the Department of Political Science at Rice University, MS-24 105 Herzstein Hall, 6100 Main St, Houston, TX 77005 (https://atsusaka.com). For helpful comments, I thank Chano Arreguin-Abarca, Wendy K. Tam Cho, Bernard Fraga, Matthew Hayes, Jonathan Homola, Eric Juenke, Melissa Marschall, Jared Oestman, Tyler Reny, Liana Reyes-Reardon, Kaitlin Senk, Randy Stevenson, Rein Taagepera, Michelle Torres, M.C. Sunny Wong, Marques Zarate, the members of the Rice Method Research Group (Gustavo Guajardo, Colin Jones, and Yui Nishimura), and the participants of the University of Houston EITM Summer Institute 2019.

Introduction

Under what conditions do racial minority candidates enter and win electoral contests? Answers to this question often provide critical normative and policy implications related to the Voting Rights Act and redistricting in the U.S. (where geographical concentration of minority voters and first-past-the-post (FPTP) elections are often met) (Grofman, Handley and Niemi, 1992; Elmendorf, Quinn and Abrajano, 2016; **?**). To date, the district racial composition has been considered as the most important determinant of minority descriptive representation in the U.S. (Swain, 1993; Cameron, Epstein and O'halloran, 1996; Canon, 1999; Lublin, 1999; Grose, 2011). To explain why, previous research has provided two competing theories, which emphasize *either* minority candidate supply (Shah, 2014; Juenke, 2014; Juenke and Shah, 2015, 2016) *or* voter demand (Barreto, Segura and Woods, 2004; Bedoya, 2005; Casellas, 2010; Lublin, 1999; Lublin et al., 2009), but the interaction between the two has not been fully explored. In contrast, this article shows that both theories can be integrated under a unified theory of minority candidate decision-making. In so doing, it offers new insight that the district racial composition is only half of the story and that previous elections also matter as minority candidates can learn about voter demand from past electoral performance of co-ethnic candidates.

This article makes several contributions. First, I derive a *quantitatively predictive logical model* — a specific type of formal model which provides quantitative predictions on the outcome of interest (exact probability in this study) — of minority candidate emergence by first applying an emerging approach in comparative politics to minority representation in the U.S. (Taagepera, 2007, 2008; Li and Shugart, 2016; Shugart and Taagepera, 2017; Taagepera and Nemčok, 2019). Substantively, the logical model provides two new important observable implications: (1) the *racial margin of victory* (the extent to which minority candidates win relative to their white counterparts) in the last elections increases the probability of minority candidate emergence and (2) its influence becomes the greatest in districts where minority and white voters are equally distributed.

Formally, I demonstrate that the probability of minority candidates running for office can be computed by a simple function of the racial margin of victory in the last elections (expressed by M) and the expected racial margin of victory based on the district racial composition (expressed by C). The logical model theorizes that minority candidates decide to run for office when the probability of winning is high, which they estimate based on what happened in the last elections (M) and what their districts look like now

(C). This way, it theorizes minority candidate emergence as an "interaction" between candidates' strategic calculation and voters' demand (Bullock III and Johnson, 1985; Bullock III and Smith, 1990; Canon, 1999; Fraga, 2014). Importantly, the model never implies that other elements such as partisanship and incumbency do not matter. Rather, it suggests that these other factors can be *summarized* into the margin of victory in previous elections by quantifying the voter demand for minority candidates (which is a function of various factors including partisanship of candidates and voters). Consequently, the logical model provides a key to understand a set of cases that previous research cannot explain, such as the absence of minority candidates in majority minority districts and the presence of minority candidates in majority white districts.

Next, I illustrate the logical model by visualizing its quantitative predictions and clarifying the two key observable implications. I then test the model predictions with a novel data set of Louisiana mayoral elections from 1986 to 2016. I demonstrate that the logical model has remarkably high predictive performance both in-sample and out-of-sample and it can correctly predict 89% of black candidate emergence. It is then shown that the *logical* model (obtained without looking at any data) dominates linear and logistic regressions (*statistical* models) in predictive performance.

Moreover, to ensure the internal validity of the analysis, I further test additional observable implications from the model (i.e., what should be true if the model properly describes the reality). As expected, I find that most elections where black candidates emerged do not feature more than one black candidate, that the model does not predict the number of black candidates or white candidate emergence, but that it can correctly predict 94% of black candidate victory. Finally, to confirm the external validity, I compile a data set of state legislative general elections in 2012 and 2014 and test the model predictions against the data. I demonstrate that the logical model can accurately predict minority candidate emergence in different levels of office in nineteen unique states and for different racial groups (black and hispanic candidates) with high predictive power (86% can be correctly predicted).

As elaborated below, the innovation of this study is to provide a parsimonious mathematical model of minority representation which (1) is based on deductive logic as opposed to statistical analysis; (2) is able to offer quantitative predictions about the probability of minority candidate emergence; and (3) ultimately helps researchers design political institutions that could promote minority representation. One of the advantages of this approach is that it tightly aligns the theoretical concepts of interest with their empirical measures as required in good model building (Granato, Lo and Wong, 2010). For example, I demonstrate that in FPTP elections (that I focus on below) C is algebraically identical to the "percentage" of minority voters. Moreover, the model includes a multiplicative form $(MC)^{1/2}$ to represent the argument that minority candidates take into account both M and C when calculating the probability of winning. As I discuss below, the multiplication is the result of taking the "geometric mean" of the two quantities. This is highly advantageous in scientific studies of politics because it can mitigate researcher degrees of freedom in measuring concepts and choosing functional forms of variables in analysis (along with the impact of sampling variability in data) and thus ease the accumulation of knowledge.

Ultimately, the proposed model offers a useful benchmark for developing logical models of minority representation more broadly. If researchers can logically connect institutional variables that are subject to change with relevant outcomes of minority representation (e.g., the number of minority representatives in a legislature) and if they can represent such relationships by simple mathematical equations, it becomes possible for them to *design* political institutions that could promote minority representation (Taagepera, 2008; Li and Shugart, 2016).

1 Voter-Demand and Candidate-Supply Theories

To explain the relative lack of minority descriptive representation, scholars have considered two competing theories, which include voter-demand and candidate-supply theories (Shah, 2014; Juenke and Shah, 2015, 2016). The demand-side theory sees the relative dearth of minority politicians as the minority candidate *defeat* problem. According to this viewpoint, minority candidates cannot win electoral contests outside of majority minority districts due to strong opposition from white voters (Abosch, Barreto and Woods, 2007; Lublin et al., 2009; Trounstine and Valdini, 2008). Thus, the relative lack of minority representatives is the result of white voters who do not vote for minority candidates. An important normative and policy implication of the voter demand theory is that creation and maintenance of majority minority districts is *the* solution to minority underrepresentation (Barreto, Segura and Woods, 2004; Bedoya, 2005; Casellas, 2010; Lublin, 1999; Lublin et al., 2009).[1]

In contrast, more recent scholarship based on the supply-side theory describes minority underrepresentation as the minority candidate *retreat* problem. Here, minority candidates do not run for office outside of majority minority districts even though they have good chances of winning once they enter electoral compe-

[1]Here, I only focus on minority descriptive representation and specifically minority candidate emergence. Thus, the proposed model does not intend to provide a direct explanation or prediction on minority symbolic or/and substantive representation. Nonetheless, such connections could be a fruitful topic to explore in future research.

titions (Shah, 2014; Juenke, 2014; Juenke and Shah, 2015, 2016; Fraga, Juenke and Shah, 2020). Therefore, the lack of minority representatives is due to potential minority candidates' miscalculation of the odds of winning. One critical implication of the candidate supply theory is that minority underrepresentation can be partly solved by providing minority potential candidates with high quality information about the (future) voting behavior of the white and minority electorates in their districts.

While each of the two competing theories seems plausible, several unsolved problems remain. First, previous research has not explicitly modeled the decision-making of minority potential candidates. Additionally, most research has treated "racial bias" by white voters as a black box or a fixed phenomenon. In other words, while the literature has recognized that white voters are less likely to vote for minority candidates, it does not fully explore (or quantify) the variance in the extent of such racial bias in different elections and how minority candidates respond to varying degrees of racial bias, if any. Finally, no single study has simultaneously incorporated important information about how districts differ from one another in important ways — specifically, what I call the *racial regime* of the district and the temporal dimension (that I describe below) into the theory of minority descriptive representation.

By the racial regime, I refer to a different type of district characterized by its racial composition, which might influence minority candidate emergence. For example, Juenke and Shah (2016) examine minority candidate emergence in white, racially mixed, and majority minority districts, respectively, and report interactive effects of racial regimes, candidate race, and candidate partisanship on minority candidate emergence. Shah (2017) also focuses on three "racial profiles" including majority white, multiracial, and majority minority cities and illustrates different power dynamics between racial groups in different types of cities. By the temporal dimension, I mean a connection between current elections and past elections in terms of electoral performance of minority candidates. For example, Marschall, Ruhil and Shah (2010, 114-15) discuss that past success in minority electoral bits may influence minority candidate victory in later years. Similarly, Shah (2014, 269, 271) examines the impact of having any history of black candidacy on the current black minority emergence. Shah (2017) also studies how the changes in the district racial composition would affect minority electoral fortune and candidate supply in later years. Below, I show that the racial regime and the temporal dimension can be seamlessly combined when focusing on the decision-making of minority candidates.

In the next section, I introduce a logical model of minority candidate emergence and illustrate how it unifies both voter-demand and candidate-supply theories of minority descriptive representation, while solving the remaining problems in the literature. Building on previous research, I only focus on FPTP elections in the rest of the article.

2 Logical Model of Minority Candidate Emergence

In this section, I present a quantitatively predictive logical model of minority candidate emergence. I then elaborate and visualize key observable implications from the model. Before formally introducing the model, I first describe my theoretical argument about when and where minority candidates decide to run for office.

In the proposed model, I attempt to understand district-level minority candidate emergence as the result of decision-making by the most viable minority politicians, whose sole agenda is to win elections in which two racial groups (i.e., minority and majority groups) are in competition. Given these scope conditions, I theorize that minority candidates decide to run for office when they think they have a high probability of winning. Minority candidates then attempt to calculate the likelihood of winning in the upcoming elections, but they have limited information about their chances. To calculate the likelihood of winning, minority candidates rely on two sources of information: the (1) electoral performance of co-ethnic candidates in the last elections and (2) district racial composition. I argue that the value of information from the last elections increases as districts become more competitive or more racially heterogeneous because the district's racial makeup is not as informative in racially balanced districts as in racially homogeneous districts.

Finally, it is worth emphasizing that while focusing on these two information, I do not mean to imply that nothing else matters. Rather, my purpose is to present a parsimonious model which reflects a simplified version of the reality (but nevertheless captures important parts of that reality) so that it can assist in *designing* political institutions that could enhance minority representation. In fact, other information such as partisanship are summarized into the electoral performance of minority candidates (Online Appendix A.5). Ultimately whether the proposed model is valid depends on whether it produces accurate predictions about minority candidate emergence — and as I show below, it does. In the rest of this section, I formalize my argument and provide a logical model of minority candidate emergence.

2.1 Quantitatively Predictive Logical Models

To theorize minority candidate emergence, I adopt an emerging approach in comparative politics to derive a *quantitatively predictive logical model* (or logical model in short) (Taagepera, 2007, 2008; Li and

Shugart, 2016; Shugart and Taagepera, 2017; Taagepera and Nemčok, 2019).[2] According to Taagepera (2008), "Quantitatively logical predictive models include all models that can be introduced without input of numerical data, based solely on broader general notions" (29). The fundamental goal of this approach is to use deductive logic to derive a parsimonious mathematical equation which can accurately predict the outcome of interest (so that it can assist in designing political institutions, like electoral systems, to produce desirable social outcomes). A canonical example is the Seat Product Model developed by Taagepera (2007), where the effective number of parties with legislative representation (N_s) is predicted by a function of the average district magnitude (M) and the assembly size (S) in a given country as $N_s = (MS)^{1/6}$. Thus, both explanation and prediction are important components of the approach.

Another prominent feature of this approach is that logical models have a tighter connection between theoretical concepts and their empirical measures, which enables researchers to provide stronger empirical tests for their theoretical arguments than conventional approaches. Moreover, the logical model approach has clearer policy implications because logical models can tell exactly *how much* outcomes change (i.e., quantitative predictions) if we vary one or more parts of their equations. Therefore, logical models can offer a considerably more information for important practical problems including electoral engineering and constitutional design, which have enormous impacts on representation and democratic process (Li and Shugart, 2016, 23). While such logical modeling has hardly appeared in the literature of minority representation, or in political science studies more generally, it appears to be more common in other scientific disciplines including physics (Colomer, 2007). I must emphasize again that *by definition* the following logical model is a parsimonious representation of the reality and does not explain every possible detail in minority representation. The essence of the approach is that even though it considers only two variables, it can correctly predict 89% of minority candidate emergence as shown in Section 4. Below, I start by introducing four assumptions and then describe the logical model.

2.2 Scope Conditions

To clarify which kinds of electoral contexts the model accounts for, I limit the scope conditions by stating four assumptions.

[2] To the best of my knowledge, this is the first attempt to apply "logical models" (as a special case of formal models) to the study of minority representation. This does not mean, however, that no other research has used formal models or game theoretic models to study minority representation and politics (e.g., **?**).

Assumption 1 (Biracial elections). *Electoral competitions are held over two racial groups (majority and minority voters).*

The first assumption states that I consider biracial elections where two racial groups compete with each other. Below, I consider districts where over 95% of the voting population is composed of two groups as biracial districts.[3]

Assumption 2 (Non-zero pool). *There is always a non-zero number of minority politicians or potential candidates who could run if conditions allow in each district.*

This assumption excludes the possibility that we do not observe any minority candidate running for office due to the lack of minority potential candidates in the candidate pool.[4]

Assumption 3 (Instrumental candidates). *Minority candidates are short-term instrumental such that their primary goal is to get elected in upcoming elections.*

This assumption rules out the possibility that minority candidates decide to run for office due to non-instrumental reasons; namely, they seek to run for office in order to obtain benefits from the action of running itself (e.g., raising voice for symbolic representation or selling names for future elections).

Assumption 4 (The most viable candidate). *Whether a district features at least one minority candidate solely depends on the strategic decision of the most viable minority politician in the district.*

This assumption enables us to model district-level minority candidate emergence as the most viable minority politician's decision-making process in the district.[5] This assumption postulates that when and where minority candidates are observed does not depend on the behavior of minority politicians whose perceived

[3]Below, I assume that the other voters (5% or less of the voting population) behave just as majority voters. Empirically, I then count the other voters as white voters. It must be emphasized that biracial elections (as defined here) are not rare in the U.S. For example, 25.5%, 28.5%, and 17.7% of state legislative general elections in 2012 and 2014 were black-white, hispanic-white, and asian-white biracial elections, respectively.

[4]For studies looking at the potential-candidate pool, see Maisel and Stone (1997) and Fox and Lawless (2004).

[5]By minority "candidates," I refer to minority candidates and their teams including party stuff and strategists. However, to simplify the term, I simply use the term candidates.

odds of winning is not the highest among co-ethnic potential candidates. In other words, less viable minority politicians would not decide to enter the electoral competition unless the most viable minority politicians do so. Along with Assumption 1, this can be considered as a racial version of the "M+1" rule (Reed, 1990; Cox, 1997).[6]

In summary, Assumptions 1-4 claim that district-level minority candidate emergence can be modeled as the decision-making of the most viable minority politicians whose sole agenda is to win biracial elections.

2.3 The Logical Model

Based on these assumptions, I derive a logical model of minority candidate emergence. The full derivation is detailed in Online Appendix A and interested readers are referred to its extensive discussion. Here, I only present the following result:

$$\text{Pr}(E) = \widehat{P} = \mathbf{F}\big((MC)^{1/2} - 50\big) \tag{1}$$

The logical model suggests that the probability of minority candidate entry, Pr(E), is equal to the estimated probability, $\widehat{P}$, that minority candidates win the electoral contest once they run for office. It further shows that such estimated probability can be computed by two variables M and C through some cumulative distribution function (CDF) **F**, which models the uncertainty around minority candidates' estimates of the probability of winning.[7]

M is what I call the racial **m**argin of victory in the last elections and it quantifies the extent to which minority candidates "safely secured" their descriptive representation relative to their white counterparts. More precisely, it is based on the difference between the vote shares that went to the top minority candidate, V^M_{t-1}, and the top white candidate, V^W_{t-1}, in the last election in each district: $M = \frac{1}{2}(V^M_{t-1} - V^W_{t-1}) + 50$.[8] Drawing on the concept of the margin of victory, it quantifies the amount of vote shares that need to be modified in order to change the "race" of the winner. Online Appendix A.4 presents a proof that this quantity is universal across any configuration of races including the contests without minority candidates (among eight possible patterns). The advantage of the proposed model, then, is that it can predict and explain minority candidate emergence *even in districts where no minority candidate appeared before* (see

[6] The "M+1" rule posits that the number of serious or viable candidates (or parties) in a district tends to be its district magnitude (M) plus one. In FPTP elections, where $M = 1$, the rule suggests that the number of serious candidates tend to be 2.

[7] While the model can be written as Pr$(E_{it}) = \widehat{P}_{it} = \mathbf{F}\big((M_{it-1}C_{it})^{1/2} - 50\big)$ to index district by i and time period by t, I suppress such indices to follow the convention in logical model building.

[8] More precisely, M is an affinely transformed racial margin of victory as $M = \tilde{M} + 50$, where $\tilde{M}$ is the raw margin of victory.

Online Appendix A.5).

Next, C denotes the *expected* racial margin of victory in upcoming elections based on the racial **c**omposition of districts. It expresses the racial margin of victory under the assumption that there are only one minority and white candidates, respectively, and that minority and white voters only vote for their co-ethnic candidates (perfect co-ethnic voting) with the same level of turnout. After some algebra, it turns out that C is equivalent to the percentage of minority voters.[9] Here, I am not implying that the above assumption is the most realistic description of the world. Rather, it defines one *extreme situation* that the model can predict in principle.

The key idea is that what happened in the past (M) and what the district looks like now (C) provide two "limit scenarios." One limit scenario is that the result of an upcoming election will converge to the result of the last election, whereas another limit case is that it converges to the result based on the district racial composition. I then theorize that *likely electoral results will be located somewhere between the two extreme situations*. To express this idea formally, I take an average of the two quantities via a geometric mean $(MC)^{1/2}$.[10] I specifically use geometric means here to reflect the idea that the impacts of the two variables are interactive (see Figure 1) in addition to the fact that geometric means have favorable properties in logical models (Taagepera, 2008). Finally, without changing any substantive meaning, 50 is subtracted from the geometric mean to adjust for the transformation applied to M and C in the derivation.[11] The resulting value will be an input of a given CDF that outputs values between 0 and 1 (which we can interpret as predicted probabilities).

To sum up, the logical model provides quantitative predictions about the probability of minority candidate emergence with only two terms M and C. Importantly, both terms are based on widely agreed-upon measurements in political science research and, thus, theoretical concepts and measurements are tightly aligned in the model as required in good model building (Granato, Lo and Wong, 2010).[12] It is also worth emphasizing that the model was derived from a deductive logic and not stemmed from any statistical analysis or even "data." Consequently, its predictions are stable and not susceptible to how researchers measure concepts and what sample or training data they use as in machine learning or other types of predictive models.

[9]To show this, let me simply state that the raw expected racial margin of victory becomes $\tilde{C}$ =%MinorityVoters - 50. After applying the same affine transformation as for M, $C = \tilde{C} + 50 =$ %MinorityVoters.

[10]A geometric mean of N numbers $(y_1, ..., y_N)$ is defined as $\sqrt[N]{y_1 *, ..., * y_N} = (\prod_{i=1}^{N} y_i)^{1/N}$.

[11]$\tilde{M}$ and $\tilde{C}$ are transformed as $M = \tilde{M} + 50$ and $C = \tilde{C} + 50$ in order to avoid the multiplication of negative numbers.

[12]While the racial margin of victory is a novel concept, it is based on a well-known concept of the margin of victory.

Substantively, the logical model indicates that minority candidate emergence is based on minority candidates' decision-making process as advocated by the supply-side theory, whereas their decision-making reflects potential vote shares that minority candidates might receive as the demand-side theory expects. In other words, the model claims that minority candidates react to past voter demand M to estimate the future electoral fortune given the district racial composition C. This way, the logical model unifies the two theoretical perspectives which have been contrasted against each other in previous research. In so doing, it provides a more comprehensive picture of minority candidate emergence, and it does so in a very parsimonious manner.

2.4 Observable Implications

To illustrate how the model works, I first provide a simple numerical example. Suppose that black voters are the minority group of interest and white voters are the majority group of reference. Suppose also that the top black candidate and the top white candidate obtained 30% and 50% of vote shares, respectively, in the last election in a given district.[13] Suppose also that 60% and 40% of the district is composed of black and white voters, respectively. This exemplifies the district where black voters are a slightly numerical majority but a white candidate won the last election.

With this information, it is easy to compute that $M = \frac{1}{2}(30-50)+50 = 40$ and $C = 60$. The model then suggests that the probability of black candidate entry becomes: $\mathbf{F}((40 * 60)^{1/2} - 50) = \mathbf{F}(48.99 - 50) = \mathbf{F}(-1.01)$. When a probit function is used for **F**, the probability of black candidate emergence becomes $\Phi(-1.01) = 0.156$ or there is about 16% chance that a black candidate appears in an upcoming election in the district.

Figure 1 visualizes the model's quantitative predictions under varying conditions. The left panel plots the probability of minority candidate emergence against $(MC)^{1/2} - 50$ with different choices for **F**. When the resulting value is negative, the probability is lower than or equal to 0.5, whereas it becomes higher than or equal to 0.5 when the term is positive. Substantively, $(MC)^{1/2} - 50$ represents an estimated margin of victory for minority candidates and the shape of **F** controls the degree of uncertainty around such estimates. The key message here is that the overall look of the S-curve does not change greatly regardless of the choice of **F**.[14] The graph also explains why -50 appears in Equation (1), which is to locate the entire function

[13] Importantly, these vote shares need not to be summed up to 100% when multiple candidates (including write-in candidates) are on ballots and they obtain some portion of the total ballots.

[14] For this reason, below, I use CDFs for N(0,5) and N(0,1) when visualizing the model prediction and computing the predictive

around 0.

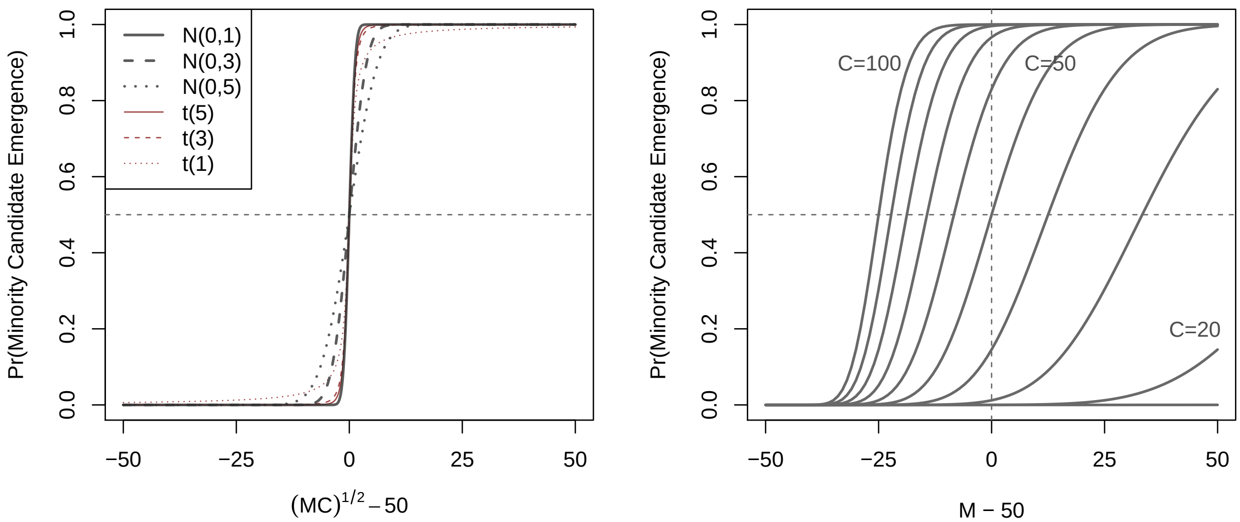

Figure 1: **Quantitative Predictions of the Logical Model**
Note: The left panel shows the probability of minority candidate emergence as a function of $(MC)^{1/2} - 50$ with various forms of **F**. The right panel visualizes the impact of M (racial margin of victory) on the probability under varying the levels of C (percentage of minority voters).

At first glance, the two terms in the logical model seem to have equal impacts on the probability of candidate emergence. Somewhat surprisingly, this is not an accurate description of the model. To see why, see the right panel of Figure 1 where I plot the probability of minority candidate emergence against $M - 50(= \tilde{M})$ with varying values of C. It illustrates that (1) increase in M raises the probability, but (2) the degree of such increase depends on the value of C. For example, when $C = 20$ (20% minority districts), the probability of minority candidate emergence is mostly 0 and it starts to marginally increase after $M - 50$ is beyond 25. Substantively, this means that minority candidate emergence is not likely in the 20% minority districts *unless* black candidates were elected with large margins (did "extremely well") in the last elections relative to their white counterparts.

The opposite case can be seen when $C \approx 100$ (predominantly minority districts). Here, the probability of minority candidate running is mostly 1 and it starts to decline when $M - 10$ is less than -10. Substantively, this indicates that minority candidates tend to be on ballots with certainty *unless* they were defeated by white candidates with great margins (did "extremely poorly") in the last elections. Finally, the influence of M on

performance, respectively.

the probability becomes the greatest when $C = 50$ (50% minority districts). The logic behind this is that, in perfectly racially balanced districts, the district racial composition provides no information about minority candidates' relative advantage, and thus what happened in the last elections becomes very informative. Thus, the logical model suggests that minority candidates as decision-makers put more weight on the information from the last elections as their districts becomes more racially balanced.[15]

Figure 1 demonstrates the property of logical models that they can provide "quantitative" predictions (i.e., exact probability in this context) as opposed to qualitative or directional predictions (i.e., the increase or decrease in the probability) (Taagepera, 2008).

3 Data

To test how closely the model can predict the emergence of minority candidates in the real world, I construct a novel candidate-level data set of mayoral elections in 313 Louisiana municipalities from 1986 to 2016. Louisiana elections serve as a great benchmark for testing the model predictions since more than 96% of voters are either African American or white (according to the official registration records with self-reported race, see Online Appendix B.5). The data contains vote totals for 5297 candidates in 2037 elections, and it is collapsed at the election level so that each contest becomes the unit of analysis. Moreover, I compiled information about candidates' race based on internet and news article searches.[16]

As the logical model predicts varying effects of the racial margin of victory conditional upon the district racial composition, I coded the racial regime (i.e., majority white, racially balanced, and majority minority (black) districts) according to the average percentage of black voting-age population (VAP) over time. To code the racial regime, I employed cutoff points of 40% and 65%, respectively, the latter of which is based on

[15]The asymmetry seen in the right panel of Figure 1 is derived from the fact that the logical model is based on the geometric mean (as opposed to arithmetic mean) of the two terms.

[16]Louisiana mayoral elections provide a great test case for the model predictions also because they use a unique electoral system called the majority run-off system, which enables us to overcome a potential problem of using general election data to study minority candidate emergence. In general elections, the absence of minority candidates stems from two possibilities: (1) no minority candidate decided to run for office to begin with and (2) minority candidates emerged but were defeated in primary elections. It must be emphasized that previous research using general election data does not (and cannot) empirically differentiate the two potential mechanisms (but state legislative general elections could be an exception as most of them do not have competitive primaries). In contrast, in the majority run-off system, all candidates participate in open-primary elections regardless of partisan affiliation, and the candidate with the majority votes becomes the winner. When no candidate obtains the majority votes, then, the top two candidates compete each other in a run-off election. I found that 10% of the initial data were such run-off elections. I drop these elections to enable a better test for the model predictions since studying minority candidate emergence in run-offs faces with the same problem we have in general elections. This enables us to eliminate the second possibility for the absence of minority candidate, and make inferences about minority candidate emergence as the direct consequence of minority candidates' decision to run for office.

empirical and legal arguments about majority minority districts (Cameron, Epstein and O'halloran, 1996). Among 2037 elections, 63.5% are from majority white, 25.6% are from racially balanced, and 11.4% are from majority minority districts, respectively. I also confirmed that very few districts experienced drastic demographic changes (see Online Appendix B.2).

The outcome variable of interest is a binary variable denoting whether an election features at least one black candidate. One or more black candidates appear in about 26% of elections in the entire data. By the racial regime, 6% of observations in majority white districts, 45.7% in racially balanced districts, and 94.2% in majority minority districts feature at least one black candidate. Figure 2 displays the distribution of the outcome variable across districts over time sorted by the average percentage of black VAP. Here, each row corresponds to a particular district and it is color-coded either as white (for no election), light pink (for no black candidate), or dark pink (for having at least one black candidate). It visualizes that it is very rare to see black candidates in majority white districts (below 40%), whereas it is very common to see black candidates in majority minority districts (over 65%).

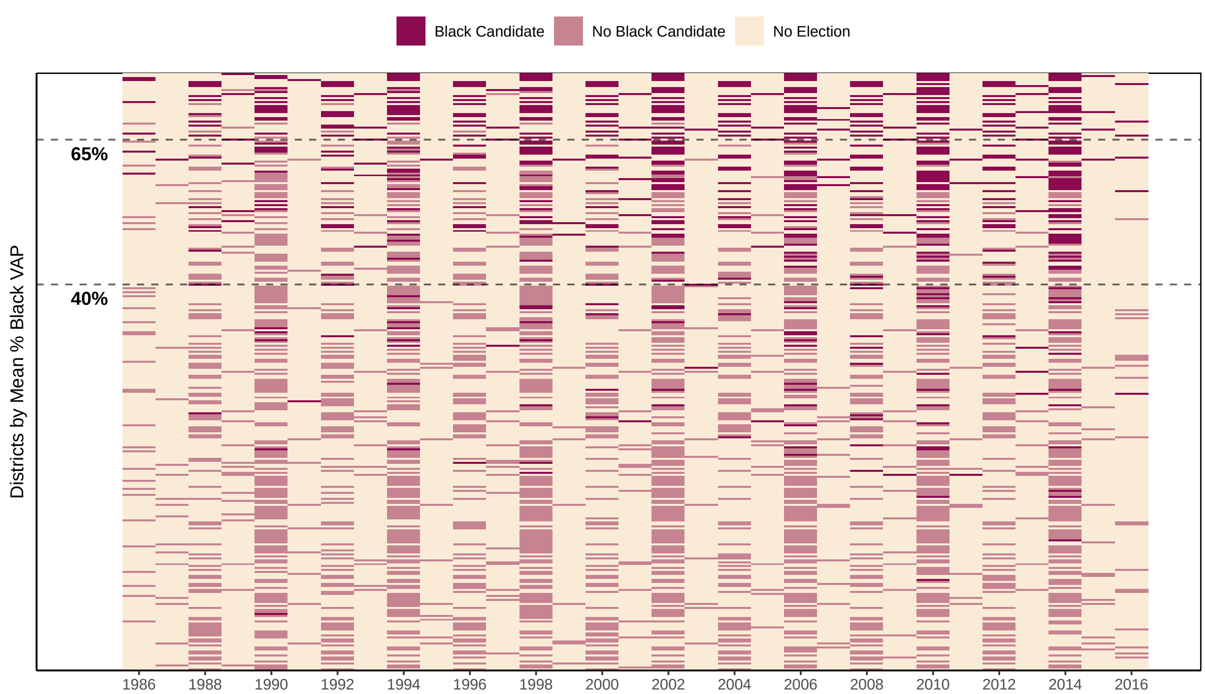

Figure 2: **Distribution of the Outcome Variable across Districts over Time**
Note: This figure portrays the distribution of the outcome variable in 303 districts from 1986 to 2016. Dark pink represents elections with one or more black candidates, light pink indicates elections without any black candidate, and off-white shows years with no election. The districts are ordered by the average percentage of black voting-age population in the entire time period from the highest (top) to the lowest (bottom).

Finally, the racial margin of victory is measured as $M = \frac{1}{2}(V_{t-1}^{M} - V_{t-1}^{W}) + 50$, where V_{t-1}^{M} and V_{t-1}^{W} are the vote shares obtained by the top minority (black) and the top white candidates in the last elections, respectively. C is measured by the percentage of black VAP. Online Appendix B presents descriptive statistics of the data.

4 Testing the Implications of the Model

4.1 Model Prediction and its Novelty

I begin by examining the predictive power of the logical model. Specifically, I compute the predicted probabilities of black candidate emergence based only on Equation (1) and compare them to the observed data points. Figure 3 visualizes the model predictions along with the observed data as a contour plot. Here, open circles ($\circ$) denote elections with at least one black candidate and crosses ($\times$) represent elections with no black candidate. The plot takes the racial margin of victory on its x-axis and the percentage of black VAP on its y-axis.

The gradation of color represents the predicted probabilities of black candidate emergence based on the logical model. Here, the redder areas show higher probabilities that black candidates run for office and whiter regions indicate lower probabilities. The figure implies that even a casual visual inspection can tell that the logical model predicts observed black candidate emergence quite accurately. Namely, most elections with black candidates ($\circ$) are located in the regions with predicted probabilities of 0.5 and higher and most elections without black candidate ($\times$) appear in the areas with predicted probabilities of 0.5 and lower.[17]

To highlight the contribution of the logical model, I mark (with squares) two observations as examples of cases that previous research cannot fully explain, but the logical model can properly account for. The left square shows the *absence* of minority candidate in a majority minority district (where over 75% are minority), whereas the right square marks the *presence* of minority candidate in a majority white district (where less than 40% are minority). Both cases are (or must be) *puzzling to prior studies* because it is believed that minority candidate emergence is heavily determined by the proportion of minority voters. Indeed, both demand and supply-side theories suggest that, given their district racial compositions, we must observe minority candidate(s) running in the first case, while we should not expect any minority candidacy

[17]The concentration of data points around -25 in the x-axis implies that many elections were fought between two viable white candidates who obtained more or less equal vote shares. For example, when there are two white candidates (and no black candidate) whose vote shares are 49 and 51 percentage points, respectively, the racial margin of victory becomes $(0 - 51)/2 = -25.5$.

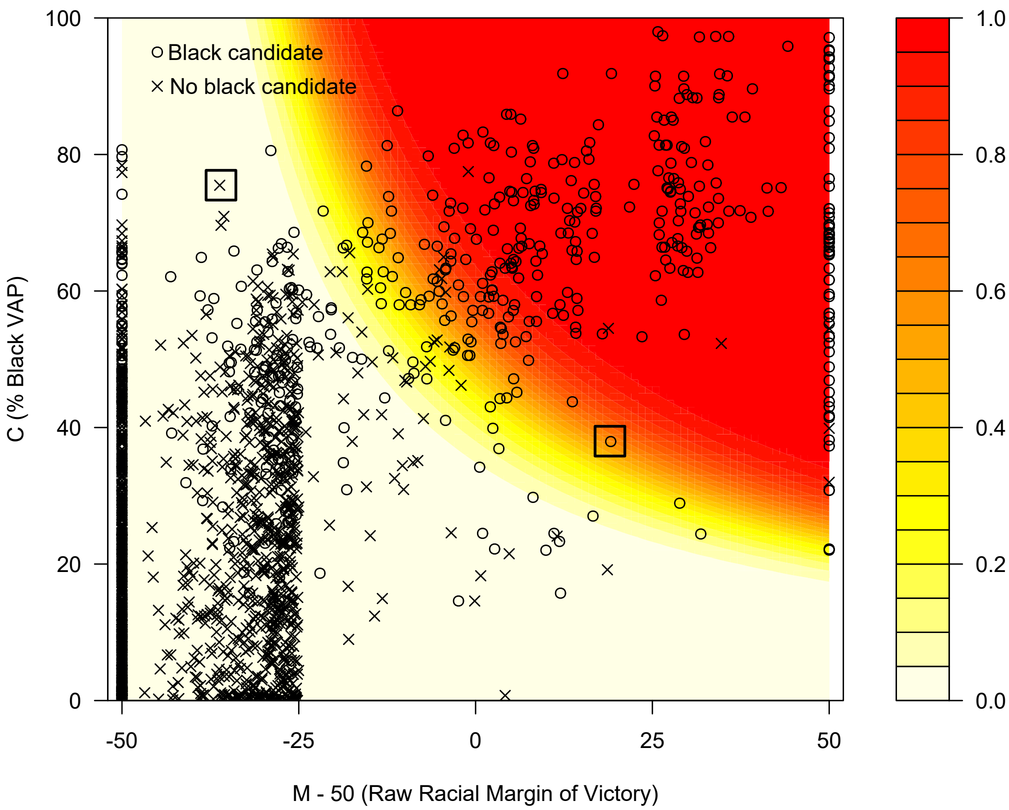

Figure 3: **Model Prediction of Black Candidate Emergence**
Note: This figure visualizes the predicted probabilities of black candidate emergence as a function of M (racial margin of victory) and C (% black VAP) as a contour plot. Elections with black candidates are shown as open circles ($\circ$), whereas elections without black candidate are represented by crosses ($\times$). Two observations marked by squares are the examples that previous research cannot explain but the logical model can.

in the second case.

In contrast, the logical model indicates that both two cases "make sense" (predicted probabilities are 0.00 and 0.56, respectively). Indeed, the logical model expects the absence of minority candidates *even in* majority minority districts when their co-ethnic candidates performed "poorly" in the last elections (relative to their white counterparts), whereas it anticipates the emergence of minority candidates *even in* majority white districts when their co-ethnic candidates won the last elections with large racial margin of victories (performed "well" relative to the white candidates). Thus, not only can the logical model explain such previously puzzling cases, but also it can provide quantitative predictions on the outcomes for these cases. Consequently, the logical model offers novel insight that the district racial composition is only half the story of minority candidate emergence.

4.2 Predictive Performance

To quantify the above finding, I compute the accuracy of model predictions based on the expected Percentage Correctly Predicted (ePCP). The ePCP provides the percentage of observations for which the model can correctly predict their values (0 or 1), while accounting for "how close" such predictions are (Herron, 1999). The results are reported in Table 1.

	In-sample Prediction (%)			Out-of-sample Prediction (%)		
	Logical Model	LPM	Logit	Logical Model	LPM	Logit
All Districts (N=2037)	**88.6**	83.2	85.0	**88.3**	82.9	84.7
Majority White Districts (N=1293)	**94.1**	90.3	90.8	**94.1**	89.7	90.1
Racially Balanced Districts (N=521)	**73.7**	66.4	67.0	**75.6**	66.9	67.7
Majority Minority Districts (N=223)	**90.9**	91.8	93.6	**91.1**	90.1	91.4

Table 1: **Predictive Performance of the Logical Model**
Note: This table reports the predictive performance of the logical model and linear and logistic regressions by ePCPs. ePCPs are calculated as $\frac{1}{N}(\sum_{y_i=1}\widehat{P}_i + \sum_{y_i=0}(1-\widehat{P}_i))$, where N is the number of units and y_i and $\widehat{P}_i$ are true and predicted values for unit i, respectively. For out-of-sample prediction, Leave-P-Out cross validation is employed where P is the number of units from the same municipality.

Column 1 of Table 1 shows that based only on the logical model we can correctly predict about 89% of cases in the data. It also reports that the logical model successfully predicts about 94%, 74%, and 91% of cases in majority white, racially balanced, and majority minority districts, respectively. These results are rather remarkable because, again, the predictions are based on the *logical* model (which is derived independent of any data) and not machine learning or other predictive models (which are derived from data).

To bolster this point, I run a series of linear probability models (LPM) with the percentage black VAP and the racial margin of victory as predictors and generate ePCPs. The results are reported in Column 2 of Table 1. The ePCP based on this "atheoretical" regression is about 83% for the full sample and this is lower than the result based on the logical model.[18] Moreover, the LPM generated lower ePCPs compared to the logical model for majority white and racially balanced districts. For majority minority districts, the LPM seems to perform slightly better than the logical model. However, since these districts consist of only about 11% of the data, the logical model performs better on average. Next, I also run a set of logistic regressions

[18] I call this regression "atheoretical" because it does not have any justification for the functional form of the predictors (the regression includes the two predictors linearly).

with the same predictors. The results are reported in Column 3 of Table 1. While logistic regressions seem to have higher ePCPs than LPMs, the overall patterns remain the same and they do not perform better than the logical model in most elections.

Moreover, I also run LMP and logistic regressions with 34 additional variables that previous research tends to include in multivariate regressions (Results are reported in Online Appendix C.4). I find that ePCPs based on theses extended regressions are still lower than the ePCPs from the logical model despite the fact that the extended regressions have 32 additional variables. Theses results demonstrate the power of logical models and suggest that parsimonious theoretical models could better describe the world than *ad hoc* statistical models (Taagepera, 2008).

Some readers may be concerned with the external validity of these findings. They might wonder to what extent the results are stemmed from the feature of our data set (or sample). To mitigate such concern, I compare the out-of-sample predictive performance of the logical model and LPM and logistic regressions based on Leave-P-Out cross validation (Celisse et al., 2014).[19] Columns 4-6 of Table 1 illustrate that the out-of-sample performance of the logical model is still quite high (88.3%) and higher than or equal to these regressions, suggesting that the logical model does not just fit this data set by chance but provides a stable prediction even outside of it.

Finally, I replicate the above analysis by only using data for elections with at least two candidates (excluding unopposed elections) and open races where incumbents did not run (Online Appendix C.1). I find that the substantive conclusion remains the same: the logical model has high predictive power both in opposed elections (82% in-sample and 84.3% out-of-sample ePCPs) and in open races (84.4% in-sample and 81.6% out-of-sample ePCPs).

5 Internal and External Validity

5.1 Internal Validity

If the logical model indeed predicts minority candidate emergence (E) through the estimated probability of winning ($\widehat{P}$), it *should not* predict other political outcomes such as the *number* of minority candidates, whereas it *must* predict minority candidate victory. This is a critical point for validating the logical model since the model is based on the strategic entry of the most viable minority candidates. To investigate this

[19] I set $P = 303$ and evaluate the predictive performance of the logical model and regressions for each district.

point and ensure the internal validity of the above analysis, I examine several key observable implications from the model. Such observable implications include: (1) the majority of elections with minority candidate emergence features only one minority candidate; (2) the logical model does not predict the number of minority candidates; (3) the number of minority candidates follows some random pattern; and (4) the logical model predicts minority candidate victory; (5) the logical model does not predict white candidate emergence. I found evidence for all of these five expectations.

First, I found that 60% of the elections where black candidates appear had only one black candidate, followed by 26% with two black candidates, 10% with three black candidates, and 4% with four or more black candidates. Second, I found that the logical model does not predict the number of black candidates. The left panel of Figure 4 plots the number of black candidates on ballots over the model predictions as a contour plot. It shows that there is no clear relationship between the number of black candidates and the racial margin of victory (x-axis). To provide more quantitative evidence, I also run a series of count models with the number of black candidates as the dependent variable and the racial margin of victory and the percentage of black VAP as two explanatory variables (The results are reported in Online Appendix C.2). I found that even though the percentage of black VAP is positively associated with the number of black candidates, the racial margin of victory has no association that is statistically (and substantively) significant.

Third, I found that the number of black candidates follows a "random pattern" (instead of model predictions) and, more specifically, a power law in its tail.[20] The power law can be observed in diverse random phenomena where many observations are clustered around some typical values (e.g., the intensity of wars, the severity of terrorist attacks, and the frequency of US family names) (Clauset, Shalizi and Newman, 2009). When empirical data follows the power law (distribution), very few values are observed with significantly high frequencies with many other values with low frequencies. The right plot of Figure 4 visualizes this pattern for the number of black candidates in our data. The hypothesis test suggests that the null hypothesis that the data follows a power law distribution cannot be rejected (p-value=0.79).

Fourth, I found that the logical model can correctly predict 94% of black candidate victory (in ePCP). Substantively, this could mean that minority candidates indeed decide whether to run for office by carefully consulting with the likelihood of winning. This result is also critical since it enables us to approximate the logical model of minority *descriptive representation* by the logical model of minority *candidate emergence*.

[20] In reality, we can never be certain that the observed quantities follow a power law due to the sampling variability in tail values. Thus, a more conservative conclusion is that, given the performed hypothesis test, it is more likely that the tail of the number of black candidates is drawn from a power-law distribution (with the minimum number is 3).

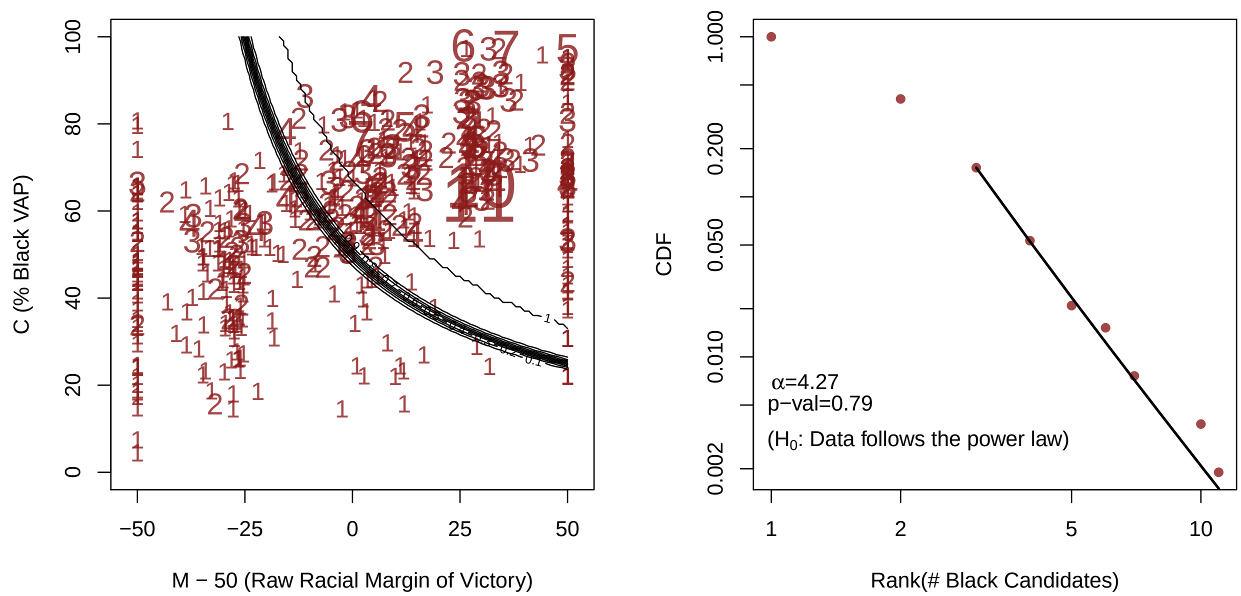

Figure 4: **The Model Does Not Predict the Number of Black Candidates**
Note:The left panel plots the number of black candidates running for office over the parameter space of the logical model with model predictions as a contour plot. The right panel shows that the number of black candidates follows a power law distribution.

Finally, I found that the logical model does *not* predict the emergence of white candidates. Figure 5 plots the observed white candidate emergence over the model predictions, where the x-axis represents the racial margin of victory for white candidates and the y-axis is the percentage of white VAP. It demonstrates that while white candidates emerge in elections where the logical model expects them to emerge, *they also seem to appear in elections where the logical model expects them to abstain*. In other words, white candidates seem to run for office in districts where the perceived probability of winning is not high enough and black candidates would not decide to run if they were in similar situations. The computed in-sample ePCPs are 0.61, 0.82, and 0.98 for majority minority, racially balanced, and majority white districts, respectively. If the logical model applies to white candidate emergence as well, the predictive power should not vary this much by the racial regime.

The strikingly low predictive power for majority minority districts (0.61 is slightly better than a coin flip) indicates that the logical model of minority candidate emergence cannot properly explain and predict white candidate emergence. A further analysis shows that this pattern applies to only white male candidate emergence. I find that the ePCPs for white male and white female candidates are 0.63 and 0.91, respectively,

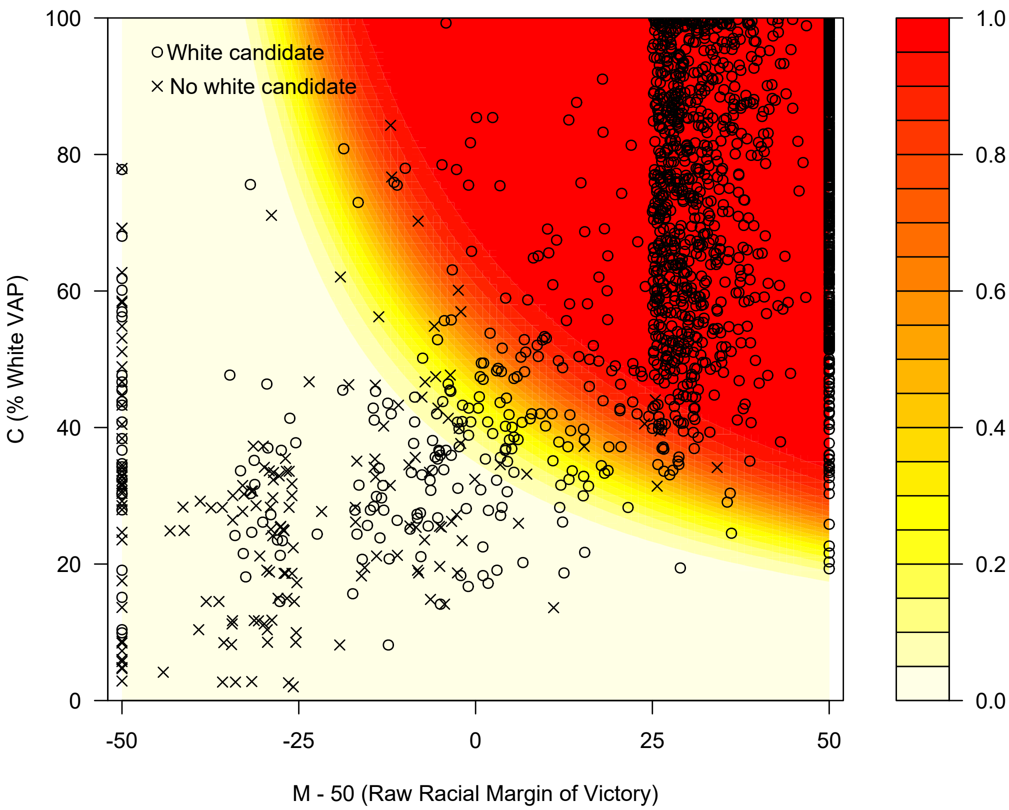

Figure 5: **Model Prediction of White Candidate Emergence**
Note:This figure visualizes the predicted probabilities of white candidate emergence as a function of M (racial margin of victory) and C (% white VAP) as a contour plot. Elections with white candidates are shown as open circles (○), whereas elections without white candidate are represented by crosses (×).

in majority minority districts. In contrast, the ePCPs for black male and black female candidates are 0.96 and 0.98, respectively, in majority white districts. This may imply that male candidates tend to overestimate the probability of winning when they run for office (thus **F** is different for them) (e.g., Fox and Lawless, 2004). Alternatively, it might be the case that they do not overestimate, but correctly assess the fact that non-white and female voters are more likely to vote for white male candidates than vice versa (and thus their racial margin of victory is different from that of black and white female candidates) (thus C is shifted toward 50 for them). While this is in and of itself an interesting finding, I leave to future research a more detailed investigation of such intertwined relationships between race and gender in descriptive representation.

Taken together, these analyses provide evidence for the claim that the proposed model is indeed the model of minority candidate emergence.[21] In Online Appendix C.5, I further show that the key observable implications from the model hold even after controlling for several variables in the multivariate regression framework.

5.2 External Validity

Next, I check the external validity of the above analysis by testing the model predictions against the data of state legislative general elections in 2012 and 2014 complied by Fraga, Juenke and Shah (2019). Out of 9716 State House and Senate elections for which they coded minority candidate emergence, I identified 150 FPTP elections where two racial groups (*either* black and white *or* hispanic and white voters) competed with each other. More specifically, I chose districts (elections) where over 97% of the citizen voting-age population (CVAP) is composed of two racial groups. Among these biracial districts, I further selected districts where minority voters (either black or hispanic voters) make up more than 10% of the CVAP (i.e., I left out extremely white districts).[22] This leaves 150 elections from nineteen unique states, where 73 are black-white biracial elections and 77 are hispanic-white biracial elections.[23] Finally, I coded the racial margin of victory for each election based on the Ballotpedia and tested the model predictions against observed minority candidate emergence coded by Fraga, Juenke and Shah (2019).

The results are shown in Figure 6, where districts with minority candidates are denoted by their capitalized state name (e.g., TX) and elections without any minority candidate are represented by their (bolded) lower-cased state name (e.g., **tx**). As in Figure 3, the color gradation represents the predicted probabilities of observing minority candidate based on the logical model. The figure provides supportive evidence that the logical model accurately predicts minority candidate emergence in different levels of office (i.e., state house and senate) in various states for different racial groups. The results for elections in South Carolina and Texas are especially illustrative of the predictive power of the model, where the percentage of minority

[21] I further conduct other placebo tests with female candidate emergence as false outcomes for the model predictions (Online Appendix C.3), finding that the logical model does not predict female candidate emergence. This suggests that the logical model is indeed a model of "racial minority" candidate emergence.

[22] This does not mean that biracial elections are rarely observed in state legislative elections. When I keep elections where over 95% of the CVAP is composed of two groups, I obtain 2473 black-white biracial elections and 2770 hispanic-white biracial elections, respectively. Thus, using 150 elections only provide a conservative test for the model predictions. Also, very few districts satisfy the second condition for asian voters, and thus asian candidate emergence is not examined here and should left to future research.

[23] These states include Alabama, Colorado, Florida, Georgia, Indiana, Illinois, Kansas, Kentucky, Maryland, Michigan, Missouri, New Mexico, Ohio, South Carolina, Tennessee, Texas, Uta, and West Virginia. While Arizona also provided multiple biracial elections, I excluded elections from Arizona because its state elections were not FPTP elections (two candidates are selected from a single district).

CVAP is more or less the same, but the elections with higher (lower) racial margin of victories featured (did not feature) minority candidates within each state. The in-sample ePCP for the entire sample is 0.91 and the ePCPs only for black and hispanic candidate emergence are 0.89 and 0.93, respectively. In short, I find that the logical model provides a generalizable knowledge about minority candidate emergence across levels of office, states, and racial groups.

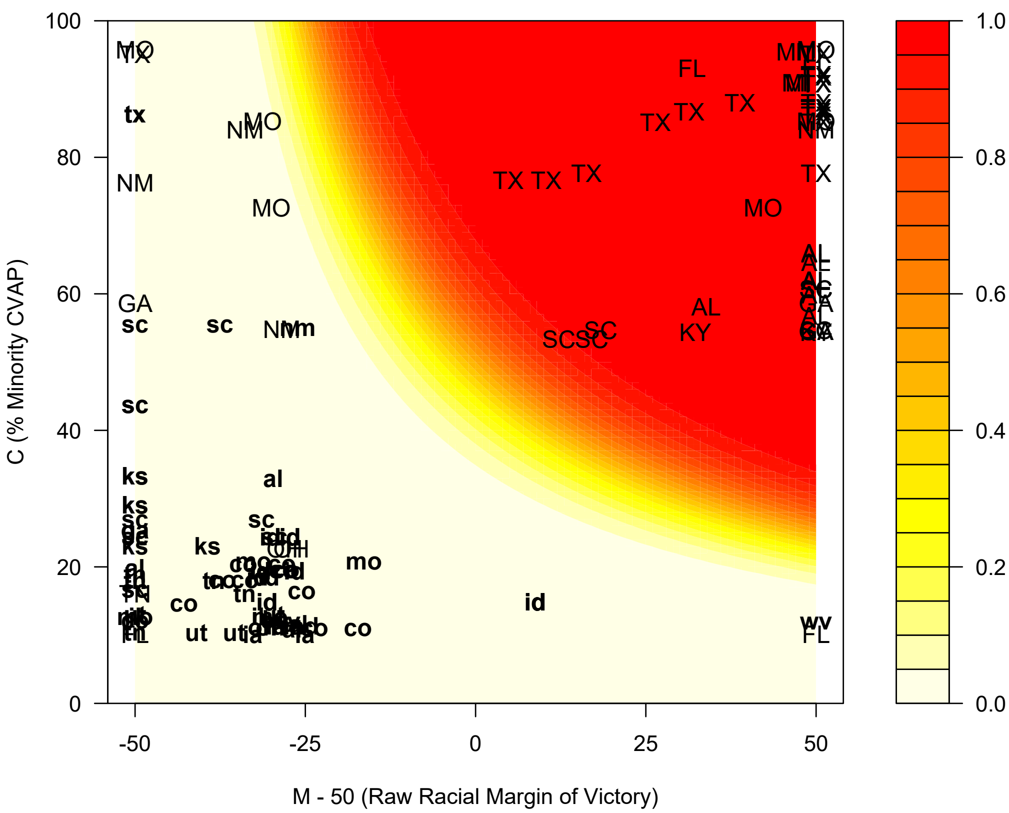

Figure 6: **Model Prediction of Minority Candidate Emergence (State Legislative Elections)**
Note: This figure visualizes the predicted probabilities of minority candidate emergence as a function of M (racial margin of victory) and C (% black or hispanic CVAP) as a contour plot. Elections with minority candidates are shown as Capitalized state names (e.g., TX), whereas elections without black candidate are represented by lower case state names (e.g., **tx**).

Concluding Remarks

Minority descriptive representation is a critical component of representative democracy and has been extensively studied by scholars of racial politics, representation, voting rights, election law, and electoral systems.

This article contributes to the literature by providing a quantitatively predictive logical model of minority candidate emergence (and by proxy minority descriptive representation), which unifies previously divided theoretical camps of voter demand and candidate supply; models minority candidate decision-making based on deductive logic and clearly stated assumptions; and offers highly accurate predictions of minority candidate emergence and victory. This article has introduced a novel concept of the racial margin of victory and demonstrated that the district racial composition is only half of the story, enabling us to explain the outcomes that previous research was not able to account for.

The logical model derived in this article offers a useful benchmark for developing logical models of minority representation more broadly. If such general models enable researchers to logically connect institutional variables that are subject to change with relevant outcomes of minority representation via simple mathematical equations, it would further enrich "applied" political science, where political science knowledge is used to design political institutions (Li and Shugart, 2016). Developing and linking such logical models makes for fruitful opportunity for future research.

To develop such a general model, future research must tackle at least three problems. First, the implications of the proposed model must be further tested against different data sets. Second, future research should relax Assumption 1 and generalize the model for multiple racial groups. Specifically, deriving M and C in multiracial FPTP elections would be a fruitful topic for future research. Finally, extending the model into non-FPTP elections would be another fruitful topic. Deriving M and C in at-large elections and (single or multi-winner) ranked-choice voting, for example, would be an important area of future study.

References

Abosch, Yishaiya, Matt A Barreto and Nathan D Woods. 2007. "An assessment of racially polarized voting for and against Latino candidates in California." *Voting Right Act Reauthorization of 2006: Perspectives on Democracy, Participation, and Power* pp. 107–131.

Barreto, Matt A, Gary M Segura and Nathan D Woods. 2004. "The mobilizing effect of majority–minority districts on Latino turnout." *American Political Science Review* 98(1):65–75.

Bedoya, Alvaro. 2005. "The Unforseen Effects of Georgia v. Ashcroft on the Latino Community." *Yale LJ* 115:2112.

Bullock III, Charles S and A Brock Smith. 1990. "Black success in local runoff elections." *The Journal of Politics* 52(4):1205–1220.

Bullock III, Charles S and Loch K Johnson. 1985. "Runoff elections in Georgia." *The Journal of Politics* 47(3):937–946.

Cameron, Charles, David Epstein and Sharyn O'halloran. 1996. "Do majority-minority districts maximize substantive black representation in Congress?" *American Political Science Review* 90(4):794–812.

Canon, David T. 1999. *Race, redistricting, and representation: The unintended consequences of black majority districts*. University of Chicago Press.

Casellas, Jason P. 2010. *Latino representation in state houses and congress*. Cambridge University Press.

Celisse, Alain et al. 2014. "Optimal cross-validation in density estimation with the $L\hat{}\{2\}$-loss." *The Annals of Statistics* 42(5):1879–1910.

Clauset, Aaron, Cosma Rohilla Shalizi and Mark EJ Newman. 2009. "Power-law distributions in empirical data." *SIAM review* 51(4):661–703.

Colomer, Josep M. 2007. "What other sciences look like." *European Political Science* 6(2):134–142.

Cox, Gary W. 1997. *Making votes count: strategic coordination in the world's electoral systems*. Cambridge University Press.

Elmendorf, Christopher S, Kevin M Quinn and Marisa A Abrajano. 2016. "Racially Polarized Voting." *U. Chi. L. Rev.* 83:587.

Fox, Richard L and Jennifer L Lawless. 2004. "Entering the arena? Gender and the decision to run for office." *American Journal of Political Science* 48(2):264–280.

Fraga, Bernard L. 2014. Race, Party, and Candidate Prospects across the Multiple Stages of Congressional Elections. Technical report Working Paper. http://www. bernardfraga. com/uploads/2/2/3/4/22341374 .

Fraga, Bernard L., Eric Gonzalez Juenke and Paru Shah. 2019. "Replication Data for: One Run Leads to Another: Minority Incumbents and the Emergence of Lower Ticket Minority Candidates.".
URL: *https://doi.org/10.7910/DVN/IUXG2I*

Fraga, Bernard L, Eric Gonzalez Juenke and Paru Shah. 2020. "One Run Leads to Another: Minority Incumbents and the Emergence of Lower Ticket Minority Candidates." *The Journal of Politics* 82(2):771–775.

Granato, Jim, Melody Lo and MC Sunny Wong. 2010. "A framework for unifying formal and empirical analysis." *American Journal of Political Science* 54(3):783–797.

Grofman, Bernard, Lisa Handley and Richard G Niemi. 1992. *Minority representation and the quest for voting equality*. Cambridge University Press.

Grose, Christian R. 2011. *Congress in black and white: Race and representation in Washington and at home*. Cambridge University Press.

Herron, Michael C. 1999. "Postestimation uncertainty in limited dependent variable models." *Political Analysis* 8(1):83–98.

Juenke, Eric Gonzalez. 2014. "Ignorance is bias: The effect of Latino Losers on models of Latino representation." *American Journal of Political Science* 58(3):593–603.

Juenke, Eric Gonzalez and Paru Shah. 2015. "Not the usual story: the effect of candidate supply on models of Latino descriptive representation." *Politics, Groups, and Identities* 3(3):438–453.

Juenke, Eric Gonzalez and Paru Shah. 2016. “Demand and supply: Racial and ethnic minority candidates in White districts.” *Journal of Race, Ethnicity and Politics* 1(1):60–90.

Li, Yuhui and Matthew S Shugart. 2016. “The seat product model of the effective number of parties: A case for applied political science.” *Electoral Studies* 41:23–34.

Lublin, David. 1999. *The paradox of representation: Racial gerrymandering and minority interests in Congress*. Princeton University Press.

Lublin, David, Thomas L Brunell, Bernard Grofman and Lisa Handley. 2009. “Has the Voting Rights Act Outlived Its Usefulness? In a Word,No.” *Legislative Studies Quarterly* 34(4):525–553.

Maisel, L Sandy and Walter J Stone. 1997. “Determinants of candidate emergence in US house elections: An exploratory study.” *Legislative Studies Quarterly* pp. 79–96.

Marschall, Melissa J, Anirudh VS Ruhil and Paru R Shah. 2010. “The new racial calculus: Electoral institutions and black representation in local legislatures.” *American Journal of Political Science* 54(1):107–124.

Reed, Steven R. 1990. “Structure and behaviour: Extending Duverger’s Law to the Japanese case.” *British Journal of Political Science* 20(3):335–356.

Shah, Paru. 2014. “It takes a Black candidate: A supply-side theory of minority representation.” *Political Research Quarterly* 67(2):266–279.

Shah, Paru. 2017. “Racial change, racial threat, and minority representation in cities.” *Urban Affairs Review* p. 1078087417737386.

Shugart, Matthew S and Rein Taagepera. 2017. *Votes from seats: Logical models of electoral systems*. Cambridge University Press.

Swain, Carol Miller. 1993. *Black faces, black interests: The representation of African Americans in Congress*. Harvard University Press Cambridge, MA.

Taagepera, Rein. 2007. *Predicting party sizes: The logic of simple electoral systems*. Oxford University Press on Demand.

Taagepera, Rein. 2008. *Making social sciences more scientific: The need for predictive models*. OUP Oxford.

Taagepera, Rein and Miroslav Nemčok. 2019. “Majoritarian and consensual democracies: The seat product connection.” *Party Politics* p. 1354068819871441.

Trounstine, Jessica and Melody E Valdini. 2008. “The context matters: The effects of single-member versus at-large districts on city council diversity.” *American Journal of Political Science* 52(3):554–569.

Online Appendix

For “Unifying the Demand and Supply-Side Theories of Minority Representation: A Logical Model”

Table of Contents

A Derivation of the Logical Model

In this section, I detail the derivation of the logical model of minority candidate emergence Pr$(E) = \widehat{P} = \mathbf{F}\Big((MC)^{1/2} - 50\Big)$. This section also accompanies with the proof of the racial margin of victory.

A.1 Candidate Entry

Based on Assumptions 1-4, I consider the logical model of minority candidate emergence. Following the literature of political ambition (Black, 1972; Lazarus, 2008; Aldrich, 1995; Jacobson and Kernell, 1983; Stone, Maisel and Maestas, 2004), I start by applying the following model of candidate entry to minority candidate emergence:

$$u_{it} = \widehat{P}_{it}B_{it} - C_{it} \tag{1}$$

Here, u_{it} is the utility that candidate i obtains from running for office, $\widehat{P}_{it}$ is the candidate's estimate of the probability of winning if she enters the race, B_{it} is the benefits of holding the office, and C_{it} is the cost of running, which is greater than zero, all at time t.[1] Black (1972) suggests that a rational office-seeking candidate enters the race if and only if $\widehat{P}_{it}B_{it} > C_{it} > 0$ and $u_{it} > u_{it}(A)$, where $u_{it}(A)$ expresses the utility that the candidate receives when staying out.[2]

While this model looks fairly simple at first glance, it is in fact a very complicated model where four variables and three inequalities collectively determine the outcome. As we often prefer parsimonious models to complex models in logical model building (Taagepera, 2008), here, I only focus on the win probability term by taking other terms out of the equation.[3] Consequently, I directly model the probability of entry E_{it} as the estimated probability of winning:

$$\text{Pr}(E_{it}) = \widehat{P}_{it} \tag{2}$$

Substantively, this implies that minority candidate emergence can be predicted by knowing about $\widehat{P}_{it}$, that is, how likely it is for minority candidates to win when they enter the race. The argument that I am making here is that *all things being equal* minority candidate emergence can be solely predicted by such estimated probability of winning. The second half of model building then involves information, with which minority candidates estimate such probability.

[1] In contrast to previous studies, I use the notation $\widehat{P}_{it}$ (as opposed to P_{it}) to make it clear that it is an estimate of the true probability of winning.

[2] Lazarus (2008) discusses that $u_{it}(A)$ is usually very different for experienced politicians and amateurs, which affect the required level of $\widehat{P}_{it}$ in the model; however, I do not distinguish the two types of politicians here.

[3] The resulting model is also obtained by assuming that $B_{it} = 1, C_{it} = 0.5$, and $u_{it}(A) \sim N(0, \sigma^2)$, where $\sigma^2 > 0$. Under these assumptions, the original model tells that minority candidates run if and only if $\widehat{P}_{it} > 0.5 + u_{it}(A)$. Since 0.5 and u_{it} are constant and random numbers, we can say that the probability of entry is a function of the win probability term as Pr$(E_{it}) = \widehat{P}_{it} + u'_{it}$, where u'_{it} combines the constant and the error term. If we ignore the random component, we obtain Equation (2). While the model can be presented both in a deterministic form (without the error term) and stochastic form (with the error term), I choose to use the deterministic form due to its simplicity.

A.2 Racial Margin of Victory

How can minority candidates calculate $\widehat{P}_{it}$? Here, I argue that the probability of winning is a function of what I call the *racial margin of victory*. In order to elaborate what it means, let me first introduce the concept of the margin of victory as follows:

DEFINITION 1 (Margin of Victory): *The margin of victory is the minimum number of vote shares that have to be modified in order to change the outcome (i.e., winner) of an election.*

In first-past-the-post (FPTP) elections, the margin of victory is computed as half the difference in vote shares between the winner and the runner-up. For example, when I have the winner who received 60% and the runner-up who obtained 40% of the total ballots, the margin of victory is $\frac{1}{2} * (60 - 40) = 10$. This means that if I remove 10% of the total ballots from the winner and allocate them to the runner-up, the election will be a tie (50% ballots vs. 50% ballots) and thus such modification changes the outcome of the election.[4]

Building on this concept, I then introduce the "racial" margin of victory as follows:

DEFINITION 2 (Racial Margin of Victory): *The racial margin of victory is the minimum number of vote shares that have to be modified in order to change the race of the winner of an election.*

To contextualize my argument, let me define African American voters as the minority group of interest and white voters as the majority group of reference. In FPTP elections, then, the racial margin of victory (assuming no ties and rounding the difference) can be computed as:

$$\text{Racial Margin of Victory} = \left| \frac{V^B - V^W}{2} \right|,$$

where V^B represents the share of **v**otes received by the top **b**lack candidate (in percentage), V^W represents the share of **v**otes received by the top **w**hite candidate, and $|\cdot|$ is an absolute value operator. In other words, the racial margin of victory is calculated as half the absolute difference in the vote share received by the top black candidates and the vote share received by the top white candidates.[5]

An important property of the racial margin of victory is that it does not distinguish the direction of racial change. Put differently, the concept is agnostic about whether the new winner is black or white after modifying the vote shares. Removing the absolute value operator from it, however, enables us to calculate the minimum number of ballots that need to be modified in order to *replace the white winner with a black winner*. I call it as a *signed racial margin of victory* and define it as follows:

[4] To compute the "tie-free" margin, one can simply add 1 to the above number (Xia, 2012; Magrino et al., 2011), while I assume that there is always a tie breaker in my argument as discussed above.

[5] Here, I do not consider the presence of run-off elections and assume that every election is a decisive election. Surprisingly, this measurement only depends on the ballots received by the top black and top white candidates and is not a function of other factors such as the number of black and white candidates, the internal distribution of ballots within the same racial group, and relative advantages of black runner-up to white runner-up. In Online Appendix A.4, I consider all eight possible scenarios (i.e., from one black candidate with one white candidate to multiple black candidates with no white candidate) and demonstrate that only vote shares of the top black and white candidates are required to calculate the racial margin of victory.

$$\text{Signed Racial Margin of Victory} = \frac{V^B - V^W}{2}$$

Substantively, the signed racial margin of victory represents the *extent to which minority candidates safely secure their descriptive representation relative to their white counterparts*. For simplicity, I hereafter use the term "margin of victory" to refer to the *signed* racial margin of victory.

Now, how is the (signed) racial margin of victory tied back to the argument of minority candidate entry? The idea is here that, under Assumptions 1-4, the probability of winning can be expressed as a deterministic function of the racial margin of victory. Thus, if black candidates are fully short-term instrumental with complete information, the probability of winning becomes a function of the vote shares of the top black and white candidates in *upcoming* elections (i.e., at time t):

$$P_{it} = \mathbb{1}\left(\frac{V_{it}^B - V_{it}^W}{2} > 0\right) \tag{3}$$

Here, $\mathbb{1}$ is an indicator function that takes 1 if the inside condition is satisfied and 0 otherwise. Substantively, this means that the probability of winning for the most viable black candidate is 1 (i.e., winning with certainty) if the top black candidate obtains higher vote shares (or one more vote) than the top white candidate and 0 (i.e., losing with certainty) otherwise.[6] Hence, **if such future vote shares are *ex ante* known, the probability of winning is a deterministic function of the racial margin of victory, and the rational minority candidate decides to run (not to run) for office with certainty**.[7]

A.3 Two Limits for the Racial Margin of Victory

Although the above argument makes logical sense, it is usually impossible for any candidate to perfectly calculate future vote shares due to the cognitive burden, lack of enough information, and time restriction. Thus, I instead argue that the probability of winning is, in practice, calculated based on an educated guess about future vote shares. I then assume the following *estimated* probability of winning:

$$\widehat{P}_{it} = \mathbf{F}\left(\frac{\widehat{V}_{it}^B - \widehat{V}_{it}^W}{2}\right), \tag{4}$$

where the "hat symbols" ($\widehat{\cdot}$) imply that the probability and vote shares are *estimated* quantities as opposed to known values. Here, **F** represents a cumulative distribution function (CDF) of any elliptical distribution (e.g., normal and t-distributions) which maps the input values onto the probabilistic space (values between 0 and 1). One example of **F** is a CDF of a normal distribution (called a probit function) with mean 0 and standard deviation 15, which maps the values between about -50 and 50 onto the probabilistic scale (i.e., numbers between 0 and 1). Here, the standard deviation controls the degree of non-zero probability that is assigned to values which are far away from the mean. *Substantively, it represents how much candidates*

[6] Here, I assume that there is always a tie-breaker and do not consider the case where the two quantities are the exactly same.

[7] Note that the argument presented so far is not surprising at all; rather, it simply clarifies a well known rational choice-type explanation that candidates decide to run for office when they see higher chances of winning. And if they can foreseen their electoral fortune, they will either enter the race or abstain. However, logical model building requires a clear definition and measurement of each concept and this section serves this purpose by connecting minority candidate emergence with a well discussed notion of the margin of victory in political science.

would allow for the possibility of miscalculating the racial margin of victory.[8]

Next, I theorize that in order for black candidates, as for any candidates, to estimate the racial margin of victory they rely on **two limits (extreme cases)**, which include (1) the district racial composition (as an expected margin of victory) and (2) the racial margin of victory in the last elections. In other words, the probability of winning in upcoming elections can be inferred from "what the district looks like now" and "what happened last time" in there. Let me explain what this means.

First, the district racial composition is informative because it can be directly translated to the expected difference in vote shares under the assumption of "perfect co-ethnic voting" (all voters vote for the candidate from their racial group). For example, if a black candidate is running against a white candidate in a district with 60% black and 40% white voters, the expected vote share will become 60% and 40% for the black and white candidates, respectively; and the "best guess" for half the difference in vote shares in the FTTP elections becomes 10% points (under the perfect co-ethnic voting assumption). Here, I call this quantity as an *expected racial margin of victory* (based on the district racial composition). This quantity is obtained by %black voters -50.

Second, the racial margin of victory in the last elections is instructive since it exactly tells how much black candidates can solicit crossover votes from white voters, while reserving co-ethnic votes from black voters, *if* the political climate remains the same from the last elections. Put differently, the past racial margin of victory quantifies a potential deviation from the perfect co-ethnic voting assumption. It must be highlighted that the racial margin of victory in the last election is a known (and thus measurable) quantity, while the racial margin of victory in an upcoming election is unknown (because nobody cannot see the future). To simplify the terminology, however, **I will simply call this quantity as the racial margin of victory (at time $t-1$) unless otherwise noted**.

Now, I theorize that the future racial margin of victory must be located somewhere between the two limit cases (i.e., the expected and past racial margin of victories). I formalize this argument by taking the "geometric mean" of the two quantities and replace the speculated racial margin of victory at time t in Equation (4) with the geometric mean. Let $\tilde{C}$ denote the *expected* racial margin of victory (based on the district racial **C**omposition). Let $\tilde{M}$ denote the racial **M**argin of victory. Without changing their substantive meaning, I then shift both quantities by adding 50 to avoid the multiplication of negative numbers as follows:

$$M = \tilde{M} + 50 \tag{5}$$

$$C = \tilde{C} + 50 \tag{6}$$

Here, the "tilde symbols" ($\tilde{.}$) indicate that these are unadjusted (or raw) quantities. From now on, I call these adjusted quantities M and C simply as the racial margin of victory and the expected racial margin of victory, respectively unless otherwise noted.

Finally, connecting Equation (2) and Equations (4)-(6) yields:

$$\Pr(E) = \widehat{P} = \mathbf{F}\Big((MC)^{1/2} - 50\Big), \tag{7}$$

where $(MC)^{1/2}$ is the geometric mean of the two quantities and -50 is there to account for the adjustment

[8] My substantive argument does not change regardless of scaling as long as the function **F** has an appropriate tuning parameter as discussed below.

shown in Equations (5)-(6). The purpose of this re-adjustment is to centralize the argument of **F** with mean 0 so that its scale does not change from Equation (4). Substantively, this ensures that positive (negative) numbers inside **F** mean higher (lower) probabilities of minority candidate emergence.

Here, $\tilde{M} = \frac{1}{2}(V_{t-1}^B - V_{t-1}^W)$, where V_{t-1}^B and V_{t-1}^M are the vote shares obtained by the top black and white candidates, respectively at time $t-1$. $\tilde{C} = \%$black voters $- 50$ (at time t). Applying the adjustments in Equations (5)-(6) yields:

$$M = \frac{1}{2}\left(V_{t-1}^B - V_{t-1}^W\right) + 50 \tag{8}$$

$$C = \%\text{black voters} \tag{9}$$

Thus, the adjusted expected racial margin of victory is algebraically equivalent to the percentage of black voters.

A.4 Proof of the Racial Margin of Victory

Here, I examine all (eight) possible different patterns of biracial elections and demonstrate that the racial margin of victory introduced in Online Appendix A.2 can be applied to any type of FPTP elections. Recall that the racial margin of victory refers to the minimum amount of vote shares one needs to modify in order to change the race of the winner in an election. Suppose that black voters are the minority group of interest and white voters are the majority group of reference. Let B denote the top black candidate and W denote the top white candidate on a ballot. Let b denote any other black candidate and w denote any other white candidate on the ballot. Let V^B and V^W be the vote shares obtained by B and W, respectively. Then, I show that the racial margin of victory is $\frac{1}{2}|V^B - V^W|$ regardless of the number and order of black and white candidates.

A.4.1 One black candidate & one white candidate

When there are only one black candidate and one white candidate, all the possible patterns of electoral results can be obtained by the following permutations.

1st	W	B
2nd	B	W

Here, 1st means the candidate in the 1st position (i.e., winner) and 2nd refers to the candidate in the 2nd position (i.e., runner-up) (and so on). In order to change the race of the winner, one needs to change the order of the winner and the runner-up. Thus, the racial margin of victory is $\frac{1}{2}|(V^B - V^W)|$.

A.4.2 One black candidate & multiple whites candidates

When there are one black candidate and two white candidates, all the possible patterns of electoral results are as follows.

1st	W	W	B
2nd	w	B	W
3rd	B	w	w

Again, to change the race of the winner, one only needs to consider the vote shares obtained by the (only) black and the top white candidates. Thus, the racial margin of victory is $\frac{1}{2}|(V^B - V^W)|$. Since this fact does not change even when there are more than two white candidates, this applies to a more general case of one black and multiple white candidates.

A.4.3 Multiple blacks candidates & one white candidate

In elections with two black candidates and only one white candidate, all possible orderings of candidates are as follows:

1st	W	B	B
2nd	B	W	b
3rd	b	b	W

Following the same logic as A.4.2, the racial margin of victory is calculated as $\frac{1}{2}|(V^B - V^W)|$. Since this fact does not change even when there are more than two black candidates, this applies to a more general case of multiple black and one white candidates.

A.4.4 Multiple black candidates & multiple white candidates

When there are two black and two white candidates on a ballot, the following covers all the possible electoral outcomes.

1st	W	W	W	B	B	B
2nd	w	B	B	W	W	b
3rd	B	w	b	b	w	W
4th	b	b	w	w	b	w

Here, it requires only $\frac{1}{2}|(V^B - V^W)|$ to change the race of the winner. This can be extended to elections with more than two black or white candidates, and thus it applies to a more general case of multiple black and multiple white candidates.

A.4.5 No black candidate & one white candidate

When elections are unopposed with a white winner, only possible pattern is as follows:

1st	W

Now, this can be considered as a special case of A.4.1. (One black candidate & one white candidate), where the black candidate received zero vote. Thus,

1st	W	$(V^W = 100)$
2nd	B	$(V^B = 0)$

Consequently, the racial margin of vote can be computed as $\frac{1}{2}|(V^B - V^W)| = \frac{1}{2}| - V^W| = 50$.

A.4.6 No black candidate & multiple white candidates

When there are multiple white candidates with no black candidate, all possible electoral outcomes can be demonstrated as follows:

1st	W
2nd	w
⋮	⋮

Now, this can be thought of as a special case of A.4.2. (one black candidate & multiple white candidates) where the black candidate received zero vote. Thus,

1st	W	(V^W)
2nd	w	
⋮	⋮	
Last	B	$(V^B = 0)$

As a result, the racial margin of vote is $\frac{1}{2}|(V^B - V^W)| = \frac{1}{2}| - V^W|$.

A.4.7 One black candidate & no white candidate

When elections are unopposed with a white winner, only possible pattern is as follows:

1st	B

Now, this can be considered as a special case of A.4.1. (One black candidate & one white candidate), where the white candidate received zero vote. Thus,

1st	B	$(V^B = 100)$
2nd	W	$(V^W = 0)$

Consequently, the racial margin of victory can be computed as $\frac{1}{2}|(V^B - V^W)| = \frac{1}{2}|V^B| = 50$.

A.4.8 Multiple black candidates, no white candidate

When there are multiple white candidates with no black candidate, all possible electoral outcomes can be demonstrated as follows:

1st	B
2nd	b
⋮	⋮

Now, this can be thought of as a special case of A.4.3. (Multiple black candidates & one white candidate) where the white candidate received zero vote. Thus,

1st	B	(V^B)
2nd	b	
⋮	⋮	
Last	W	$(V^W = 0)$

As a result, the racial margin of victory is $\frac{1}{2}|(V^B - V^W)| = \frac{1}{2}|V^B|$.

A.5 Additional Discussion on the Racial Margin of Victory

Some readers may find the above discussion about elections with no black (minority) candidate hard to digest. How can the absence of minority candidates be informative in terms of future electoral odds for minority candidates? In fact, this question is very important and contains a critical element to understand the proposed logical model. The key to understand this point is that minority representation is partly determined by the "strategic coordination" both *among minority candidates* and *among white candidates*.

First, recall that the logical model describes elections as political competitions between minority and majority groups. Recall also that the racial margin of victory is defined as the minimum number of vote shares that one needs to modify in order to change the "race" of the winner. As reflected in these assumption and definition, the logical model only considers whether the winner would be a minority politician or a white politician.

Next, the racial margin of victory quantifies how safely minority candidates secure seats *relative to their white counterparts*. The part "relative to their white counterparts" is critical to understand the concept of the racial margin of victory. This part indicates that the racial margin of victory is *not* about how many votes minority candidates "collectively" received independent of white candidates, but about how many votes the minority candidate with the highest chance of winning received compared to the white candidate with the highest chance of winning. This implies that the racial margin of victory reflects how well both minority candidates and white candidates coordinate among themselves so that there would be *less vote splitting*.

To illustrate this point, let me consider the following two extreme scenarios. Suppose that there are only 100 voters and they all turned out. In the first scenario, there was only one white candidate and no black candidate. By definition, the white candidate received 100 votes and won, whereas the "black candidate" received 0 vote and lost. Based on this result, the racial margin of victory in this election is $\frac{1}{2}(0 - 100) = -50$. This means that if a black candidate were running for office in this election she would have won if she had received at least 50+1 votes.

In the second scenario, there were eleven white candidates and no black candidate, and the winner (white politician) received 10 votes and the other ten white candidates equally received 9 votes. This is an extreme example for vote splitting (and the lack of strategic coordination). Based on this result, the racial margin of victory is $\frac{1}{2}(0-10) = -5$. This implies that if a black candidate were running for office in this election she would have won if she had received at least 5+1 votes.

These two scenarios show that, even though both elections featured no black candidate, the implications for how competitive these elections were (in terms of minority descriptive representation) are completely different. In the first case, the black candidate must receive more than 51 votes (50% of the total ballots + 1) to get elected if the same election happens again. In the second case, however, she must receive only 6 votes (5% of the total ballots + 1) to be the winner if the upcoming election looks exactly the same as the last election. Thus, it would be natural to consider that there is a higher chance for minority candidates to secure their seats relative to their white counterparts in the second election (with a worse strategic coordination among white candidates) than in the first contest (with a better strategic coordination among white candidates).

Although I have focused on the coordination among white candidates so far, the same reasoning applies to the coordination among minority candidates. If there are multiple minority candidates who split a set of ballots that could have been concentrated on a single minority candidate, one could argue that the top minority candidate did poorly relative to her white counterparts. This illustrates the code idea of the racial margin of victory and the logical model: previous elections provide information about how well the minority candidate with the highest chance of winning did compared to the white candidate with the highest chance of winning, which are a function of both individual performance and group coordination.

Two other points are worth emphasizing here. First, while discussing the above point, I have not discussed the district racial composition at all. This is because I have only concerned with how many votes both the top minority and the top white candidates received, and *not who voted for these candidates*. In computing the racial margin of victory, indeed, who voted for the top minority or the top white candidate provides no information. This does not mean, however, that the district racial composition is not correlated with the racial margin of victory at all. One could imagine that, in the absolute sense, the top white candidate might receive fewer votes relative to the top minority candidate in a majority minority district, and vise versa, due to the presence of the moderate or strong level of co-ethnic voting. Nevertheless, the top minority candidate's vote share is also affected by the level of strategic coordination among minority candidates as seen above. This suggests that while the district racial composition may influence the sign and magnitude of the racial margin of victory, it does not have a deterministic relationship with it. In fact, it is C that quantifies the racial margin of victory when it is a deterministic function of the district racial composition.

Second, the logical model does not directly talk about other seemingly important information such as the district "partisan" composition and the incumbency status. The model does not directly consider these elements because, again, they will be well summarized into the racial margin of victory. For example, when the district is heavily white and democratic and the majority of voters voted for the top minority candidate (who eventually won the race), the racial margin of victory could be close to the maximum value (50). In fact, the racial margin of victory is more informative than the district partisan composition *per se* because the former contains information about how actual elections went rather than how they would go as speculated by the district partisan composition.

To conclude, the racial margin of victory is a novel concept that summarizes and quantifies multiple important information about electoral performance of both minority candidates and white candidates, district racial and partisan composition, incumbency status, and strategic coordination among minority candidates and among white candidates.

B Louisiana Mayoral Elections

This section provides various descriptive statistics for Louisiana mayoral elections.

B.1 Distributions of M and C

In the Louisiana mayoral election data, the racial margin of victory is distributed somewhat irregularly, while the percentage of black voters seems to follow an exponential-type distribution. The Pearson's correlation coefficient between the two variables is 0.64, which can be visually confirmed as a strong positive correlation in the right panel below.

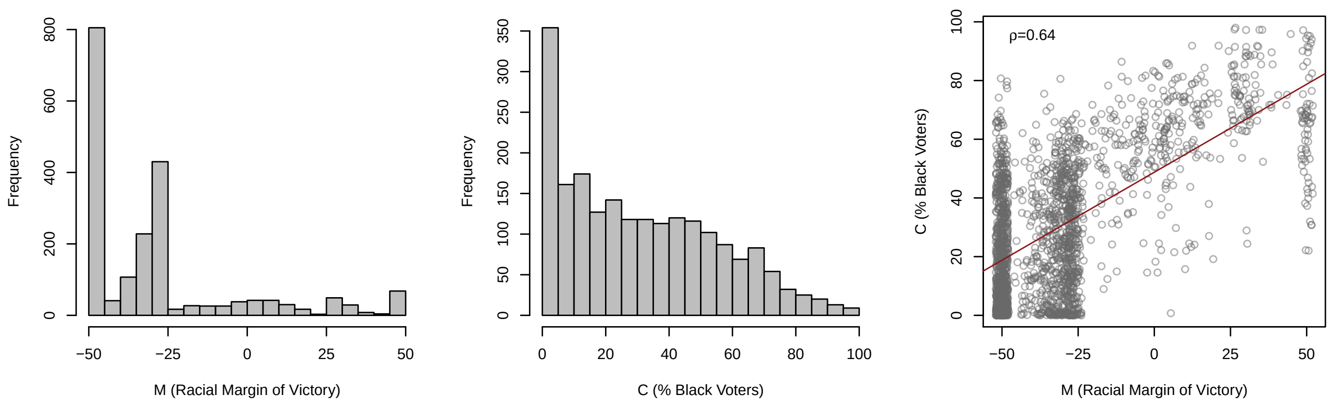

Figure B.1: **The Distributions of M and C with the Scatter Plot**
Note: The x-axis is jittered in the scatter plot.

B.2 Stability of the Racial Regime

To code the racial regime, I use the cutoff points of 40% and 65% (for the average black VAP). Figure B.2 shows the stability of such coding scheme. It portrays that very few municipalities "cross" the cutoff points within the observed periods.

B.3 Overtime Variation of Black Candidate Emergence

To investigate whether there is any temporal pattern in black candidate emergence, Figure B.3 plots the proportion of elections with black candidates over time. It also accounts for the racial regime. I find that there is almost no temporal pattern except for the fact that black candidate emergence has been slightly more common in racially balanced districts over time.

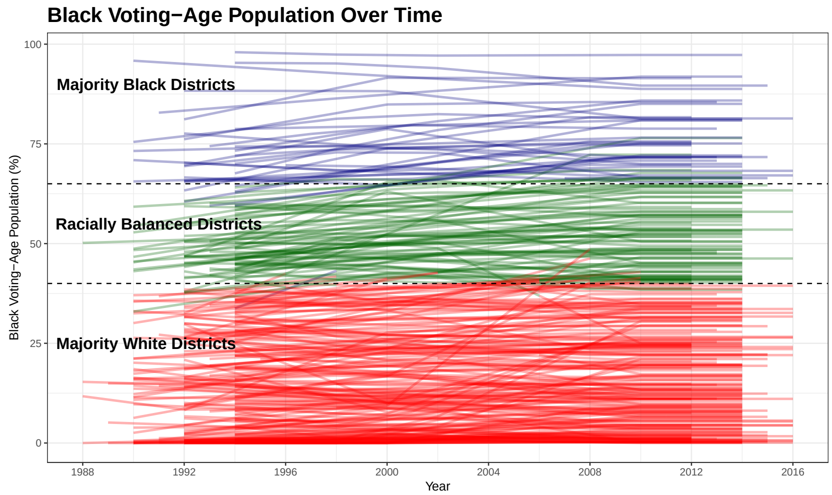

Figure B.2: **The changes in black voting-age population over time by racial regime**
Note: This graph visualizes the changes in black voting-age population for all districts over time as well as their corresponding racial regime. It demonstrates that very few districts experience a dramatic demographic change such that the percentage of black voting-age population crosses the boundaries of racial regime (i.e., 40% and 65%).

Figure B.3: **Overtime Variation of Black Candidate Emergence**
Note: This graph portrays the proportion of elections with black candidates over time by the racial regime.

B.4 Overtime Variation of the Racial Margin of Victory

To study whether there is any temporal pattern in the racial margin of victory, Figure B.4 plots the average values of the racial margin of victory over time. It also accounts for the racial regime. The figure illustrates that there is a clear temporal pattern for majority minority districts (black candidates have increasingly performed better relative to their white counterparts over time), whereas no such pattern can be observed for racially balanced and majority white districts.

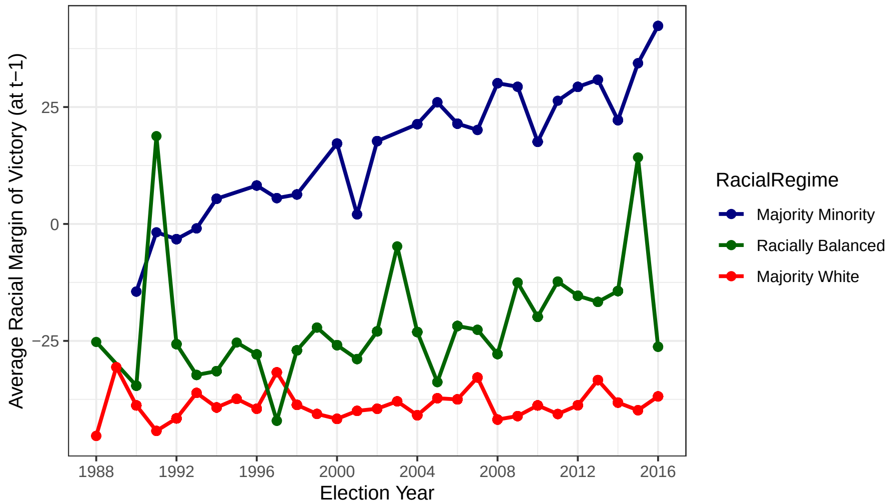

Figure B.4: **Overtime Variation of the Racial Margin of Victory**
Note: This graph portrays the proportion of elections with black candidates over time by the racial regime.

B.5 Registration by Race

In the State of Louisiana, the proportion of registered voters who identify themselves as neither white or black ranges from about 0.04 to 0.05 (Figure B.5). This implies that Louisiana mayoral elections are appropriate cases for validating the logical model based on Assumption 1.

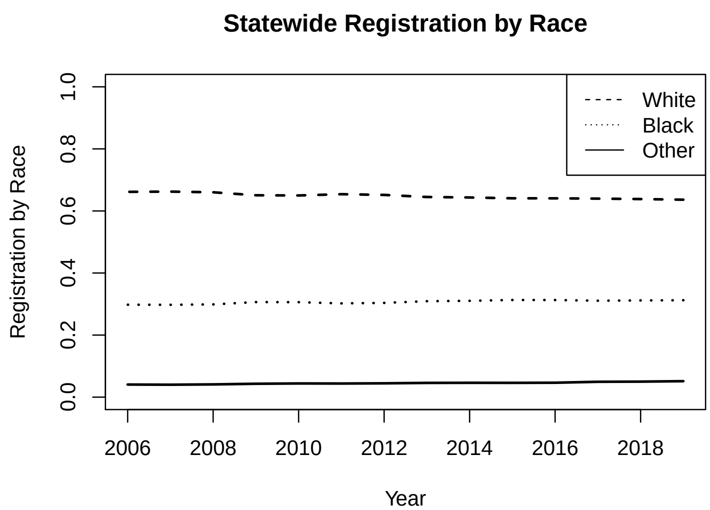

Figure B.5: **Registered Voters by Self-Reported Race**
Note: This plot represents the statewide voter registration record by self-reported race in Louisiana from 2006 to 2019. The data was collected from Louisiana Secretary of State website. It demonstrates that the proportion of registered voters who identify themselves neither as white or black ranges from about 0.0401 to 0.0515, giving a justification for considering non-partisan mayoral elections as biracial elections.

C Additional Findings

This section reports the results of additional analyses that support the claim in the article.

C.1 Unopposed Elections and Open Races

In this section, I replicate my analysis of predictive performance in Section 4 by limiting the sample to opposed elections (not-unopposed elections) and open races (where incumbents are not running). Among 2037 elections, 1167 (57.3%) are unopposed and 584 (28.7%) feature incumbents. Figure C.1 visualizes the distribution of opposed elections and open races by the racial regime. In Figure C.2 and Table C.1, I visualize the model predictions over observed data points and report the predictive performance in ePCP for opposed elections. For open races, I present the same results in Figure C.3 and Table C.2. Taken together, these evidence further validate the logical model.

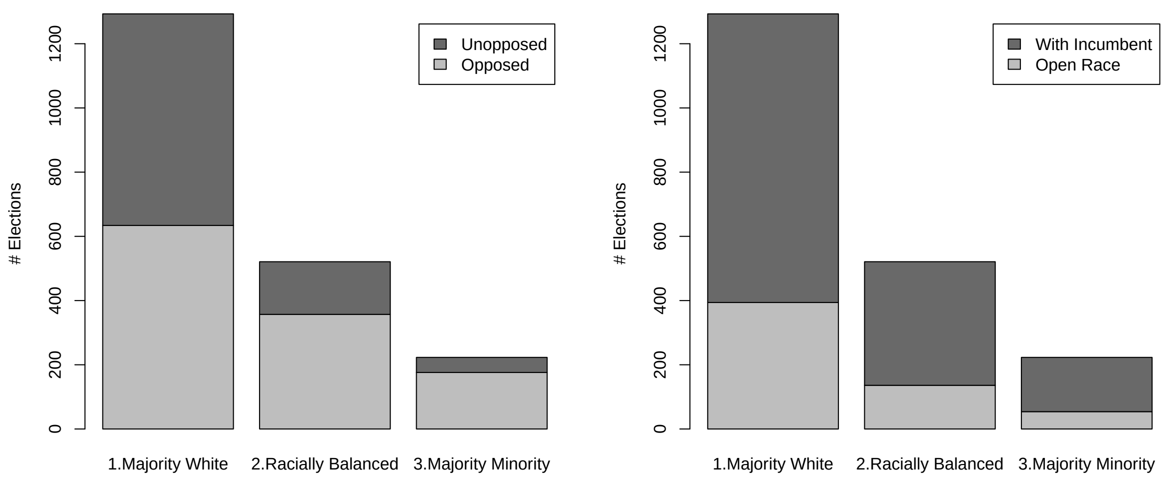

Figure C.1: **Frequency of Opposed Elections and Open Races**
Note: This figure portrays the distribution of opposed elections and open races.

	In-sample Prediction (%)			Out-of-sample Prediction (%)		
	Logical Model	LPM	Logit	Logical Model	LPM	Logit
All Districts	**82.0**	78.1	79.6	**84.3**	80.2	81.2
Majority White Districts	**89.8**	83.9	84.6	**91.4**	85.5	85.4
Racially Balanced Districts	**64.9**	63.2	63.5	**68.1**	62.7	63.0
Majority Minority Districts	**89.1**	96.1	97.1	**89.3**	94.0	95.0

Table C.1: **Predictive Performance of the Logical Model (Only Opposed Elections)**
Note: This table reports the weighted percentage of observations for which the model correctly predicts their values (ePCP), where the weight reflects the distance between the true value and predicted value.

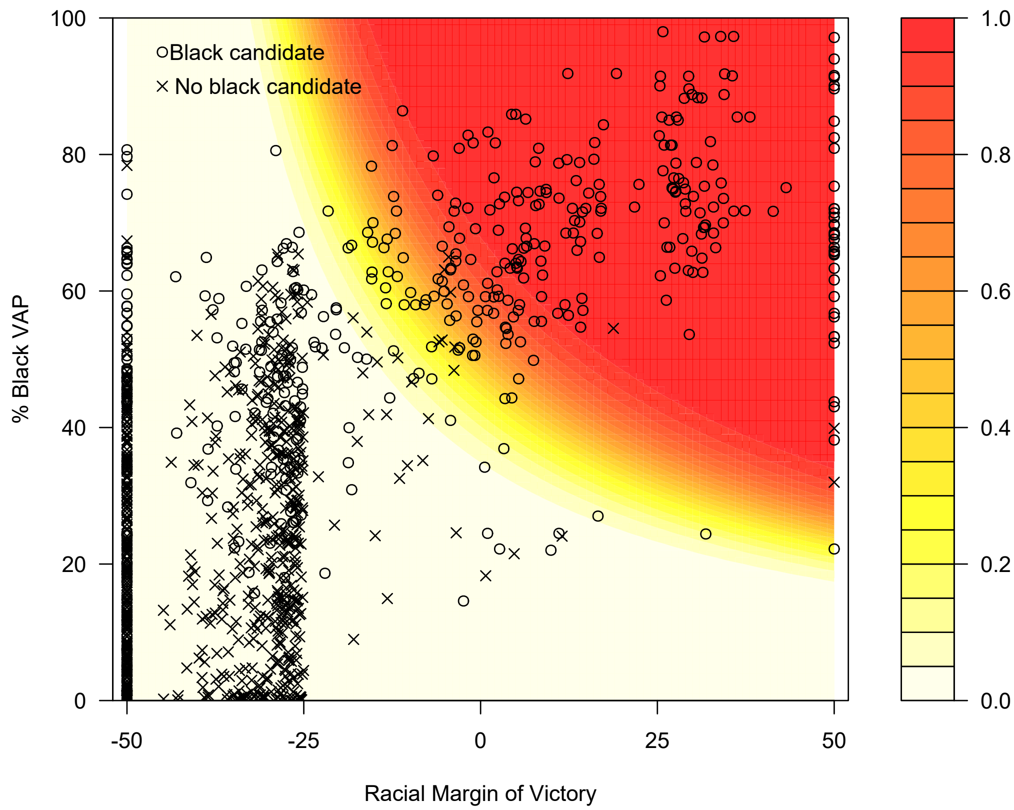

Figure C.2: **Model Prediction of Black Candidate Emergence (Excluding Unopposd Elections)**
Note: This figure visualizes the predicted probabilities of black candidate emergence as a function of racial electoral performance and % black VAP in a contour plot. Elections with black candidates are shown as open circles (○), whereas elections without black candidate are represented by crosses (×). Elections where incumbents were unopposed are excluded.

	In-sample Prediction (%)			Out-of-sample Prediction (%)		
	Logical Model	LPM	Logit	Logical Model	LPM	Logit
All Districts	**84.4**	80.4	81.6	**81.6**	78.9	80.5
Majority White Districts	**91.6**	86.8	87.0	**90.9**	85.2	85.2
Racially Balanced Districts	**60.7**	60.0	60.0	**57.3**	60.0	60.0
Majority Minority Districts	**91.5**	94.5	96.5	**92.2**	91.4	91.8

Table C.2: **Predictive Performance of the Logical Model (Only Open Races)**
Note: This table reports the weighted percentage of observations for which the model correctly predicts their values (ePCP), where the weight reflects the distance between the true value and predicted value.

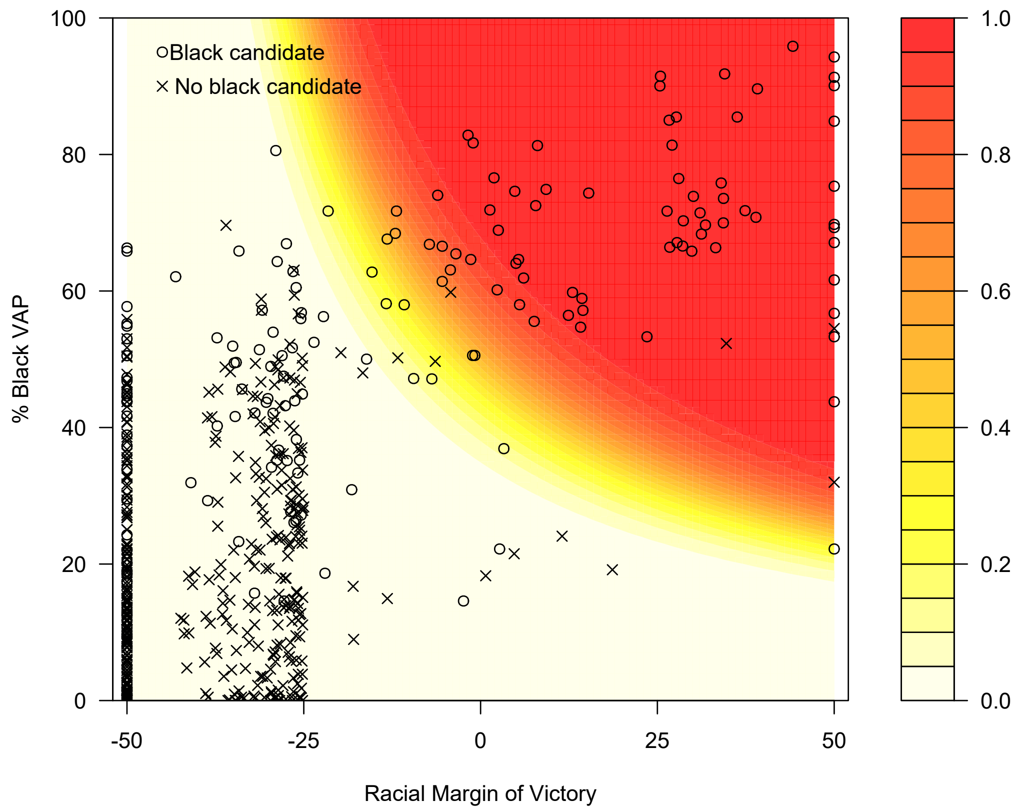

Figure C.3: **Model Prediction of Black Candidate Emergence (Only Open Races)**
Note: This figure visualizes the predicted probabilities of black candidate emergence as a function of racial electoral performance and % black VAP in a contour plot. Elections with black candidates are shown as open circles (○), whereas elections without black candidate are represented by crosses (×). Elections where incumbents were running for re-election are excluded.

C.2 Count Model Regression Results

To buttress the claim that the racial margin of victory (and thus the logical model) does not predict the number of black candidates running for office, I run a series of count regressions (count models) with the number of black candidates as the dependent variable. Figure C.3 report the regression results. In Regressions 3-4, I dropped two observations which have 10 and 11 black candidates as outliers. I find that regardless of model specification the racial margin of victory has no statistically significant association with the number of black candidates (at the 0.05 level).

	Regression 1	Regression 2	Regression 3	Regression 4
Intercepts	-.238	-.238	-.293	-.293
	(.145)	(.145)	(.147)	(.147)
Racial Margin	.002	.002	.002	.002
	(.001)	(.001)	(.001)	(.001)
% Blacks	.012	.012	.013	.013
	(.002)	(.002)	(.002)	(.002)
Specification	Poisson	Negative Binomial	Poisson	Negative Binomial
Outliers			dropped	dropped
N	526	526	524	524

Table C.3: **Regression Results**
Note: This table reports the results of count regressions. Regardless of model specification, the racial margin of victory does not have a statistically significant association with the number of black candidates.

C.3 Model Prediction with Placebo Outcomes

To bolster the above findings, I performed a placebo test where female candidate emergence was used as a false outcome. I found that the ePCP for female candidate emergence is **51.9**. This is a remarkable result because an unbised coin flip would yield 50. This implies that the logical model does not do any better than a coin flip in terms of predicting the emergence of female candidate, which it is what we expect if the logical model actually is a model for (racial) *minority* candidate emergence.

C.4 Extended Regression Results

To further highlight the predictive power of the logical model, I compare the ePCPs obtained from the logical model with those from LPM and logistic regressions with 34 additional variables. These variables include the percentage of blacks and whites with B.A., a dummy variable for election cycle, a dummy variable for open races, the percentage of whites who are 65 and over, a measure for human density, dummy variables for election yaers, and dummy variables for municipalities.

Table C.4 shows that although these heavily parameterized regressions have higher predictive power for racially balanced and majority minority districts **in-sample**, their predictions got significantly worse **out-of-sample**. This is because regressions with such large numbers of covariates overfit the data (fit well in this particular data, but has a lower generalizability). Thus, even with the help of additional variables, statistical models cannot dominate the logical model in terms of predictive power.

C.5 Regression Analysis

The above analysis demonstrates that, *all things being equal*, the logical model can predict minority candidate emergence as a function of the two variables with high predictive performance. It is important to acknowledge, however, that such conditional statement with the logical model is somewhat different from conditional statements used in more conventional regression analysis and causal inference, respectively (Clark and Golder, 2015). Indeed, some readers may wonder if the observable implications from the logical

	In-sample Prediction (%)			Out-of-sample Prediction (%)		
	Logical Model	LPM	Logit	Logical Model	LPM	Logit
All Districts	**88.6**	84.3	87.0	**88.3**	83.7	86.1
Majority White Districts	**94.1**	90.7	91.8	**94.1**	89.9	90.4
Racially Balanced Districts	**73.7**	74.1	75.7	**75.6**	69.1	71.6
Majority Minority Districts	**90.9**	93.7	1	**92.5**	91.1	90.4

Table C.4: **Predictive Performance of the Logical Model (with 34 More Variables)**
Note: This table reports the weighted percentage of observations for which the model correctly predicts their values (ePCP), where the weight reflects the distance between the true value and predicted value.

model hold even after controlling for other variables as high predictive performance does not necessarily mean the presence of causal mechanisms (Breiman et al., 2001; Shmueli et al., 2010).

Here, I tackle this problem by defining the racial margin of victory as the independent variable of interest and black candidate emergence as the dependent variable in the mayoral election data.[9] I choose to focus on the racial margin of victory because it is a novel concept in the literature, whereas it is widely known that the proportion of minority voters is strongly correlated with minority candidate emergence and victory (e.g., Cameron, Epstein and O'halloran, 1996; Shah, 2014). I then run a series of Bayesian logistic regressions and the detailed findings are discussed in Appendix C.6.

The main finding is shown in Figure C.4. It plots the predicted (posterior) probabilities of minority candidate emergence against the racial margin of victory by the racial regime after controlling for several variables (based on Column 3 of Table C.5). Here, I calculate the predicted probabilities by setting all covariates to their median values for each racial regime. If the logical model has a valid explanatory power, the predicted probabilities must look like the right panel of Figure 1. I find that this is exactly the case.

The figure illustrates that the substantive "effect" of the racial margin of victory is larger in racially balanced districts than in majority white or majority minority districts. In racially balanced districts, the change in the racial margin of victory from its lowest to highest values seems to boost the predicted probability by about 0.8, whereas such change seems to be more modest in majority white (0.2) and majority minority (0.4) districts. This provides supportive evidence for the key observable implications of the model. I performed various types of robustness checks to examine the interval validity of the above analysis (Online Appendix C.6), finding that the substantive conclusion stays the same.

C.6 Bayesian Logistic Regressions with Robustness Checks

Here, I report the results of Bayesian logistic regressions with multiple robustness checks. I first consider the bivariate relationship between the racial margin of victory and minority candidate emergence using a simple Bayesian logistic regression. The results are shown in Regression 1 of Table C.5 (the posterior mean is shown with 95% credible intervals). The results represent that a higher racial margin of victory is associated with a greater chance of observing minority candidates. To account for a set of potential confounders which affect both the racial margin of victory (at time $t-1$) and minority candidate emergence

[9] Here, I employ a more conventional regression analysis rather than the causal inference approach based on the potential-outcomes framework (Hernan and Robins, 2010; Morgan and Winship, 2015). Indeed, integrating the logical model approach with the potential-outcomes framework is one of the promising areas of inquiry (but beyond the scope of this article) and I leave to future research the development of such a framework.

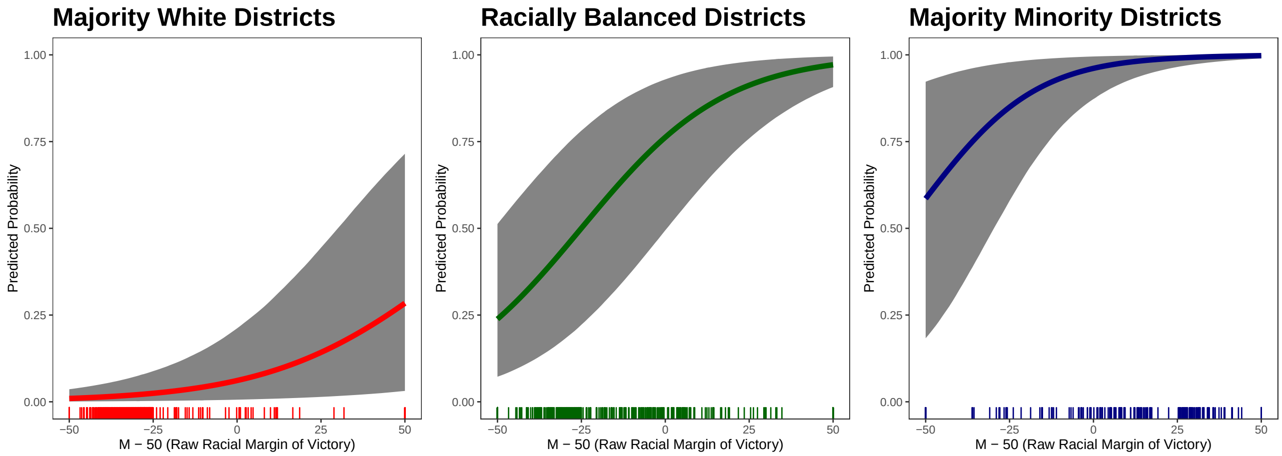

Figure C.4: Substantive Effects of the Racial Margin of Victory by the Racial Regime
This figure displays predicted probabilities (bold curves) that black candidates running for office in three racial regimes with 95% credible intervals (gray shades). Rug plots show the empirical distributions of the racial margin of victory. Predicted probabilities are computed based on typical covariate values for each type of district (i.e., each regime has a different baseline).

(at time t), I next include several covariates including the proportion of black VAP, the proportion of black and white population with college degrees, and election cycle and apply district and year specific random intercepts to account for other unobserved potential confounders which are specific to each district and year. The results shown in Regression 2 indicate that the relationship between the racial margin of victory and minority candidate emergence still holds.

To test the second half of the observable implications, I also estimate a set of Bayesian logistic regressions featuring random coefficients for the racial margin of victory by the racial regime (i.e., majority white, racially balanced, and majority minority districts). Regression 3 reports the results, demonstrating that in all racial regimes the racial margin of victory is positively associated with the likelihood of observing minority candidate. Regression 4 further shows that the conclusion holds for the subset of the data where elections were open-races.

To ensure the internal validity of the regression results, I performed multiple robustness checks by excluding all unopposed elections (column 4 in Table C.5) as well as employing different cutoff points for racial regime and subsetting data before and after 2005 (see Table C.6). I apply these robustness checks since my original results could stem from a particular definition of racial regime, racial margin of victory based on unopposed elections, and data within particular time periods. Moreover, I conducted placebo tests using the presence of female candidates as a placebo outcome to check if the variable of interest only affects outcome values related to racial politics. The idea is that if the proposed model of *minority* candidate emergence is theoretically sound, it should *not* predict female candidate emergence. Figure C.5 portrays the results of the placebo tests. The results demonstrate that both the black population size and the racial margin of victory do not have any bivariate relationship with female candidate emergence. One exception is that in majority minority districts there seems to be a slight association between the racial margin of victory and female candidate emergence. I suspect that this may be due to a potential correlation between race and gender of candidates. Taken together, these evidence show that my substantive results are not susceptible to differences in measurements and the presence of unopposed elections; and thus they are not a product of

	Regression 1	Regression 2	Regression 3	Regression 4
Racial margin	.085	.048		
	(.077, .093)	(.039, .058)		
Racial margin (majority white)			.039	.065
			(.025, .053)	(.042, .089)
Racial margin (racially balanced)			.051	.078
			(.039, .063)	(.059, .098)
Racial margin (majority black)			.062	.082
			(.039, .087)	(.040, .128)
Covariates		✓	✓	✓
Random effects (28 years)		✓	✓	✓
Random effects (301 units)		✓	✓	✓
Only open races				✓
N	2037	2037	2037	1169

Table C.5: **Results of Bayesian Logistic Regressions**
Note: This table reports the posterior estimates for the coefficients of interest in Bayesian logistic regression. 95% credible intervals are shown inside parentheses. Rows 2-4 report that random slopes for the racial margin of victory are estimated by the types of districts.

mere chances.

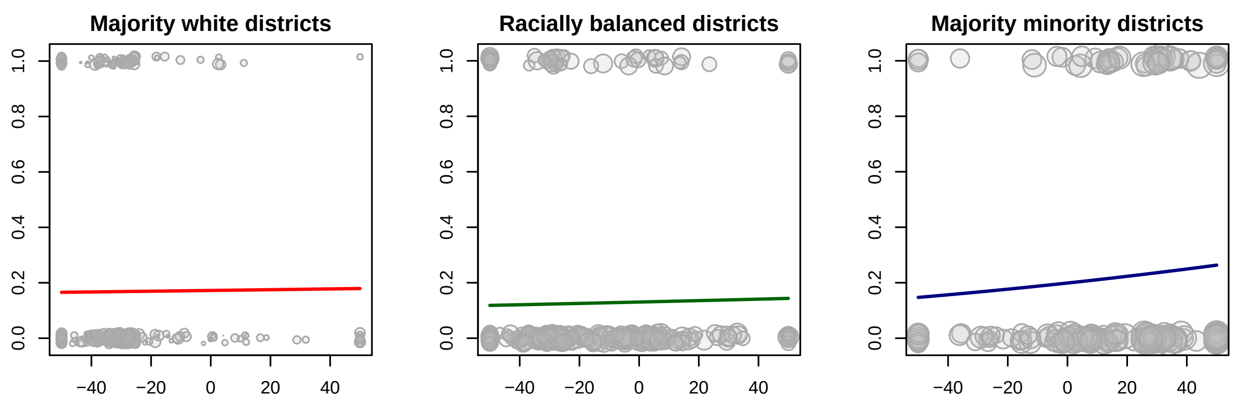

Figure C.5: **Placebo Tests with Female Candidate Emergence as the False Outcome Variable**
Note: This figure visualizes the results of placebo tests showing the bivariate relationships between racial electoral performance and the presence of female candidates. If the proposed model is theoretically sound, it must not predict female candidate emergence, and I confirmed this point in all racial regime. The size of open circles is proportional to the size of black voting-age population.

Finally, I report multiple robustness checks by employing different cutoff points for racial regime, excluding all unopposed elections, and subsetting data before and after 2005. The estimated posterior means with lower and upper credible intervals are shown in Table C.6.

	40/65%	35/65%	30/65%	25/65%
REP (majority minority)	.039 (.025, .053)	.039 (.024, .054)	.042 (.025, .059)	.040 (.022, .058)
REP (racially balanced)	.051 (.039, .063)	.048 (.037, .060)	.047 (.036, .057)	.047 (.036, .057)
REP (majority white)	.062 (.039, .087)	.068 (.044, .094)	.067 (.042, .093)	.067 (.042, .092)
Covariates	✓	✓	✓	✓
RE (Years)	✓	✓	✓	✓
RE (Municipalities)	✓	✓	✓	✓
N	2037	2037	2037	2037

	40/70%	40/75%	40/80%	35/70%
REP (majority minority)	.039 (.026, .053)	.039 (.025, .052)	.040 (.027, .054)	.040 (.026, .055)
REP (racially balanced)	.051 (.040, .063)	.051 (.040, .063)	.053 (.042, .064)	.050 (.039, .060)
REP (majority white)	.075 (.039, .113)	.085 (.041, .132)	.038 (-.028, .110)	.073 (.037, .111)
Covariates	✓	✓	✓	✓
RE (Years)	✓	✓	✓	✓
RE (Municipalities)	✓	✓	✓	✓
N	2037	2037	2037	2037

	30/75%	25/80%	Before 2005	After 2005
REP (majority minority)	.043 (.026, .060)	.043 (.026, .061)	.043 (.021, .065)	.037 (.016, .057)
REP (racially balanced)	.048 (.038, .058)	.049 (.040, .060)	.049 (.032, .066)	.061 (.043, .081)
REP (majority white)	.082 (.038, .129)	.036 (-.027, .106)	.059 (.031, 087)	.086 (.006, .201)
Covariates	✓	✓		
RE (Years)	✓	✓	✓	✓
RE (Municipalities)	✓	✓	✓	✓
N	2037	2037	1166	871

Table C.6: **Estimated Results with Different Cutoff Points and Subset of Data**
Note: This table shows the posterior estimates of the effect of racial electoral performance on minority candidate emergence using different cutoff points for racial regime and subsets of data. The results demonstrate that the original result is not susceptible to these changes in the cutoff points and time periods for analysis.

References for Online Appendix

Aldrich, John H. 1995. *Why parties?: The origin and transformation of political parties in America*. University of Chicago Press.

Black, Gordon S. 1972. "A theory of political ambition: Career choices and the role of structural incentives." *American political science review* 66(1):144–159.

Breiman, Leo et al. 2001. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." *Statistical science* 16(3):199–231.

Cameron, Charles, David Epstein and Sharyn O'halloran. 1996. "Do majority-minority districts maximize substantive black representation in Congress?" *American Political Science Review* 90(4):794–812.

Clark, William Roberts and Matt Golder. 2015. "Big data, causal inference, and formal theory: Contradictory trends in political science?: Introduction." *PS: Political Science & Politics* 48(1):65–70.

Hernan, Miguel A and James M Robins. 2010. "Causal inference.".

Jacobson, Gary C and Samuel Kernell. 1983. "Strategy and choice in congressional elections.".

Lazarus, Jeffrey. 2008. "Buying in: Testing the rational model of candidate entry." *The Journal of Politics* 70(3):837–850.

Magrino, Thomas R, Ronald L Rivest, Emily Shen and David Wagner. 2011. Computing the Margin of Victory in IRV Elections. In *EVT/WOTE*.

Morgan, Stephen L and Christopher Winship. 2015. *Counterfactuals and causal inference*. Cambridge University Press.

Shah, Paru. 2014. "It takes a Black candidate: A supply-side theory of minority representation." *Political Research Quarterly* 67(2):266–279.

Shmueli, Galit et al. 2010. "To explain or to predict?" *Statistical science* 25(3):289–310.

Stone, Walter J, L Sandy Maisel and Cherie D Maestas. 2004. "Quality counts: Extending the strategic politician model of incumbent deterrence." *American Journal of Political Science* 48(3):479–495.

Taagepera, Rein. 2008. *Making social sciences more scientific: The need for predictive models*. OUP Oxford.

Xia, Lirong. 2012. Computing the margin of victory for various voting rules. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM pp. 982–999.