

CONGRESSIONAL PRIMARIES: A STUDY OF CANDIDATE EMERGENCE AND
PRIMARY ELECTION OUTCOMES

by

STEPHEN PETTIGREW

(Under the Direction of Jamie Carson)

ABSTRACT

Research about congressional elections has almost exclusively focused on congressional general elections. The purpose of this thesis is to begin to address the gap in the literature by analyzing trends related to congressional primaries. The empirical analysis relies on an original dataset that comprises all Democratic and Republican candidates for the U.S. House of Representatives from 2000 to 2010. The first half of the empirical analysis considers the factors that influence the emergence of different types of candidates in primaries. The second half discusses the variables that play a part in the outcome of congressional primaries. After estimating several statistical models, I demonstrate the importance of incumbency and district partisanship in the dynamics of primaries. The thesis serves as a launching-off point for a broader research agenda related to primary elections and the impact of actions taken by members of Congress.

INDEX WORDS: Congress; House of Representatives; primary elections; candidate emergence; strategic politicians

CONGRESSIONAL PRIMARIES: A STUDY OF CANDIDATE EMERGENCE AND
PRIMARY ELECTION OUTCOMES

by:

STEPHEN PETTIGREW

Bachelor of Arts, University of Georgia, 2011

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2011

© 2011

Stephen Pettigrew

All Rights Reserved

CONGRESSIONAL PRIMARIES: A STUDY OF CANDIDATE EMERGENCE AND
PRIMARY ELECTION OUTCOMES

by

STEPHEN PETTIGREW

Major Professor:	Jamie Carson
Committee:	Ryan Bakker
	Michael Crespín

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2011

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER	
1 Introduction	1
2 Literature Review	4
The incumbency advantage	6
Strategic decision making by challengers	8
Other factors contributing to candidate emergence and vote outcomes	10
Primaries and the general election	12
3 Theory	14
4 Data Collection and Description	26
Descriptive statistics	29
Quality candidate emergence.....	33
Primary election winners.....	37
5 Modeling Candidate Emergence in Congressional Primary Elections	39
Modeling overall candidate emergence	41
Modeling quality candidate emergence	46
6 Modeling Outcomes of Congressional Primary Elections	50
Modeling incumbent vote share.....	51

Modeling challenger vote share	54
Modeling election competitiveness.....	57
7 Discussion	61
Revisiting the candidate emergence hypotheses.....	61
Revisiting the election outcome hypotheses.....	65
8 Limitations and Prospects for Future Research.....	69
9 Conclusion.....	71
REFERENCES	73
APPENDICES	
A Tables.....	78
B Figures.....	90
C Derivation of Herfindahl-related equations	98

LIST OF TABLES

	Page
Table 1: Primaries with zero candidates versus primaries with at least one candidate	78
Table 2: Candidates by year and party	79
Table 3: Quality candidate emergence by open/ non-open seat	80
Table 4: District partisanship by quality candidate emergence	81
Table 5: Percentage of primary elections won by candidate type	82
Table 6: Winning percentages by candidate type	83
Table 7: Variable summaries for district-level analyses	84
Table 8: Poisson model of candidate emergence in congressional primaries	85
Table 9: Logit model of quality candidate emergence in congressional primaries	86
Table 10: OLS model of incumbent vote percentage	87
Table 11: OLS model of challenger vote percentage in contested races	88
Table 12: OLS model of primary competitiveness	89

LIST OF FIGURES

	Page
Figure 1: Average number of candidates in contested primaries.....	90
Figure 2: Expected number of candidates by presidential vote	91
Figure 3: Expected number of candidates by congressional vote.....	92
Figure 4: Number of candidates by percent unemployed (presidential in and out parties).....	93
Figure 5: Probability of quality candidate emergence by presidential vote	94
Figure 6: Probability of quality candidate emergence by congressional vote.....	95
Figure 7: Effect of explanatory variables on competitiveness	96
Figure 8: Effect of explanatory variables on vote proportions	97

CHAPTER 1

INTRODUCTION

In 2010 there were almost 1,700 different individuals who vied for the opportunity to represent about 700,000 of their peers in the U.S. House of Representatives. For months in the winter and spring of the election year these potential members of Congress raised money, spoke with constituents, and distributed campaign literature. By the end of the summer the pool of 1,700 candidates had narrowed to about 800. These 800 politicians, who had received the blessing of the Republican or Democratic Party in their district, were scrutinized by the media and voters throughout the fall campaign. Little attention was paid to the 900 candidates who failed to make it beyond the primary stage of the midterms. These losing primary election candidates are often neglected by political science research, which tends to focus almost entirely on general elections.

To be sure, candidates in congressional general elections represent a wide variety of backgrounds and belief systems, but candidates in primaries take that variety to a different level. There are losing candidates in primaries that are lawyers and members of the business community. Truck drivers, convicted sex offenders, relatives of sports icons¹, a professional wrestler named Jon Stewart (IL-10 in 2000), and even one future President of the United States (Obama in IL-1 in 2000) have also run unsuccessfully in the primaries. Of course there are large numbers of serious candidates that run in the primaries, but each year there is also a group of primary candidates that is best described as less than serious. In 2000 a primary candidate in

¹ Barbara Dooley, the wife of former Georgia football coach Vince Dooley's, ran in Georgia's 12th district in 2002, and Brian Rooney, the grandson of former Steelers owner Art Rooney, ran in Michigan's 7th district in 2010.

Arkansas' second congressional district promised to jump out of an airplane if he won the election. Another eccentric primary loser (Glenn Coggeshell in WA-2 in 2004) ran a *Lord of the Rings*-themed campaign. This thesis attempts to expand the understanding of congressional elections by telling the stories of all candidates that run, not just those who win the primaries.

In contrast to presidential candidates, candidates for congressional seats have been nominated by direct primary in most states for more than a century. Some of the earliest direct primary elections in American history occurred during the 1870s (Ware 2002). By the 1890s parties in most states used primary elections as the principle mechanism for nominating candidates for Congress. By 1917 almost every state required that a primary system be used for choosing candidates for the highest statewide office. When the United States became involved in World War II in 1941, all but three states used direct primaries to nominate candidates for House elections (Galderisi and Ezra 2001:17). In the decades since, the rules that govern primary election administration have been tweaked, as has the decision of when to hold the primary during the election year. Yet despite the long standing that primary elections have in American history, relatively little political science research has been dedicated to their study.

The purpose of this thesis is to enhance the literature on congressional primary elections. My goal is to explore and analyze the factors which influence both the emergence of candidates and the outcome of primaries. After a review of the relevant literature on candidate emergence and congressional elections, I develop some theoretical expectations and hypotheses about primary elections for the U.S. House of Representatives. I then detail the collection of a new dataset of Democratic and Republican congressional candidates from 2000 to 2010. After reporting descriptive statistics derived from the data, I turn to empirically testing my hypotheses. I utilize a Poisson count model to examine the emergence of all candidates in primary elections.

The factors influencing quality candidate emergence are then modeled by logit regression. I also feature three models that test the hypotheses related to individual candidates' vote shares in primary elections. The thesis ends with a final consideration of the implications of the findings, limitations of the study, possibilities for future research, and concluding remarks.

CHAPTER 2

LITERATURE REVIEW

When compared to the wealth of political science research devoted to congressional general elections, congressional primaries have received little attention. One scholar notes that most of the work conducted on challenger emergence has focused only on “candidates who...win the primary election, are not members of the incumbent’s party, challenge an incumbent, and have political experience” (Lazarus 2008: 838). Most researchers analyze primary election data to gain leverage over broad questions of electoral behavior that are not directly related to primary elections. For example, Lawless and Pearson (2008) consider data from House primary elections from 1958 to 2004 to answer the question of whether gender affects a candidate’s vote share or level of electoral competition faced. They show that because primaries are low visibility affairs with limited turnout, voters will tend to use heuristics like gender to choose between candidates. Lawless and Pearson’s work relies upon a dataset of primary election candidates, but the analysis suffers from potential omitted variable bias from excluding candidate experience from their models. That would not be an issue if general election data were used because candidate experience data in congressional general elections is readily available for analysis. The same data for primary candidates have not been collected, in part contributing to and in part resulting from a scanty research base on primary elections.

Other research has considered the impact that a primary campaign or election can have on the outcome of a general election. Primary election divisiveness has been found to have little influence on a party’s chance of winning in the general election, at least when the effect of

incumbent marginality is controlled (Born 1981). However, primary elections do have an effect on the fall campaign, insofar as primaries allow candidates to establish a campaign structure which helps them succeed in November (Ezra 2001: 48). Primaries give candidates a head start on piecing together the organizational structure of their campaign, hiring staff, increasing name recognition, building legitimacy, honing in on an effective campaign message, and solidifying their party's electoral base. Lazarus (2005) describes the endogenous relationship between primary election competitiveness and general election outcome. By analyzing election data from 1976 to 1998, he shows that primaries tend to be more competitive because there is a high electoral expectation for a particular party.

This small portion of literature specifically dedicated to primary elections takes as a given what other research often ignores: that most partisan American elections, and almost all congressional elections, are best thought of as a two-stage process. Under most circumstances, a candidate cannot win an elected office in a general election without first having received the endorsement of his or her party through a primary. Adams and Merrill (2008) consider the implications of two-stage elections by constructing a theoretical model of campaign strategy. Pulling from research on nascent primary systems in South America (see Carey and Polga-Hecimovich 2006), the authors demonstrate that vote maximizing candidates will pursue disparate campaign strategies depending on whether their party holds a primary. They also show that, although primary elections may tend to pull candidates away from the median voter of the general electorate, parties benefit from primaries by finding stronger candidates for the general election who are more effective campaigners. Primaries also increase general election competition because they allow weak party candidates to position themselves closer to the median voter for the general election.

The incumbency advantage

As with most work on congressional elections, the sentinel piece upon which this thesis is based is David Mayhew's 1974 book *Congress: The Electoral Connection*. In it, Mayhew argues that members of Congress are best thought of as single-minded seekers of reelection. He suggests that representatives are strategic decision makers and that reelection must be their proximate goal if they are to achieve any of their other ambitions. The book introduces advertising, credit claiming, and position taking as the three main tools that incumbents use to build electoral security. This electoral security for incumbents has been the focal point of a large portion of congressional election studies in the years since. Jacobson (2009) has synthesized much of the last 25 years of this research in his book *The Politics of Congressional Elections*. He shows that one of the biggest predictive factors of whether a candidate chooses to pursue a House seat is whether the incumbent is seeking reelection. Strong challengers, especially those with previous electoral experience, place incumbency at the center of their strategic decision making prior to entering a race.

Detering strong candidates from emerging is one of the strongest contributors to the incumbency advantage. Early work by Kazee (1983) used interviews of potential congressional candidates to provide preliminary, individual-level evidence that incumbents deter electoral competition. The interviews indicate that challengers can be deterred by the presence of an incumbent on the ballot, even when the challenger perceives the incumbent to be vulnerable. However, the deterrent impact that incumbency may produce has different impacts on different types of candidates. Banks and Kiewiet (1989) argue that potentially challenging an incumbent does not deter weak candidates from running for a House seat in the same way that it does for strong candidates. Through formal modeling and empirical analysis, Banks and Kiewiet are able

to show that non-quality candidates have a better chance of winning against an incumbent than in an open seat district. This is because open seat races tend to attract strong challengers in both the primary and general election stages, compared to races with an incumbent which are less likely to cultivate strong primary competition. Utilizing primary election data from 1980 to 1984, the authors show that candidates with no elective experience tend to have more electoral success when they run in a non-open seat.

Banks and Kiewiets' findings have not gone unchallenged. David Canon (1993) suggests that it would be irrational for any politician to challenge an incumbent. In focusing exclusively on inexperienced candidates, Canon shows that Banks and Kiewiets' findings are influenced by the omission of incumbent vulnerability, and political newcomers are four times more likely to win an open seat than against an incumbent. Canon also divides the sample of candidates with no electoral experience into amateur challengers, who have zero political experience, and ambitious challengers, who may not have held elected office but have campaigned or been involved in politics in the past. He demonstrates that the latter group of challengers tends to be more strategic in their behavior and, in doing so, act more like experienced politicians than amateur challengers.

More recently scholars have been interested in the specific actions that incumbents take to try to deter electoral competition. Most incumbents maintain a skeletal campaign organization between elections to discourage other politicians from running for their seat (Herrnson 2008). Likewise members of Congress use sizable financial war chests to detract opposition (Box-Steffensmeier 1996; Goodliffe 2001). The research on congressional elections seems to be at a consensus that the biggest predictor of candidate emergence is whether an incumbent is seeking reelection or if the seat is open.

This trend does not mean, however, that incumbents are easily reelected every two years; they are regularly challenged in both the primaries and general elections. In every election there are losing incumbents who faced little or no competition two years prior. Maisel and Stone (2001: 29) try to explain why this is the case by seeking to answer the counterfactual: why are there elections in which nobody challenges the incumbent? They argue that aside from having no interest in running for office or their chance of winning in the general election being low, many candidates are turned off by the two-stage nature of congressional elections. Candidates are deterred by both the possibility of strong primary competition and by poor party organization in the district, which would not bode well for the candidate's general election prospects.

Strategic decision making by challengers

Receiving an almost equal amount of attention in the literature as incumbency behavior, strategic decision making by challengers is a healthy research base to which Maisel and Stone's work has been a contribution. Probably the earliest piece that specifically considers the notion of political ambition by politicians was written by Gordon Black (1972). He introduces the idea of politicians as rational actors that will maximize utility when evaluating whether to run for a higher office. Based on interviews of San Francisco area city council members, Black formally models how a rational politician would act when deciding whether to seek a more prominent elected office. He also introduces the notion of progressive commitment (now referred to as progressive ambition) and positional commitment (static ambition). Politicians with progressive ambition have a desire to attain higher elected offices, whereas those with static ambition are content remaining in their current political institution.

Sandy Maisel's 1986 book *From Obscurity to Oblivion* provides some explanations for why an individual with progressive ambition, in a general sense, would decide to run in a

particular election year. Pulling from interviews of congressional candidates, as well as his own experience running for Congress, Maisel lays out the basic thought process of a potential candidate. First, the candidate must evaluate his or her own personal characteristics to determine whether he or she is capable of winning the election, even in the most favorable circumstances. They then assess the field of potential contenders in the primary and general elections. Maisel notes that most candidates had little knowledge about whom they would be running against, but he does emphasize that incumbency is the best predictor of whether a candidate emerges. Other work has parsed out more specific factors which contribute to challenger emergence decisions. Bond, Covington, and Fleisher (1985) analyze how incumbent behavior, local partisan conditions, district diversity, and national tides affect candidate emergence. By analyzing all congressional races featuring an incumbent in 1980, they find that partisanship and aggregate partisan trends impact the patterns of candidate emergence, while incumbent behavior, such as roll call votes and constituency service, and district demographic heterogeneity do not.

Research on challenger emergence has taken off in recent years, in part because of the collection of a strong candidate-level dataset. Maisel and Stone (1997; also Stone and Maisel 2003) provide an analysis of survey data from 200 random congressional districts in 1997. The survey of political leaders in these districts sought their opinions related to potential congressional candidates. The authors show that, aside from whether or not the seat is open, district partisanship and candidate experience play a strong role in an individual's decision to run for the House. Because these elections are a two-step process, the chance of winning the seat is treated as a conditional probability of winning the nomination and winning the general election: "a strong chance of winning the nomination can be offset by a weak chance of winning the general election, and vice versa" (Stone and Maisel 2003: 952). Maisel and Stone find strong

support for Black's rational choice framework for candidate emergence, demonstrating that it is a function of the costs and benefits of winning the seat, as well as the probability of winning the nomination and general election. Other recent work under the rational choice paradigm has furthered its application to congressional elections. Lazarus (2008) shows that challenger entry is strongly correlated with the probability of winning the seat. By analyzing primary elections from 1989 to 1998, the article shows that candidates enter a race when the probability of them winning is high.

Other factors contributing to candidate emergence and vote outcomes

Another branch of congressional election literature highlights the impact that national conditions and electoral expectations have on strategic decision making by politicians. The foundational work in this area is Jacobson and Kernell's (1983) Strategy and Choice in Congressional Elections. In this book, political elites are posited to be the essential link between turning national political and economic trends into electoral outcomes. Jacobson and Kernell argue that "the more extreme the electoral climate, the greater will be the divergence between the parties in the overall quality of their candidates" (1983: 23). They show that there is a strong correlation between national conditions and the strength of the field of House candidates for each party. Their analysis reiterates that incumbency plays an important role in candidate emergence, showing that experienced candidates are much more likely to run for an open seat, regardless of the other electoral circumstances (see Jacobson 1989). Also emphasized is that the decision to run for the U.S. House is ultimately an individual level decision, left up to the candidate. They point out that if a party could recruit a strong set of candidates, they could potentially counteract poor national tides and mitigate potential losses. This does not occur, however, because individual politicians do not want to go against their own self-interest by running in unfavorable

electoral conditions. The book also briefly explores primary election competition, showing that Democrats had more competitive primaries between 1972 and 1978. They speculate that this could be evidence that the congressional in-party will tend to have more intra-party competition, but they settle on the explanation that the Democrats have a more diverse party coalition, thus generating more competition.

Shortly after the publication of Jacobson and Kernell's book, William Bianco (1984) presented further evidence about the effects of national conditions on elections. His analysis of 108 congressional elections from 1974 to 1980 shows that economic conditions affect quality challenger emergence, but that the effect disappears for open seat elections. District partisanship also has a significant impact on emergence for both closed and open seat races. He theorizes that the dynamics of candidate emergence are different for open seat races because challengers cannot easily assess their chance of winning when their general election opponent is not clear. This information gap may explain why open seat races often tend to attract more candidates.

Redistricting has also been considered as a factor that influences candidate emergence and electoral competition. Hetherington, Larson, and Globetti (2003) find that as elections get further away from the prior redistricting cycle and closer to the next one, quality challengers are less likely to emerge for a House seat. They also show that as an incumbent becomes more secure, in terms of their electoral margin, quality candidates run less often, and this effect becomes magnified through the redistricting cycle.² The theoretical explanation that underlies this trend is that the redrawing of district boundaries disrupts an incumbent's connection with his or her constituents. A strong challenger feels that his or her best opportunity to defeat the incumbent comes when the incumbent has not yet had an opportunity to establish personal connections with any of his or her new constituents. In terms of redistricting affecting vote

² Carson, Engstrom, and Roberts (2006) find a similar redistricting effect in the nineteenth century.

proportions, the literature is at less of a consensus. Abramowitz, Alexander, and Gunning (2006) find that redistricting does not affect incumbent security or marginality. They do not find any large jumps in the number of safe races following a redrawing of district boundaries. Friedman and Holden (2009), on the other hand, maintain that redistricting hurts incumbents' electoral margins.³

Primaries and the general election

Students of congressional elections have also considered the implications that political ideology may have on primary and general election outcomes. Given that primary electorates tend to have a different ideological median point than general election voters (Geer 1986, 1988; Norrander 1989), candidates facing competition in both stages of the election must strike an ideological balance. Brady, Han, and Pope (2007) hypothesize that candidates who are close to the median point of the general election constituency will perform worse in the primary, and vice versa. An analysis of election outcomes from 1956 to 1998 shows that extreme candidates fare worse in the general election but better in the primaries. Also, a content analysis of the issue positions of candidates who defeated incumbents shows that 71 percent of these primary challengers positioned themselves at a more extreme ideological position. Similar work on U.S. Senate primary elections finds that extreme roll call voting by incumbents helps in the primaries, but hurts in the general election (Hirano, Snyder, Ansolabehere, and Hansen 2008).

Much of the remaining work related to primary elections considers how specific rules and laws affect outcomes. Many states, particularly those in the South, feature a runoff primary system in which a candidate must receive a majority of the vote, lest he or she face the recipient of the second most votes in a runoff. Turnout in these runoff elections almost always drops off

³ The findings of this paper, however, are potentially driven by the omission of challenger quality and campaign spending from the empirical models.

from the level of participation in the initial primary election, with the relationship conditioned on the amount of money being spent during the runoff election period (Bullock and Johnson 1992; Bullock, Gaddie, and Ferrington 2002). Primary elections are also distinct from general elections based on who is allowed to vote. States with closed congressional primary systems, in which only registered partisans are allowed to cast a ballot, tend to nominate more ideologically extreme candidates (Kanthak and Morton 2001: 116). Research on presidential primaries also finds an effect of the election structure on the representativeness of the electorate. Open presidential primary states tend to have more moderate and younger primary electorates (Kaufmann, Gimpel, and Hoffman 2003).

CHAPTER 3

THEORY

Given the relatively small amount of political science research dedicated to analyzing congressional primary elections, one may question why it is important to understand the dynamics of primaries. Given the depth of research that has already answered a litany of questions about general elections, why is it important to also understand primaries, if no legislators are directly chosen as a consequence of winning a primary? I contend that there are several important answers to these questions. As previously discussed, congressional elections are best thought of as a two-stage process. Under almost all circumstances, a winning candidate in the general election could not have become a public official without first having been vetted by the nominating mechanism of his or her party. Often, scholars identify trends that exist within the context of general elections and then generalize the conclusions without having explored their theories in the context of primaries.

Neglecting the role of primaries when researching congressional elections can create serious methodological concerns for an analysis. Research about candidate emergence and strategic decision making is mired by selection bias if the data analyzed only samples from the pool of candidates who ran in general elections. To exclude from the sample all candidates who lost in the primaries is akin to explaining what causes war to break out by omitting instances of international strife that did not result in armed conflict. Take for example, Hetherington, Larson, and Globetti's (2003) paper on redistricting and candidate emergence, which makes strong and persuasive conclusions but does not consider primary election losers. This omission could have a

serious impact on their findings, given that the number of primary candidates tends to spike following a redistricting (Maisel 1986). This is not to say that the omission of primary election losers has definitely caused inaccurate conclusions to be drawn, only that a more complete understanding can be reached by analyzing primaries. More likely, the inclusion of primary candidates into an analysis of elections would bolster the conclusions of prior research. Primary elections are a hotbed of non-quality challenger candidates who fail in their bid for their party's nomination. Omitting them from an analysis only biases the study against the research by underestimating the importance of elected experience in running for Congress. Even outside the context of candidate emergence, congressional election research would be better served to consider more fully the implications of primaries. Take, for example, the growing literature on the role that gender plays in American elections. Herrnson, Lay, and Stokes (2003) develop a theory of gender issue ownership and find that female candidates tend to be more successful electorally when they campaign on issues that are traditionally associated with males. Their research design however only surveyed general election candidates and, in doing so, artificially censored their data to exclude women candidates that failed to win their primary election.

It is also important from a normative standpoint to understand the dynamics of candidate emergence and electoral outcomes in primary elections. To accurately describe what motivates individuals to run for office, it is particularly important to consider the candidates who have not been especially politically active in the past. Most of these candidates emerge during the primary stage, but quickly fade away after losing the nomination bid. Their decision making process, as well as that of candidates who surrender a current elected office in favor of seeking a seat in Congress, are important to understand in a democratic system that is built on political and electoral competition. Most would agree that we should strive to be represented by those who are

most fit to represent. Narrowing the gap that many Americans believe exists between this lofty ideal and the present day reality requires comprehension of the electoral mechanisms, such as primary elections, that have contributed to the development of the gap.

Research on primary elections also becomes increasingly relevant in today's context of strong party polarization. As polarization continues to grow, inter-party competition in general election races may diminish because of the shrinking number of swing and independent voters who tend to hold strong influence on electoral outcomes. If partisan polarization decreases the number of marginal seats in Congress, primary elections will become the principal mechanism by which members of Congress are selected. During the time in which the Democratic Party dominated politics in the South, primaries were the only way in which constituents were able to hold their representatives accountable for their actions. Polarization could create a similar trend in many modern day districts. Sticking to an analysis of general elections, then, misses a key piece of the electoral puzzle.

Given these methodological, empirical, and normative justifications for studying primaries, one may ask why a significant amount of research has not already been dedicated to the topic. The best explanation is that the lack of research results from a lack of collected data. Thanks to extensive database construction by Gary Jacobson and other scholars, general elections for House seats have become remarkably easy to study because of the fruitful datasets from which to work. Primary elections data, on the other hand, are few and far between. When much of the research on primaries was conducted in the 1980s and 1990s it was an almost impossible task to collect individual-level data on all primary candidates. Newspapers cannot, and do not, publish profiles about every individual on the ballot of each House primary. In most circumstances, coverage of primaries is limited to an official endorsement of one candidate by

the periodical and a brief mention of their strongest electoral competitor. Only since the advent of the internet has searching for candidate level data on primaries become feasible. And even then, it has only been in the last couple years that Google and Lexis Nexus have developed easily-searched databases of print media. The analysis in this paper is built from the creation of an individual-level primary election dataset which utilized these relatively new technologies. The analysis of these data will be divided into two sections: a set of statistical models that consider district-level candidate emergence and a set of analyses that look at electoral outcomes. As such, several hypotheses will be applicable to an analysis of candidate emergence patterns and others to an analysis of primary election outcomes.

Many of the expectations about the dynamics of primary congressional elections are derived from the comparable literature about general elections. The expected similarities between primaries and general elections are numerous. Incumbents are expected to have an ability to deter candidate emergence in the primaries just as they are in the general election. Vulnerable incumbents should face stiffer primary competition, and safe incumbents should face less. Also, in terms of the influence that candidate experience plays on electoral outcomes, challengers that have previously served in an elected public office are expected to have more success at the polls in primaries, just as has been demonstrated in the general election literature.

It is important, as well, to consider the differences between primaries and general elections for Congress because, if we are to expect the contests to transpire in the same way, then there would be little reason to study primaries. It is important to emphasize, though, that although candidates in both elections seek the same prize, primaries are not the same as general elections. The first thing to note, as should be obvious, is that primary elections are ultimately a single-party affair. Because party constituencies have become increasingly homogenous,

candidates in primaries would tend to have less ideological differences with each other than they would with opposition from the other party. This reality suggests that there may be an increased role played by district partisanship on primary election candidate emergence. Whereas general election candidate emergence may be more attributable to national economic conditions, rather than party strength within the district, candidate decision-making in the primaries may be more strongly tied with their party's expected electoral fortune in the district.

There are two other features of primary election that distinguish them from general elections. Rules dictating ballot access are remarkably different for the two stages of the election. Federal law mandates that citizens cannot be denied the right to cast a ballot in their district's congressional general election. Primaries, though, are a party affair and, as a result, most states leave ballot access up to the desires of the parties. Parties in some states choose to allow anybody to vote in their congressional primary elections, but many have restrictive rules that require a voter to have some sort of formal allegiance to the party. The potential result of this difference is that certain types of primaries may result in different types of candidates choosing to run and win. The implications of this difference are further developed later in this chapter.

A final substantial difference between primaries and general elections is the influence that incumbent deterrence can play. General elections are appropriately classified dichotomously. Some general elections are open, with the incumbent not seeking reelection, while most have the incumbent running on the ballot. In primary elections there is an added third category that is situated somewhere in the middle of these two categories. A large number of primary elections are in districts in which the incumbent is seeking reelection, but the incumbent is a member of the other party. The effect of incumbent deterrence is expected to be felt in these primaries, but the magnitude of the effect should be somewhere in between the analogous effect for open seat

primaries and primaries in the incumbent's party. Candidates in the incumbent's opposition party must expect to face the incumbent in the general election, but they can at least feel slightly more comfortable that they can win the primary.

Perhaps the biggest factor that influences candidate emergence, for challengers with elected experience and for those without, is whether a seat's current occupant is seeking reelection. Prior research has identified this single variable as one of the strongest predictors of challenger emergence (Rohde 1979; Jacobson 1989). The reasons for the deterrent role that incumbents play are numerous. Incumbents have a wealth of institutional resources at their disposal (Cover and Brumberg 1982). They also tend to be more seasoned politicians who have established a personal vote among their constituents (Fenno 1978; Cox and Katz 1996). Challengers almost always find it difficult to eclipse the fundraising abilities of an incumbent and, more often than not, cannot even come close in spending (Abramowitz 1991; Epstein and Zemsky 1995). All of these factors make it so that almost any candidate would prefer to run in an open seat congressional district than against an incumbent. This leads me to my first two hypotheses:

H1: Open seat congressional district races will tend to attract more candidates than districts in which the incumbent is seeking reelection. Quality challengers will also be more likely to run in an open seat district.

H2: Among seats in which the incumbent is seeking reelection, the incumbent's party primary will have fewer candidates, and attract fewer quality challengers, than the primary of the opposing party.

The length of time in which the incumbent has served in the House should also have an impact on the patterns of candidate emergence (Hernnson 2008). Members of Congress who

have served in the House for a long period of time have been able to do so for a reason. Often citizens who advocate for term limits lament the fact that many representatives have served for upwards of five or more terms. Especially of concern to these citizens are incumbents who espouse issue positions that are on the ideological fringes. What is forgotten is that incumbents do not *automatically* get reelected; they are vetted every two years just like every other candidate and, although they certainly do have some advantages over challengers, can be booted from office through electoral means. This suggests that incumbents who have served in Congress for a long period of time have won reelection because their constituents are happy with their Representative's issue positions and personal characteristics.

On the other hand, members of Congress who have not been in office for a long period of time have likewise not been vetted and held up to the electoral scrutiny of their experienced peers. Freshman members of Congress in particular, those who have not yet been up for reelection, are perhaps the most vulnerable incumbents. This may seem to run counter to some of the congressional general election literature which focuses on the electoral bump that these incumbents receive in their first election, the so-called "sophomore surge" (Gelman and King 1990). But these earlier findings are based on comparing the bump received between the incumbent's first and second elections to the vote differences between all other back-to-back election cycles. This does not mean that sophomores are electorally safer than tenured incumbents, rather that sophomores simply tend to make larger electoral gains when compared to their last election, in which they may have faced an incumbent or ran in an open seat race.

Furthermore, I believe that freshmen incumbents are actually more electorally vulnerable than longer-tenured incumbents. Freshmen have not established a strong personal vote among their constituents. They may not have totally mastered the art of electioneering, whether through

resource allocation, fundraising, messaging, or any other aspect of a congressional campaign.

The sophomore incumbent may also be ideologically out of touch with their district's constituents and was only elected two years prior on the heels of a strong national tide in favor of their party (Canes-Wrone, Brady, and Cogan 2002). All of these possibilities lead me to my next hypotheses:

H3: Incumbents seeking reelection for the first time will face more challengers in the primaries than more tenured incumbents, both in terms of overall number of challengers and number of experienced candidates. This trend will exist for both the incumbent's party and the challenger party.

H4: A freshman incumbent candidate will receive a lower vote proportion in the primaries than an incumbent that has been in Congress for multiple terms, ceteris paribus.

The degree to which an incumbent has established an incumbency advantage and, conversely, is perceived to be electorally vulnerable, may play a role in the decision making process of a potential primary challenger. Jacobson (2009: 42) demonstrates that "experienced challengers are more likely to run against incumbents who had closer contests in the last election." This effect of incumbent vulnerability has been found in other types of candidate emergence studies as well (see Maisel 1986 and Lazarus 2005). The reason for this tendency is quite easy to see when it is phrased more simply: an incumbent who is perceived to be easier to beat will attract a larger number of candidates who wish to be the one to deliver that defeat.

H5: The more electorally vulnerable the incumbent, the more likely both quality and non-quality candidates are to run against the incumbent, for both the incumbent's party and the challenger party.

H6: Weak electoral performances by the incumbent in a general election will translate into weaker electoral performances in the next primary election. There will be a positive correlation between vote proportions received by an incumbent in a general election and the primary two years later.

Another factor that may contribute to the perception of incumbent vulnerability is redistricting. This vulnerability stems from the fact that when the lines of an incumbent's district are reconfigured, there is no longer congruence between the incumbent's previous district and his or her new district. The individuals whom the incumbent has not represented before have not received particularized benefits from the incumbent, may not recognize the incumbent's name, and possibly have never even seen the incumbent before, whether on television or at a political event. An incumbency advantage derived from the personal vote is non-existent among many of these constituents. Herrnson (2008) argues that redistricting affects incumbents to the point that many strategically retire, rather than face tough electoral circumstances. As such, redistricting years are perceived by many potential challengers as opportune times to seek a congressional seat. Even if the incumbent is seeking reelection, a challenger has a better chance than usual to accumulate votes among the district's new constituents. This leads me to my next hypothesis:

H7: Congressional primaries in states that have redrawn the district boundaries will tend to attract more candidates overall, and more quality candidates, than primaries in states that have not been redistricted.

H8: An incumbent that has been subjected to redistricting will receive a lower proportion of votes in the primary than an incumbent whose district lines were unchanged.

On some level, almost all individuals that run for Congress do so because they want to win the right to represent their district in Washington. This may seem to be a questionable

assumption, given that there are candidates that run for office on all levels of government that have been driven to campaign by ulterior motives, such as building focus on a particular issue or establishing business connections. But even among those individuals, it is difficult to argue that, given the choice between winning or losing the election, any candidate would chose to lose. Thus district partisanship likely plays an important role in the decision making of a potential congressional candidate.⁴ If a party receives 80% or 90% of constituent loyalty in the district, that party's nominee is all but guaranteed to be the general election victor; thus the probability of winning the seat for a member of that party is nearly equal to the probability of winning the primary election. On the other hand, if a district consistently votes 90% to 10% against a candidate's party, their probability of winning the seat in Congress is substantially lower than their probability of winning the primary. Candidates' emergence decisions are strongly tied to their probability of victory (Black 1972), and thus a candidate is more likely to run when her probability of defeat in the general election is minimized. Candidates are less likely to run when district partisanship suggests a strong possibility of general election defeat.

H9: District partisanship and candidate emergence are positively correlated. That is, a district that heavily favors one party will tend to feature more candidates and more quality challengers. Likewise, the worse a candidate's party stands within the district, the less likely that candidate is to emerge.

Even having formulated the desire to seek a seat in the U.S. House, many politicians wait for the opportune time to run for the office. Whether the seat in their district opens up or whether

⁴ That being said, inclusion of the people who run regardless of district partisanship will tend to minimize any potential impact that partisanship is playing on candidate emergence. Presence of these candidates that run for alternative reasons, regardless of the electoral circumstances, will actually make it more difficult to identify a statistically significant relationship between district partisanship and candidate emergence in the analysis. Any relationship identified by the statistical regression may indeed be stronger in reality for candidates whose principal motivation for running is to win the seat.

redistricting yields favorable electoral circumstances likely have strong influences on candidate emergence trends. Another factor may also influence a candidate's decision to run in a particular election year: national electoral tides. I conceptualize the idea of national tides as intricately tied to the current economic conditions. An unusually weak economy, such as that of 2010, would tend to drive up the number of candidates in the primaries of the party that does not control the White House. The reason for this expectation is that a weak economy breeds political discontent, and the presidential out-party would be able to best turn that discontent into electoral gain. The president becomes a potential lightning rod for criticism about the economy, levied by members of the opposing party. In times of strong economic fortune, the presidential out-party may not anticipate strong electoral gains. Potential candidates of the out-party may foresee a difficulty in building a strong campaign message. Candidates of the in-party, under these circumstances, would attempt to use the strong economy for their own political gain. This theory leads me to my next hypothesis:

H10: Candidates of the presidential out-party are more likely to emerge when economic conditions are weak. Candidates of the presidential in-party are more likely to emerge when economic conditions are strong.

The general election literature establishes that one of the strongest predictors of incumbent electoral success is whether an experienced challenger runs in the election. Quality challengers are conceptualized as non-incumbents that have held an elective public office. These candidates have been found to receive substantial bumps in their vote total over non-quality candidates, and the former are several times more likely to win a general election (Jacobson 1989, 2009). Some reasons for this include more developed abilities to allocate resources and establish campaign messages. Quality candidates are also more likely to already have a pool of

campaign donors from which to seek initial funds. Individuals with elected experience would also be expected to have at least slightly stronger name recognition than a candidate who has never run for public office.

H11: In congressional primary elections, challengers that have held an elected public office will outperform those that have not.

The final variable to be considered is the type of primary structure a state utilizes. By this I mean the rules that a state sets with regard to ballot access in primaries. Open primary systems allow for any registered voter to participate in the primary election of any party. Closed primary rules dictate that an individual may only vote in a party's primary if they are a registered member of that party. Some states have systems that are somewhere in between completely open and completely closed. Semi-open primaries are ones in which an individual that is not affiliated with a party, or is registered as an independent, may vote in the primary of either party. Semi-closed primaries allow for independents or non-registered voters to vote in either party, but must register with that party at the time of the election.⁵ I expect that as ballot access in a primary system becomes more closed candidates will be less likely to run. An explanation for why this trend lies in ideology. Moderate candidates may be more turned away from running in a closed primary than an open one because fewer independent and cross-over voters would be able to vote in the closed primaries. These moderate candidates may assess this chance of winning the election in a closed primary to be lower than their chance in an open primary.

H12: As a primary election system that a particular state utilizes becomes more closed, the number of candidates and quality candidates for congressional seats in that state will decrease.

⁵ In this analysis, the second two groups will be treated the same because they are so similar in a practical sense.

CHAPTER 4

DATA COLLECTION AND DESCRIPTION

Given these hypothesized expectations, I will spend the remainder of the thesis analyzing the data from congressional primary elections during the 2000s. The first portion of this chapter discusses the process of collecting the data. I then turn to an initial look at the basic trends in the data. In this portion of the thesis I provide an array of descriptive statistics, as well as analyze numerous bivariate relationships. After the descriptive statistics, I empirically test my hypotheses. This chapter is broken up into a discussion of what factors influence the number of candidates running in a primary and an analysis of the emergence of challengers with elected experience. Conclusions about these two topics will be drawn from a Poisson model and a logit model, respectively. I then discuss the dynamics of electoral outcomes in congressional primaries, using several models of vote outcomes.

To investigate my theories, I constructed an original dataset of congressional primary election results. The data comprise candidates that filed to run in one of the 435 U.S. House of Representatives districts between 2000 and 2010.⁶ Only individuals that were featured on the ballot in either the Democratic or Republican primary are included; third party, independent, and write-in candidates are omitted from this analysis.⁷ The data also include candidates who

⁶ Special elections that occurred in the middle of a session of Congress are not analyzed. Most states do not hold party primaries for special elections. Runoff elections are also excluded because they would bias my dataset in the direction of having a higher number of competitive primaries. For a discussion of primary runoffs see Bullock, Gaddie, and Ferrington (2002).

⁷ Although data about write-in candidates are available through the *America Votes* book series and there are a handful of cases (<10) in which a write-in candidate won a primary election, these individuals were excluded from the data. The main reason for this decision is that write-in candidates do not officially file paperwork with their

formally backed out of the race, but had their name featured on the ballot because they had already filed paperwork with their Secretary of State's office or State Board of Elections. Candidates who were on the election ballot but did not run an active campaign are treated similarly. The reason for these inclusions is that a systematic study of whether each candidate was actively campaigning at the time of the primary election is nearly impossible. Given that there have been several thousand unique candidates that ran for Congress in the six elections being studied, there is no effective method of determining whether each candidate was definitely running an active campaign on the Tuesday of the primary election. Including these individuals in the dataset actually works against finding evidence in support of several of my hypotheses, given that their exclusion would bias the analysis in the direction of finding a larger proportion of strong, quality candidates than actually vied for the seat.

Several states in particular election years are thrown out of the dataset because of unique electoral features of their candidate selection process. Rather than holding a primary election in which party members have the opportunity to vote, three states use a party convention to choose candidates for federal and statewide offices. Under some circumstances, primary elections are held in these states following the party convention, but these races almost always feature two candidates and function as more of a runoff mechanism than a true primary. Including them could bias the sample because these elections tend to be more competitive when the party convention could not generate a consensus. Given these concerns, Connecticut from 2000 to 2008, Utah from 2000 to 2010, and Virginia from 2000 to 2010 are left out of the dataset. Two other states were excluded because they had unusual rules regulating primary elections. Louisiana is omitted from the dataset from 2000 to 2006 because it used a nonpartisan blanket,

Secretary of State. They cannot truly be considered to have 'emerged' as a candidate in the election and thus should be excluded from research concerned with candidate emergence.

or jungle, primary system in which the November Election Day ballot lists every candidate from all parties. In this primary system, if one individual receives a majority of the vote, they are elected; otherwise, the top two candidates, regardless of party, move on to a runoff election. Washington adopted an identical system in 2008 and 2010, and as such, is excluded from the dataset in these years.

Most of the information incorporated into the dataset was retrieved from volumes 24 through 28 of the *America Votes* book series.⁸ These books contain profiles of each congressional primary election, including candidate names, parties, vote totals, and the type of primary. Because the 2010 edition of *America Votes* has not yet been published, the data for this set of primaries were found on the *New York Times*' Election Results page.⁹ Because of some differences in the coding methods between the district level and candidate level analyses, the coding of these variables accompany the discussion of the models. Candidate quality, however, is handled consistently across all models. I adopt Jacobson's coding scheme for this variable, in which quality is operationalized as a dichotomous variable of whether the candidate had held an elected public office at the time of the congressional primary (see Jacobson 1989; Box-Steffensmeier 1996). Other researchers have used more nuanced scales, with varying results. Banks and Kiewiet (1989) code candidate quality on a trichotomous scale, with the extra group comprising congressional staffers, party officials, and nonelected public office holders (also see Herrnson 2008). Green and Krasno (1988) operationalize the variable as an eight point scale constructed from data about type of electoral and campaign experience, celebrity status, and other personal attributes. Bond, Covington, and Fleisher (1985) utilize a scale that includes political experience, personal characteristics, and fundraising capabilities. Jacobson's

⁸ Cook, Rhodes, Alice McGillivray, and Richard Scammon. 2001, 2003, 2005, 2007, and 2009. *America Votes: Election Returns by State*. Vols. 24-28. Washington, DC: CQ Press.

⁹ Available at: <http://elections.nytimes.com/2010/results/house> or archived data by state available upon request.

dichotomous coding scheme is utilized for my analysis because, aside from gaining general acceptance in much of the congressional election literature, it is parsimonious, provides for easier data collection, and does not sacrifice explanatory power. These data for approximately 85 percent of the candidates were coded by searching for their name and state in Google News and restricting the search to the election year. Another 11 percent were coded from a second search of Lexis Nexus Academic, similarly restricting the search to the election year. Coding for two percent of candidates came as a result of a third search of Google, in which the campaign websites of several candidates yielded the necessary information. I was not able to conclusively determine whether the remaining candidates, fewer than two percent of the entire pool, had held an elected public office at any point. Because there were only about 135 such candidates out of a pool of almost 8000, these observations are listwise deleted from the individual-level analyses data. In the district-level analysis these candidates are considered to be non-quality candidates.¹⁰

Descriptive statistics

The data that were collected yielded a total of 7721 Democratic and Republican candidates contesting 4541 House primary elections between 2000 and 2010. The figure of 4541 elections is comprised of only elections in which at least one candidate was featured on the primary ballot. In order to avoid bias in the district-level models, 349 “elections” with zero candidates were added to the dataset to yield 4890 cases.¹¹ These extra observations ensure that each party primary can be matched with a primary of the other party within the dataset. Omitting the cases with no candidates would systematically bias the district-level models.

¹⁰ Chance that one of these 135 individuals had held elected office, given that I was unable to find any information about them after three database searches, is very low. They are really only included in the district level analysis which uses a Poisson model to estimate candidate emergence counts. Although I do not have background information on these individuals, I do know that they were on the ballot, so they are included in this model.

¹¹ The total number of potential primary elections that could be included in the dataset is 5220 (435 districts x 2 parties x 6 election years). The number of 4890 was reached as a consequence of omitting the primary elections of a handful of states in particular election years, as described earlier.

[Table 1 about here]

Table 1 compares the group of primaries with at least one candidate to the potential district primaries in which no candidate emerged. As can be seen, the number of each of these two groups of elections is relatively even across the 2000s. There is more variation in the number of elections across time for the contested primaries (range=53) than the primaries that did not occur (range=47). The subsets of races are also almost equally split by party. The primaries with zero candidates is comprised of 45.0% Democratic races; the primaries with at least one candidate is 50.4% Democratic.¹²

The subsets of the data show tremendous differences when other variables are taken into account. There were only 6 zero-candidate primary elections, compared to 384 contested elections, for open seat districts. This means that in percentages, 1.72% of non-contested primaries were in an open seat district, whereas open seat districts make up 8.46% of the contested primaries subset. Open seat elections tend to be among the most competitive in general elections, and these preliminary findings provide some evidence to support hypothesis 1, which suggests that open races will attract more candidates.

Perhaps the most fundamental difference between these two groups of district primaries is district partisanship. When measured by both prior presidential vote proportion in the district and the district's prior congressional general election vote, district partisanship goes heavily against the parties that did not feature a single candidate on their primary ballot. In this subset, the parties' presidential candidates' vote share in the prior election was only 36%; their prior House vote percentage (even when omitting uncontested previous general election races) is just 30.44%. These low percentages are staggering when compared to the values for primaries that

¹² The figure for the former group, 45%, does not reflect an imbalance in party make-up of these non-elections. The low percentage, in comparison to the baseline of 50%, is merely a reflection of the smaller number of cases in this subset of the data.

featured at least one candidate, where the average presidential vote was 51.11% for the party holding the primary, and the average congressional vote was 50.79%. The differences in proportion between the two groups for both measures of partisanship are statistically significant ($p < 0.001$). This statistic provides compelling support for hypothesis 9, which argues that the reason candidates fail to emerge for certain parties in particular districts is that probability of winning in the general election is extremely low. Quality candidates would not be likely to risk almost certain electoral defeat, and inexperienced challengers similarly avoid expending the costs of running a campaign that would have almost zero electoral payoff.

Given these trends regarding district partisanship, one variable that is surprisingly similar for each subset of the data is the tenure of the incumbent currently holding the seat. The mean tenure for the two groups varies by only .3 years, and the median is equal. It seems logical to hypothesize that average tenure for the subset that did not hold a primary would be substantially higher, given the extreme differences in district partisanship. A district that is so dramatically favorable to one party would not be likely to have competitive general elections, and thus would have less seat turnover. There are two possible explanations for this apparent paradox. First, it could be that incumbents in these districts are most likely to use the seat as a jumping off point for higher office. If these incumbents in minimally competitive districts are more likely to run in Senate and gubernatorial races, the tenure length would be diminished. Second, these districts may highlight the importance of studying primary elections in the first place. Although there is little inter-party competition, the tenure of these incumbents could be diminished as a result of increased competitiveness in the primaries.

[Table 2 about here]

Table 2 breaks down the total number of candidates running in House primaries in each election year. As is obvious, candidates are almost evenly dispersed across the first 5 election cycles. The 2010 primary season, however, has a substantially higher number of candidates: a full 34% more challengers than the next highest election year, 2008. The jump in the number of candidates did not occur uniformly between both parties. The number of Democratic candidates in 2010, 602, was only the third most in the dataset. But with 1084 individuals seeking Republican nominations for Congress, 2010 saw 74% more Republican candidates running than the next highest election, 2002.

[Figure 1 about here]

This increase in candidate emergence can still be found when the number of total contested races is accounted for. Figure 1 is a histogram that charts the average number of individuals running in a contested primary election for either party. Throughout the entire timeframe of the dataset, the average number of candidates running in an election is 1.70. When the data are divided by year and party, this average ranges from 1.51 to 1.71.¹³ The median number of candidates for the entire dataset is 1, and only 36% of all races in which an election was held featured more than one candidate on the ballot. This means that almost two-thirds of candidates who sought their party's endorsement for Congress received it without opposition. The impact of open primaries on candidate emergence is also clear. Whereas districts with an incumbent seeking reelection average 1.53 candidates per race, seats that are completely open have a mean of 3.50 individuals on each party's ballot. This difference is statistically significant ($p < .001$) and provides evidence that open seat primaries are more competitive than those with an incumbent.

¹³ This range omits the average number of candidates for Republican primaries in 2010.

Contested Republican primaries in 2010 featured a staggering 2.69 candidates per election. A more in depth assessment of what is driving the unusually high number of Republican candidates in 2010 accompanies the statistical models in the next chapter. This unusual case does provide evidence of the importance of national tides in influencing candidate emergence.¹⁴ The economic situation in 2010 was by far the worst of any election year since 2000. There was not a single state with unemployment above 9% in January of any election year between 2000 and 2008; there were 29 such states in 2010. Strong discontent with the economic situation may have been prompting both political veterans and rookies in the Republican Party to run for Congress.

Quality candidate emergence

Table 2 also describes the total number of candidates with electoral experience that emerged for each party. In total there were 1064 such challengers, out of 5459 total non-incumbent candidates in the dataset. When divided by party, the proportion of quality candidates to non-quality candidates is almost even. For the GOP, 19.10% of all challengers had elected experience; for the Democrats this figure was 19.98%.¹⁵ There were also not significant differences in the number of state legislators that the two parties recruited. Republicans had 290 quality candidates on the ballot; Democrats had 214.¹⁶ It is perhaps more important to analyze the raw numbers of quality candidates that emerged, rather than simply the proportion. This is because each addition of a quality candidate to a party's candidate field potentially increases the number of competitive general election races. In terms of the actual number of challengers, Republicans recruited a larger number of quality candidates for primary elections than did

¹⁴ 2006 is widely considered to have been a strong national-tide election for Democrats. The average number of Democratic candidates in 2006 was the second highest in the dataset, 1.71, yet this figure is dwarfed by the number from the GOP in 2010.

¹⁵ This difference is not statistically significant, even at low thresholds of significance ($p=.416$).

¹⁶ This difference is also not statistically significant ($p=.278$).

Democrats in four of the six election years studied.¹⁷ Even 2008, when Democrats made large gains in the House general elections, Republicans were able to put 102 quality candidates on their primary election ballots, whereas Democrats only had 90.

It is interesting to note the variation in recruiting success across elections for each party. The Democrats show little variability in the number of quality challengers that emerge during the primary elections. The numbers of quality candidates for the party range from 68 in 2004 to 90 in 2006 and 2008. Similarly, there is almost no variation in the success of recruiting state legislators for the Democrats. Small spikes in the number of quality candidates can be seen in 2006 and 2008, years in which the Democrats expected strong electoral tides, but these jumps are not seen in the number of state legislators being recruited.

The Republican Party, on the other hand, has had tremendous variation in the number of quality candidates emerging in the primaries. The fewest number of quality candidates the GOP recruited was 74 in 2000. In 2010 this number was almost double, 142. Republican recruitment efforts also seem to track national trends more consistently than do the Democrats' recruitment. Republicans' strongest year in recruiting, as well as in the November elections, was 2010. Likewise, 2002 was a strong year for Republicans in both areas. The one anomaly, however, comes in 2008. Republicans recruited their strongest candidate field for the entire decade to that point, yet it was Democrats who were able to make big gains in the November election. This seemingly paradoxical situation may be in part attributable to the timing of emergence decisions. Most candidates would have been required to file for the primary in early 2008, when the national tides did not drastically disfavor the Republicans. By September though, when the

¹⁷ The exceptions to this were 2000 and 2006, and even in those years the Democratic recruitment advantage was modest, 3 and 10 respectively.

financial market collapsed and the stock market took substantial hits, the political momentum had swung heavily toward the Democrats.

[Table 3 about here]

The presence of an incumbent in a primary election seems to have substantial implications on the quality candidate emergence patterns of both the incumbent's party, as well as the challenger's party. Table 3 shows number of instances in which a quality candidate emerged for open seats versus non-open seats. In terms of proportions, it is clear that quality candidates enter primary elections with a much higher frequency when there is no incumbent seeking reelection. Usually about 70% of primaries for an open House seat attract at least one quality challenger, compared to about 12% of primaries for an occupied House seat. These data provide descriptive evidence in support of the hypothesis that open seat primaries will be more likely to attract quality challengers.

[Table 4 about here]

Table 4 indicates that the relationship between candidate emergence and open seats is further moderated by district partisanship. The table separates the data between elections in which an experienced challenger did and did not run. It then separates these groups by whether the seat was open or closed and reports the party's average previous presidential and congressional vote percentage within the particular district. The data in the table show the role that district partisanship plays in candidate emergence patterns. Primaries for open seats in which no quality candidate runs are in districts that vote strongly against the potential quality challenger's party. On the other hand, quality candidates seek the nomination for an open seat in districts that have a more favorable district partisanship. These data support the theory of experienced challengers as strategic decision makers. When a seat in the House opens up, a

quality candidate is much more likely to run for that seat when the voters of the district are more favorable for the candidate. Also noteworthy is that district partisanship seems to have little effect on quality candidate decision making in non-open seats. The most likely explanation for this is that, for the incumbent's party, quality candidates are unlikely to challenge their fellow partisans, and for the challenger party, experienced politicians would rather wait for the seat to become open.

Prior research has suggested that non-quality candidates attempt to maximize their potential of being elected by running against incumbents because these races are least likely to attract strong competition in the primaries (Bank and Kiewiet 1989). If the authors are accurate in their conclusions then we should find that non-quality candidates enter primaries at lower rates when electoral conditions are sufficiently favorable for a quality candidate to emerge. The evidence from the elections since 2000, however, indicates that the opposite may be true. There are 3759 elections in which there were no quality candidates. In these races, 3250 non-quality candidates ran for their party's nomination, an average of .86 non-experienced candidates per race. Of the 782 races with at least one quality challenger, there were 1105 inexperienced candidates, averaging 1.41 per race. The difference between the two groups of .55 candidates per race is statistically significant ($p < .001$). These data, of course, represent only bivariate relationships and do not systematically control for other factors, such as district partisanship and national conditions. The descriptive data do seem to suggest however that the conclusions reached from an analysis of the 1980 through 1984 congressional primaries do not hold up when the 2000 to 2010 elections are analyzed. Certainly a more extensive analysis, which includes data from a longer time period, would yield more certain conclusions about the strategic behavior of non-quality candidates.

Primary election winners

[Table 5 about here]

I now turn my descriptive analysis away from candidate emergence and toward a consideration of who wins and who loses primary elections for the U.S. House. Table 5 reports the percent of primary elections that are won by various classifications of candidates. As should be obvious, the percentage of primaries won by incumbents and challengers is nearly equal. The value for incumbents is slightly lower, but this is to be expected given that fewer than half of all primaries feature an incumbent on the party's ballot. Quality challengers win about 13% of all primary elections, and state legislators make up about 6% of these wins.

[Table 6 about here]

Table 6 shows the percentage of time that each type of candidate wins their primary. These numbers are distinct from those in the previous table because the data in the latter table omit instances in which there are no candidates of a particular classification on a primary ballot. The percentages in Table 6 reported can be thought of as 'winning percentages' of each type of candidate: the percentage of time the candidate type wins, given that they are on the ballot. By far the group with the highest winning percentage is incumbents; never in the last six elections have more than 2% of incumbents seeking reelection been defeated in the primaries. The next strongest group is state legislators, of which 55% won their primary. Non-quality challengers fare the worst, with only 41% winning. The success rate of quality challengers and state legislators grows when you consider that many of these candidates that lose do so to another quality challenger or state legislator. Of the races that feature one or more quality challenger, an experienced candidate wins 75% of the time. The same figure for state legislators is 72%.

Also important to specifically consider are primary elections in which an incumbent is defeated, guaranteeing changes in congressional representation. Between 2000 and 2010 there were 21 incumbents who sought reelection, but failed to win their party's nomination. A majority of these defeats occurred following a redistricting: eight defeats came in 2002 after the decennial reapportionment, two came in 2004 after Texas's mid-decade redistricting, and the final occurred in 2006 after Georgia redrew its congressional district lines. All but five incumbents were defeated either by another incumbent or a quality challenger. The most interesting characteristic of the defeated incumbents is that almost all of them would be considered electorally safe, by most standard general election metrics. On average they received 66% of the two-party vote in the previous general election, and their party's previous presidential candidate garnered 63% of voters' support.

These figures provide evidence of an oversight of much of the general election literature. These instances of incumbent primary defeat would often be treated the same as an incumbent retirement in almost all general election research. Similarly, districts that are so heavily partisan in one direction would not be expected to have much general election turnover. But electoral accountability remains, as a result of primary election competition. Take, for example, Carolyn Cheeks-Kilpatrick's district in Michigan in 2010. The Representative had received 79.5% of the vote in the previous general election, and Barack Obama had received 85% of the votes from Cheeks-Kilpatrick's constituents. It would have been almost impossible for a Republican to defeat the incumbent in the general election. Yet she received only 40% of the vote in the 2010 primary, losing to a state legislator who ran a campaign focused on ethics. Important narratives such as this one are often lost when research ignores primary elections.

CHAPTER 5

MODELING CANDIDATE EMERGENCE IN CONGRESSIONAL PRIMARY ELECTIONS

This chapter utilizes several statistical analyses to model the factors that influence various patterns of candidate emergence in congressional primaries. The unit of analysis for these regressions is the party primary nested in congressional districts. I first use a Poisson regression to model the emergence of all candidates for a primary race. I then specifically focus on quality candidate emergence, utilizing a logit model. These models will allow me to test several of my hypotheses. I begin first, however, by outlining the specification of the models and the operationalization of the key variables.

[Table 7 about here]

Table 7 provides a summary of each of the variables used in the district level analyses. The first two hypotheses suggest that incumbency will serve as a deterrent to the emergence of candidates. I test this hypothesis by including two dichotomous variables: one (*incum*) indicates whether the incumbent is seeking renomination in this party's primary and the other (*incotherparty*) indicates whether the incumbent is seeking renomination in the other party's primary. The reference category for these variables is open seat races. For each district with a non-open seat, the primary of one of the parties is coded as a 1 for *incum* while the other party is coded as a 1 for *incotherparty*.

The next two variables are redistricting (*redist*) and freshman (*fr*). A value of 1 for the redistricting regressor indicates that district boundaries have been redrawn since the last election. A 1 for the freshman regressor indicates that the incumbent is in his or her first term in Congress.

The rules under which the election is administered are also operationalized as dummy variables. If a primary is closed to only voters that are registered with a particular party, the *lawclosed* regressor is coded 1. If the primary is run under either semi-open or semi-closed rules, the *lawsemi* regressor is coded 1. Open primary election systems are left as the reference category. The final dummy variable in these models is *partyinout*. This regressor indicates whether the primary is being held for the party that currently holds the White House (0) or for the party that does not (1).

District partisan makeup is operationalized as the party's two-party share of the presidential vote in the previous election, within the particular congressional district (*prez*). Because there is an observation for both parties for each district included in the sample, the value of this variable for one party is p and the value for the other party is $1-p$. Although also operationalized as the results from a previous election, incumbent vulnerability is measured slightly differently from district partisan makeup. Previous congressional vote (*voteprior*) is used as the measure of incumbent vulnerability. While *prez* ranges from 0 to 100, *voteprior* ranges from 50 to 100, and the variable is the same for both parties within a congressional district. The reason for this is that the effect of an electorally vulnerable incumbent, one that only received about 50% of the vote in the previous election, is expected to increase candidate emergence in both parties (Maisel 1986). Also, allowing the variable to range from 0 to 100 would decrease the efficiency of the model estimation because of colinearity issues with presidential vote in the district.

The final variable in these models is the national economic tide variable. This is operationalized as the rate of unemployment (*unemploy*) in the state in which the election

occurred. This measurement was taken in January of the election year.¹⁸ This measure is by no means perfect. Because it is state-level unemployment data, it does not fluctuate across congressional districts within a state. This was necessary because unemployment numbers by congressional district are not available. Other measures of national tides, such as change in GDP or presidential approval were not chosen for similar reasons. District-level GDP change data do not exist, and there is likewise no polling data on presidential approval for every congressional district in every election year studied. In the model, *unemploy* is interacted with *partyinout* to model the differential effect of unemployment by party that I hypothesized in H10.

Modeling overall candidate emergence

In order to test the hypotheses related to the emergence of any candidates for Congress, I use a Poisson regression model with fixed effects for election year. The variables in the model are operationalized as described in the above section and the formula being modeled is:

$$\begin{aligned} \text{candnumber} = & \alpha + \text{incum}\beta_1 + \text{incotherparty}\beta_2 + \text{redist}\beta_3 + \text{fr}\beta_4 + \text{lawclosed}\beta_5 + \\ & \text{lawsemi}\beta_6 + \text{prez}\beta_7 + \text{voteprior}\beta_8 + \text{partyinout}\beta_9 + \text{unemploy}\beta_{10} + \\ & \text{unemployXpartyinout}\beta_{11} + \text{y02}\beta_{12} + \text{y04}\beta_{13} + \text{y06}\beta_{14} + \text{y08}\beta_{15} + \text{y10}\beta_{16} + \varepsilon \end{aligned}$$

where α is the intercept of the model and ε is a stochastic error term. Dummy variables for each election year (with 2000 as a reference category) are used to control for fixed effects of any year-specific idiosyncrasies.¹⁹

[Table 8 about here]

¹⁸ Maisel's (1986) work indicates that over 70% of candidates decide to run late in the year before the election or early in the election year. January unemployment was picked as a reasonable estimation of the economic conditions upon which candidates based their emergence decision.

¹⁹ An obvious omission in this model is a variable that controls for the incumbent's financial warchest. A more complete discussion of warchest and fundraising variables and a rationale for their exclusion comes in later chapters.

Table 8 provides the MLE for the Poisson model. The first thing to note in this table is that the coefficients with the largest statistically significant magnitudes are the dummies indicating whether the incumbent sought reelection. Each of these variables is significant in the negative direction, suggesting that when an incumbent runs fewer candidates are likely to enter the race. There is also a positive, significant relationship between district partisanship and candidate emergence. The regressor measuring incumbent vulnerability, *voteprior*, seems to provide evidence in support of hypothesis 6. The negative sign on that coefficient indicates that the better the incumbent fared in the prior general election, the fewer candidates are likely to challenge him or her in the next primary. The negative coefficients for the two *law-* variables also give evidence for hypothesis 12, which posits that races will be more competitive as they become more open.

Support is also found for the expectations posited in hypothesis 10 about the influence of national economic conditions. Taking into account the *unemploy* and the *partyinout* variables, as well as their interaction, I find that when the unemployment rate is at zero percent primaries of the president's party are expected to have higher numbers of candidates than primaries of the opposing party. As the level of unemployment increases, however, the number of candidates in primaries for the in-party drop, as indicated by the coefficient on the *unemploy* variable. The positive value of the coefficient for the interaction term indicates that, for the presidential out-party, higher numbers of candidates are expected to run as unemployment in their state increases.

The Poisson regression results also provide evidence that runs counter to some expectations laid out in my hypotheses. The effects of redistricting, *redist*, and incumbent tenure, *fr*, were not found to have an effect on candidate emergence that is significantly different from zero. The count model identifies what has already been established in the descriptive statistics

section: that the 2010 election featured an unusually high number of candidates. The significant positive value of the coefficient for *y10* indicates that the number of candidates running in a primary in 2010 was significantly higher than the number in the reference category, 2000, *ceteris paribus*.²⁰ This suggests that there were unique dynamics at play, in terms of candidate emergence decisions, in the 2010 midterm elections that go beyond what is readily explainable by just national economic trends. This could be initial evidence that these patterns in 2010 were the consequence of some sort of Tea Party effect, although more rigorous testing of this theory would be necessary to provide any conclusive assessments.

To consider more specifically the effects of each of the variables in the Poisson model, Table 8 also provides the change in predicted number of candidates as each variable moves from its minimum to its maximum, while holding all other variables at their mean or mode. The last column of the table reports the first differences for the dummy variables in the model and the change in the expected number of candidates when continuous variables jump from their minimum to their maximum. The values in the first two rows of the table provide evidence that supports hypotheses 1 and 2. The expected decrease in candidate number for the incumbent's party when there's an incumbent running is 1.64; the similar figure for the opposing party is 1.32. These estimates provide support for hypothesis 1, that an incumbent seeking reelection will diminish candidate emergence compared to open seat races. Also, the smaller coefficient for the *incumotherparty*, compared to the coefficient for *incum*, provides some evidence to support hypothesis 2.²¹ It is also important to remember the substantive meanings of these variables

²⁰ When the model is respecified to use 2010 as the reference category, the coefficients for all five of the election year dummies are statistically significant. This indicates that candidate emergence in 2010 was significantly higher than every other election in the dataset, not just 2000.

²¹ Whether this difference is significantly different from zero is difficult to establish. Doing so would require predicting the distribution of expected candidate numbers for every possible configuration of all independent variables and looking for overlap in the distributions when *incum* equals one and *incumotherparty* equals one.

when they are interpreted as well. For the *incum* regressor, a value of one signifies that there is necessarily one more candidate running in the primary, since the incumbent is included on the ballot. The expected change of -1.64 candidates for these races does not take this substantive reality into account. Because of this, the number of expected candidates in a primary that features an incumbent is actually closer to 2.64 candidates less than an open seat race.²²

The magnitudes of change for the two *law*- variables suggest that if open, semi-open, and closed primaries are thought of as a continuum, then the relationship between primary type and candidate emergence is unidirectional and monotonic. That is to say, the decrease in the expected number of candidates is larger for a completely closed primary system than for a system that is between partially closed and partially open. Thus as ballot access laws become more restrictive, some potential candidates seem to turn away from entering the primary election.

[Figure 2 about here]

Figure 2 shows the impact that changes in district partisanship have on candidate emergence, divided out by whether the seat is open or whether the incumbent is seeking reelection in either party. The bands represent 95% confidence intervals around the means of simulated data at each level of presidential vote, holding all other variables at their respective measure of central tendency. The trend of the graph offers support for hypothesis 9, which posits that district partisanship and candidate emergence are positively correlated. Party primaries in districts that vote heavily against the party almost never attract more than one or two candidates, regardless of whether the seat is open or whether there is an incumbent running. As district partisanship evens out between the parties, the effect on candidate emergence for open seat races

²² To demonstrate this point, consider a hypothetical example of an open seat race that is predicted to have 3 non-incumbent candidates running. If the variable *incum* is changed from 0 to 1, the expected number of candidates changes from 3 to 1.36. One of these 1.36 candidates is the incumbent herself, meaning that the change in the number of challengers that are predicted to emerge decreases by 2.64.

begins to diverge from the effect in non-open seat districts. Open seat races for 50-50 districts are expected to attract almost twice as many candidates as non-open seat races for either party. When a district provides near unanimous support for a party, candidates within that district run much more often. In these cases, open seats attract between five and seven challengers; races with an incumbent in the other party attract two or three candidates, and incumbents usually face one challenger in their own party's primary.

[Figure 3 about here]

Evidence that provides some support for hypothesis 5 can be found in Figure 3, which is similar in configuration to Figure 2, except that the variable plotted on the x-axis is previous general election vote within the district. The three bands in this chart indicate that the relationship between incumbent vulnerability, as measured by prior electoral success, and candidate emergence is negative, almost perfectly linear, and homoscedastic. The support for the hypothesis is not overwhelming by any means. Although the graphs indicate a statistically significant difference between the various types of elections, the substantive interpretation does not yield strong empirical conclusions. Vulnerable incumbents seeking reelection can only expect to face about .1 more challengers in their own primary than incumbents that won in general election landslides. The effect on the opposing party's primary is only slightly larger for these districts; only about .25 more challengers emerge to run in districts with a marginal incumbent than seats with a safe incumbent.

[Figure 4 about here]

The interactive effects of national economic conditions and party are displayed in the two graphs of Figure 4. The first graph, which reflects the effect of increased unemployment on candidate emergence for the party that occupies the White House, shows that higher levels of

unemployment seem to discourage candidates from running for their party's nomination for Congress. The effect is most dramatic for districts in which the incumbent is not seeking reelection, where increases in the unemployment rate significantly drive down candidate emergence. The effect also exists for non-closed districts, although the drop off is smaller, and the differences between the incumbent's party and the opposing party are not significantly different at very high levels of unemployment. For the presidential out-party the impact of rises in unemployment is an increase in candidate numbers. The increases, however, do not seem to be especially large. For all three types of elections, the 95% confidence intervals at the lowest level of unemployment include values that are also inside the confidence intervals at the highest level of unemployment. It is important to note that the relationships mapped by these graphs may be wholly an artifact of the electoral circumstances of 2010. There is distinct heteroscedacity in each of the six confidence bands. The bands are most narrow between unemployment values of 3% and 7%, which encompass most observations between 2000 and 2008. When the rate of unemployment exceeds ten percent, which is only observed for the 2010 elections, the predicted number of candidates becomes much less certain. The trends identified in the graph may also be consequences of alternative idiosyncrasies of 2010, an election in which the Republicans, who were the presidential out-party, ran more candidates than average, and the Democrats ran fewer.

Modeling quality candidate emergence

The second portion of the candidate emergence analysis will consider the factors that influence the decision making processes of politicians that have experience running and winning public elections in the past. The dependent variable for this analysis is a dummy, measuring whether at least one quality candidate emerged to run in a party's primary election. Because the

response variable is binary, I use a logit model with fixed effects for election year that has the same explanatory variable specification as the previous Poisson model.

[Table 9 about here]

The results from the logit regression can be found in Table 9. These results provide further evidence of the deterrent power of incumbency. The probability of a quality candidate emerging is significantly lower when the district's seat is being defended by an incumbent than when it is open. The third column of Table 9 reflects the change in predicted probability of quality candidate emergence when each variable moves from its minimum to its maximum, holding all other explanatory variables at their mean or mode. Just like the Poisson model indicates, and consistent with hypothesis 2, the deterrent effect is larger for the incumbent's own party than it is for the opposition party. The predicted probability of a quality candidate running is decreased by 68.75% when an incumbent runs in the same party. For the opposing party's primary, the corresponding decrease is 51.12%.

In contrast to the Poisson model, the logit model estimates that freshman members of Congress are significantly more likely to face quality candidates than more tenured members of Congress. The probability of a quality candidate running against a freshman is 9.31% higher than an experienced politician running against a representative further along in his or her career in Congress. Also unlike the Poisson model, the primary rules, as well as national conditions, are not found to have an impact on quality candidate emergence. Neither of the dummy variables for ballot access rules are statistically significant in the logit model, nor are the *partyinout* and *unemploy* regressor, as well as their interaction. The finding with regard to primary rules is not particularly surprising. Experienced politicians probably recognize that the laws dictating who can vote in the primaries of a state are static, and thus they do not take them into account when

deciding to contest an election for Congress. The null findings with regard to national conditions, on the other hand, run counter to hypothesized expectations. Not only do these coefficients not meet the standard statistical significant level of $p < .05$, they do not even meet the incredibly generous threshold of $p < .25$. This perhaps suggests that quality candidates do not even consider national economic conditions when deciding whether to run. If this is indeed the case it would run counter to the findings of much of the previous literature. A more developed consideration of these unusual findings can be found in the discussion chapter.

[Figure 5 about here]

District partisanship is found to have a significant effect on quality candidate emergence; these individuals are more likely to contest a primary when the district's voters favor the potential candidate's party. Figure 5 illustrates this strong effect by graphing 95% confidence intervals of simulated data for each value of prior presidential vote. The effect of district partisanship is found to be most dramatic in open seat races, in which the probability of quality candidate emergence exceeds 50% when a districted voted just 40% for the candidate's party in the prior presidential election. Also of interest is that, for open seats, the confidence interval narrows tremendously as presidential vote exceeds 60%. This indicates, with strong confidence, that a quality candidate is almost certain to emerge when a seat in a heavily partisan district opens up. The effect is almost as dramatic for primaries with the incumbent running in the other party, although a quality candidate is not expected to emerge until district partisanship reaches about 75-25. The effects in the incumbent's own primary also begin to increase around 75%, although the predicted probability of quality candidate emergence never exceeds 50%.

[Figure 6 about here]

The logit model demonstrates that a heightened incumbency advantage also discourages quality candidates from emerging. Figure 6 shows this effect across all levels of the congressional vote variable. For the incumbent's party, the predicted probability of quality candidate emergence is almost zero across the entire range of prior general election vote share. The effect for the opposing party in non-open seats is that the predicted probabilities begin near 30% and diminish to about 15%. The positive concavity of the graph indicates that the marginal effect of congressional vote diminishes across the range of the variable. Unlike the analogous graph for the Poisson model, the confidence bands for the logit model do not overlap for all values of the independent variable. This suggests that quality candidate emergence patterns against the most marginal of incumbents are significantly different than the patterns against the most electorally safe incumbents.

CHAPTER 6

MODELING OUTCOMES OF CONGRESSIONAL PRIMARY ELECTIONS

To test the remaining four hypotheses related to primary election outcomes, several models of primary election outcome will be analyzed. When considering these individual-level data, the biggest statistical problem that is encountered is that of compositional data. Within each party's primary election, the vote proportions for each candidate add up to 100%. Katz and King (1999) show that estimating an OLS model with compositional dependent variable data yields biased parameter estimates. Typically research on American elections avoids this problem by only estimating the vote share of one of the two parties, using the other as a reference category. But because primary elections feature any number of candidates, I must employ alternative methods of analysis.

The issue of compositional data that I encounter is similar to the one encountered by scholars that study district-level vote outcomes in multiparty electoral systems. The main way that the compositional data problem has been remedied in this context is through transformation of the dependent variable and then utilization of seemingly unrelated regression (SUR) (Zellner 1962). Vote proportions are recalculated as a log ratio of the party's vote proportion to the vote proportion of another party that serves as a reference category. SUR then treats each dependent variable observation as independent of all other observations. Tomz, Tucker, and Wittenberg (2002) show that this methodology produces efficient and unbiased estimates of the population parameters of multiparty electoral outcomes.²³ For the purpose of this analysis, however, there is

²³ For more discussion of dealing with compositional data, see also Honaker, Katz, and King 2002 and Jackson 2002.

no obvious reference candidate that can be used in all elections, so Tomz, Tucker, and Wittenbergs' methodology cannot be used to test the hypotheses of this thesis. To work around this problem, I model primary election vote outcomes in three different ways. I first estimate a simple OLS regression to model incumbent vote share. I then estimate a similar model, but with challenger vote shares as the dependent variable. I conclude with a third OLS model, this time using as a dependent variable a Herfindahl-Hirschman index of electoral competition.

Modeling incumbent vote share

In order to avoid the problem of compositional data, I first use OLS estimation to test the hypotheses on the vote share of only incumbents. Each dependent variable is not dependent on the values of other variables because vote outcomes for incumbents are not a zero sum games.²⁴ The regressors *fr*, *redist*, and *voteprior* are operationalized in the exact same way as the district-level analyses. The *cand* regressor is the number of opponents that each candidate faced in the primary and controls for the natural depression in vote percentages a candidate receives as the number of competitors in an election increases. The variable takes into account the fact that a candidate that earns 55% of the vote in an election with ten candidates was much more successful than a candidate that earned 55% against just one competitor. The variable's actual range in the sample is 0 to 9 in the model of all primaries with an incumbent and 1 to 9 when uncontested races are omitted. The final variable, *vsqual*, measures the total number of quality candidates that are running against the incumbent in his or her party's primary. This variable ranges from 0 to 4.²⁵

²⁴ Primaries that featured more than one incumbent, as a consequence of redistricting, are omitted from this sample (n=6).

²⁵ Operationalizing this variable as a dichotomy of whether a single quality candidate emerged does not affect the substantive interpretation of the model results.

An important variable that is omitted from this regression, as well as all other electoral outcome analyses, is the amount of money spent in the primary by each candidate. The omission comes as a consequence of the difficulty of collecting these data for primary elections. The challenge with collecting them is that FEC reports do not easily break down fundraising numbers by election stage, only by quarter. This does not pose a significant concern for candidates that lose in the primaries, because their total yearly expenditure equals the total they spent in the primaries. For primary election victors the total spent in the primaries can only be obtained by going through the arduous processes of crudely matching the FEC quarterly data with the date on which the primary occurred. Further complicating the data collection process is that some primary election winners, especially those facing weak electoral competition, begin campaigning for the general election before the primary has even occurred. This biases the influence of a primary spending variable by confounding it with expected general election competitiveness. I will attempt in future research to address these problems and include candidate spending, but the current analysis excludes the variable. I do not, however, expect that this omission will tremendously impact the substantive interpretation of my results. In the early 1980s fundraising data were not regularly included in statistical models of general election outcomes, and yet researchers were able to come to conclusions related to the incumbency advantage and other phenomena that withstood the test of future empirical scrutiny.

[Table 10 about here]

The results from the OLS regressions can be found in Table 10. The first two columns of regression output are for the sample of incumbents that includes contested and uncontested races. The third and forth columns provide the results when uncontested incumbents are removed from the sample. The first and third columns use the incumbents' vote share as the dependent variable.

The second and fourth columns take the square root the dependent variable in order to make the distribution closer to normal. As can be seen, this transformation does not affect the substantive interpretation of the model. For ease of interpretation, I will only interpret the coefficients from the first and third models.

Across all four model specifications, the impact of *voteprior* is statistically significant in the expected direction. As was hypothesized, incumbents that were more successful in the previous general election will tend to receive higher vote percentages. The magnitude of the impact varies between the two samples. For the sample that includes all incumbent races, a 1% increase in prior general election margin corresponds with a .027% increase in primary vote share. Because the variable ranges from 50 to 100, the potential impact of this variable in the model ranges from 1.35% to 2.7%. Changes this small cannot be thought of as especially noteworthy. The impact in the model of only contested races has slightly more substantively interesting results. For every increase in percentage point in general election vote share, incumbents receive .107% more votes in the next primary. An incumbent that received close to 100% of votes in his or her past general election would receive about 5.35% more primary votes than an incumbent that just squeaked by in his or her past general election. This may not seem like an enormous difference on its own, but in a competitive election with several candidates, a boost of 5% for any candidate cannot be taken lightly.

The *vsqual* variable provides evidence in support of hypothesis 11. For each model, the emergence of a quality candidate decreases the vote share of an incumbent by several percentage points. The coefficient from the full dataset indicates that an incumbent that faces a quality challenger receives 7.21% fewer votes than an incumbent facing a non-quality challenger. The model with only contested races shows this value is 10.55%. Accounting for the *cand* variable

coefficient as well, an incumbent's vote percentage drops by about 13.9% when the first quality candidate emerges.²⁶

These models do not provide supporting evidence for hypotheses 4 and 8, which posit an impact of incumbent tenure and redistricting on vote outcomes. All four models yield null results for the relationship between redistricting and votes. This may be a consequence of the crude, dichotomous measure, which does not capture the extent or magnitude of the redistricting. The null result may also be an artifact from the time period being studied, which includes only one redistricting cycle. The estimates may have been different if 2002 had been as strong an anti-incumbent year as 2010, for example. The null hypothesis cannot be rejected for the *fr* variable in all but one model. The exception is the model of the entire sample that has a transformed dependent variable. In this instance, the positive coefficient actually indicates that freshman incumbents do better in primaries than non-freshman, evidence that goes directly against the hypothesized relationship. A more in-depth consideration of these findings on the hypotheses is provided after the presentation of each model in this chapter.

Modeling challenger vote share

The next set of models that I will discuss uses OLS to model challenger vote share in races that feature an incumbent and in races that do not. The discussion in this section should be taken with a grain of salt. For elections that feature an incumbent, the dependent variable of vote proportion avoids the concern of compositional data because all values do not add up to 100% within a particular primary. Yet vote proportions within primaries are still correlated with each other. For the subset of elections that do not feature an incumbent, I still encounter the problem of compositional data. Cubing the dependent variable transforms its distribution to become more normal, and this possibly diminishes the bias associated with compositional data. I am not able to

²⁶ This figure is the sum of the *cand* and *vsqual* coefficients.

use a dependent variable transformation and SUR to model the variable because observation sets are not identical across districts.²⁷ Yet with no applicable statistical methodology to serve as a remedy, I keep these potentially serious methodological concerns in mind, nonetheless allowing the results from the models to provide some preliminary evidence regarding the hypothesized relationships.

These models have a similar specification to the previous models of incumbent vote share. The one exception is that *vsqual* has been replaced with *qualadv*. The *qualadv* variable captures the advantage or disadvantage an individual candidate is expected to have when they are a quality candidate or they are running against a quality candidate (Cox and Katz 1996). When a challenger has electoral experience and is running against no other quality candidates or incumbents, the variable is coded as 1. A challenger that is not quality and running against a quality candidate or an incumbent has a -1 for *qualadv*. If a candidate is not quality and they are not facing quality candidates or an incumbent in the primary, or if a quality challenger is running against an incumbent or other quality candidate, the variable is coded 0.

[Table 11 about here]

The regression results for challenger vote share are presented in Table 11. The table presents robust standard errors that were computed by clustering the sample by individual elections. This accounts for expected heteroscedasticity in the variation of vote totals across primaries. The first two columns of results are for the sample of challengers that are running in an election against an incumbent. The samples for the last two models are challengers that are running in primaries for an open seat or primaries in the incumbent's opposition party.

Subsetting the data based on whether a challenger is facing an incumbent allows me to search for

²⁷ This is to say that SUR effectively models multiparty district level outcomes because the same set of parties are on the ballot in each district. I do not have this luxury with congressional primaries, because no individual candidate runs in more than one election.

evidence of differential effects between incumbent races and non-incumbent races, without necessitating triple or quadruple interaction terms.

The model results indicate that challenger experience significantly influences vote outcome. The coefficient for *qualadv* in all four models is statistically significant in the positive direction, indicating that an experienced politician that does not face experienced challengers receives a higher proportion of votes. Likewise, non-quality candidates that face an incumbent or quality candidate do worse in the primary. For races that feature incumbents, the magnitude of this relationship is 12.87. Candidates with a quality advantage receive 12.87% more votes, and candidates with a quality disadvantage receive 12.87% fewer votes than candidates who do not have a quality advantage or disadvantage. The value of this coefficient in the model that excludes incumbent races is 11.79. These results provide strong evidence for the hypothesis that quality candidates will outperform non-quality candidates in primaries.

Although it may not be readily obvious, the coefficients for *voteprior* in all four models also provide evidence that supports hypothesis 6. The hypothesis suggests that an incumbency advantage and incumbent vulnerability will significantly impact vote shares in elections featuring an incumbent. The significant and negative coefficients for the first two models demonstrate this. As the incumbent's previous vote share increases, the vote proportion received by challengers in the next primary diminish. The range of this effect is between 6.98 percentage points (when *voteprior* is 50) and 13.97 points (when *voteprior* is 100) for each individual candidate within an incumbent-challenged race. The fact that the null hypothesis cannot be rejected for the *voteprior* variable in races that do not feature an incumbent indicates that incumbency advantage or vulnerability does not impact the outcome of these elections. A significant effect in these models would have seriously cast doubt as to the validity of the

findings in the first two models, but because this is not the case, we can more firmly say that these models provide supporting evidence for hypothesis 6.

These models do not yield significant results for the *fr* or *redist* variables. Like the previous analysis of incumbent vote share, the findings related to challenger vote outcomes indicate that the hypothesized relationship with regard to these variables may not exist. If the coefficients for these two variables were positive and significant, it would show that challengers running against freshman and redistricted incumbents have stronger electoral performances. But because the effects of these variables are indistinguishable from zero, no evidence is found to support my hypotheses.

Modeling election competitiveness

Because of the difficulties in estimating a model of electoral outcome that samples from all primary candidates, this section uses a measure of electoral competition to test the applicable hypotheses. The Herfindahl-Hirschman index is used by economists to calculate competition within a market (Herfindahl 1950; Hirschman 1945). The index value for a particular market scales the proportion of the market controlled by each firm by calculating the sum of squares of the firms' market shares:

$$\text{herfindahl}_j = \sum_{i=1}^N p_i^2$$

where j is the market, N is the number of firms within the market, and p_i^2 is the squared value of the proportion of the market controlled by each firm. In the context of this thesis, each primary election is analogous to a market, and the vote proportion of each candidate represents a firm's market share. Other political scientists have used a Herfindahl index to operationalize competition in various contexts: congressional issue focus (Koford 1989), dividedness of a party system (Molinar 1991), and national ethnic fractionalization (Cederamn and Girardin 2007).

Lower values on the index represent higher levels of primary competition; higher values represent low competition. The index is theoretically bounded at the low end by 0, and it is bounded by 1 at the high end. Each primary that features just one candidate receives an index value of 1.

Although the dependent variable for this model is derived from individual-level vote outcomes, the unit of analysis for this model is party primaries within a congressional district, since electoral competition is a district-level measure. The model includes three district-level dummy variables: *incum*, *fr*, and *redist*. Each regressor is operationalized exactly the same as in the candidate emergence models. The *voteprior* and *prez* variables are also operationalized as they were in the candidate emergence models. I interact *voteprior* with *incum* to model the expectation that incumbent vulnerability or incumbency advantage only affects elections in which the incumbent is actually on the ballot. The final variable included in the model is *totalqual* which is the total number of challengers with elected political experience (minus the incumbent) in a primary. This variable ranges from 0 to 8, with a mean of .2379 and a standard deviation of .6222. I exclude a control variable for candidate number because this information is implicitly included in the calculation of the dependent variable. The lower bound of the index is dictated by the number of candidates within the primary. The lower bound of the Herfindahl index with N candidates is $1/N$.²⁸ A primary with two candidates cannot have a calculated Herfindahl value of less than .5. With four candidates the lower bound is .25.

[Table 12 about here]

The results from the OLS regression with district Herfindahl index values as the dependent variable are found in Table 12. The betas in this table are interpreted in the scale of

²⁸ A market is most divided when each firm controls in equal proportion of the market, $\frac{1}{N}$ %. Summing these market shares across the range of N, we find that $\sum (\frac{1}{N})^2 = \frac{1}{N}$.

the Herfindahl index. For ease of interpretation, I take the change in the Herfindahl index associated with an $X\beta$ change in a dependent variable and transform it back into a vote proportion metric. For an election featuring two candidates²⁹, the change in the winning candidate's vote proportion is calculated by the equation³⁰:

$$\Delta p = \frac{\sqrt{4p_0^2 - 4p_0 + 1 + \gamma}}{2} - \frac{\sqrt{4p_0^2 - 4p_0 + 1}}{2}$$

in which X is the unit change in the independent variable of interest, β is the coefficient associated with that independent variable, and p_0 is the vote proportion received by the winning candidate before the addition of the effect of the independent variable. An γ change $X\beta$ in an explanatory variable has a linear effect on the value of the Herfindahl index value, but this equation demonstrates that the effect on the change in vote proportion is conditioned on the level of the vote proportion without the γ change. This non-linear relationship necessitates a graphical display of the model's results to glean a better understanding of the impact of the independent variables.

[Figures 7 and 8 about here]

As should be expected, races that feature incumbents are significantly less competitive than those without incumbents. The dashed line in Figure 7 demonstrates that although races that feature an incumbent are less competitive, *ceteris paribus*, the effect of incumbency diminishes across the range of p_0 . Figure 8 plots the effect on a winning candidate's vote proportion for each variable moving from a value of zero to one. On average, the winner of a two-candidate election with an incumbent receives between 10 and 20% more votes than the winner of a similar election that does not feature an incumbent. The results in Table 12 also reiterate the findings of the

²⁹ Because of increases in dimensionality, interpretation of cases in which the size of the candidate pool is larger than two becomes incredibly difficult.

³⁰ A derivation of this equation is found in Appendix 1.

candidate emergence models that district partisanship has a significant positive effect on the level of primary competition. The more strongly in favor of one party a district is, the more competitive that party's primaries will be in the district.

The *totalqual* variable's coefficient provides some evidence that supports hypothesis 11. The negative significant coefficient value suggests that an increase in the number of quality candidates in a primary corresponds to an increase in the competitiveness of that election. This does not explicitly reflect that hypothesized relationship between a candidate's political experience and their electoral success, but it does provide evidence by proxy. The green lines³¹ in Figures 7 and 8 show that the effect of the entrance of a quality candidate decreases the vote proportion received by the winning candidate, but this effect has diminishing returns. Winners of races with a quality candidate receive between 10 and 30% fewer votes than winners of races without one. This finding may seem counterintuitive because quality candidates are expected to receive more votes, but most races with one quality candidate also have either an incumbent or second quality candidate, which would increase overall competition. In addition, the results of the OLS model provide evidence that runs counter to the expectation of hypothesis 8. Redistricting actually seems to drive town electoral competition. Figures 7 and 8 show that redistricting races have winners that receive 5 to 10% more votes than primaries in districts that were not redistricted. The model also tells us that incumbent tenure and incumbent vulnerability do not significantly impact electoral competition.

³¹ The range of these graphs is bounded by about .75 and 1, rather than .5 and 1 like the other graphs. The lower bound of the Herfindahl index for a two candidate election is .5. Because the coefficient's value is negative and the Herfindahl index cannot take a value lower than .5, the graph of the variable's effect is bounded.

CHAPTER 7

DISCUSSION

Revisiting the candidate emergence hypotheses

This section will discuss the support, or lack thereof provided by the statistical analyses for the hypotheses laid out earlier in the thesis. The two hypotheses for which the strongest supporting evidence was found were the first two. In both of the candidate emergence analyses, the largest significant betas were for the two dummy variables which measured whether the incumbent sought reelection. For the Poisson model, the only two variables with a larger change in the predicted number of candidates than the *incum* and *incotherparty* were *pres* and the *unemployXpartyinout* interaction term. The change across the range of this second set of variables does not represent a realistically possible jump in each independent variable's value. District presidential vote would never suddenly jump from 0 to 100% between election cycles; likewise a surge in unemployment from 2.3 to 14.9% is not impossible, but it is extremely unlikely to occur. Thus the two incumbency variables had the largest realistic impact on overall candidate emergence. In the logit model, the only variable that had a larger change in predicted probability than *incum* or *incotherparty*, moving from the variable's minimum to its maximum, was presidential vote, but again a change in district presidential vote from 0 to 100% is unrealistic. Not only do these models provide strong evidence in support of hypothesis 1, the data reaffirm the contention of a substantial body of previous literature – that incumbency is the strongest predictive factor of candidate emergence.

The logit analysis also provides evidence to support the second hypothesis. A 95% confidence interval constructed around the coefficient for each of the two variables does not overlap with the 95% confidence interval of the other variable's coefficient. In terms of quality candidate emergence, incumbents seeking reelection have a significantly strong deterrent effect on their own party's primary than the opposing party. The Poisson model, however, does not provide such clear evidence. The 95% confidence intervals overlap for the two variables' coefficients. There is no overlap between the two incumbent groups across the range of prior presidential and congressional vote in Figures 2 and 3, but Figure 4 shows that at high levels of unemployment, the deterrent effect of incumbency is not significantly different between the two parties. These mixed pieces of information provide some evidence to reject the null hypothesis in favor of hypothesis 2, although the evidence is not especially strong.

The evidence related to hypothesis 3 is similarly mixed. The Poisson model shows that the presence of a freshman member of Congress in a seat does not significantly impact overall candidate emergence in that district. In contrast, the logit regression demonstrates that quality challengers are 9.31% more likely to run when a freshman Representative occupies the district's seat. These divergent results may indicate something about the strategic decision making processes by experienced and inexperienced politicians. Perhaps quality candidates are more astute in assessing when an incumbent may best be defeated and thus run in open seat races and when the incumbent is early in his or her tenure in Congress. Non-quality candidates, on the other hand, may not make this discrimination in their decision making process, yielding a result that is not statistically significant in the event count model.

Hypothesis 5 posits that more vulnerable incumbents will tend to face more primary competition. The coefficients for this variable, *voteprior*, in both candidate emergence models

are significant, although the substantive significance of this variable is questionable. The predicted number of candidates that occurs as an incumbent performs better in the previous elections decreases for both parties, but the actual change is quite minimal. For the party opposing the incumbent the number of expected candidates drops from 1.5 to about 1.4; for the incumbent's party the average decreases from 1.1 to 1. The effects on quality candidate emergence are by no means overwhelming either. For the opposition party, the probability of quality candidate emergence drops from 30% to 15% and the 95% confidence bands at the two ends of the variable's range do not overlap. But for the incumbent's party the predicted probability changes less than 3% across the range of *voteprior*. These findings suggest that the hypothesized relationship between incumbent vulnerability and candidate emergence is a weak one at best. As a consequence, however, these findings strengthen the support for the first two hypotheses because they indicate that incumbents are good at deterring electoral competition, no matter how well or poorly they did in the previous election.

There was no evidence found in the candidate emergence analyses to reject the null hypothesis related to redistricting (hypothesis 7). Neither model had a statistically significant coefficient for the *redist* variable, suggesting that being in a redistricting year does not impact overall patterns of candidate emergence. These findings may just be an artifact of the time period studied though. Because there are only six elections in the dataset, there may indeed be a relationship between the two variables, but it does not show up because of other idiosyncrasies of elections in the 2000s. The null findings related to redistricting run counter to the conclusions of Hetherington, Larson, and Globetti (2003) who find that temporal proximity to redistricting significantly impacts candidate emergence patterns. The findings of this thesis do not negate

their contributions though; they only suggest that a broader time period ought to be analyzed to fully reassess the conclusions related to redistricting.

Aside from the first two hypotheses, the hypothesis that received the strongest support from the statistical models was hypothesis 9 which conjectures that favorable district partisanship will translate into increases numbers of candidates running. District partisanship, as measured by previous presidential vote, seems to be the strongest predictor of candidate emergence in congressional primaries outside of incumbency. Figure 5 is perhaps the most telling. It shows the huge jump in the predicted probability of a quality candidate emerging as district partisanship becomes more favorable. It also charts this trend across open races and races with the incumbent seeking reelection, demonstrating that for all three types of elections, district partisanship plays an important role in the decision making of candidates.

Hypothesis 10 predicts that presidential out-party candidates will be more likely to emerge when economic conditions are weak, and in-party candidates will emerge when national tides are strong. There is some support for this hypothesis, although future research will be necessary to offer more definite conclusions. The variables are statistically significant in the event count model, but the graphs across the range of unemployment rates in Figure 4 only shows a significant impact for the president's party. The confidence bands for the presidential out-party do not spread out enough to show statistical significance for this variable. This suggests that if there is an impact that national conditions have on candidate emergence, it is only for the president's party. When the economy is strong, voters in the president's party may tend to grow more complacent. The impact of national tides does not seem to impact quality candidate emergence at the primary stage. The null findings of the logit model runs counter to previous research which has shown that experienced candidates make strategic decisions when

deciding to run for Congress (see Jacobson and Kernell 1981). Rather than this analysis tearing down a significant piece of established congressional election literature, it is more likely that my crude operationalization of national conditions as state level unemployment rate could be muting the true relationship. Future work could refine this measure, collecting unemployment or GDP change data on the district level.

The final hypothesis that necessitated a district level analysis was hypothesis 12. The Poisson model provided evidence to support the idea that fewer candidates will emerge when ballot access in a primary system is more limited. In contrast, the null results of the logit model could indicate a more nuanced relationship. Quality candidates may not consider this to be a factor when they run because ballot access laws are generally static across election cycles, so the type of primary law does not make any one election year more or less favorable for an individual candidate.

Revisiting the election outcome hypotheses

The remaining four hypotheses necessitated empirical testing using dependent variables that captured the outcome of congressional primary elections. The first of these hypotheses was hypothesis 4 which posited that freshman incumbents would receive a smaller vote proportion than more tenured incumbents. No evidence to support this hypothesis was found in any of the three models of electoral outcomes. As I discussed in the theory section, previous research has found that most incumbents seeking reelection for the first time receive a boost in their general vote share over their initial election. The main explanation for this is that incumbents tend to face weaker challengers in their first reelection bid, compared to their competition in their first election. The models suggest that the magnitude of the impact of having a freshman incumbent

in a race is not different from zero. These null findings indicate that freshman fare neither better nor worse in their first renomination bid.

Hypothesis 6 argues that weak prior general election performances for incumbents will translate into weak primary performances and vice versa. There is some evidence found to support this hypothesis, although it is not particularly strong. The analysis of incumbent vote percentages shows that as their previous vote share increases, the incumbent does better in the primaries. The actual effect in terms of percentages is quite small though: only about 2 or 3 percentage points. The OLS model of challenger vote outcomes also provides positive evidence to support the hypothesis. This model of vote proportions of challengers against incumbents tells us that as incumbents do better in the past general election, challengers do worse in the next primary election. The magnitude of this effect is between 7 and 13% for an individual challenger. Also, the variable did not have a significant impact on the OLS models that excluded races with an incumbent indicating that the variable is indeed capturing incumbency advantage and incumbent vulnerability, rather than district partisanship. The OLS regression of the electoral competition index does not yield significant results for this variable, indicating that primaries are no more or less competitive based on the political strength or weakness of the incumbent member of Congress. The implication of this finding is difficult to discern, although it may be a consequence of the operationalization of the dependent variable in the model. Weak incumbents may indeed do worse in the primaries when they are vulnerable, but the overall district-level competitiveness of the election may not be substantially changed. Taken together, however, the findings across the three models seem to suggest that primary election outcomes are in part dictated by the strength or weakness of the incumbent, although the magnitude of this impact should not be considered especially strong.

Just like the incumbent tenure hypothesis, the redistricting hypothesis was found to have little support from the empirical models. The coefficient for the redistricting dummy variable was not significant for the model of incumbent vote share nor for the model of challenger vote share. The model that utilized the Herfindahl index actually provided evidence contrary to the hypothesis. The positive and significant coefficient in this analysis indicates that electoral competition is actually depressed in redistricting years. Thus it is safe to conclude that these analyses provide no support for hypothesis 8. This lack of support could be a result of the crude measure of redistricting. Unlike Hetherington, Larson and Globetti's 2003 piece on redistricting, the variable in this analysis is coded as a mere dichotomous variable, rather than a more continuous measure of how many years it has been since the previous district redrawing.³² The null results also potentially come as a consequence of a relatively small time period analyzed, at least in terms of the number of redistricting cycles that are included. The only instances of redistricting captured by the dataset are the decennial reapportionment of 2002 and mid-decade district redraws in Texas in 2004 and Georgia in 2006.

The final hypothesis that is tested as hypothesis 11, that quality candidates will outperform non-quality candidates, *ceteris paribus*. Strong empirical evidence to support this hypothesis was found across the board. The regression of incumbent vote shares shows that running against a quality challenger decreases the incumbent's vote proportion by almost ten points. Likewise, the similar regression of challenger votes indicates that quality challengers receive about 12% more vote than non-quality candidates. The electoral competition model also shows that the number of quality candidates running in an election drives up the competitiveness of that primary. The results of these models yield strong evidence to suggest that the trends

³² See also Crespin 2005 which uses GIS to approximate the extent of redistricting that occurs.

identified by the congressional general election literature regarding the impact of challenger quality also exist in congressional primary elections.

CHAPTER 8

LIMITATIONS AND PROSPECTS FOR FUTURE RESEARCH

Without a doubt, the biggest room for improvement for the analyses conducted in this thesis is in the modeling of individual vote outcomes. As described earlier, these models potentially suffer from omitted variable bias because they exclude any measure of candidate spending. Although a significant undertaking, the collection of these data will be a necessary step in more thoroughly describing the outcomes of congressional primaries. The bigger endeavor for improving the tests of hypotheses related to vote outcomes is generating an unbiased estimator for the dependent variable of vote proportion. The statistical analyses in this thesis provide initial indications of the validity of the hypothesized trends, but in order to be more conclusive in my assessments of the theories, I would need to use a more sophisticated methodology. Perhaps the solution to the problem of compositional data lies in the creation of a new MLE optimizer that constrains the predicted values of vote outcomes to be bounded by 0 and 1 and to add up to 1 within a primary. An alternative route could be simply estimating a conditional logit model of individual candidate success. The problem with this approach is that some of the variables that are of interest, in terms of hypothesis testing, do not vary within primary elections and thus are dropped by a conditional logit model. A similar model that allows for a multi-level structure to the data could potentially provide fruitful conclusions.

The other major area for improvement of the research would be to refine the operationalizations of some of the explanatory variables. Many of these possibilities have already been discussed, but the biggest opportunity for a strengthened independent variable

measure is with the national trend variable. The candidate emergence analyses had this variable operationalized as the rate of unemployment within the state during the year of the election. This may not generate significant problems for economically homogenous states with small populations like Alaska, Vermont, or Wyoming, but it could potentially be a problem for larger states with diverse economies. It is not particularly reasonable to assume that the rate of unemployment for every congressional district in a large state like California is equal to the unemployment rate in state as a whole. However, avoiding this problem necessitates collection of district-level economic data for each election year. Unfortunately, these data are not readily available and the process of collecting them would be tedious and imprecise.³³

Another way in which this analysis could be improved would be by expanding the timeframe beyond just the primaries in the 2000s. Working backward into the 1990s and 1980s could mitigate any year-specific effects, such as that of 2010, that may be driving statistical results when only six elections are studied. Expansion of the dataset would also allow for the inclusion of a wider variety of electoral circumstances. Broadening into the 1990s would allow for inclusion of the Republican revolution elections of 1992 and 1994, for example. The dataset could also be enhanced by more in-depth analyses of year-specific idiosyncrasies, rather than an exclusively longitudinal expansion. The primaries data from 2010, accompanied with general election results from the same year, could be harnessed to assess the electoral implications of the Tea Party in that election year. The data from other years could also be used to explore the impacts of specific roll call votes in order to see whether moderate voting records negatively affect incumbents in the primaries.

³³ The imprecision comes as a consequence of the fact that unemployment data can be collected on the county level, but congressional districts often split counties into multiple pieces. Likewise, many rural districts encompass several counties. Because unemployment rate is a percentage, it would be also necessary to collect population data by county to ensure an accurate measure.

CHAPTER 9

CONCLUSION

Perhaps the biggest takeaway message from this thesis is that, more than anything, incumbency matters in the context of congressional primary elections. This finding should come as no surprise to anyone familiar with the literature on the incumbency advantage. Just as incumbent-contested seats tend to be less competitive than open-seat races in the general election, this thesis provides evidence that incumbents tend to run in significantly less competitive primary elections, as quantified by election outcomes. This thesis also provides evidence that incumbents are able to deter candidate emergence not only in their own party, but also in the primary election of the opposing party. Aside from incumbency, district partisanship is a factor that has a significant impact on primary elections under almost any electoral circumstance. District partisanship affects candidates' emergence decisions because they recognize the implications that weak partisan circumstances have on their chances of winning the seat in the general election. The overall competitiveness of a primary election also increases with favorable district partisanship.

This analysis also demonstrates that some of the factors that influence the dynamic of primary elections have a much more muted effect than either incumbency or district partisanship. The candidate emergence models show that incumbent vulnerability has an impact on the decision making process of politicians, both experienced and inexperienced politically, but the magnitude of the effects are quite minor. Vulnerability and incumbency advantage also impacts the actual outcome of primaries. A higher level of success in a general election typically

translates into more success in the primary two years later. Incumbent tenure in Congress is only found to affect the decision making of experienced politicians. Quality candidates run more often against first term incumbents. This trend related in incumbent tenure, however, does not translate into increased electoral vulnerability for freshman members of Congress.

Likewise, redistricting is not found to impact vote outcomes. Redistricted incumbents are no more vulnerable to electoral defeat than incumbents with the same district boundaries. In addition, redistricting is not found to significantly affect candidate emergence decisions. Hetherington, Larson, and Globetti (2003) find that redistricting can influence candidates' decisions to run for a seat, but these conclusions are not substantiated by the analysis in this thesis. The thesis also demonstrates the robustness of the conclusions related to candidate experience reached by scholars that only explored general elections. The quality candidate emergence models show that those with elected experience tend to be strategic in their decision of when to run for Congress. Once these candidates choose to run, they are substantially more effective at garnering support among the electorate.

REFERENCES

- Abramowitz, Alan I. 1991. "Incumbency, Campaign Spending, and the Decline of Competition in U.S. House Elections." *The Journal of Politics* 53 (1):34-56.
- Abramowitz, Alan I., Brad Alexander, and Matthew Gunning. 2006. "Incumbency, Redistricting, and the Decline of Competition in U.S. House Elections." *The Journal of Politics* 68 (1):75-88.
- Adams, James, and Samuel Merrill III. 2008. "Candidate and Party Strategies in Two-Stage Elections Beginning with a Primary." *American Journal of Political Science* 52 (2):344-59.
- Banks, Jeffrey S., and Roderick Kiewiet. 1989. "Explaining Patterns of Candidate Competition in Congressional Elections." *American Journal of Political Science* 33 (4):997-1015.
- Bianco, William T. 1984. "Strategic Decisions on Candidacy in US Congressional Districts." *Legislative Studies Quarterly* 9 (2):351-64.
- Black, Gordon S. 1972. "A Theory of Political Ambition: Career Choices and the Role of Structural Incentives." *The American Political Science Review* 66 (1):144-59.
- Bond, Jon R., Cary Covington, and Richard Fleisher. 1985. "Explaining Challenger Quality in Congressional Elections." *The Journal of Politics* 47 (2):510-29.
- Born, Richard. 1981. "The Influence of House Primary Election Divisiveness on General Election Margins, 1962-76." *The Journal of Politics* 43 (3):640-61.
- Box-Steffensmeier, Janet M. 1996. "A Dynamic Analysis of The Role of War Chests in Campaign Strategy." *American Journal of Political Science* 40 (2):352-71.
- Brady, David W., Hanhrie Han, and Jeremy C. Pope. 2007. "Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32:79-105.
- Bullock, Charles S., Ronald Keith Gaddie, and Anders Ferrington. 2002. "System Structure, Campaign Stimuli, and Voter Falloff in Runoff Primaries." *The Journal of Politics* 64 (4):1210-24.
- Bullock, Charles S., and Loch K. Johnson. 1992. *Runoff Elections in the United States*. Chapel Hill: University of North Carolina Press.

- Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *The American Political Science Review* 96 (1):127-40.
- Canon, David T. 1993. "Sacrificial Lambs or Strategic Politicians? Political Amateurs in U.S. House Elections." *American Journal of Political Science* 37 (4):1119-41.
- Carey, John M., and John Polga-Hecimovich. 2006. "Primary Elections and Candidate Strength in Latin America." *The Journal of Politics* 68 (3):530-43.
- Carson, Jamie L., Erik J. Engstrom, and M. Roberts Jason. 2006. "Redistricting, Candidate Entry, and the Politics of Nineteenth-Century U.S. House Elections." *American Journal of Political Science* 50 (2):283-93.
- Cederman, Lars-Erik, and Luc Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." *The American Political Science Review* 101 (1):173-85.
- Cover, Albert D., and Bruce S. Brumberg. 1982. "Baby Books and Ballots: The Impact of Congressional Mail on Constituent Opinion." *The American Political Science Review* 76 (2):347-59.
- Cox, Gary W., and Jonathan N. Katz. 1996. "Why Did the Incumbency Advantage in U.S. House Elections Grow?" *American Journal of Political Science* 40 (2):478-97.
- Crespin, Mike. 2005 "Using Geographic Information Systems to Measure District Change, 2000-02." *Political Analysis* 13 (13): 253-260.
- Epstein, David, and Peter Zemsky. 1995. "Money Talks: Deterring Quality Challengers in Congressional Elections." *The American Political Science Review* 89 (2):295-308.
- Ezra, Marni. 2001. "The Benefits and Burdens of Congressional Primary Elections." In *Congressional Primaries and the Politics of Representation*. eds. Peter F. Galderisi, Marni Ezra, and Michael Lyons. Lanham, MD: Rowman & Littlefield Publishers, pp. 48-61.
- Fenno, Richard F. 1978. *Home Style : House Members in Their Districts*. New York: Harper Collins.
- Friedman, John N., and Richard T. Holden. 2009. "The Rising Incumbent Reelection Rate: What's Gerrymandering Got to Do With It?" *The Journal of Politics* 71 (2):593-611.
- Galerisi, Peter F., and Marni Ezra. 2001. "Congressional Primaries in Historical and Theoretical Context." In *Congressional Primaries and the Politics of Representation*. eds. Peter F. Galderisi, Marni Ezra, and Michael Lyons. Lanham, MD: Rowman & Littlefield Publishers, pp. 11-26.
- Geer, John G. 1986. "Rules Governing Presidential Primaries." *The Journal of Politics* 48 (4):1006-25.

- . 1988. "Assessing the Representativeness of Electorates in Presidential Primaries." *American Journal of Political Science* 32 (4):929-45.
- Gelman, Andrew, and Gary King. 1990. "Estimating Incumbency Advantage without Bias." *American Journal of Political Science* 34 (4):1142-64.
- Goodliffe, Jay. 2001. "The Effect of War Chests on Challenger Entry in U.S. House Elections." *American Journal of Political Science* 45 (4):830-44.
- Green, Donald Philip, and Jonathan S. Krasno. 1988. "Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections." *American Journal of Political Science* 32 (4):884-907.
- Herfindahl, Orris. 1950. "Concentration in the Steel Industry." PhD dissertation. Columbia University.
- Herrnson, Paul S. 2008. *Congressional Elections: Campaigning at Home and in Washington*. 5th ed. Washington, D.C.: CQ Press.
- Herrnson, Paul S., J. Celeste Lay, and Atiya Kai Stokes. 2003. "Women Running "as Women": Candidate Gender, Campaign Issues, and Voter-Targeting Strategies." *The Journal of Politics* 65 (1):244-55.
- Hetherington, Marc J., Bruce Larson, and Suzanne Globetti. 2003. "The Redistricting Cycle and Strategic Candidate Decisions in U.S. House Races." *The Journal of Politics* 65 (4):1221-34.
- Hirano, Shigeo, James Snyder, Stephen Ansolabehere, and John Mark Hansen. 2010. "Primary Elections and Partisan Polarization in the U.S. Congress." *Quarterly Journal of Political Science* 5 (2):169-91.
- Hirschman, Albert. 1945. *National Power and the Structure of Foreign Trade*. Berkeley: University of California Press.
- Honaker, James, Jonathan N. Katz, and Gary King. 2002. "A Fast, Easy, and Efficient Estimator for Multiparty Electoral Data." *Political Analysis* 10 (1):84-100.
- Jackson, John E. 2002. "A Seemingly Unrelated Regression Model for Analyzing Multiparty Elections." *Political Analysis* 10 (1):49-65.
- Jacobson, Gary C. 1989. "Strategic Politicians and the Dynamics of U.S. House Elections, 1946-86." *The American Political Science Review* 83 (3):773-93.
- . 2009. *The Politics of Congressional Elections*. 7th ed. New York: Pearson/Longman.
- Jacobson, Gary C., and Samuel Kernell. 1981. *Strategy and Choice in Congressional Elections*. New Haven: Yale University Press.

- Kanthak, Kristin and Rebecca Morton. 2001. "The Effects of Electoral Rules on Congressional Primaries." In *Congressional Primaries and the Politics of Representation*. eds. Peter F. Galderisi, Marni Ezra, and Michael Lyons. Lanham, MD: Rowman & Littlefield Publishers, pp. 116-131.
- Katz, Jonathan N., and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data." *The American Political Science Review* 93 (1):15-32.
- Kaufmann, Karen M., James G. Gimpel, and Adam H. Hoffman. 2003. "A Promise Fulfilled? Open Primaries and Representation." *The Journal of Politics* 65 (2):457-76.
- Kazee, Thomas A. 1983. "The Deterrent Effect of Incumbency on Recruiting Challengers in U. S. House Elections." *Legislative Studies Quarterly* 8 (3):469-80.
- Koford, Kenneth. 1989. "Dimensions in Congressional Voting." *The American Political Science Review* 83 (3):949-62.
- Lawless, Jennifer L., and Kathryn Pearson. 2008. "The Primary Reason for Women's Underrepresentation? Reevaluating the Conventional Wisdom." *The Journal of Politics* 70 (01):67-82.
- Lazarus, Jeffrey. 2005. "Unintended Consequences: Anticipation of General Election Outcomes And Primary Election Divisiveness." *Legislative Studies Quarterly* 30:435-61.
- . 2008. "Buying In: Testing the Rational Model of Candidate Entry." *The Journal of Politics* 70 (03):837-50.
- Maisel, Louis Sandy. 1986. *From Obscurity to Oblivion: Running in the Congressional Primary*. 2nd ed. Knoxville: University of Tennessee Press.
- Maisel, Sandy, and Walter J. Stone. 1997. "Determinants of Candidate Emergence in U. S. House Elections: An Exploratory Study." *Legislative Studies Quarterly* 22 (1):79-96.
- . 2001. "Primary Elections as a Deterrence to Candidacy for the U.S. House of Representatives." In *Congressional Primaries and the Politics of Representation*. eds. Peter F. Galderisi, Marni Ezra, and Michael Lyons. Lanham, MD: Rowman & Littlefield Publishers, pp. 29-47.
- Mayhew, David R. 2004. *Congress: The Electoral Connection*. 2nd ed. New Haven: Yale University Press.
- Molinar, Juan. 1991. "Counting the Number of Parties: An Alternative Index." *The American Political Science Review* 85 (4):1383-91.
- Norrander, Barbara. 1989. "Ideological Representativeness of Presidential Primary Voters." *American Journal of Political Science* 33 (3):570-87.

- Rohde, David W. 1979. "Risk-Bearing and Progressive Ambition: The Case of Members of the United States House of Representatives." *American Journal of Political Science* 23 (1):1-26.
- Stone, Walter J., and L. Sandy Maisel. 2003. "The Not-so-Simple Calculus of Winning: Potential U.S. House Candidates' Nomination and General Election Prospects." *The Journal of Politics* 65 (4):951-77.
- Tomz, Michael, Joshua A. Tucker, and Jason Wittenberg. 2002. "An Easy and Accurate Regression Model for Multiparty Electoral Data." *Political Analysis* 10 (1):66-83.
- Ware, Alan. 2002. *The American Direct Primary: Party Institutionalization and the Transformatino in the North*. Cambridge: Cambridge University Press.
- Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57 (298):348-68.

Table 1: Primaries with zero candidates versus primaries with at least one candidate

	Potential primary elections with zero candidates	Primary elections with at least one candidate
Total number	349	4541
2000	57	741
2002	80	738
2004	66	752
2006	53	765
2008	60	754
2010	33	791
Democrats	157	2288
Republicans	192	2253
Districts with an open seat	6	384
Mean district presidential vote for the party in previous election	35.59%	51.11%
Mean congressional vote for the party in previous election	30.44%	50.79%
Mean tenure of House incumbent	10.8 years	11.1 years
Median	9 years	9 years

Table 2: Candidates by year and party

Year	Candidates	Average Number of Candidates	Incumbents	Challengers	Quality Challengers	State Legislators
2000	1144	1.54	368	776	151	75
2002	1196	1.62	377	819	178	84
2004	1190	1.58	385	805	151	74
2006	1249	1.63	383	866	170	72
2008	1256	1.67	374	882	192	79
2010	1686	2.13	375	1311	222	120
Total	7721	1.70	2262	5459	1064	504
GOP						
2000	588	1.58	180	408	74	40
2002	624	1.67	185	439	95	48
2004	607	1.61	195	412	83	42
2006	562	1.54	198	364	80	37
2008	616	1.70	160	456	102	43
2010	1084	2.68	147	937	142	80
Total	4081	1.81	1065	3016	576	290
Dems						
2000	556	1.51	188	368	77	35
2002	572	1.57	192	380	83	36
2004	583	1.55	190	393	68	32
2006	687	1.71	185	502	90	35
2008	640	1.64	214	426	90	36
2010	602	1.56	228	374	80	40
Total	3640	1.59	1197	2443	488	214

Table 3: Quality candidate emergence by open/ non-open seat

Year	Total number races	Primaries with no quality challengers	Primaries with at least one quality challenger
<i>Open seats</i>	384	114 (29.7%)	270 (70.3%)
2000	56	14 (25.0%)	42 (75.0%)
2002	89	25 (28.1%)	64 (71.9%)
2004	52	20 (38.5%)	32 (61.5%)
2006	54	13 (24.1%)	41 (75.9%)
2008	62	17 (27.4%)	45 (72.6%)
2010	71	25 (35.2%)	46 (64.8%)
<i>Non-open seats</i>	4157	3645 (87.7%)	512 (12.3%)
2000	685	609 (88.9%)	76 (11.1%)
2002	649	585 (90.1%)	64 (9.9%)
2004	700	619 (88.4%)	81 (11.6%)
2006	711	629 (88.5%)	82 (11.5%)
2008	692	596 (86.1%)	96 (13.9%)
2010	720	607 (84.3%)	113 (15.7%)

Table 4: District partisanship by quality candidate emergence

	Races with no quality candidates		Races with a quality candidate	
	Open seat	Non-open seat	Open seat	Non-open seat
Presidential vote in the district	42.03%	49.96%	57.15%	50.32%
Congressional vote in the district	33.91%	50.72%	53.54%	44.43%

Table 5: Percentage of primary elections won by candidate type

	Number of races	Incumbents	Challengers	Quality challengers	Non- quality challengers	State legislators
2000	740	49.1	50.9	13.8	37.1	7.2
2002	738	49.3	50.7	12.2	38.5	6.2
2004	752	50.7	49.3	11.4	37.9	5.6
2006	765	49.8	50.2	11.8	38.4	4.4
2008	753	49	51	14.1	36.9	6.5
2010	790	46.6	53.4	14.2	39.2	8.1
Total	4538	49.1	50.9	12.9	38	6.3

Table 6: Winning percentages by candidate type

	Incumbents	Challengers	Quality challengers	Non-quality challengers	State legislators
Total	98.8%	42.3%	53.5%	40.6%	55.1%
2000	99.2%	48.5%	65.4%	45.9%	71.6%
2002	98.1%	45.4%	49.7%	45.5%	52.3%
2004	99.2%	46.0%	54.8%	45.7%	53.2%
2006	99.5%	44.3%	52.0%	44.3%	44.7%
2008	98.7%	43.5%	53.8%	41.4%	59.8%
2010	98.4%	32.1%	48.5%	29.1%	51.6%

Table 7: Variable summaries district-level analyses

Variable abbreviation	What the variable measures	Range	Mean (for continuous) or Mode (for dichotomous)
incum	Is the incumbent running in this party's primary	0: open seat 1: incumbent in primary	0
incotherparty	Is the incumbent running in the other party's primary	0: open seat 1: incumbent in other primary	0
redist	Were the district boundaries redrawn since the last election	0: no redistricting 1: redistricting	0
fr	Is the incumbent seeking reelection for the first time	0: not a freshman 1: freshman	0
lawclosed	Is the primary administered under closed primary rules	0: open primary 1: closed primary	0
lawsemi	Is the primary administered under semi-closed or semi-open rules	0: open primary 1: semi-open/closed primary	0
prez	This party's two-party share of this district's presidential vote in the last election	0-100	50
voteprior	Incumbent's two-party share of the previous general election vote	50.01-100	71.64
partyinout	Is this primary for the presidential in- or out-party	0: in-party 1: out-party	0
unemploy	Unemployment rate for the district's state	2.3-14.9	6.45
y02, y04, y06, y08, y10	Dummy variables for year	0: year 2000 1: other year	0
candnumber	Number of candidates running in this party's primary	0-15	1.58
qual	Is a quality candidate running in this primary	0: no quality candidates 1: at least one quality candidate	0

Table 8: Poisson model of candidate emergence in congressional primaries

Dependent variable: Count of number of candidates running in a party's primary

	Coefficient	Standard error	Change in candidate number prediction
Constant	1.01	.10	-
Incum	-1.04*	.034	-1.64
incotherparty	-.74*	.34	-1.32
Redist	.019	.085	-.048
Fr	.044	.036	.11
lawclosed	-.12*	.028	-.28
Lawsemi	-.091*	.0289	-.22
Prez	.015*	.0010	9.93
voteprior	-.0048*	.00080	-2.01
partyinout	-.25*	.065	-.56
unemploy	-.052*	.011	-1.51
unemploy*partyinout	.067*	.0092	4.30
y02	.021	.095	.053
y04	.048	.046	.13
y06	.074	.042	.19
y08	.062	.042	.16
y10	.42*	.071	1.32

n=4890

Log likelihood:-6847.90

AIC: 2.81

LR test: $\chi^2(16) = 1225.78$ ($p < .0001$)Coefficients with a * indicate statistical significance at $p < .05$ (two-tailed)

Table 9: Logit model of quality candidate emergence in congressional primaries

Dependent variable: Quality candidate emergence dummy

	Coefficient	Standard error	Change in predicted probability from min to max
constant	-.18	.44	-
incum	-4.57*	.17	-.69
incotherparty	-2.29*	.14	-.51
redist	.30	.33	.058
fr	.51*	.13	.093
lawclosed	-.057	.11	-.012
lawsemi	-.11	.12	-.022
prez	.059*	.0045	.87
voteprior	-.026*	.0035	-.27
partyinout	.10	.26	.020
unemploy	-.00088	.044	-.0027
unemploy*partyinout	.022	.038	.077
y02	-.47	.37	-.11
y04	-.18	.18	-.038
y06	-.012	.17	-.0024
y08	.056	.17	.011
y10	.15	.30	.030

n=4890

Log likelihood: -1600.96

AIC= .66

Percent correctly classified: 87.53%

PRE: 21.99%

Coefficients with a * are statistically significant at $p < .05$ (two-tailed)

Table 10: OLS model of incumbent vote percentage

	With uncontested races		Without uncontested races	
Response variable	candpct	candpct ³	candpct	candpct ³
Constant	96.64* (.77)	930638.5* (14892.33)	76.99* (.24)	471105.2s* (38268.19)
Fr	.91 (.46)	21694.73* (9007.93)	.99 (1.43)	16908.78 (22381.82)
Redist	-.72 (.40)	-12255.16 (7698.52)	-1.68 (1.23)	-19955.15 (19212.32)
Voteprior	.027* (.010)	436.15* (198.65)	.11* (.032)	1791.79* (495.10)
Cand	-10.88* (.23)	-239850.4* (4548.11)	-3.36* (.58)	-45096.01* (9094.84)
Vsqual	-7.21* (.63)	-53903.84* (12316.38)	-10.55* (1.01)	-142920 (15875.7)
N	2250	2250	580	580
Adjusted R2	.6612	.6739	.3356	.2826
Omnibus F-test	p<.0001	p<.0001	p<.0001	p<.0001

Cells contain betas with standard errors in parentheses

Coefficients with a * are statistically significant at p<.05 (two-tailed)

Table 11: OLS model of challenger vote percentage in contested races

	Races with incumbents		Races without incumbents	
Dependent var.	candpct	candpct ^{1/2}	candpct	candpct ^{1/2}
Constant	43.28* (2.87)	7.053* (.28)	46.91* (.86)	6.87* (.076)
fr	-.78 (1.04)	-.12 (.13)	-.015 (.067)	-.0051 (.0062)
redist	.12 (1.04)	-.0084 (.12)	.73 (.57)	.055 (.052)
voteprior	-.14* (.03)	-.016* (.0031)	.017 (.011)	.00024 (.00097)
cands	-3.05* (1.23)	-.45* (.044)	-4.40* (.34)	-.46* (.030)
qualadv	12.87* (1.23)	1.50* (.13)	11.79* (.66)	1.13* (.060)
N	836	836	3259	3259
Clusters	560	560	1048	1048
Omnibus f-test	p<.0001	p<.0001	p<.0001	p<.0001
Adj. R2	.2545	.3129	.4175	.4590

Cells contain betas with robust standard errors in parentheses

Coefficients with a * are statistically significant at p<.05 (two-tailed)

Table 12: OLS model of primary competitiveness

Dependent variable: Herfindahl index of primary competition

Independent variable	Beta (standard error)
constant	.87* (.028)
incum	.085* (.032)
fr	.015 (.0098)
redist	.022* (.0082)
voteprior	.00019 (.00031)
incracexvoteprior	.00075 (.00046)
prez	-.0017* (.00031)
totalqual	-.15*(.0055)
N	4539
Adjusted R2	.2707
Omnibus F-test	p<.0001

Coefficients with a * are statistically significant at p<.05 (two-tailed)

Figure 1: Average number of candidates in contested primaries

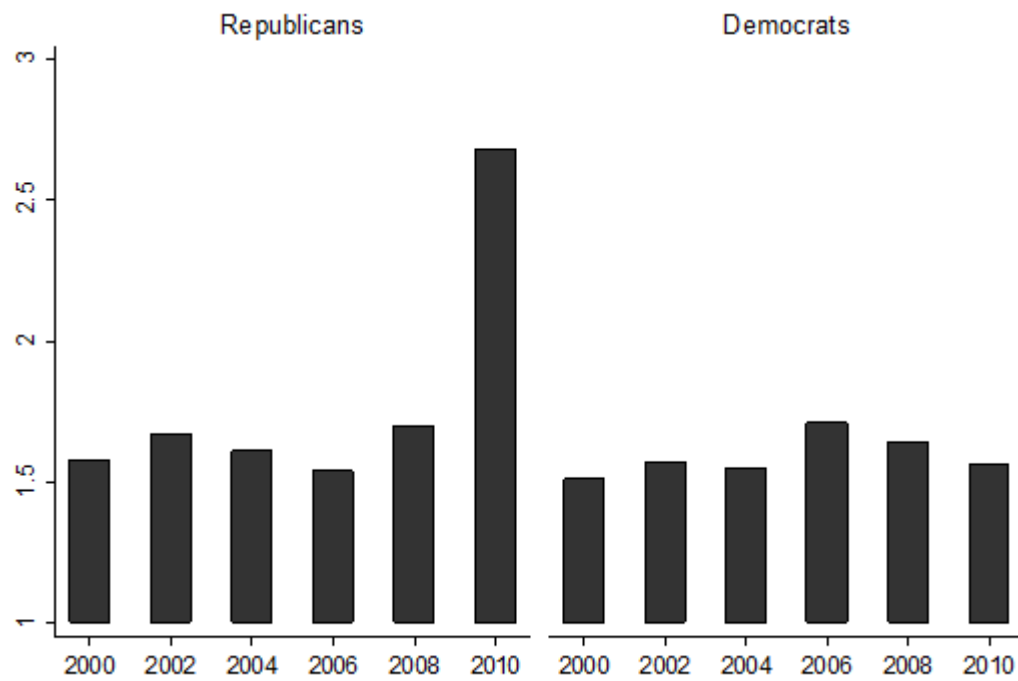


Figure 2: Expected number of candidates by presidential vote

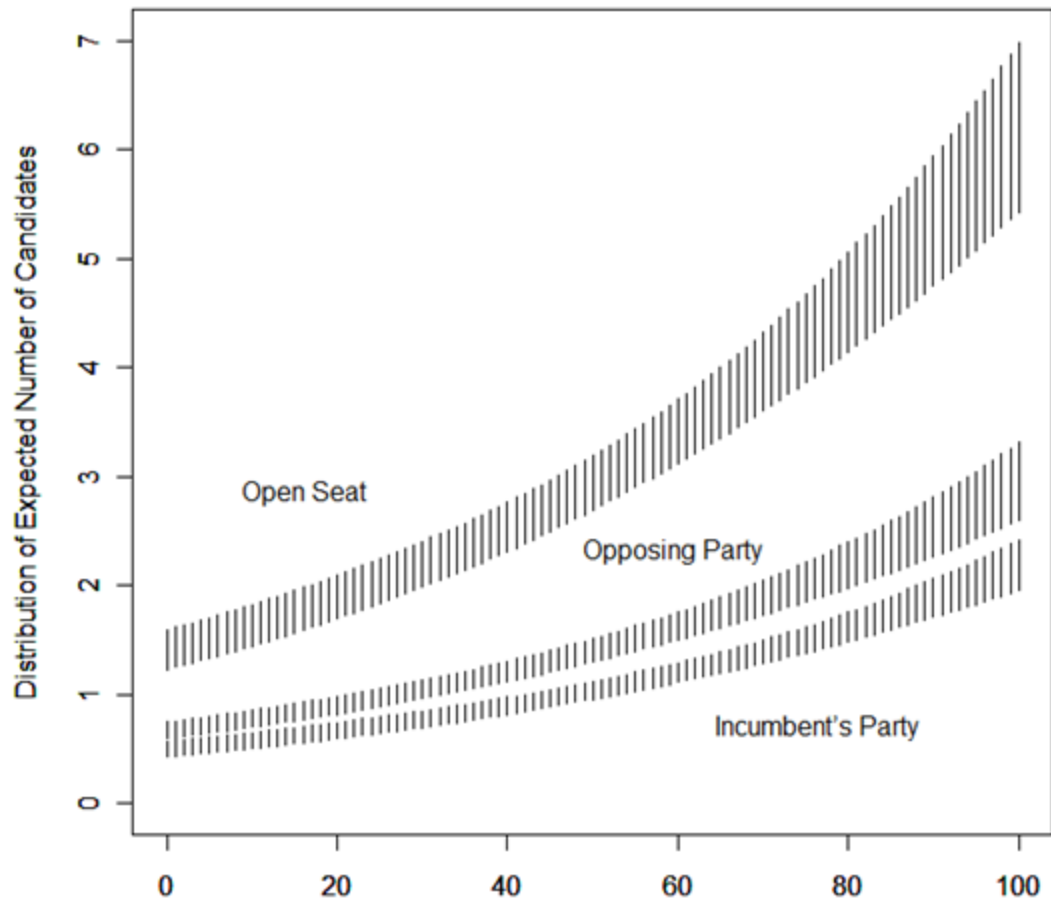


Figure 3: Expected number of candidates by congressional vote

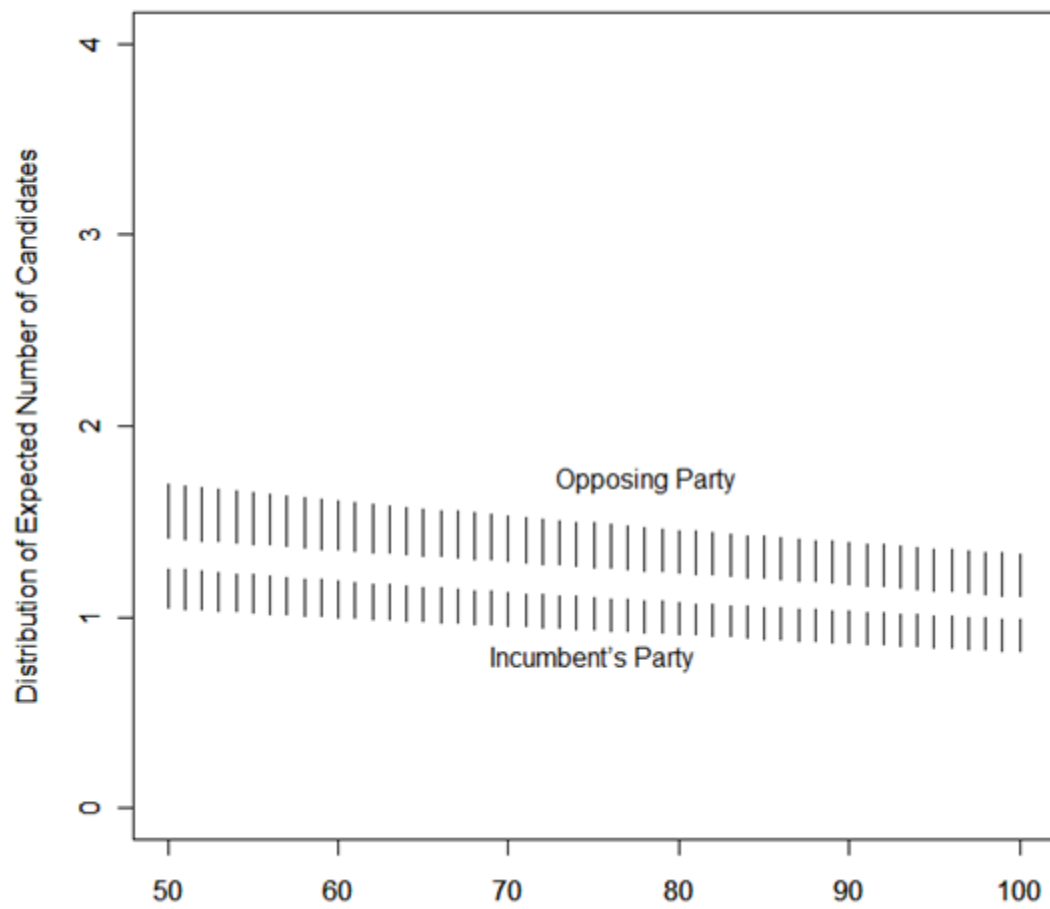


Figure 4: Number of candidates by percent unemployed (presidential in and out parties)

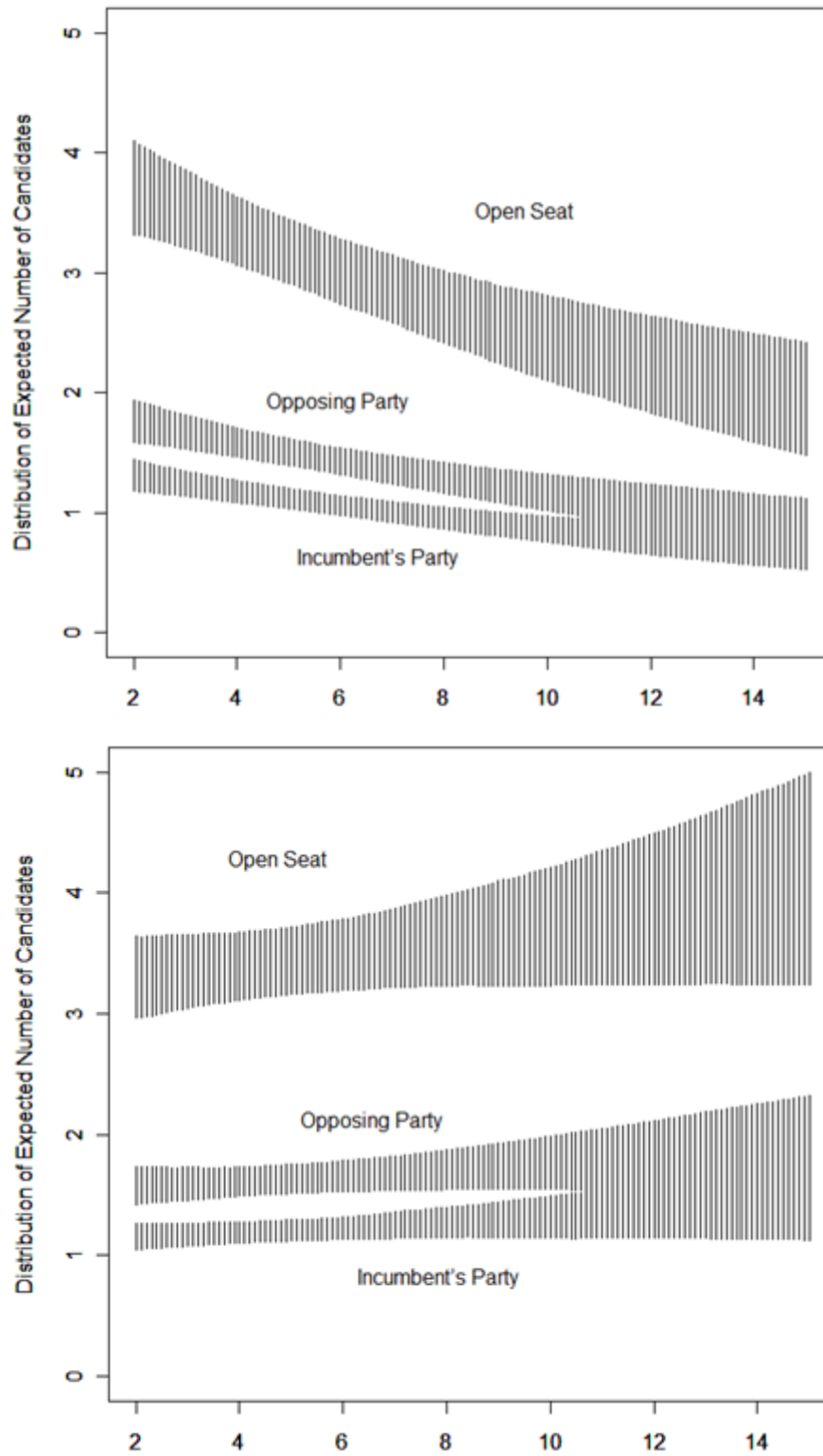


Figure 5: Probability of quality candidate emergence by presidential vote

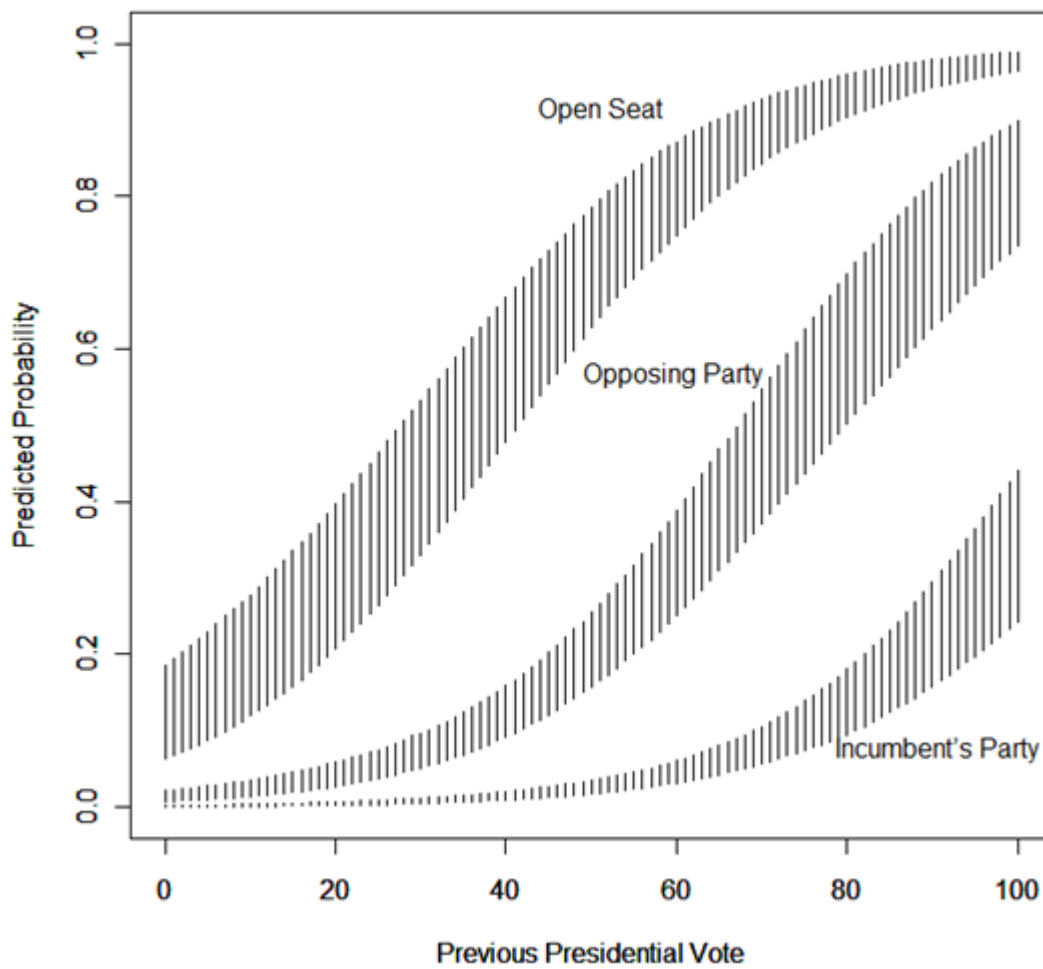


Figure 6: Probability of quality candidate emergence by congressional vote

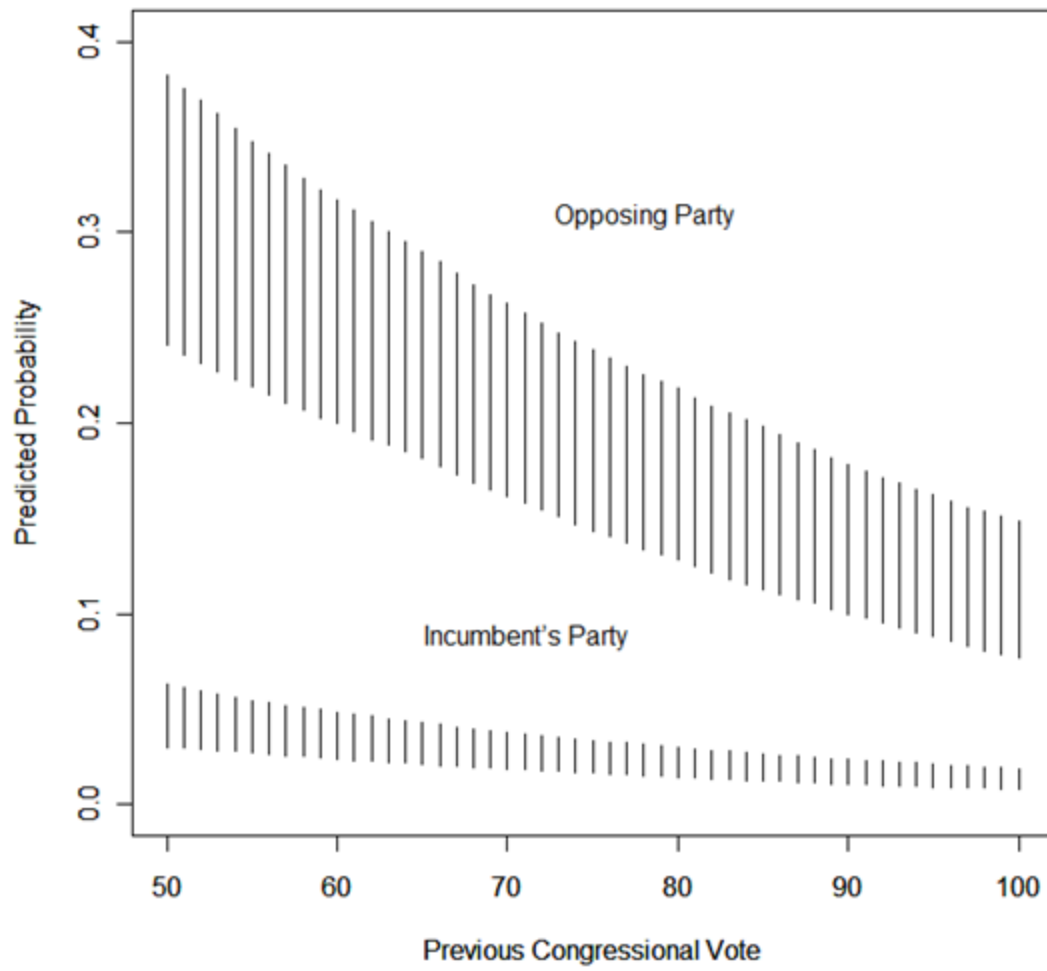


Figure 7: Effect of explanatory variables on competitiveness

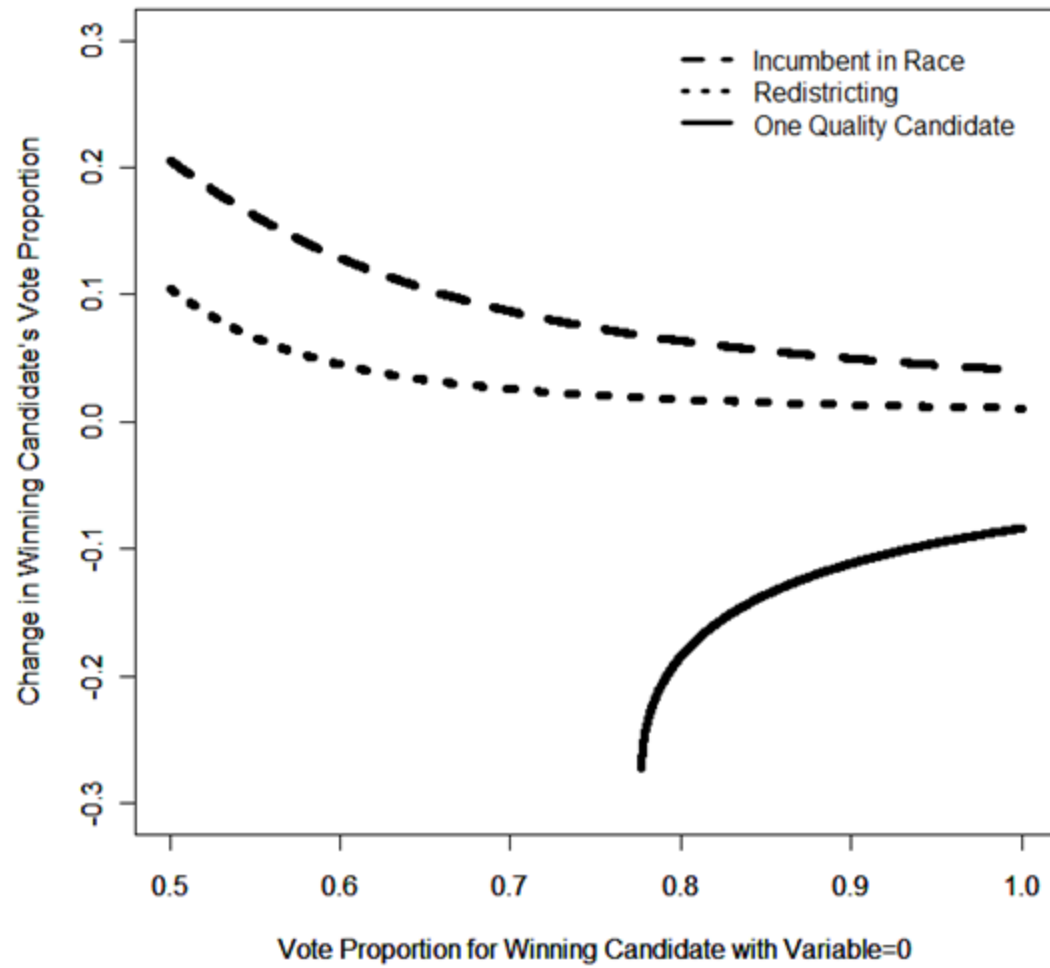
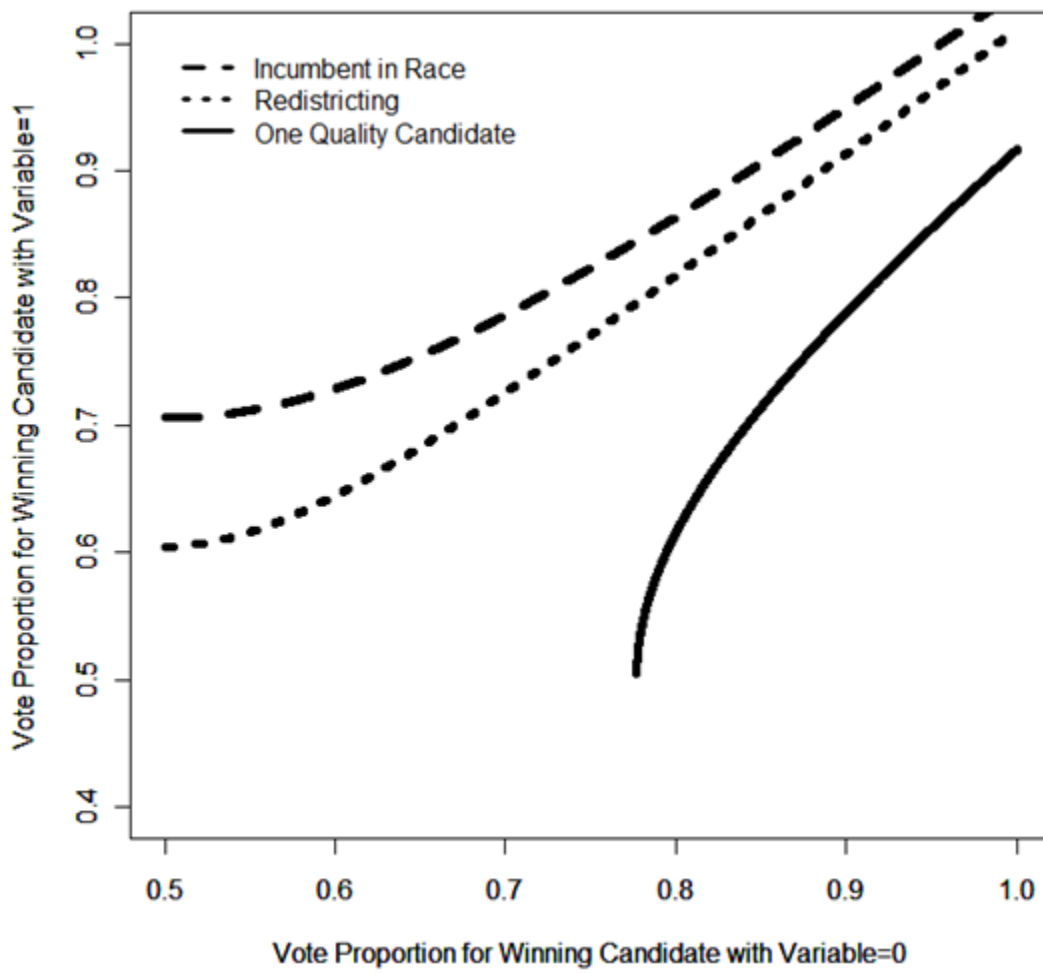


Figure 8: Effect of explanatory variables on vote proportions



Appendix C: Derivation of Herfindahl-related equations

The value of the Herfindahl index of district j is:

$$H_j = \sum_{i=1}^N p_i^2$$

For a district that has two candidates running in the primary, each with vote proportion p :

$$H = p_1^2 + p_2^2$$

Because $p_2 = 1 - p_1$:

$$H = p^2 + (p - 1)^2$$

which simplifies to:

$$H = 2p^2 - 2p + 1$$

Solving for p we get:

$$p = \frac{1}{2} + \frac{\sqrt{2H - 1}}{2}$$

To calculate the change of p associated with a γ change $X\beta$ in the Herfindahl index dependent variable:

$$\Delta p = \frac{\sqrt{2(H_0 + \gamma) - 1}}{2} - \frac{\sqrt{2H_0 - 1}}{2}$$

Substituting $p^2 + (p - 1)^2$ for H and simplifying we get:

$$\Delta p = \frac{\sqrt{4p_0^2 - 4p_0 + 1 + \gamma}}{2} - \frac{\sqrt{4p_0^2 - 4p_0 + 1}}{2}$$