PRIMARY ELECTIONS AND POLITICAL POLARIZATION:
EXPLORINGIDEOLOGICAL HETEROGENEITY IN PRIMARY ELECTORATES

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Government

By

Alexander R. Podkul, M.A.

Washington, DC
June 20, 2019
Primary Elections and Political Polarization: Exploring Ideological Heterogeneity in Primary Electorates

Alexander R. Podkul, M.A.

Dissertation Advisor: Michael A. Bailey, Ph.D.

Abstract

Are congressional primary electorates ideologically representative of nonvoters? Previous literature exploring the policy preferences held by primary voters neglects both cross-cycle changing electoral conditions and cross-constituency political contours. This dissertation proposes a different approach for studying the ideological representativeness of primary electorates by considering heterogeneity across both cycles and constituencies. Leveraging variation in this representativeness, the project further tests the association between primary electorate extremism and political polarization.

To examine this question, this dissertation analyzes both primary voters’ ideological preferences and voting behavior. Studying primary voter ideology, the first data set in this project generates constituency-level measures of primary electorates ideological scores for partisans within states and congressional districts. The measures are developed using ideal points estimated from a battery of policy preference items asked by the Cooperative Congressional Election Study (CCES), which are then modeled at the constituency level using national voter file data. Studying voter behavior, the second data set explores primary voters’ vote choices using an expansive series of district-level exit polls.

The results from this project indicate significant primary voter ideological heterogeneity across electorates of both major political parties at both the state and congressional district levels. Further findings demonstrate that primary electorates in
extreme districts are asymmetrically polarized with Republican primary voters being more extreme under more conservative conditions than Democratic primary voters in more liberal districts. Additional results, however, suggest a lack of association between extremist primary electorates and extremism among party nominees and elected officials, contrary to conventional wisdom.

This dissertation suggests a new perspective for considering the ideological representativeness of primary voters. From this perspective, the project’s findings cast doubt upon claims of primary voter ideological unrepresentativeness as contributing to contemporary polarization.

INDEX WORDS: Primary elections, Political ideology, Voting behavior, Representation, Congressional nominations, Ideal points
The process of writing this dissertation started early in my graduate school career. Without guidance from mentors and support from colleagues, however, this project would never have grown from the seminar paper it once was.

I am grateful for the support from my dissertation committee. They challenged me to work on a thesis that was more cohesive than the one I originally intended to write. I especially thank committee chair Mike Bailey for his mentorship, advice, and enthusiasm. Jon Ladd and Michele Swers provided helpful perspectives for how to shape this project, especially in its nascent stage, and I am appreciative of their guidance.

Without the generosity that provided the data introduced in Chapters 2 and 3, it would have been impossible to complete this dissertation. For voter file data, I thank the team at Data Trust, especially Dishant Rana for approving my data request and Dorothy Chesley for writing SQL code on my behalf. Additional thanks to David Seawright for connecting me with the Data Trust team. For the exit polls, I thank the entire Primaries Project Team at Brookings. I especially thank Elaine Kamarck for her mentorship, which has been influential in facilitating my intrigue with primary elections and electorates. Additional thanks to Nick Zepos, Christine Stenglein, and Grace Wallack.

Earlier versions of Chapters 2 through 4 benefited greatly from suggestions by discussants and others at conferences and meetings hosted by the following: the Midwest
Political Science Association (MPSA), the Society for Political Methodology (Pol-Meth), the Department of Politics at the University of Virginia, and the Graduate Working Group at Georgetown University. These interactions helped me critically consider fundamental pieces of the project.

Throughout graduate school, my Ph.D. cohort has always been the support system I needed. Erum Haider, Nikhil Kalyanpur, Inu Manak, Caitlin McGowan, DongJoon Park, and Shubha Kamala Prasad have been especially supportive of my work, often pushing me to think about the forest instead of how best to estimate the branches. Other department colleagues like Rachel Blum, Paula Ganga, Karin Kitchens, and Chris Schorr have been helpful guides throughout graduate school. Hoya Saxa.

Finally, I am indebted to my family and friends for their endless love and support. My parents, especially, have been a regular source of encouragement and, oftentimes, nourishment. My parents, sister, and grandmother have motivated me to be confident in my work and I think of them first when I count my blessings.

All remaining errors are, of course, my own.
### Table of Contents

#### Chapter

1. **Introduction** ........................................... 1
   1.1 Plan of the Dissertation ............................. 7

2. **The Ideology of Primary Electorates** .............. 10
   2.1 Who Votes In Party Primaries? ...................... 13
   2.2 The Ideology of the “National Primary Voter” .... 18
   2.3 Heterogeneous Ideological Positioning by Primary Electorates 28
   2.4 Conclusions and Discussion ......................... 41

3. **The (Non-)Polarizing Effects of Primary Voter Rules** 44
   3.1 Primary Rules ........................................ 47
   3.2 Data and Methods ...................................... 54
   3.3 The Null Effects of Primary Rules on Polarization 66
   3.4 Discussion ........................................... 73

4. **Asymmetric Primary Voter Extremism in Extreme Districts** 75
   4.1 Two Electorates ...................................... 77
   4.2 Data and Methods ...................................... 81
   4.3 Republican Extremism and Democratic Representativeness in Safe Seats 86
   4.4 Discussion ........................................... 93

5. **Primary Electorates and Representation** .......... 95
   5.1 Voter Extremism and Polarization in Representation 97
   5.2 Voter Extremism and Ideological Primary Challenges 110
   5.3 Conclusion and Discussion ......................... 120

6. **Conclusions and Discussion** ....................... 123
   6.1 Dissertation Conclusions ........................... 123
   6.2 Suggestions for Future Research .................... 126
   6.3 Final Thought ....................................... 128
### List of Figures

1.1 Ideological Self-Placement of Partisan Primary Voters and Nonvoters, 2008 – 2016 .................................................. 4

1.2 Increased Partisan Polarization Among Candidates and Legislators, 1980 – 2016 .......................................................... 6

2.1 Estimated CCES Ideal Points by Party and Ideological Self-Identification, 2006-2016 ............................................ 25

2.2 Estimated CCES Ideal Points by Primary Vote Status by Election Cycle, 2008-2016 ....................................................... 27

2.3 State-Level MRP Estimates for Primary Voters and Copartisan Non-voters, 2010-2016 ............................................................. 37

2.4 State-Level MRP Estimates Density Plots for Sample Constituencies, 2016 ........................................................................ 38

2.5 District-Level MRP Estimates as Relative Extremism, 2010-2016 .............................................................................. 40


3.2 Primary Voter Ideology by Election Rules (Fitted Values) .... 70

4.1 Congressional Exit Polls Races: Distribution of Cook PVI ................................................................. 85

4.2 Primary Voter Ideology by District Partisanship (Fitted Values) ............................................................................. 89

4.3 Extreme Vote Choice by District Partisanship (Predicted Probabilities) ......................................................... 92

5.1 Median Voter Schematic ................................................................. 102

5.2 Primary Voter Extremism Relative to the Median Voter .............................................................................. 103

5.3 Comparing Primary Voter and Primary Winner Ideologies ........................................................................... 106

5.4 Tracking Ideological Primary Challenges, 1970–2012 .................................................................................. 114

5.5 Ideological Primary Challenges by Relative Voter Extremism Across Parties, 2010–2012 (Predicted Probabilities) .................................................................................. 119

A1 Unweighted and Weighted MRP Estimates, 2010-2016 ................................................................. 132

A2 Sample Excerpt from Exit Polls Questionnaire (2018, Iowa-3, Democratic Primary) ................................................................. 134

A3 Map of Exit Polls Sampling Frame versus Sample, 2016 .............................................................................. 136

A4 Map of Exit Polls Sampling Frame versus Sample, 2018 .............................................................................. 136

A5 Candidate Primary Vote Share by Unweighted Exit Polls Estimated Vote Share ................................................................. 137

A6 Extreme Candidate Top Two Vote Shares When Facing Moderate Candidate .............................................................................. 140

A7 District Median Voter Ideology by District Presidential Vote Share, 2010–2016 ................................................................. 142
List of Tables

2.1 Cooperative Congressional Election Study Sample Sizes by Year . . . 21
2.2 Ideology of Primary Voters, General Election Voters, and Party Fol-
lowers, 2010 ................................................................. 26
3.1 Primary Election Rules, Summarized ........................................... 48
3.2 Congressional Exit Polls Sample Sizes by District-Primary, 2016–2018 58
3.3 Exit Poll Respondents’ Vote Choices In Ideological Races, 2016-2018 . 64
3.4 Primary Rules Linear Model: District Ideology Estimates ............... 68
3.5 Primary Rules Logistic Model: Exit Poll Vote Choices ................. 72
4.1 Relative Primary Voter Extremism by PVI Grouping, 2010-2016 .... 79
4.2 District Partisanship Linear Model: District Ideology Estimates .... 87
4.3 District Partisanship Logistic Model: Exit Poll Vote Choices ......... 91
5.1 Polarization by Primary Voters Linear Model, 2010 – 2016 .......... 108
5.2 Polarized Voting Patterns by Primary Voters Linear Model, 2010 – 2016109
5.3 Ideological Primary Challenge Logistic Model by Party, 2010–2012 . 118
A1 CCES Survey Items Used as Bridge Observations ....................... 130
A2 Mean Ideological Score for Democratic Voters (Ideal Point and Sym-
bolic), 2008-2016 ............................................................. 131
A3 Mean Ideological Score for Republican Voters (Ideal Point and Sym-
bolic), 2008-2016 ............................................................. 131
A4 Comparing Party Registration Details, State-Level Estimates ....... 133
A5 Comparing Party Registration Details, District-Level Estimates ..... 133
A6 Exit Poll Summary Information in Accordance with AAPOR Code of
Professional Ethics & Practices Section III-A. .............................. 135
A7 List of Moderate and Extreme Candidates in Exit Polls Sample, 2016–
2018 ................................................................................. 139
A8 Primary Context Model: District Ideology Estimates ................. 141
A9 Primary Challenge Model (Raw Data) .......................................... 143
Chapter 1

Introduction

“The people who show up for primary elections tend to be much more extreme, much more of the activist wings of the political parties... so primary elections tend to be one of the ways in which we end up with either an extreme right or an extreme left candidate.” – Pildes (2010)

“So long as primary voters remain ideologically polarized, we expect candidates will benefit from staking out increasingly liberal or conservative positions.” – King, Orlando and Sparks (2016)

Congressional primary elections often attract little attention. Despite their significant role in helping political parties nominate general election candidates, primary races are regularly low turnout affairs. For example, during the 2014 midterm House primaries, only 13.7% of registered voters turned out to vote (DeSilver, 2018). Although this voter apathy is alarming, it is nevertheless understandable considering that congressional primaries are rarely competitive or, in some cases, even contested (Kamarck and Podkul, 2018a). Further, these primaries often take place over a calendar of six or seven months,\(^1\) facilitating significant elections fatigue among even most the politically interested. For these reasons, the median congressional primary often escapes critical media coverage and is routinely understudied in elections research (Kamarck and Podkul, 2018b).

\(^1\)For example, the 2018 midterm congressional primaries began on March 6th (Texas) and ended on September 11th (New Hampshire and Rhode Island) (Federal Election Commission, 2017).
Since primary elections traditionally include relatively few voters, many sources—
including pundits and researchers—have speculated about the effects caused by the
ideological unrepresentativeness of primary voters. As the claims in the epigraph of
this chapter note, the conventional argument takes the following form: first, primary
voters are identified as ideologically extreme and, second, due to these extreme pref-
erences, primary voters nominate polarizing candidates. As another academic claim
argues, “For primary voters, ideology is often the candidate’s litmus test” (Burden,
2001). Pairing this premise with the low turnout rates, ideologues can supposedly
goast to their parties’ nomination with the support of a relatively small number of
voters. Notably, this argument is understood as political law in the halls of Congress.
Quoting a member of the legislature, Richard Fenno recounts this dynamic, writing,
“You can win a primary with 25,000 zealots” (Fenno, 1978).

Conventional wisdom claims by pundits routinely lament the polarizing effects
of “extremist” primary voters. In one account of the Republican primaries in the
2018 midterms, Republican voters were labeled “hardline conservative primary voters”
(Ollstein, 2018). In another account of New York’s Democratic primaries, The New
York Times argued explicitly, “A lower turnout tends to favor candidates who appeal
to more progressive voters” (Mays, 2018). Proposing a potential solution for this issue,
David Wasserman of the Cook Political Report explained how to begin to correct
political polarization with Michael Smerconish on CNN:

WASSERMAN: But I would also like to see more turnout in these pri-
maries. If people want to disincentivize extremism –
SMERCONISH: Go vote.
WASSERMAN: – if people don’t like the fringes, vote in every election,
not just the ones that occur in November (CNN, 2017).
These broad claims continue to suggest that the ideologically extreme views held by primary voters are exacerbating polarization in Washington.

Recent political science scholarship regarding ideological representativeness of primary voters, however, is not nearly as clear. While many accounts argue the ideological priorities of primary voters (such as, e.g., Burden 2001; Kaufmann, Gimpel and Hoffman 2003), others find that these voters are not ideologically distinct from primary nonvoters (see, e.g., Norrander 1989; Abramowitz 2008; Boatright 2014; Sides et al. 2018). Summarizing these trends, some scholars have concluded that the results from this research agenda is mixed (Hirano et al., 2010) or, at best, unsettled.

This present dissertation aims to better understand the ideological representativeness of primary voters. Most political science research focusing on this research agenda isolates primary voters at the national level. For example, such approaches broadly look at the ideological placement of primary voters and compare that placement to those who did not vote in a primary election. Figure 1.1 demonstrates this general approach by comparing primary voting and non-voting partisans according to their ideological self-identification using Cooperative Congressional Election Study (CCES) data from 2008 to 2016. As the figure indicates, primary voters are slightly more extreme than non-voters although the extent of this extremism varies by cycle. (For example, in 2008 Democratic primary voters are less extreme compared to self-identified Democratic primary non-voters compared to the 2016 cycle.) Although not an exact replication of earlier evidence, this approach to analyzing primary voter ideological representativeness is very similar to previous methods used for exploring this

---

2A more comprehensive literature review explores this earlier research in more detail below in Chapter 2.

3Figure 1.1 plots self-identified ideology across election cycles by self-identified partisans (including leaners) for vote validated primary voters and non-voters.
To determine primary voter ideological representativeness by focusing on nationally aggregated primary voters introduces a few methodological assumptions that may be substantively significant. First, by looking at national primary voter ideology by averaging voters, one is essentially estimating an ideological position of an electorate that does not exist. Even in presidential elections, delegates are allocated by a series...
of state-level contests where voter pools may be heavily influenced by the primary calendar and other campaign effects. Second, analyzing the ideology of a national primary voter could possibly average out meaningful differences across states and congressional districts. For example, if primary electorates are ideologically extreme in Wyoming, Montana, and Idaho but not in California, then the average ideological distance (weighted by population) would cover up meaningful heterogeneity. (This issue may be compounded when trying to understand the effects of primary electorate ideology in institutions like the U.S. Senate.) Further, additional methodological considerations such as how to measure voter ideology need to be considered as well.

In an attempt to correct for these improper assumptions, this dissertation project begins with a novel approach for estimating primary voter ideology. Instead of focusing on the estimation of a national primary voter’s ideology, this project analyzes congressional primary voters within their own constituencies and estimates primary voter ideology at the constituency-level for partisans within both states and congressional districts. These estimates allow for the investigation of heterogeneity in ideological representativeness across meaningful sub-units. The first part of this project reports these estimates for voters and non-voters and describes the finding that relative extremism exhibited by primary voters differs significantly depending upon the constituency. The second part of this project explores different causes and effects of primary voter ideology as they may relate to the political polarization summarized in Figure 1.2. Figure 1.2 plots the increasing distance between the mean ideological scores for partisan candidates to the U.S. House (top) and legislators within the U.S. House (bottom) from 1980 to 2018 as evidence of the contemporary trends of polarization.
By leveraging variation in primary voter ideology (within sub-units across election cycles), this project explores potential causes of extremism by focusing on primary participation rules as well as district partisanship (i.e. seat safety). Supplementing this agenda, this project also utilizes a series of exit polls to further explore extreme voting behavior by primary voters. Lastly, this dissertation seeks to uncover any influences by primary voters on candidate or legislator ideology to assert whether primary voters should, in fact, be suspects in the current search for the causes of polarization.

Figure 1.2: Increased Partisan Polarization Among Candidates and Legislators, 1980 – 2016
1.1 Plan of the Dissertation

The remainder of this dissertation explores this primary voter ideological representativeness across states and congressional districts. By using a variety of data sources, this project first outlines a unique estimation strategy for identifying constituency-level measurements of primary voter extremism for exploring ideological heterogeneity and then explores potential causes and effects of this extremism. The project next introduces novel evidence from a series of congressional primaries exit polls to explore ideological vote decisions by primary voters.

Chapter 2 explores the average relative primary voter ideological profile among voters in the major political parties at the national- and constituency-level. This chapter introduces the large-scale survey data used to estimate voter ideology to understand the national primary voter profile. Next, these national estimates are then paired with voter file data using a multilevel regression and poststratification framework to uncover estimates for primary voters within states and congressional districts. The chapter then explores these estimates and provides a descriptive account of primary voters and their co-partisan primary non-voters. This account demonstrates that primary voter extremism differs across constituencies, noting that the constituency-level approach improves our understanding of Republican and Democratic primary voters.

Chapter 3 examines the relationship of primary participation rules and primary voter ideology. After describing the differences between these participation rules – for example, closed and open primaries – this chapter conducts a pair of analyses to explore whether more restrictive primary systems are associated with greater levels of primary voter extremism. Building upon previous work that studies the relationship between primary rules and candidate ideology, the first analysis uses the estimates
developed in the previous chapter and finds little substantive differences between constituencies using different primary rules controlling for other demographic and electoral factors. The second analysis introduces the congressional primaries exit polls (2016, 2018) and similarly finds no association between primary rules and extremist voting behavior by voters.

Following a comparable outline as Chapter 3, Chapter 4 investigates the relationship between electoral context and primary voter extremism with respect to “safe seats” (and district partisanship). Using a pair of analyses leveraging the voter scores from Chapter 2 and the exit polls introduced in Chapter 3, this chapter finds asymmetric polarization by district partisanship: Republicans in hyper-Republican districts are more likely to be ideologically extreme but Democrats in hyper-Democratic districts are not different than their co-partisan voters in more competitive toss-up seats. However, the second analysis shows that this finding runs counter to voting behavior, which uncovers a pattern of more moderate voting choices by voters in more extreme districts.

Chapter 5 connects voters to political actors and explores whether extremist primary voters are associated with more ideologically extreme candidates and legislators. Through the use of ideological scoring from previous research using roll call votes and campaign finance, ideological party nominees and elected officials are not associated with more ideologically unrepresentative primary electorates despite suggestions from conventional wisdom. Further, I also study explore whether extremist electorates produce more ideological primary challenges to incumbent members of Congress and find a null result.

Chapter 6 concludes. After a brief summary of the dissertation’s empirical conclusions and substantive findings, this chapter closes with a discussion of suggested future research by noting some the limitations of the work presented in this project.
Further, it suggests areas for future data collection to help better explore how the findings from this dissertation may relate to broader time-variant trends of partisan polarization.

Are primary voters ideologically extreme? Does this relative extremism differ across states and congressional districts? Does it matter?
“It was basically a miracle, you know, from God straight through the people who worked so hard for me,” explained Dave Brat on the morning following his surprise victory against Majority Leader Eric Cantor for the Republican nomination to represent Virginia’s 7th congressional district in the U.S. House of Representatives (Sandelhmn, 2014). Cantor, who many saw as a successor to the Speakership following John Boehner’s tenure, was the highest ranking member in House leadership ever to lose a primary election, leading The New York Times to label it “one of the most stunning primary election upsets in congressional history” (Martin, 2014). Brat’s victory was that of a conservative outsider over the establishment, the Tea Party over Republican leadership.

As primary election night approached on June 10, 2014, most election prognosticators predicted Cantor would easily coast to another victory, just as the incumbent had done in every election since his first U.S. House of Representatives race in 2000. Cantor outspent his competitor nearly 25:1 and maintained a consistent polling double-digit polling lead. One public poll from May 2014 had Cantor up thirty-four points, and an internal poll from the Cantor campaign in the final week of the nomination reported the Majority Leader up over thirty points as well (Shepard, 2014). Nevertheless, on primary night, Cantor was defeated by Brat 55.5% to 44.5% in a record-turnout election (Virginia Department of Elections, 2014).
Although there were many hypothesized causes for Cantor’s historic loss by the popular press – such as Dave Brat’s endorsements by local tea party grassroots organizations (Martin, 2014) and support from conservative radio figures like Laura Ingraham and Mark Levin (Peters, 2014) – most accounts of the race repeatedly pointed to the ideologically extreme primary electorate as playing a major role in Brat’s nomination. To explain the results in Virginia’s 7th congressional district, *NPR*’s Peter Sagal claimed, “Eric Cantor was voted out by conservative primary voters.”¹ In a report titled “Primarily Extreme,” *Slate*’s William Saletan compared the policy positions and ideological self-identification from the 2012 national exit polls to a survey of VA-7 “active GOP primary voters” and concluded: “The people who vote in [these] primaries are wildly unrepresentative of the general population” (Saletan, 2014). Further still, David Wasserman best captured the conventional wisdom of how to interpret the primary’s results:

> “Cantor privately chastised tea partiers in his conference who fomented the 2013 government shutdown, came out in favor of restoring parts of the Voting Rights Act, and helped craft a watered-down DREAM Act that would provide a path to legalization for immigrants who came to the U.S. illegally as children. But he seemed to do so with fairly little regard to how rock-ribbed conservative primary voters back home would react to these pragmatic gestures” (emphasis added). (Wasserman, 2014)

In the primary postmortems that followed, the ideological extremism – or *conservatism* – of Republican primary voters was persistently perceived as a main cause for Cantor’s unexpected loss.

¹“Who’s Bill This Time.” *NPR.* [https://www.npr.org/2014/06/14/321791559/whos-bill-this-time](https://www.npr.org/2014/06/14/321791559/whos-bill-this-time)
However, Eric Cantor was not the only “establishment Republican” on the ballot during the 2014 congressional primary cycle. If all Republican primary voters were as extreme as they were suggested to be in Virginia’s 7th, then expectations might dictate that more than just two other Republican incumbents would lose their renominations that cycle (Kamarck and Podkul, 2014). Further, compared to his colleagues in the chambers of the House of Representatives, Cantor was more conservative than 55% of his fellow Republicans in the 113th Congress (Lewis et al., 2017). Yet, despite the popular press accounts labelling primary voters as “extreme,” there remains a puzzle surrounding how members who were more moderate than Cantor were easily renominated by their own extreme primary electorates. Are primary voters always more ideologically extreme when compared to broader populations, as suggested by the popular press? Or are some primary electorates more ideologically extreme (like VA-7 during the 2014 cycle) and others not? How might political science scholarship explain and estimate these dynamics?

To better explore the question of ideological representativeness of primary election voters, the remainder of this chapter develops constituency-specific measures of ideology for primary partisans. Using these measures, we can better explore: a) whether primary electorates are, in fact, relatively ideologically extreme and b) whether this relative extremism is consistent across different federal constituencies. Towards this end, the remainder of this chapter is broken into three sections. First, I will review the previous literature and accounts surrounding primary electorate ideology and demonstrate how this chapter’s estimates relate to earlier scholarship. Second, I estimate a “national primary voter” ideology and compare this estimate to previous work.

To provide a counter example opposite Cantor: Rodney Frelinghuysen (R-NJ-11) had a much more moderate DW-NOMINATE score (0.292) than Eric Cantor (0.518) in the 113th Congress. Nevertheless, in 2014, Frelinghuysen summarily defeated his conservative challenger Rick Van Glahn 66.7%-33.3%.
Third, using the ideal points estimated in the previous section, I separately estimate constituency-level primary voter ideological scores to explore whether there is heterogeneous ideological positioning by primary electorates across states and congressional districts. After exploring the results from the estimation process, I discuss how these results better explain primary voter ideology and how these estimates will be employed in the chapters that follow.

2.1 Who Votes In Party Primaries?

Discussions of contemporary political polarization often point to the primary nominating system as driving ideological extremism within both major political parties today. While there are a number of hypothesized possible causal mechanisms for primaries’ role in polarization, among the most common remain those indicting the “ideological electorate.” This ideological electorate supposedly nominates extreme candidates\(^3\) who subsequently pull the Republican Party further to the right and the Democratic Party further to the left. “Primary elections,” writes Sam Wang, “when turnout is low and dominated by the most motivated partisans, determine who gets elected” (Wang, 2016). Building on this contention, others have explained this dynamic by writing, “The enormous pressure to please narrow, extreme and grossly unrepresentative bases of primary voters has straitjacketed members...” (Wasserman, 2016).

This view of primary electorates is not novel. Writing in 1956, political scientist V. O. Key explained, “the fact seems to be... that those who vote in primaries do not make up miniatures of the party membership” (Key, 1956, p. 145). Although Key does not empirically test this claim, it is significant in that it established an important

\(^3\)Or, alternatively, requires more moderate candidates to appeal to this ideological electorate.
line of scholarship for those studying American elections by forwarding the research question: are primary voters unrepresentative of other voters/partisans? Testing this thesis a decade later, Austin Ranney found that – when analyzing Wisconsin gubernatorial primary voters in 1966 – primary voters were not actually unrepresentative when it came to issue positions or candidate preferences (Ranney, 1968). Ranney’s methodology introduces an important limitation that existed in the political science research on this topic in the decades that followed, specifically regarding the external validity of his test. Were Ranney’s findings generalizable or were they unique to Wisconsin gubernatorial primary voters in 1966?

In the many studies that followed since this preliminary test, the evidence of an ideologically unrepresentative primary electorate can at best be described as mixed. Arguing that primary voters are ideological, much of the focus has been on presidential primaries. In these studies, however, scholars have disagreed about how to estimate unrepresentativeness of primary voters. One study of the 1968 and 1972 Democratic primaries compared primary voters to all Democratic identifiers (i.e. including primary voters) and found primary voters to be more self-identified liberal (Lengle, 1981). Another study looking at the 1972 Democratic race compiled a number of different comparative groups, finding primary voters were more liberal when compared to nonvoters (Kritzer, 1980).4

Agreeing with these findings, more contemporary work also contends that voters are ideologically extreme. Such studies often argue primary voters cast their ballots on ideology alone, stating, “For primary voters, ideology is often the candidate’s litmus test” (Burden, 2001). Others accept the premise of ideological extremity as part of broader arguments about the strategic dilemma presented to candidates (see, e.g., Brady, Han and Pope 2007; Burden 2004), echoing earlier political science scholar-

ship (such as in Mayhew 1974; Fenno 1978). This strain of literature can best be summarized by political scientist Gary Jacobson in his *The Politics of Congressional Elections* where he explains, “Primary electorates are much more partisan and prone to ideological extremity, and the need to please them is one force behind party polarization in Congress” (Jacobson, 2009, p. 20).

A parallel strain of literature studies primary voter ideological representativeness but reaches the opposite conclusion. Writing about a series of presidential primary races, Geer compared primary voters to those of the “party following,” a group which includes identifiers “of the party who turned out, regardless of whom they voted for” as well as independents and crossover voters. Making this comparison, Geer found presidential primary voters from twenty-three state presidential primaries more moderate (Geer, 1989). Similarly, in a widely-cited study of presidential primary voters in 1988, primary voters in both parties are no different than general election voters with respect to ideological self-identification or extremism in issue positions (Norrander, 1989). Summarizing her work years later, Norrander explains that primary electorates are distinct in other ways: “Rather than being a more ideologically extreme proportion of the electorate, presidential primary voters are more aptly described as the slightly more interested and more knowledgable segment of the electorate” (Norrander, 2015, p. 73). More recently using estimated policy preference ideal points and ideological self-identification using vote-validated surveys from 2008-2014, Sides et al. (2018) agree with Norrander, arguing:

“Our evidence does not confirm repeated claims that the primary electorate is ideologically extreme or otherwise distinctive – even in the con-

---

5Mayhew (1974) discusses “a November electorate and a primary electorate nested inside it but not a representative sample of it” (p. 45) while Fenno (1978) famously characterizes the primary constituency as the second-most inner concentric circle among personal-primary-reelection-geographic constituencies.
evoters were *ideologically representative subsets of the broader party fol-
lowing* (emphasis added) (Sides et al., 2018, p. 8).

With the introduction of this scholarship, the evidence of primary electorate repre-
sentativeness is further complicated.

There are a number of reasons for these different findings. First, as noted above,
the comparative groups are not always constant. For example, Lengle, Kritzer, and
Geer make comparisons between primary voters and different “control groups” (see,
e.g., Buell 2004). Second, each study has a slightly different scope than the next.
Summarizing this trend, Hirano et al. write, “whether one observes differences in the
ideological positions of these two electorates depends in part upon the states and
the elections investigated” (Hirano et al., 2010). Third, while most scholarship rely
upon some sort of survey instrument to measure voter attitudes, there is significant
methodological variation across these studies. From ANES studies (Norrander, 1989)
to exit poll analyses (Geer, 1989; Kaufmann, Gimpel and Hoffman, 2003) to ideal
point estimation using vote-validated surveys (Sides et al., 2018), there are many
different approaches for tackling this question.

Despite this methodological diversity, none of the above noted studies provide
extensive *constituency-level estimates* of primary voter ideology. Those that capture
constituency-level findings (such as Ranney’s study of 1968 gubernatorial primary
voters in Wisconsin) only do so for a limited number of constituencies. Others (like
Sides et al.’s ideal point estimation of national surveys) provide *national* estimates
but fail to shed any light on voters within particular constituencies. To the best
knowledge of this author, only two previous studies attempt to estimate constituency-
level ideology. Hill (2015) utilizes a similar framework than this present chapter but
differs in its estimation strategy and the number of elections that it studies. Surminsky
(2018), however, uses a different estimation framework but works with same unit of analysis. (Details below explicitly contrast this project from these two earlier studies.) Uncovering constituency-level estimates of primary voter ideology will help advance contemporary understandings about primary voters within a meaningful context. Considering the great variability of characteristics associated with primary races, estimating the ideology of particular constituencies can help outline which factors may produce more polarizing outcomes (see, e.g., Chapter 5) or how campaigns might exploit these differences for their own objectives (see “Future Research” noted in Chapter 6). For example, in any given cycle of House primaries, there is a great difference across races in different districts. In some races, primaries are “open” seeking to fill the seat of a retiring or resigning incumbent while, in others, partisans consider incumbents and challengers. Additional contextual information – such as district partisanship and primary participation rules – may also influence the profile of a district’s average primary voter. Further, estimating primary voter ideology by constituency better approximates federal political processes. Unlike some previous research that averages across primary voters regardless of district, political context, rules, demographic profiles, and other district characteristics, this chapter hopes to contribute to growing scholarship by developing ideological estimates by each constituency to tell a more detailed story about primary voter polarization.

In the sections that follow, I first engage with the previous literature that estimates ideological scores for “national primary voters” and then, following, I re-weight these scores in a multilevel regression and poststratification framework to find constituency-level ideological estimates for primary voters and primary non-voters belonging to each of the two major political parties.
2.2 The Ideology of the “National Primary Voter”

In order to estimate constituency-level estimates, this project must first estimate national-level primary voter and non-voter ideology. Even though there is no such thing as a “national primary voter” – considering that even presidential nominations are decided through a series of state-level contests – estimating this hypothetical voters’ ideological score is advantageous because it allows this project to first replicate previous findings and then provide a foundation for the constituency-level re-weighting that will occur in the following section.

2.2.1 Previous Work

To estimate national primary voter ideology this project builds upon work put forth by both Boatright (2014) and Sides et al. (2018). Both projects seek – in part – to estimate the relative ideological placements of primary electorates. Similarly, both works rely (in part) on the Cooperative Congressional Election Study, while the former focuses on the 2010 iteration and the latter on 2008 through 2014. The projects diverge in estimating this national ideology, however, with Boatright analyzing ideological self-identification – viz. survey responses on a 7-point ordinal ideological scale arranged from very liberal to very conservative – and Sides et al. incorporating ideological ideal points estimated from policy preference questions. Finally, the projects compare different subgroups of the electorate. In Boatright’s project, the author compared primary voters to: 1) general election voters, 2) registered voters, and 3) the full survey sample; whereas, Sides et al. compare primary voters to: 1) general election voters, and 2) party followers (cf. Geer 1989).

Borrowing from Sides et al. (2018) this project refers to self-reported ideology via survey responses as symbolic ideology. See, e.g., Table 2.2 below.

Akin to the differences between Geer, Lengle, Kritzer, and others studying primary voters in the late 1960’s and early 1970’s.
In Boatright (2014)’s comparison, the results demonstrate that according to ideological self-placement, primary voters in 2010 are not that much different from general election voters, registered voters, or American adults (i.e. the full CCES sample). For example, 50.5% of primary voters identify as “conservative” compared to 48.7% of general election voters. Similarly, 27.8% of primary voters label themselves “liberal” compared to just 28.1% of general election voters and 27.4% of registered voters (see Boatright 2014, Table 3.1). According to ideological self-identification, there is not much difference. By finding little dissimilarities in 2010 – which the author labels an “anomalous year” – he concludes, “We should, then, consider claims about the unrepresentativeness of primary electorates with caution” (pg. 90).

Sides et al. (2018) summarize their findings by stating, “In 2008, 2010, 2012, and 2014, primary voters were ideologically representative subsets of the broader party following” (pg. 8). While the study neglects to include statistical significance tests because “each survey contains tens of thousands of respondents and... are very precisely estimated,” there are some ideological differences in the ideal points worth noting in their Table 2 (as they mention on the following page). Among Democratic voters, primary voters are extreme when compared to party followers in 2010 (difference in ideal points: 0.09), 2012 (0.15), and 2014 (0.1). Among Republican voters, primary voters are extreme in each of the elections covered by CCES data: 2008 (0.12), 2010 (0.12), 2012 (0.20), and 2014 (0.13). By the authors’ own logic, the ideal points therefore reveal primary voter extremism (in a “statistical” sense) for seven of eight electorates. While placing these findings in substantive language, the project compares the standard deviations of voter ideal points to the standard deviations of DW-NOMINATE scores and find these are substantively small effects. For example, the authors compare the difference in Democratic primary voters to Democratic party followers in 2012 as akin to the difference between Oregon Senators Jeff Merkley and
Ron Wyden on the DW-NOMINATE scale. While this is perhaps a subjective exercise, it appears worth noting that in the 113th Congress (2013–2015) Merkley was more liberal than 91% of Democrats and Wyden only more liberal than 53% of Democrats (Lewis et al., 2017). Although the ideal point story is only part of the Sides et al. project (considering they also explore symbolic ideology as well as particular policy preferences), the ideal points tell a story quite different than the articles’ overall conclusion.

The remainder of this section will borrow from these two previous approaches and aim to replicate the conclusions from these studies. By doing so, this process will ensure data that has consistent findings will be used in the constituency-estimation outlined in the next step.

2.2.2 Estimating the National Primary Voter Ideology

Following a similar methodological approach as previous work, this section replicates the estimation of survey-based ideal points using a series of national studies asking respondents (i.e. voters) a battery of policy preference questions, which form the basis of the constituency-level estimates developed later. First, I will introduce the data used in this project and, next, I will explain how ideal points for each respondent were estimated.

National-Level Data

The main data source for estimating “national primary voter” ideology in this project is the Cooperative Congressional Election Survey (CCES). The CCES is an online survey conducted by YouGov during the autumn of each year asking respondents a

Specifically, this project relies upon the 2006-2012 CCES Cumulative Common Content (Ansolabehere and Pettigrew, 2014), the 2014 CCES Common Content (Ansolabehere and Schaffner, 2015), the 2016 CCES Common Content (Ansolabehere and Schaffner, 2017), and
Table 2.1: Cooperative Congressional Election Study Sample Sizes by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>36421</td>
</tr>
<tr>
<td>2007</td>
<td>10000</td>
</tr>
<tr>
<td>2008</td>
<td>32800</td>
</tr>
<tr>
<td>2009</td>
<td>13800</td>
</tr>
<tr>
<td>2010</td>
<td>55400</td>
</tr>
<tr>
<td>2011</td>
<td>20150</td>
</tr>
<tr>
<td>2012</td>
<td>54535</td>
</tr>
<tr>
<td>2014</td>
<td>56200</td>
</tr>
<tr>
<td>2016</td>
<td>64600</td>
</tr>
<tr>
<td></td>
<td>Obs. 343906</td>
</tr>
</tbody>
</table>

A litany of questions spanning the areas of policy preference, political knowledge, voting behavior, and demographics. The data collected as part of the CCES is impressive for both its size and scope. For example, the 2016 CCES interviewed 64,600 respondents and asked nearly 120 questions (Ansolabehere and Schaffner, 2017). To take full advantage of all of the policy preference questions asked by the CCES, this project uses all of the available surveys from 2006 to 2016. Table 2.1 displays the sample sizes for each of these surveys. The large state and congressional districts sample sizes will also be useful for the estimation procedure in the next section of this chapter. (The mean number of respondents per congressional district-year is 75 and per state-year is 642.)

To prepare the data for the ideal point estimation, the CCES survey responses were re-shaped to imitate a “roll call matrix” with respondents taking the place of legislators (rows) and policy preference questions taking the place of bills (columns). All policy preference questions that included a clear support/oppose answer structure the CCES Cumulative Common Content (2006 - 2017) (Kuriwaki, 2018). The first three files are necessary because they include the policy preference questions used below.
were included in the matrix with all responses (cells) taking on one of three values: support (for), oppose (against), or missing (skip, do not know).\(^9\) In order to combine CCES surveys from different years and election cycles, questions that were repeated verbatim across surveys served as “bridge observations” (Bailey and Chang, 2001; Bailey, 2007). A list of all the bridge observations (and the corresponding surveys) is reproduced in Table A1. Ultimately, this data reconfiguration yielded a matrix of 343,906 respondents with preferences across 100 unique policies including gay marriage, global warming, firearm rights, immigration, social security, foreign policy, and many others.

Further, the CCES is a useful data source for this project because voting behavior has been validated using a voter file. Rather than needing to rely upon self-reported survey instruments for tracking primary voting (which survey respondents routinely overestimate; Belli, Traugott and Beckmann 2001; Ansolabehere and Hersh 2012), CCES data provides each respondent’s validated vote information. Working with the Democratic microtargeting firm Catalist, the CCES data links respondents anonymously to their voter file records (which are made available to Catalist from state secretaries of state offices and other election administrators). This additional information is very useful for comparing voters and nonvoters, especially in primary elections. Although all respondents aren’t matched, the match rate is relatively high. (For example, in the 2016 CCES nearly two in three respondents were matched to a voter file.) From the 2016 CCES codebook:

\(^9\)Questions with ordinal or scaled responses were collapsed into the for/oppose framework. For example, from 2007 to 2012 respondents were asked their opinion on affirmative action and provided four responses: “1 - strongly support, 2 - somewhat support, 3 - somewhat oppose, and 4 - strongly oppose” (Ansolabehere and Pettigrew, 2014). For the present study, 1’s and 2’s were categorized as support and 3’s and 4’s were categorized as oppose.
“It should be noted that a record may not be matched either because the individual is not registered to vote or because of incomplete or inaccurate information that prevented a match. Matches are made only with records for which there is a high level of confidence that the respondent is being assigned to the correct record.”

This vote validated data is necessary for this chapter considering the central comparison in this project is between primary voters and primary nonvoters.

**Ideal Point Estimation Method**

Using the roll call matrix created from aggregated CCES files (2006 to 2016), this project used a survey-based ideal point estimation procedure to uncover a unidimensional estimated latent ideological score for each CCES respondent. This approach for measuring ideological ideal point scores has been applied to many contexts, such as legislators (Poole and Rosenthal, 2006; Clinton, Jackman and Rivers, 2004; Shor and McCarty, 2011), candidates (Bonica, 2014), and interest group organizations (McKay, 2008; Bonica, 2013). This method has also been introduced to measuring the ideological positions of survey respondents (treating them as “legislators” in the classic set-up) by leveraging policy position survey items (treating them as “roll call votes”) (Jessee, 2009; Bafumi and Herron, 2010; Jessee, 2012). Using the CCES (2006 to 2016), this project relies upon a sample of over 300,000 respondents (see Table 2.1) across over 100 policy position questions. Although identical policy position questions are not asked in each iteration of the CCES survey, respondents were asked a number of identical questions each cycle (for example, on abortion and immigration policy) that serve as “bridge questions” to connect respondents across years and place them

---

10For more details, see: Ansolabehere and Schaffner (2017)
in a “common preference scale” (see, Bailey and Chang 2001; Bailey 2007). A two-parameter binomial item response model is used to estimate these ideal points in a single dimension:

$$y_{ij}^* = \alpha_j + \theta_i^T \beta_j + \epsilon_{ij}$$

(2.1)

where $i$ is respondent with $i = 1, ..., N$ of $N$ respondents, $j$ is policy preference with $j = 1, ..., M$ of $M$ policy preferences, $y_{ij}^*$ is the latent propensity for policy preference to be 1 (or “support”) for respondent $i$ on policy $j$, $\alpha_j$ is the item difficulty parameter for policy preference item $j$, $\beta_j$ is the item discrimination parameter for policy preference $j$, $\theta_i$ is the estimated ideal point for respondent $i$, and $\epsilon$ is independently and identically distributed assumed to follow a standard normal distribution (following Imai, Lo and Olmsted 2016). This estimation results in $N$ number of ideal point estimates from $M$ policy preference questions for a decade of CCES respondents.\textsuperscript{11}

2.2.3 The Polarized Primary Voter

Results from this estimation of a “national primary voter” indicate an overall finding that primary voters are often ideologically unrepresentative of all partisans; however, the extent of this unrepresentativeness differs across election cycles.

Results

The results in Figure 2.1 seem to indicate face validity to the ideal points measured from the 2006 to 2016 CCES. In the top panel, it finds that Democrats hold more negative (liberal) policy preferences in a unidimensional ideal point model, Independent and unaffiliated voters have middle of the road preferences (with wide variation), and

\textsuperscript{11} For computational purposes, the model was estimated using the expectation-maximization (EM) framework of Imai, Lo and Olmsted (2016) using the \texttt{emIRT} package in R.
Republicans hold more positive (conservative) policy preferences. In the lower panel, we see an ordinal arrangement of the density masses of each ideological subgroup with mean estimated ideal points of very liberal self-identified voters to the left of liberals which is to the left of moderates which is to the left of conservatives which is to the left of very conservatives.

Figure 2.1: Estimated CCES Ideal Points by Party and Ideological Self-Identification, 2006-2016

Next, to begin to explore the “national primary voter” replication of earlier work, Table 2.2 shows the average ideal point and average symbolic ideological score for primary voters, general election voters, and party followers for 2010.\(^\text{12}\) Here, the data demonstrate consistent findings across both parties: primary voters are more extreme

\(^{12}\)Symbolic ideology measures a five-point scale self-identification of a respondent’s ideological position, from 1 (Very liberal) to 5 (Very conservative). This measure does not include respondents who answered “Not sure.”
Table 2.2: Ideology of Primary Voters, General Election Voters, and Party Followers, 2010

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voted in primary</td>
<td>Voted only in general</td>
</tr>
<tr>
<td></td>
<td>Party following (self-id)</td>
<td></td>
</tr>
<tr>
<td>Avg. Ideal Point</td>
<td>-0.69</td>
<td>-0.61</td>
</tr>
<tr>
<td>Avg. Symbolic Ideology</td>
<td>2.41</td>
<td>2.45</td>
</tr>
</tbody>
</table>

than general election voters and party followers on both measures. Also, according to both measures, Republican primary voters are relatively more extreme than general election voters and party followers compared to Democratic primary voters. These findings agree with those from Sides et al. when comparing primary voters and general election-only voters.¹³

Following from the earlier literature, the decision of how to define a particular subset of the population is very important. While it is easy to define primary voters, it is far more difficult to determine what comparison to make in determining their unrepresentativeness. Even comparing these primary voters to general election voters is difficult: should the comparison be between primary voters and all general election voters (see Boatright) or just those general election voters who did not vote in the primary (see Sides et al.)? Further, how should one operationalize copartisan when

¹³Further, these estimates highly correlate with those from Sides et al. Among the CCES respondents used by both of these projects – viz. respondents from 2008 to 2014 – ideal points correlate at $r = 0.94$. The findings presented in Table 2.2 further confirm the direction and relative distances of these ideal points from earlier research.
Figure 2.2: Estimated CCES Ideal Points by Primary Vote Status by Election Cycle, 2008-2016

attempting to compare co-partisan primary and general voters? Seeking to combine previous approaches, this present project compares co-partisan primary voters and primary non-voters (irrespective of whether they voted in the general election). This decision combines and averages those who voted in the general (but not the primary) and party followers to provide a conservative estimate that specifically focuses on the subpopulation of interest, viz. primary voters. Specifically, this section of the project identifies partisanship using self-reported party identification via the CCES cumulative file (Kuriwaki, 2018); that is, for example, a non-voting Republican is a respondent who was determined as not voting in either party’s congressional primary election but self-identified as a Republican.

Demonstrating this difference, Figure 2.2 compares the distribution of ideological scores from co-partisan primary voters to non-voters by each election cycle from 2008
There are a number of noteworthy observations from this chart. First, the national primary voter has been getting relatively more extreme (tracking with secular trends of political polarization). However, this polarization appears asymmetrical. While Democratic primary voters have been moving to the left relative to co-partisan primary nonvoters, Republican primary voters have been fairly consistently conservative in the years covered by this data. Nevertheless, these trends in the national primary voter only begin to explore the ideological representativeness of these voters. In the following section, this project builds upon these survey-based ideal points to estimate constituency-level scores for primary voters and non-voters to explore whether some constituencies have more ideologically unrepresentative primary voters than others.

2.3 Heterogeneous Ideological Positioning by Primary Electorates

The evidence surrounding the ideological representativeness of primary voters – from previous research and the previous section in this chapter – at the national level is mixed at best (Hirano et al., 2010). Seeking to better understand how the ideology of this active subset of primary voters relates to primary nonvoters, this project explores the possibility of heterogeneous ideological representativeness across different electoral contexts. While earlier research assumes a national electorate by comparing all U.S. primary voters to all U.S. primary nonvoters, this project instead estimates the ideology of primary voters and primary nonvoters for each federal constituency (i.e. each state and each congressional district) (cf. Hill 2015; Surminsky 2018). This present section first explains how these constituency-level ideological scores are gener-

\footnote{Mean ideological scores are presented by cycle in Table A2 for Democratic Party voters and Table A3 for Republican Party voters.}
ated and then describes these estimates for both states and districts. In the chapters that follow, these estimates will be used in a hypothesis-testing framework.

2.3.1 Estimating Constituency-Level Primary Voter Ideology

Using the ideal points estimated from bridged CCES survey items (2006-2016) as described in the previous section, this section manipulates those estimates by incorporating them in a multilevel regression and poststratification (MRP) context to uncover constituency-level estimates for primary voters and their co-partisan primary nonvoters. As its name suggests, MRP is a two-step estimation process for uncovering sub-unit estimates by using unit data paired with sub-unit characteristics. The first step – multilevel regression – estimates a multilevel model with some characteristic (in this case, ideological score) modeled as a function of individual and sub-unit variables. The second step – poststratification – uses predictions produced in the first-step and weights those predictions according to sub-unit population characteristic frequency tables (Kastellec, Lax and Phillips, 2016; Lax and Phillips, 2009).

In most political science applications of MRP, sub-unit population characteristic frequency tables are borrowed from the U.S. Census; however, since this project is focused on voting status, voter file data that was made available to the author is used instead.\(^{15}\) Nevertheless, U.S. Census data is also among the sources incorporated for calculating geographic effects (details below). For the congressional-district effects, data was collected from the 1-year estimates from the American Community Survey (U.S. Census Bureau, 2006-2016) and the Decennial Data from the U.S. Census

\(^{15}\)U.S. Census data does not capture primary voting data. Further, the general election voting data that it does provide relies upon respondent self-report (for more details, see McDonald 2019). The advantage to using the voter file is primary voting status is validated by each voter’s state secretary of state’s office so respondents cannot inflate reported vote participation.
(U.S. Census Bureau, 2000, 2010). For the state effects, data was collected from the Decennial Census (U.S. Census Bureau, 2000, 2010), the Current Population Survey (via Hirsch, Macpherson and Vroman 2018), and the U.S. Religion Census: Religious Congregations and Membership Study (Association of Religion Data Archives, 2000, 2010).

These two steps – multilevel regression and poststratification – are described in further detail below.

**Step One: Multilevel Regression**

In the multilevel regression stage, two separate models were estimated: one for state-level estimates and another for district-level estimates. Both models were built around the individual-level model described in Equation 2.2. The individual-level variables incorporate standard demographics of sex, race, age-range, and education (cf. Kastellec, Lax and Phillips 2016) and also politically-related terms such as party-year, party-sex, and party-primary voting status. The final term represents the geographic effects explained on the following page. The individual model was thus estimated as:

$$
y_i = \beta_0 + \alpha_{Sex}^i + \alpha_{Race}^i + \alpha_{Age}^i + \alpha_{Educ}^i + \alpha_{Party.Year}^i + \alpha_{Party.Sex}^i + \alpha_{Party.Primary}^i + \alpha_{Geo.Year}^i$$

(2.2)

for each respondent $i$ with $y$ representing each respondent’s estimated ideal point where each of the terms following $\beta_0$ are modeled effects (varying-intercepts) such that:
For each of these demographic variables, there are a number of possible values (represented by the for... statement). For each of these terms, the variables take on the following values: sex includes male/female; race includes white/black/asian/hispanic/other; age includes bins of 18-24/25-34/35-44/45-54/55-64/65+; education includes numbered categories representing completed high school/completed college/completed graduate or professional school/attended vocational or technical school or other; party includes Republican/Democratic/Independent, unaffiliated or other; primary voting includes voted/not voted/record not matched; and year is the year of the CCES survey. The party and year terms are combined to account for secular trends of polarization.

The final term in Equation 2.2 represents the geographic effects ($\alpha_{Geo}$), which were modeled separately for state- and district-level. For the state-level model, these effects are modeled as:

$$\alpha_{Geo.Year}^{c,n} \sim N(\alpha_{Region}^{c,n} + \beta^{Union}^{c,n} + \beta^{Relig}^{c,n}, \sigma_{Geo.Year}^{2})$$
where $\textit{Region}$ is the Census Region modeled independently below, $\textit{Union}$ is the percent of each state’s nonagricultural wage and salary employees who are union members, and $\textit{Relig}$ is the Percent of Evangelical or Mormons for each of the fifty states $c$ across the four election cycles $n$ (cf. Tausanovitch and Warshaw 2013). The $\textit{Region}$ variable is then modeled as:

$$\alpha_d^{\textit{Region}} \sim N(0, \sigma_{\textit{Region}}^2), \text{ for } d = 1, \ldots, 4$$  \hspace{1cm} (2.5)$$

where states fall in a $\textit{Region}$ of Northeast, Midwest, South, or West based on their Census classification.

For the congressional district-level model, these geographic effects are modeled as:

$$\alpha_{c,n}^{\textit{Geo.Year}} \sim N(\alpha_{d[c,n]}^{\textit{State}} + \beta^{\textit{Vet}} \textit{Vet}_{c,n} + \beta^{\textit{Income}} \textit{Income}_{c,n} + \beta^{\textit{Urban}} \textit{Urban}_{c,n}, \sigma_{\textit{Geo.Year}}^2)$$ \hspace{1cm} (2.6)$$

where $\textit{State}$ is the State modeled independently below, $\textit{Vet}$ is the percent of a district who are veterans, $\textit{Income}$ is the district’s median income, and $\textit{Urban}$ is the percent of a district’s population living in an urban area for each of the 435 congressional districts $c$ across the four election cycles $n$ (cf. Tausanovitch and Warshaw 2013). The $\textit{State}$ variable is then modeled as:

$$\alpha_d^{\textit{State}} \sim N(0, \sigma_{\textit{State}}^2), \text{ for } d = 1, \ldots, 50$$ \hspace{1cm} (2.7)$$

as above.\footnote{This multilevel regression stage was estimated in R using the \texttt{lme4} package. For computational purposes, models were estimated using the \texttt{nloptwrap} optimizer. More information can be found: \url{https://cran.r-project.org/web/packages/lme4/vignettes/lmerperf.html}}
Step Two: Poststratification

Sub-unit Data. As introduced above, the poststratification step in this estimation requires sub-unit population characteristic frequency tables. Unlike earlier studies that rely upon Census data for post-stratification, this project uses voter file data because (in addition to the typical demographic variables necessary) this project also requires the incorporation of voting data. The voter file data used in this project was made available exclusively to the author from the Republican voter file firm Data Trust, L.L.C. (DT).<sup>17</sup> The information provided by DT provided the number of individuals for each combination of demographic, party, geographic, and voting status variables provided. For example, the raw 2016 voter file included over 1.3 million rows of data combinations covering over 208 million American citizens in the United States. One example from these rows identifies that in 2016 there were 54 Republican Hispanic Female College Graduates aged 35–44 who voted in the New Jersey 11th congressional district primary. Voter file data was provided to the author for each election cycle from 2010 to 2016.

These data were prepared to be used for the poststratification step by two data management steps that reshaped the data. First, DT provided a number of break-downs when it came to primary vote status – among voters we know whether one voted by mail, poll, or early and, in some states, in which party’s primary they voted. For the purposes of this project, voting methods were collapsed into “primary voters” and “primary nonvoters” and tracking the party of the primary when available. Second, for many states, the data provided a variable for partisanship. In other states, however, the data provided was based on DT’s Modeled Partisanship (<i>cf. Hersh 2015, p. 73–75</i>). For this project, modeled partisanship and actual partisanship were collapsed.

<sup>17</sup>For more information, see: http://thedatatrust.com
Evidence from scholarship identifies this should not be problematic, as commercial voter files’ modeled partisanship tend to perform well, even doing “a better job of correctly classifying the partisan affiliation [of voters] against their self-reported partisanship” (Pew Research Center, 2018).

**Poststratifying the Individual Model.** In the second (post-stratification) step of the model, predicted values were created for each row of the aggregated voter file (where rows incorporated the combination of individual-level characteristics). These predicted values were then averaged weighted by the number of voters in each row for each constituency and primary voting status such that:

\[
\theta_{g}^{MRP} = \frac{\sum_{p \in g} N_p \cdot y_p}{\sum_{p \in g} N_p}
\]  

(2.8)

where \(\theta_{g}^{MRP}\) represents the estimated ideology for group \(g\) (voting status by party by constituency by cycle) with \(p\) representing all rows of the aggregated voter file belonging to group \(g\). A comparison of the unweighted group ideal points (“cell predictions”) and the weighted group scores is presented in Figure A1, where the left panel shows a histogram of \(y_p\) (when \(g\) includes primary voters) and the right panels plots a histogram of \(\theta_{g}^{MRP}\) (when \(g\) includes primary voters).\(^{18}\) This step-two, particularly due to its reliance on an external voter file, makes it quite different than previous estimation strategies.\(^{19}\)

\(^{18}\)The left panel shows a mode near the ideal point estimate of 0. That mode disappears in the right panel because the ideal points are weighted by the number of voters at each ideal point.

\(^{19}\)As noted above, according to this author, Hill (2015) and Surminsky (2018) are the only previous works estimating relative primary electorate ideology. The Hill (2015) strategy is different because in that project’s “step 2” the author recreates the aggregated population file via CCES post-stratification weights from validated voters. The Surminsky (2018) strategy instead relies upon a variation of MRP known as multilevel regression with synthetic post-stratification (MrsP) by using a random sample of the Catalist Validated Voter Database.
2.3.2 The Difference in Electorates

The following sections explore the results of the two sets of MRP estimates (state and congressional district) in turn. Both sections confirm the necessity of estimating constituency-level ideological estimates by demonstrating that the relative primary voter ideological extremism differs across cases.

State-Level Results

The state level ideological scores for each constituency are displayed in Figure 2.3 for each of the years available (2010, 2012, 2014, and 2016). The filled dots represent the ideological score for primary voters and the hollow dots show the scores for those candidates who are co-partisan primary nonvoters. The navy blue points represent Democrats, and the dark red dots display the Republicans. The black dots show those primary electorates that utilize a top-two blanket primary system. Finally, the vertical black bar running from Arkansas to Wyoming subsets the states where state Secretaries of State and state elections divisions provide the party of a voters’ primary ballot.

Figure 2.3 displays both across-state and across-election cycle observations. First, primary voters among Republicans are more consistently extreme when compared

\[\text{Washington state used a top-two primary in each of these four cycles and California introduced the top-two in 2011. Louisiana is wholly excluded from the data due to the state’s reliance upon a two-stage nonpartisan blanket “jungle” primary. Due to a lack of data availability provided by DT, primary voter estimates are missing for Alaska in 2016.} \]

\[\text{For example, the New Jersey Department of State Division of Elections publicly provides the information of the party in which a primary voter has cast his ballot; however the Oklahoma State Election Board only provides whether a registered voter cast a primary ballot or not. Table A4 and Table A5 compare non-voter ideal points using simple difference of means tests across party registration rules and finds minor or no differences among the district- and state-level estimates, respectively.}\]
to copartisan nonvoters. For example, in 2014 there are more extreme primary electorates among Republicans than Democrats: in 2014 there are 3 more extremist electorates for Republicans than Democratic electorates. Second, the relative ideological extremism of a primary electorate changes across election cycles – in some cases, very significantly. For example in 2016, Ohio Democratic primary voters were over twice as more relatively extreme than they were in the 2014 primary. On the other side of the aisle, Rhode Island provides an interesting example; in 2014, Rhode Island Republican primary voters were actually more moderate than Republican nonvoters but in 2016 the voters and nonvoters nearly swapped ideological positions.

Third, there are some clear differences regarding the ideological distance between primary voters and nonvoters. Looking at Republicans in 2016, both Indiana and West Virginia primary voters are ideologically to the right of primary nonvoters; however, in Indiana this extremism is relatively minor where in West Virginia there is a very wide gap. To put these differences in substantive terms, the difference between West Virginia primary voters (0.31) is just shy of the difference between the weighted average ideal point between a 2016 CCES respondent who identifies as “Very Conservative” and a respondent who identifies as “Conservative” (0.36) on a five point ideological self-identification scale. The difference between Indiana Republican primary voters and nonvoters is just one fourth of this difference (0.08).
Figure 2.3: State-Level MRP Estimates for Primary Voters and Copartisan Nonvoters, 2010-2016
Figure 2.4: State-Level MRP Estimates Density Plots for Sample Constituencies, 2016.

Figure 2.4 displays the same data as Figure 2.3 but from a different perspective (where red and blue shadings represent primary voters and gray shading represents co-partisan primary non-voters). While Figure 2.3 displayed the means for these groups, Figure 2.4 shows density plots of the estimated ideal points weighted by frequency. All four panels represent 2016 voters but show quite different shapes of the data (the four states were randomly chosen from the group of states that provides ballot partisanship). Looking at the right panels, we see that Iowa Democratic primary voters
are ideologically monolithic whereas New Hampshire Democratic primary voters are much more ideologically diverse.

There are two additional notes that must be made about these estimates. First, there is much greater variability of ideological distance in the states where state secretaries provide the partisanship of a ballot. In the states where ballot party is unknown an assumption is being made regarding the vote (e.g. if a candidate is a primary voter and is a Republican then the voter is considered a Republican primary voter). This difference is meaningful because in cases with ballot partisanship independents/unaffiliated voters and opposing partisans that might vote in a party’s primary are properly considered. Second, a brief note must be made about the Arkansas estimates (especially looking at the Democratic primary voter estimates). This statistical anomaly is produced mostly as a function of the record percent of independent/unaffiliated voters in the state. As of 2016, independents make up nearly 90% of the Arkansas electorate (see, e.g., Independent Voter Project 2016). This, paired with the relatively uncompetitive Democratic primaries in the years of this data set, has helped produce this seemingly outlying estimate.

As mentioned above, these state-level results will be used alongside other data in the following chapters to explore the conditions that bring about primary voter extremism.

District-Level Results

Estimated separately from the state-level results, constituency-level estimates were also modeled for congressional districts (see Equation 2.6 and Equation 2.7). Consid-
Figure 2.5: District-Level MRP Estimates as Relative Extremism, 2010-2016.

Considering there are thousands of constituency-levels that were estimated, these findings are summarized differently than the state findings.

Figure 2.5 shows a stacked histogram plotting the relative primary voter extremism for each constituency from 2010 to 2016. The scale can be described as:

\[
Extremism_{c,d,p} = \theta_{c,d,p,\text{voter}} - \theta_{c,d,p,\text{nonvoter}}
\]

where \( Extremism \) represents the measure of primary voter extremism, \( \theta \) is the MRP constituency-level estimate in cycle \( c \) for district \( d \) in party \( p \) for primary voters and non-voters. Therefore, \( Extremism \) scores above 0 demonstrate constituency-level cases of primary voter extremism, scores at 0 find no difference in voter and non-voter

\[22\text{Number of estimates} = \frac{\text{Number of Election Cycles}}{\text{Number of congressional districts}} \cdot \frac{\text{Number of parties}}{\text{Vote Status outcomes}}\]
ideology, and scores less than 0 find primary voters to be more ideologically moderate than nonvoters.

Here we find the mode for both parties near 0.2 and a clear majority of cases where primary voters are extreme. However, for many electorates there is nearly no ideological difference between voters and nonvoters and, in rare cases, primary voters have been found to actually have been more moderate. Across the parties, Republican primary electorates are on average more ideologically extreme than Democratic primary electorates. Looking more specifically, one of the most extreme House race electorates in the country was Nebraska’s third congressional district in 2014. This district, which is represented by Rep. Adrian Smith (R), is among the most Republican districts in the nation (according Cook PVI) but data from this project would suggest that among the Republicans in this district there is substantial ideological variation – which is seen at the primary polls.

2.4 CONCLUSIONS AND DISCUSSION

This chapter and the findings presented above outline why constituency-level estimation of primary electorate ideology is necessary, explain how these estimates clearly build on the results of earlier work, and demonstrate how primary voter polarization is much more nuanced than popular press and most political science literature seem to suggest. The estimates generated by the methodology described above serve two purposes. First, they provide a descriptive account of state- and congressional district-level relative ideological extremism of this unique subset of voters. Second, they lay the foundation for tests in later chapters, viz. how these estimates might be influenced by contextual factors (such as primary type, laws, and political context).
Heterogeneity in the relative ideological representativeness of primary electorates is significant for a number of reasons. Considering there are no “national” elections in the United States, most approximations of primary voter ideology average across voters from Hawaii, New Hampshire, and Texas. While these estimates are perhaps helpful to understand sweeping claims of ideology, they tell us nothing about whether primary voters in Hawaii are ideologically distinct from nonvoters in Hawaii. In order to properly test V. O. Key’s claim of primary voters failing to “make up miniatures of the party membership” (Key, 1956), it is necessary to isolate the proper comparison. As demonstrated by the estimates presented in this chapter, this added layer to the comparison provides more nuanced conclusions. Rather than arguing all U.S. primary voters are ideologically extreme or not, we can state that, in 2010 for example, Democratic primary voters in Hawaii and Texas were ideologically extreme compared to their nonvoting co-partisans while those in New Hampshire were not.

Further, this finding of ideological heterogeneity among primary electorates introduces new hypotheses related to contemporary political polarization. Might polarization be driven by some constituencies but not others? This chapter’s descriptive findings provide a starting point for tackling such research questions.

As stated earlier, the estimates developed in this chapter are going to be used alongside additional data sources in the chapters that follow. In Chapters 3 and 4, these estimates will serve as dependent variables as different predictors related to electoral contexts will be tested to see if certain primary contexts – such as the existence of “closed primaries” where only registered partisans are permitted to vote – are associated with increased ideological extremism. In Chapter 5, these estimates will be connected with other data to see if primary electorate ideological extremism
is related to legislator or candidate extremism as a means to weigh-in on the question of primaries’ contribution to polarization.

***

Were Republican primary voters in and around Nottoway, Virginia ideological extremists in 2014 as popular press accounts might suggest? Was it, in fact, the case that “rock-ribbed conservative primary voters” upset Majority Leader Eric Cantor to put local Tea Party-favorite Dave Brat on the ballot in November? In 2014, the average ideological score for Virginia-7th Republican primary voter was 0.77 whereas co-partisan nonvoters were more moderate at 0.68. However, while these voters were ideologically extreme, the difference between primary voters and nonvoters was much less extreme (0.09) than the average Republican primary electorate (0.20). Press accounts seemed to be correct about primary voters in this race – nevertheless, there were many districts with more conservative primary voters who did not topple so-called “establishment” Members of Congress. The following chapters aim to shed more light on why this was the case.
The 2018 congressional midterm primaries were not filled with many surprises. Only four incumbents lost their bids for renomination (compared to five in 2016 and four in 2014) and only 45% of House incumbents even faced primary challengers at all (among Republicans, this was near the decade-average and, for Democrats, this was a slight uptick)\cite{Kamarck2018}. In most U.S. congressional districts, these party primaries nominated fairly conventional candidates—ranging on the Democratic side, from progressives like Rashida Tlaib (MI-13) to more traditional establishment candidates like Mikie Sherill (NJ-11) and, on the Republican side, from moderate candidates like John McCann (NJ-5) to more conservative candidates such as Steve Watkins (KS-2).

Two races that captured a great deal of public attention throughout the cycle were the Democratic House primary races in Illinois’s 3rd district and New York’s 14th district. In Illinois, the 3rd congressional district featured a contest between blue-dog moderate Rep. Dan Lipinski and Bernie Sanders-endorsed progressive Marie Newman. The ideological lines in the race were so clear that *The New York Times* described the contest as a “proxy between moderate and progressive Democrats” \cite{Bosman2018}. Even though Newman garnered a great deal of media attention and grassroots support, Rep. Lipinski won his renomination bid 51.1% to 48.9% in the March primary.
In New York, the 14th congressional district’s Democratic primary featured similar ideological contours. Congressman Joe Crowley – the Chair of the House Democratic Caucus – faced a primary challenge from progressive activist Alexandria Ocasio-Cortez. Unlike Lipinski, however, Crowley was relatively much more progressive; described by The Washington Post, Crowley “was the first member of Democratic leadership to endorse the House’s universal Medicare bill, and he joined protests against the Trump administration’s travel ban” (Weigel, 2018). Nevertheless, Ocasio-Cortez was markedly further to the left and campaigned as such. In the biggest surprise of the congressional primaries cycle, Ocasio-Cortez defeated Crowley 56.7% to 43.3% among the district’s nearly 30,000 Democratic primary voters.

While there are of course a number of different variables differentiating these two races – geography, candidate characteristics, campaign effects like endorsements and fundraising, and countless other factors – primary electoral rules must be considered for exploring the different primary outcomes. According to the National Conference of State Legislatures, Illinois operates their primary elections according to a “partially open” system whereas New York is a “closed” primary state. This difference means that while in Illinois any voter can participate in the Democratic primary so long as they state an affiliation before voting, in New York only registered Democrats can vote in the Democratic primary. In the closed primary of NY-14, the progressive candidate won; in the much more open primary of IL-3, the progressive candidate lost. Might these primary rules contribute to these differing ideological outcomes?

According to popular accounts, the answer is a resounding yes. Members of Congress often point to primary rules as contributing to contemporary polarization. In an oft-cited op-ed, Sen. Chuck Schumer (D-NY) argues that the California top-two

primary (which he refers to as an “open primary”) be a model for the nation: “This would prevent a hard-right or hard-left candidate from gaining office with the support of just a sliver of the vastly diminished primary electorate” (Schumer, 2014). In addition to Schumer, many former members have also argued for more open primaries. Summarizing these views, former Rep. Jason Altmire (D-PA) has argued:

“Instead of continuing with the system of closed primary elections that has given great advantage to the ideological extremes, some states have switched to open primaries, thereby expanding the electorate to a much wider spectrum of viewpoints. Instead of appealing only to the extremes, candidates running for election in open primaries – where all candidates for an office appear on the same ballot – must tailor their message to appeal to centrists and even voters from the opposing party” (Altmire, 2017).

Altmire’s argument has been echoed by other former representatives such as David Jolly (R-FL), Patrick Murphy (D-FL) (Peters, 2018), and Glenn Nye (D-VA) (Nye and Kitfield, 2018).

Even a number of political reform organizations have expressly advocated for opening up primary elections. No Labels – an organization committed to solving “partisan dysfunction” – has argued, “open primaries would diminish congressional partisanship by depolarizing elections, improving the abilities of moderate candidates to win primary elections and limiting the domination of partisans within the political system” (Hudson, 2011). Other reform organizations have even formed for the sole purpose of expanding the use of open primaries. For example, the group Open Primaries – which is an umbrella organization for the 501(c)(3) Open Primaries Education Fund and the 501(c)(4) Open Primaries Advocacy – claims, “Closed primaries
are a fetter... [they] make it more difficult for the American people – voters and elected officials alike – to come together across ideological lines” (Open Primaries, 2019). While these groups also cite additional arguments (such as additional voter participation concerns), one of their central arguments routinely links more closed primaries with more extremism and, subsequently, more political polarization.

The remainder of this chapter explicitly explores the connection between primary system rules and political polarization. The first section introduces previous research that studies the connection between primary rules and the ideology of political candidates. The second section looks at the literature exploring primary rules and the ideology of voters and explores a gap in the literature, which will be analyzed in the rest of the chapter. The following section introduces the data and methods for the two analyses performed in this chapter: the first test reintroduces the constituency-specific ideological estimates from the previous chapter and the second test introduces a set of congressional exit primary exit polls. The fourth section reviews the findings and results from these analyses. Finally, the chapter provides a brief discussion of primary rules and how they relate to voters’ ideology.

3.1 Primary Rules

Since the introduction of the direct primary as the mechanism for selecting party nominees for federal office, state party committees and state legislatures have experimented with a number of rules for regulating voter participation. While some primary elections are “open” to all voters to participate, others are “closed” for only registered partisans. For example, in the 2018 midterms, Florida observed a closed primary while Texas followed an open primary system. In Florida, that meant only registered partisans were eligible to vote in a party primary (i.e. only registered Democratic voters
Table 3.1: Primary Election Rules, Summarized

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Semi-closed</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Sometimes</td>
<td>Yes</td>
</tr>
<tr>
<td>Semi-open</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Sometimes</td>
<td>Yes</td>
</tr>
<tr>
<td>Open</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: Excerpted from McGhee et al. (2014) and Kanthak and Loepp (2018).

were eligible to participate in Florida’s Democratic primary) but in Texas voters were able to participate in whichever primary they chose regardless of party registration (i.e. registered Democrats and unaffiliated voters were eligible along with registered Republicans to vote in the Texas Republican primary). While a number of states have historically observed “closed” or “open” rules, there is variation in rules observance both across states as well as within states across time. For example, the Wisconsin Democratic Party has experimented with opening and closing its primaries dating back to the McGovern-Fraser reforms.²

Table 3.1 excerpts a table summarizing a number of these rules variations from McGhee et al. (2014).³ There are generally five categories considered for determining the rules categories associated with a particular primary system, summarized by each of the columns above. To paraphrase McGhee et al. (2014):

²For one dramatic retelling of closing the Democratic primary with respect to presidential nominations, see Wekkin (1984).
³Replicated again in Kanthak and Loepp (2018). As I note below, this present chapter focuses on variations of open and closed primaries. For this reason, this table does not include the rules for “nonpartisan” or top-two primary systems (such as those presently used in the states of California, Louisiana, and Washington).
- **Crossover Allowed?**: reports whether the system allows any members not belonging to the party to participate

- **Independents Only?**: reports whether independents are the only group of voters permitted to “cross over”

- **Public Decision?**: reports whether crossovers must “declare their crossover decision publicly”

- **Registration?**: reports if voters need to re-register as a party member in order to participate (e.g. via same-day party re-registration)

- **Choose Parties?**: reports if voters must stick with the same party for their entire ballot

Although there is variation even within the cells of Table 3.1, this general categorization schema is accepted among academics (see, e.g., Gerber and Morton 1998; Boatright 2013, etc.) as well as political practitioners (e.g. Ballotpedia, FairVote, NCSL, etc.).

### 3.1.1 Rules, Candidates, and Polarization

Even though conventional wisdom clearly suggests that more closed primaries produce more extreme candidates, political research – exploring the effects of primary systems on both electing legislators and nominated candidates – produces mixed findings regarding how these primary rules exacerbate polarization.

---

4For the categories presented, there is no variation for choosing parties. In “nonpartisan” primaries, however, voters are able to switch across parties

Many studies that examine the polarizing effects of primary rules on legislators produce different findings based upon the scope of the populations being studied. Studying state legislators, McGhee et al. (2014) explore how primary types may influence the ideological makeup of state legislatures (by using NPAT scores produced in Shor and McCarty 2011). The authors ultimately conclude from their difference-in-difference research design that there are no clear patterns across these primary systems. Studies of federal legislators (using ADA scores as a measure of legislator ideology) have found conflicting results. On the one hand, Gerber and Morton (1998) finds that closed primaries are associated with the most ideologically extreme candidates while semi-closed are actually the most moderate. On the other hand, Kanthak and Morton (2001) second these “nonlinear effects” but alternatively conclude more open primaries as the most extreme systems. While these findings may be driven by data scope or methodological decisions, they nevertheless report mixed findings.

Looking at party nominees – rather than just those candidates who win their general elections – paints a similar picture, although the preponderance of evidence suggests little relationship between primary type and nominee extremism. Studies that explore federal candidates using campaign finance ideological scores (by using cfScores produced in Bonica 2014) identify little differences across primary type (Rogowski and Langella, 2014). In the one study that yields statistically significant findings across these primary rules types in the expected direction (Kanthak and Loepp, 2018), these results mostly diminish when controlling for primary type (viz. whether the race is an incumbent primary, challenger primary, or open seat primary) (Podkul, N.d.). Further, an additional study that analyzes the ideological influence of nonpartisan (blanket) primaries (such as the kind used in California) via campaign

\[ ^6 \text{With linearity implied by the ordinal arrangement of primary types (from closed} - \text{ > semi-closed} - \text{ > semi-open} - \text{ > open).} \]
finance scores finds little differences across closed and blanket systems. Much of this research agenda can be summarized how Rogowski and Langella (2014) concludes: “Contrary to claims made by proponents of primary reform, the candidates running for office in closed primary systems do not appear to be substantially more ideologically extreme than candidates running for offices in open primary systems.”

When exploring candidate entrance, Robert Boatright finds in his *Getting Primaried* a lack of association between primary type and the increase in “ideological challengers” (Boatright, 2013, Table 2.7). In his look at 8,224 primary races only 1.2% of those closed primaries were considered “ideological” compared to 0.8% of those open primaries, a difference that was not found to be statistically significant.

Further, despite mixed findings with respect to the relationship of primary rules to the type of candidates they produce, scholars nevertheless control for primary rules in other study of primary elections. When studying the number of candidates in race (Herrnson and Gimpel, 1995) or fractionalization (Boatright, 2014), open primaries tend to be associated with a greater number of candidates. Other studies looking at incumbent primary vote share (Brady, Han and Pope, 2007) or other research covering candidate ideology (Nielson and Visalvanich, 2017) also incorporate primary rules and find they tend to be statistically significant controls with respect to these dependent variables.

Much of the previous literature provides little evidence for how these systems might influence primary voters, even though they are often considered along the causal mechanism connecting systems to candidates. For example, Kanthak and Morton (2001) notes in its conclusion the following theoretical explanation: “semiclosed and semiopen primaries attract independents and moderates who vote sincerely for moderate candidates (thus advantaging them), and pure-open primaries attract extremist voters (thus advantaging extreme candidates” (p. 130). This chapter picks up where
this literature has left off, by exploring both the ideological proclivities of primary voters as well as their vote choices.

3.1.2 Rules, Voters, and Polarization

Although a greater focus of scholarship has focused on the relationship between rules and primary candidates, a handful of studies have explored the relationship between rules and primary voters. Using ideal points estimated from the CCES, Sides et al. (2018) report null effects between the types of primary voters in closed, semi-closed, and open primaries when comparing them to “party followers.” As the authors of that study explain, “The differences... are not much greater in closed primaries than in open primaries, even though closed primaries thought to create larger differences by limiting the primary electorate to registered partisans” (p. 7). This finding is notably different than the one produced by Hill (2015), which concludes with counterintuitive findings, viz. that closed primaries produce more centrist primary electorates. Other research has reached different conclusions. Relying upon a different methodological approach, Kaufmann, Gimpel and Hoffman (2003) analyzes congressional exit polls from 1988 to 2000 across 113 primaries and finds that “open primaries can result in a more ideologically moderate, and a more ideologically convergent, electorate” (pg. 471).

This present chapter will draw upon methodological approaches very similar to the previous work – by using both ideal points as well as exit polls – and attempt to shed additional light on the relationship of primary rules with polarization among voters.
Hypotheses

Considering the mixed findings of previous research, this chapter continues to explore the influence that different state primary rules may have on the relative ideological representativeness of that state’s voters. Although the previous research has found a number of inconsistent findings, this present chapter borrows from the intuitive hypothesis presented in prior work and shared by the conventional wisdom by exploring the intuition that more restrictive primary rules help facilitate more ideologically extreme primary electorates.

Borrowing from works cited above that examine the relationship between primary rules and candidate ideology, this chapter tests the contention of more open systems having a moderating effect. This hypothesis, as outlined in Gerber and Morton (1998) explains, “closed primary systems will produce more extreme general election winners... than more open primary systems” (p. 312). Framed in the inverse language, Podkul (N.d.) tests, “open primary elections are associated with winning candidates who are more ideologically moderate (compared to winning candidates in closed primary elections).” Although these projects look at candidate ideology, they seem to imply more moderate candidates are the effect of more moderate electorates. In his article “The Limits of Electoral and Legislative Reform in Addressing Polarization,” Nolan McCarty summarizes this idea by writing: “It seems almost a logical certainty that opening primary elections to more nonpartisan and independent voters should have a moderating effect on politics by increasing the chance that moderate candidates get nominated” (McCarty, 2011).

Therefore, the hypotheses tested in the remainder of this chapter are:

For an additional overview, see Barber and McCarty (2013), p. 29.
$H_1$: More open primaries will be associated with lower levels of ideological extremism among primary electorates, *ceteris paribus*.

$H_2$: The relative ideological extremism associated with primary rules will follow an ordinal pattern with more open primaries being associated with more moderate primary electorates, *ceteris paribus* (i.e. from moderate to extreme: Open, Semi-Open, Semi-Closed, Closed)

The following sections will explore the data and methods used in this chapter to test these hypotheses.

3.2 Data and Methods

To study any polarizing effects produced among primary electorates by different primary rules systems, this chapter performs two related analyses to investigate, first, whether these rules are associated with more extremist primary electorates and, second, whether the rules are associated with more ideological vote patterns by these electorates. The first study explores these trends using the district-level ideological estimates developed in the last chapter. The second study introduces a second data source, the Congressional Primaries Exit Polls sponsored by The Primaries Project at The Brookings Institution from the 2016 and 2018 congressional primaries cycles. This section introduces these data and their related methods in turn.

3.2.1 Analysis 1: Using District-Level Ideological Scores

Using the district-level ideological estimates developed in the previous chapter, this chapter explores whether more closed primary systems are associated with more extreme primary electorates. This section will explain the primary dependent and
independent variables used in this analysis; next, it will introduce the equation estimated below; and, finally, it will cite additional data conditions and considerations.

Drawing upon the ideological scores estimated in the last section, this analysis measures primary voter extremism using the relative extremism of an electorate (cf. Hill 2015; Sides et al. 2018). Using the formula from Equation 2.9 – which explains extremism is measured at the district-cycle-party level by looking at the difference of ideological scores between the average voter and the average nonvoter – this chapter studies the relative extremism of primary voters. Greater values of the extremism score above zero therefore represent further ideological distance between primary voters and nonvoters while lesser values (above zero) show less ideological distance. Values below zero represent cases where primary voters are actually more moderate than nonvoters. In the tests from this analysis below, primary voter extremism serves as the main dependent variable.

The main independent variable of interest is the primary system rules, viz. whether a primary system is closed, semi-closed, semi-open, or open. Following the rules outlined in Table 3.1 above, this independent variable is measured according to The Primaries Timing Project (Boatright and Moscardelli, 2017).^8

Additional variables are considered as control variables to account for alternative explanations of the associations provided. To account for economic and demographic explanations, district-level unemployment and the percent nonwhite residing in a district are considered, respectively. Additional politically-related factors are also considered. To consider primary competitiveness as a potential explanatory cause of ideological unrepresentativeness, this project considers the effective number of candidates (ENC) in a race as well as primary turnout percent (measured as the primary

---

^8 According to that project’s codebook, this variable is: “Sourced from L. Sandy Maisel, Robert Boatright, and the National Conference of State Legislators.”
turnout divided by the district population) (Boatright, 2014). Lastly, this project also considers primary type (i.e. whether a primary is incumbent, challenger, or open seat). Podkul (N.d.) demonstrates previous findings in ideological unrepresentativeness of closed primaries with respect to candidates disappear when one considers primary type, so this project also considers this substantively meaningful category.

Using these measures, variations of the following two-way fixed effects linear model are estimated below:

\[
Extr_{c,d} = \beta_0 + \beta_{1:3} Rules_{r[c,d]} + \beta_{4:5} Type_{t[c,d]} + \beta_5 Unemploy_{c,d} + \beta_6 Nonwhite_{c,d} + \beta_7 ENC_{c,d} + \beta_8 Turnout_{c,d} + \alpha_d + \tau_c
\]  

(3.1)

where for each congressional district \(d\) (for \(d = 1, \ldots, 435\)) per cycle \(c\) (for \(c \in \{2010, 2012, 2014, 2016\}\)): \(Extr\) is the measured primary voter relative extremism, \(Rules\) represents a series of dummy variables (indexed by \(r\)) for Open, Semi-Open, and Semi-Closed primary systems (with Closed serving as the reference category), \(Type\) represents a series of dummy variables (indexed by \(t\)) for Challenger and Open Seat primaries (with Incumbent primaries serving as the reference category), \(Unemploy\) is the district unemployment rate, \(NonWhite\) is the percent of a district that is non-white, \(ENC\) is the Effective Number of Primary Candidates in the race, \(Turnout\) is the primary percent turnout, \(\alpha\) is a series of state-level dummy variables to capture time invariant state-level characteristics and \(\tau\) is a series of election cycle dummy variables. The model is estimated separately for Republicans and Democrats (cf. McGhee et al. 2014; Kanthak and Loepp 2018).

Two additional considerations were made before estimating the model. First, the association of primary voter extremism and closed primaries was only considered in
cases where an actual primary took place. That is, “races” where only one candidate appeared on the ballot (or in some cases, no candidates) were excluded from the analysis. Second, races that utilize blanket or nonpartisan primaries were also excluded from the analysis (all races in Louisiana and Washington and post-2012 races in California). Following Nielson and Visalvanich (2017), these primaries were excluded because blanket primaries have slightly different mechanisms than more traditional primaries. In a top-two nonpartisan blanket primary, for example, inter-partisan mechanisms must be considered. Further, in Louisiana’s jungle primary, candidates can outright win the election in the “first round” race. For these reasons, this chapter focuses on traditional systems.

3.2.2 Analysis 2: Exploring Vote Choice in Congressional Primary Exit Polls

The second analysis in this project draws upon a new data source: the Congressional Primaries Exit Polls. These polls, sponsored by The Primaries Project at The Brookings Institution and conducted by Edison Research, conducted a series of in-person interviews of primary voters in the 2016 and 2018 congressional primaries in randomly sampled competitive primaries across the country (cf. Geer 1989; Kaufmann, Gimpel and Hoffman 2003). In 2016, over 9,000 primary voters were surveyed and, in 2018, over 12,000 voters were surveyed. Although these surveys suffer from the regular pitfalls of exit poll research, they nevertheless provide a unique view of 21st century voters in Republican, Democratic, and Top-Two primaries across the country. This section will first introduce the data – including its sampling frame, weighting,
Table 3.2: Congressional Exit Polls Sample Sizes by District-Primary, 2016–2018

<table>
<thead>
<tr>
<th>District</th>
<th>2016</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL-4</td>
<td>373</td>
<td>416</td>
</tr>
<tr>
<td>CA-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL-6</td>
<td>114</td>
<td>204</td>
</tr>
<tr>
<td>CA-49</td>
<td></td>
<td>423</td>
</tr>
<tr>
<td>GA-9</td>
<td>375</td>
<td></td>
</tr>
<tr>
<td>FL-10</td>
<td>678</td>
<td></td>
</tr>
<tr>
<td>HI-1</td>
<td>566</td>
<td>166</td>
</tr>
<tr>
<td>FL-27</td>
<td></td>
<td>116</td>
</tr>
<tr>
<td>IA-3</td>
<td>209</td>
<td>83</td>
</tr>
<tr>
<td>ID-1</td>
<td>239</td>
<td>682</td>
</tr>
<tr>
<td>IN-1</td>
<td>653</td>
<td></td>
</tr>
<tr>
<td>IL-3</td>
<td>309</td>
<td></td>
</tr>
<tr>
<td>IN-3</td>
<td>333</td>
<td>596</td>
</tr>
<tr>
<td>MD-6</td>
<td>661</td>
<td>268</td>
</tr>
<tr>
<td>MD-1</td>
<td>381</td>
<td>614</td>
</tr>
<tr>
<td>MI-6</td>
<td>541</td>
<td>231</td>
</tr>
<tr>
<td>MD-5</td>
<td>795</td>
<td>294</td>
</tr>
<tr>
<td>MI-10</td>
<td>309</td>
<td>383</td>
</tr>
<tr>
<td>MO-1</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>MN-6</td>
<td>166</td>
<td>202</td>
</tr>
<tr>
<td>NC-9</td>
<td>271</td>
<td>262</td>
</tr>
<tr>
<td>MS-3</td>
<td>106</td>
<td>294</td>
</tr>
<tr>
<td>NJ-5</td>
<td>358</td>
<td></td>
</tr>
<tr>
<td>NC-10</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>OH-16</td>
<td>478</td>
<td>543</td>
</tr>
<tr>
<td>NY-19</td>
<td>213</td>
<td>113</td>
</tr>
<tr>
<td>OK-4</td>
<td>577</td>
<td>624</td>
</tr>
<tr>
<td>OH-14</td>
<td>273</td>
<td>402</td>
</tr>
<tr>
<td>PA-3</td>
<td>476</td>
<td></td>
</tr>
<tr>
<td>OK-2</td>
<td>107</td>
<td>115</td>
</tr>
<tr>
<td>TN-3</td>
<td>321</td>
<td></td>
</tr>
<tr>
<td>PA-2</td>
<td>459</td>
<td></td>
</tr>
<tr>
<td>TX-2</td>
<td>308</td>
<td>425</td>
</tr>
<tr>
<td>TN-5</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>TX-26</td>
<td>392</td>
<td>660</td>
</tr>
<tr>
<td>TX-4</td>
<td>135</td>
<td>473</td>
</tr>
<tr>
<td>VA-10</td>
<td>269</td>
<td></td>
</tr>
</tbody>
</table>

Data

Sample. The data for this project leverages in-person interviews of primary voters from 61 primary races across 40 congressional districts from the 2016 and 2018 electoral cycles. The polls surveyed election day voters at randomly-selected polling sites.
within sampled districts and gathered the opinions of over 22,000 voters from over 200 precincts. Table 3.2 shows the within-race sample sizes for the districts gathered. Within the sampled districts, there was a response rate of 44% and 33% for the 2016 and 2018 primaries, respectively (cf. 42% in 2014 congressional primary exit polls conducted in Bawn et al. 2017). The margin of error for the polls is reported in 2016 as ±2% for Republicans and Democrats and in 2018 as ±2% for Republicans and ±1% for Democrats. (A full summary table of the polling data in accordance with the AAPOR Code of Professional Ethics & Practices Section III-3 is reproduced in Appendix Table ??.)

Exit poll data collection began in March of each cycle (in states like Texas) and were completed as late as August (in states like Florida). Overall, the surveys used in this analysis covered contests involving 226 congressional primary candidates including 24 incumbents (with tenures ranging from five freshmen members to four members first elected before the 2000 election cycle) and 202 non-incumbents (46 challengers and 156 candidates in open seat races). The sample included two races where incumbents lost their renominations (Rep. Chaka Fattah D-PA-2 in 2016 and Rep. Robert Pittenger R-NC-9 in 2018).

Sampling Frame. The exit polls analyzed in this project studied voters in “competitive” congressional primaries. Following a clustered sampling procedure by congressional district, all congressional districts that listed two or more candidates in either the Democratic or Republican House primary were included in the districts sampling frame. This frame was then filtered to exclude states with exclusive by-mail voting (Colorado, Oregon, and Washington) as well as states with nonpartisan Top-

11 An excerpt from one of the surveys is available in appendix Figure A2.
Two primary systems (California, Louisiana, and Washington). The sample was chosen from this frame randomly, “stratified by past presidential vote and proportional to the size of the most recent presidential vote” (see Table ??). The geographic distribution of the sampling frame is reproduced for 2016 in Appendix Figure A3 and for 2018 in Appendix Figure A4. For example, if a sampled district had three Democratic candidates for Congress and one Republican candidate, only the Democratic primary voters in that district would be included in the final dataset. Across the two cycles, nine Democratic-only districts, eight Republican-only, and twenty-one both-party districts were sampled among the partisan primaries.

**Weighting.** The survey weights used in the analysis below were calculated by Edison Research to account for nonresponse bias, actual election results, and the representativeness of the sampling frame. To account for nonresponse bias, interviewers estimated the age, race, and gender of those who declined the survey and weighted respondents accordingly. To weight the responses according to the actual election results, respondents were re-weighted by their congressional vote choice in light of the final results of the election. However, this step of the weighting was minimal as the unweighted responses track rather closely with the final election results (See Figure A5 in the Appendix). Finally, to weight the responses so they are representative of the sampling frame, the data were weighted “to reflect the total number of votes [the] congressional district primary represented as a share of all competitive congressional districts” (see Table ??). The survey weights are included in all of the descriptive and modeled analyses below. Considering this weighting procedure (as well as the sampling noted in the previous section), the survey should be interpreted as representative of voters in 2016 and 2018 competitive congressional primary elections.

\(^{12}\)In the 2018 exit polls, however, two districts were randomly sampled from California and automatically included in the final survey.
Limitations. Despite the many benefits to using to this multi-race competitive congressional primary exit poll, the survey suffers from the typical drawbacks of traditional exit polls. Among the disadvantages of using exit polls in this project are nonresponse biases (see, e.g., Merkle and Edelman 2002) and voter sample issues (see, e.g., McDonald and Thornburg 2012) as well as other standard shortcomings associated with using exit poll data.\textsuperscript{13} Regarding nonresponse bias, efforts were made during the weighting described above to account for those who refused the survey (by considering the estimated demographic characteristics of the voters who refused); nevertheless, nonresponse bias of course remains among unobservable characteristics. Regarding the voter sample issue, this remains a serious issue especially in light of electorates who are more likely to vote early, by mail, or absentee. This study recognizes these issues but nevertheless presents the exit poll data with these flaws following other studies using contemporary exit polls as evidence (e.g. Donovan and Bowler 2018; Kennedy et al. 2018; Panagopoulos, Endres and Weinschenk 2018). Future iterations of this survey should carefully consider how to remedy these setbacks.

Method

Using the data introduced in the previous section, this section will explain the research method used for analyzing these data below. While Analysis 1 focuses on the relative ideological positions of primary electorates, the exit polls data can help us explore primary vote choice. To that end, this section seeks to look at the associations of extreme primary voting with different rules types.

Dependent Variable. Although the exit polls asked respondents for their positions on a handful of policy preference issues (such as trade and immigration), the

\textsuperscript{13}For a comprehensive review of exit poll measurement errors, see Chapter 2 of Bautista (2015).
surveys do not lend themselves to the same type of ideological estimation as Chapter 2. Instead, we can learn about the ideology from self-reported scores on a five-point scale. As displayed in Figure 3.1, there are somewhat asymmetric ideological positionings by voters in each party’s electorates. A plurality of Republicans, for example, identify as “very conservative” in both 2016 and in 2018 among voters in competitive primary elections from those cycles. Democratic primary voters, however, were more likely to identify as “moderate” and “somewhat liberal” in the 2016 and 2018
cycles, respectively. While Republicans are more likely to select the ideologically far right option, Democrats are much more evenly split, a finding that is consistent with scholarship on asymmetric polarization (Mann and Ornstein, 2012; Grossmann and Hopkins, 2016). However, measuring ideology among survey respondents is complicated. As explored in the last chapter, ideological self-identification – labeled in some literatures as “symbolic ideology” – may not be representative of policy preferences or other political behaviors (see, e.g., Conover and Feldman 1981; Norrander 1989; Erikson, Wright and McIver 1993).

Since this section of the chapter focuses on ideological decision-making in the ballot box this analysis studies the House candidate vote choice of primary voters. Rather than exploring at symbolic ideology, this analysis measures whether an exit poll respondent cast her vote for a moderate or extreme primary candidate. Borrowing from the selection strategy of Hall (2015), identifying “moderate” and “extreme” primary candidates followed three steps. First, I isolated all the primaries from 1980 to 2018 where there were at least two candidates. Second, these candidates were then matched with their campaign finance-estimated ideological scores using DIME data scores (Bonica, 2018). Third, following Hall (2015), races that featured first- and second-place candidate ideological score distances greater than the median distance between all first- and second-place candidates were determined as between an “extremist” and a “moderate” with the further left Democrat and further right Republican considered the “extreme candidate.” One of the benefits to using this metric is that the measure of ideological distance as determining extreme and moderate candidates allows for across-district ideological differences; a candidate who may be

---

14 For this data, I relied upon Boatright and Moscardelli (2017) as well as personal collection of 2018 returns.

15 For more details, see (Hall, 2015, pg. 21–22).
Table 3.3: Exit Poll Respondents’ Vote Choices In Ideological Races, 2016-2018

<table>
<thead>
<tr>
<th></th>
<th>Voted Extreme Candidate</th>
<th>Voted Moderate Candidate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic Primary</td>
<td>1,648</td>
<td>1,489</td>
</tr>
<tr>
<td>Voters</td>
<td>(52.5%)</td>
<td>(47.5%)</td>
</tr>
<tr>
<td>Republican Primary</td>
<td>1,899</td>
<td>2,869</td>
</tr>
<tr>
<td>Voters</td>
<td>(39.8%)</td>
<td>(60.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>3,547</td>
<td>4,358</td>
</tr>
<tr>
<td></td>
<td>(44.9%)</td>
<td>(55.1%)</td>
</tr>
</tbody>
</table>

extreme in one congressional district might be moderate in another. (This process is outlined in greater detail in this chapter’s Appendix, page 138).

Across the 2016 and 2018 congressional primaries exit polls, there were 25 races of 59 partisan primaries with top-two vote getters categorized as “extreme” and “moderate.” Pooling these voters across cycles, Table 3.3 describes the number and percent of partisan primary voters voting for each type of candidate. Contrary to the asymmetric symbolic ideology presented in Figure 3.1, this chart on vote choices finds that Republican primary voters in these districts were slightly more willing to pull the lever for “moderate” candidates while Democratic primary voters gave a slight edge to “extreme” candidates. A full list of the races and candidates that enter this sub-sample is included in Appendix Table A7. Two such candidates on this list are those who ran in 2018’s Democratic primary in Illinois’s Third Congressional District, with Rep. Dan Lipinski as the moderate and Marie Newman as the extreme candidate. Taking a closer look at the extreme candidates in this sub-sample, Appendix Figure A6 reports how these candidates performed at the polls.
**Independent Variable.** The key independent variable of interest is the primary rules, again broken down into the four categories described above: open, semi-open, semi-closed, and closed. For these election cycles, states were categorized as following these rules systems according to measures provided by the National Conference of State Legislatures. This independent variable was interacted with a partisan variable to uncover any heterogeneous effects of primary rules on extreme vote choice by political party primary.

**Control Variables.** Following the same motivations for the control variables listed in Analysis 1 above, this project looks at respondent-level characteristics as well as race-level features. The respondent-level considerations include political and ideological self-identification as well as the demographic indicators of whether a respondent was nonwhite and if she was a college graduate. The race-level features focused on primary type (incumbent, challenger, and open seat) with additional model specifications including region (Northeast, Midwest, South, and West) and whether the primary was held concurrently with the presidential primary (such as those concurrently-held races during the 2016 cycle).

**Model.** Following the structure outlined above, the following logistic regression was estimated:

---

16 The NCSL code “Partially Closed” was coded as “Semi-Closed” and the codes “Partially Open” and “Open to Unaffiliated” were coded as “Semi-Open.” These categories were collapsed to make the results from this analysis comparable with the results from Analysis 1.
\[ Pr(Extr_i = 1) = \logit^{-1}(\beta_0 + \beta_{1:3} Rules_{r[i]} + \beta_4 Republican_i + \beta_{5:7} Rules_{r[i]} \ast Republican_i + \beta_8 PID_i + \beta_9 IID_i + \beta_{10} Nonwhite_i + \beta_{11} College_i + \beta_{12:13} Type_{t[i]} ) \] 

(3.2)

where for each respondent \( i \) the probability of voting for the extremist candidate (\( Extr \)) is modeled from \( Rules \) which represents a series of dummy variables indexed by \( r \) (where \( r \in \{ \text{Semi-Closed}, \text{Semi-Open}, \text{Open} \} \) with Closed serving as the reference category), \( Republican \) is a dummy variable indicating if the voter is participating a Republican primary, \( PID \) is the partisan self-ID (arranged where higher numbers represent more Stronger Partisanship in the party primary), \( IID \) is the ideological self-ID (arranged where higher values are more conservative/liberal in Republican/Democratic primaries), \( Nonwhite \) is a dummy variable if the respondent is nonwhite, \( College \) is a dummy variable if the respondent graduated college, and \( Type \) represents the primary type \( t \) that respondent \( i \) participates in (where \( t \in \{ \text{Challenger}, \text{Open Seat} \} \) with Incumbent serving as the reference category). The model is estimated with the survey weights described above and with robust clustered standard errors by congressional district.

### 3.3 The Null Effects of Primary Rules on Polarization

This section outlines the results from each of the two analyses. The first section outlines the findings from the study using the district-level primary electorate extremism scores (2010–2016) while the second uses the data provided by the congressional primary exit polls (2016–2018).
3.3.1 Analysis 1: Results

The results from Analysis 1 are presented in Table 3.4. The first two columns use the dependent variable of primary voter extremism (described above as the estimated ideological distance between primary voters and primary nonvoters for a district) while the second two columns use the primary voter score (irrespective of how that primary voter score relates to primary nonvoters in a district). The four models presented only include races where a primary election was held (from 2010 to 2016) and are broken out by Republican and Democratic electorates.

Focusing first on the coefficients for primary rules, each coefficient is compared to Closed primaries, which acts as the reference group. Among Democratic electorates, only Open primaries are associated with a statistically significant higher level of extreme primary electorates, ceteris paribus. The findings from the primary voter score model confirm the direction of each of the Rules coefficients. From these models, it appears that Open rules are associated with more extreme electorates among Democratic electorates.

\[\text{Democratic primary voter scores have been multiplied by (-1) so cross-party comparisons can be made. For both partisan electorates, higher scores represent more extreme electorates.}\]
Among Republican electorates, however, there is more complicated “nonlinear”\(^\text{18}\) conclusion. For Republicans, open rules are associated with more moderate primary electorates. Semi-closed primaries are also associated with more moderate electorates whereas semi-open primaries are associated with more extreme electorates. Just as in the Democratic models, the direction of the coefficients is also confirmed in

\(^{18}\)To borrow the phrase from Kanthak and Morton (2001).
the fourth column. From the Republican models, it appears Open rules are associated with more moderate electorates and Semi-open rules are associated with more extreme electorates. Nevertheless, the magnitude of these coefficients is relatively small. For example, the coefficient estimated for Rules: Semi-Closed in the Republican Extremism Score model is -0.02. This means, controlling for other factors, semi-closed primary systems are associated with a shift of 0.02 in the moderate direction for primary extremism. This shift represents less than 10% of the average distance of Republican primary voters compared to primary nonvoters.

To further explore the patterns of primary rules and primary vote extremism, an additional set of models interacting primary rules and cycle (run separately for Democratic and Republican electorates) was estimated. This additional set of models provides the opportunity to look at the relationship of primary systems in light of otherwise secular trends of polarization. Predicted Primary Vote Scores are displayed in Figure 3.2 for each of the four primary rules for the election cycles covered in the data with control variables held at their mean for numeric variables and their mode for categorical variables. Although point values from this model slightly differ across panels, the trends within each panel are nearly identical across the four primary types for both Republican and Democratic primary voters. The findings from this figure confirm the lack of substantively meaningful results displayed in Table 3.4.

\[^{19}\text{cf. McGhee et al. (2014), Figure 3.}\]
Although Analysis 1 finds some statistically significant findings the substantive meanings of these differences are minor at best. For example, the largest coefficient presented in Table 3.4 finds that Democratic electorates are more extreme by a value of 0.07 in Open primaries when compared to closed primaries. As explained in Chapter 2, this is a minor substantive difference. As charted in Figure 3.2, the differences across primary systems are minor and (according to that set of models) rarely statistically
significant. In the debate between Hill (2015) and Sides et al. (2018), this analysis
sides with Sides et al. (2018).

3.3.2 Analysis 2: Results

Next, the exit polls studied in Analysis 2 produce similar findings as Analysis 1.
Table 3.5 displays these findings. The dependent variable for each model remains
whether the exit poll respondent voted for the extreme candidate over the moderate
candidate in her chosen primary election. The two columns represent similar model
specifications with the second column adding geographic region and primary con-
currence with the presidential election. Just as in Analysis 1, the reference group is
Closed primary rules. As noted the model is weighted according to the survey weights
and the model uses robust standard errors clustered by congressional district.
Table 3.5: Primary Rules Logistic Model: Exit Poll Vote Choices

<table>
<thead>
<tr>
<th></th>
<th>Voting for Extremist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Rules: Open</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td>(1.05)</td>
</tr>
<tr>
<td>Rules: Semi-Open</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
</tr>
<tr>
<td>Rules: Semi-Closed</td>
<td>−0.64</td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
</tr>
<tr>
<td>Republican</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>(1.32)</td>
</tr>
<tr>
<td>Rules: Open*Republican</td>
<td>−3.98**</td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
</tr>
<tr>
<td>Rules: Semi-Open*Republican</td>
<td>−1.99</td>
</tr>
<tr>
<td></td>
<td>(1.41)</td>
</tr>
<tr>
<td>Rules: Semi-Closed*Republican</td>
<td>−1.11</td>
</tr>
<tr>
<td></td>
<td>(1.50)</td>
</tr>
<tr>
<td>Respondent-Level</td>
<td></td>
</tr>
<tr>
<td>Party-ID</td>
<td>−0.12</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Ideological-ID</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
</tr>
<tr>
<td>College Grad</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Race-Level</td>
<td></td>
</tr>
<tr>
<td>Type: Challenger</td>
<td>−0.59</td>
</tr>
<tr>
<td></td>
<td>(0.84)</td>
</tr>
<tr>
<td>Type: Open Seat</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
</tr>
<tr>
<td>Region: Northeast</td>
<td>−1.05</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Region: South</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Region: West</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Concurrent w/ Pres.</td>
<td>−0.39</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>−0.36</td>
</tr>
<tr>
<td></td>
<td>(1.06)</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>7128</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.21</td>
</tr>
<tr>
<td>L.R.</td>
<td>1246.40</td>
</tr>
</tbody>
</table>

***p < 0.001, **p < 0.01, *p < 0.05.
Robust standard errors clustered by congressional district.
The findings from Table 3.5 fail to find a statistically significant coefficient for any of the primary rules indicators or political party. For the interacted terms (Rules * Republican), only one of the terms is significant: Rules : Open * Republican. This term is statistically significant in the negative direction indicating that for voters in the Republican primary, Open primaries are associated with higher levels of voting for more moderate candidates. However, in a joint significance test of the interaction terms in this model, this effect is not statistically significant.

The null findings from this exit polls analysis suggest little relationship between primary participation rules and primary electorate vote choice. That is, when primary voters are provided the opportunity to vote between a moderate candidate and an extreme candidate, this choice does not appear to be related to primary system, controlling for respondent- and race-level factors.

Analysis 2 generally confirms the findings from Analysis 1: Primary rules are not associated with polarizing effects.

3.4 DISCUSSION

The main finding from this chapter is that primary rules regarding voter participation do not appear to have an effect on either the relative ideology of primary electorates or the vote choices made by primary voters at the ballot box. Using a pair of analyses looking at primary voters from the last eight years, this chapter concludes the following: 1) there are no significant substantive effects of primary rules on the electorates that show up to the polls with respect to their ideological positioning; and 2) there are no clear effects of primary rules on voting for ideological extremists. This look at congressional primaries identifies a null finding. Whether open or closed or somewhere in between, congressional primary rules on participation do not appear to
have an influence on either the ideological composition of primary voters or primary voters’ voting patterns.

Earlier, this chapter outlined the mixed findings present in the political science literature on this topic. Among others, this chapter agrees with the findings from Sides et al. (2018) and Rogowski and Langella (2014). To quote Kaufmann, Gimpel and Hoffman (2003) (as cited in Sides et al. 2018): “... the key to greater ideological representativeness is not the rules alone.” This chapter concurs with this analysis and finds little difference across primary rules.

Ceteris paribus, Rep. Lipinski would have likely been renominated in a closed primary and then-candidate Alexandria Ocasio-Cortez would have probably defeat Rep. Crowley in an open primary anyway. As far as ideological policy preferences and vote choice are concerned, primary rules do not make much of a difference.
For decades, the number of congressional seats safely held by each of the two major parties has skyrocketed. According to one measure from The Cook Political Report, following the 2016 election there were only 72 congressional districts that were “swing seats” (measured as being between D+5 and R+5 by the Cook Partisan Voting Index). Accounting for less than one-sixth of the U.S. House Chamber, these 72 districts in 2016 represent a 56% decline since 1997 and a 20% decline since 2012 (Wasserman and Flinn, 2017). As a consequence of the increasing number of safe seats, the ideological make-up of Congress changes a great deal, especially when members in safe seats are replaced by more ideological co-partisans (Theriault, 2006).

Further, the expansion of safe seats also has an impact on congressional primary elections. Despite their relatively low turnout – fewer than one in five Americans voted in the “high voter turnout” House primaries of 2018 (DeSilver, 2018) – primaries are increasingly becoming the central forum for voters to make decisions about who represents them in Congress. When the November general election becomes somewhat of a foregone conclusion, primary elections become the voters’ main mechanism for deciding upon their representation in Washington, DC. Considering this fact, understanding the conditions that bring about primary voter extremism becomes all the more significant for understanding voter behavior, ideology, and polarization among voters, candidates, and elected representatives.
Take, for example, the 2010 Republican primary in South Carolina’s 4th Congressional District. According to the district ideological estimates from Chapter 2, in 2010 Republican voters in SC-4 were significantly more extreme relative to Republican nonvoters compared to the average Republican primary constituency. Additionally, South Carolina’s 4th Congressional District was also a very safe Republican seat; in 2010, the district was considered an R+15 following McCain’s dominant performance in the district during the 2008 presidential contest beating Obama 60.6% to 37.7%.

Beginning in 2005, the district was represented by conservative Republican Rep. Bob Inglis.\(^1\) For most of his congressional career, Inglis was a reliable conservative vote. He routinely scored in the 90’s in the American Conservative Union scorecard (American Conservative Union, 2018) and, according to Vote View, was more conservative than 65% of Republicans in the 111th Congress (with a NOMINATE score of 0.52).\(^2\) However, on a number of key votes Inglis demonstrated some moderate inclinations. His support for Wall Street bailouts (WYFF, 2010) and man-made climate change (Corn, 2010) as well as his opposition to the Iraq War troop surge of 2007 caused many in his district to think he had, according to Politico, “gone wobbly on conservative values” (Isenstadt, 2010).

Inglis’s perceived moderation on key issues caused four different candidates to primary him.\(^3\) Among those challengers was 7th Circuit Solicitor Trey Gowdy. Gowdy, supported by many conservatives in the district including those associated with the tea party movement, forced Inglis to a primary runoff where Gowdy ultimately won in a landslide, 70.7% to 29.3% (South Carolina State Election Commission, 2010).

\(^1\)Inglis previously represented the district from 1993 to 1999 after stepping down to run for the U.S. Senate, which he lost.

\(^2\)See, Lewis et al. (2017).

\(^3\)According to Boatright (2013), these challenges were coded as “ideological” in nature (data accessed from Boatright and Moscardelli 2017). More details on ideological primary challenges are presented in Chapter 5.
Undoubtedly, Gowdy’s presence in the U.S. House of Representatives was a step to the right for the district as he was more conservative than 90% of Republicans in the 112th Congress (with a NOMINATE score of 0.66). Gowdy went on to have a high-profile congressional career, which included stints as Chairman for both the House Oversight Committee and the Select Committee formed to investigate the Benghazi Terrorist Attack of 2012. This career – and the rightward pull in South Carolina 4th district’s representation – began with a primary.

Seeking to better understand the dynamic in the 2010 South Carolina Republican primary, this present chapter explores whether safe seats (“extreme districts”) are associated with more extreme primary electorates. To do so, this chapter unfolds in four sections. First, a brief review of the literature looking at the relationships of safe seats to primary elections is explored. Second, the data and methods are introduced. Following the outline of the last chapter, this section first explains the data and methods for an analysis using the district-level ideological estimates introduced in Chapter 2 and then explains the data and methods for a second analysis using the congressional primary exit polls introduced in Chapter 3. Third, the results from these analyses are explored. Lastly, a discussion of these findings – and how they prepare for the next chapter’s look at primaries and representation – concludes.

4.1 Two Electorates

Given the two-stage nature of contemporary American elections, political scientists have raised questions about the separate overlapping constituencies: the primary and the general electorates. Some scholarship emphasizes the overlapping nature of these electorates. Fenno (1978), for example, expressly acknowledges these constituencies as “concentric,” with the primary constituency embedded within the broader “the

---

4See, Lewis et al. (2017).
reelection constituency.” Nevertheless, his *Home Style* acknowledges the necessity of members of Congress appealing to their primary voters. Quoting one of the members he interviews, Fenno writes, “Everybody needs some group which is strongly for him – especially in a primary. You can win a primary with 25,000 zealots.” Other researchers, however, emphasize this dissimilarities between these groups. For example, Owen and Grofman (2006) explore how additional considerations in the first round (such as general election viability) influence polarization.

In today’s politics, however, we must consider another variable when studying the divergence of primary and general electorates: district partisanship. As explained in this chapter’s introduction, today there are fewer and fewer “swing seats” with many districts’ incumbent partisans nearly guaranteed victory in the November election. Writing about the politics of California, Seth Masket explains this dynamic of contemporary elections:

“A few decades ago... it was a challenge for officeholders to please both the median voter in the primary and the median voter in the general election.... In a ‘safe’ district, however, the median general election voter is a partisan rather than an unattached independent; the tension between pleasing both the primary and general median voters is largely alleviated” (Masket, 2011, p. 49).

While it is certainly the case that the median voter in a safe district is a partisan, it is a unique empirical question as to whether primary voters are even more extreme under these conditions. This chapter explores this question, viz. whether primary voters are even more extreme in “safe seats.”

Previous literature has explored the relationship of “safe districts” (via district partisanship) to ideology in primaries. For example, district partisanship has been found
Table 4.1: Relative Primary Voter Extremism by PVI Grouping, 2010-2016

<table>
<thead>
<tr>
<th></th>
<th>≥ D+20</th>
<th>D+6 to D+19</th>
<th>D+5 to R+5</th>
<th>R+6 to R+19</th>
<th>≥ R+20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic</td>
<td>97.6</td>
<td>84.2</td>
<td>88.5</td>
<td>88.2</td>
<td>84.1</td>
</tr>
<tr>
<td>Republican</td>
<td>92.9</td>
<td>83.6</td>
<td>97.5</td>
<td>96.3</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: Cells reflect % of primary constituencies where voters are more ideologically extreme compared to co-partisan primary non-voters. Only constituencies in contested primaries included.

to be related with incumbent performance (Herrnson and Gimpel, 1995; Thomsen, 2018) as well as candidate extremism (Nielson and Visalvanich, 2017). It has not, however, been associated with how primary participation rules (see Chapter 3) influence candidate ideology (McGhee et al., 2014; Kanthak and Loepp, 2018). While this literature mostly explores the effects on candidates, this chapter attempts to fill a gap in the literature regarding the relationship of district partisanship with ideology by focusing on the attitude and behaviors of primary voting electorates.

Précising the district-level constituency estimates developed in this project’s Chapter 2, Table 4.1 summarizes the prevalence of primary voter extremism by district partisanship scores for congressional district primary constituencies when a primary election was held from 2010 to 2016. Each column of the table represents different bins of Cook PVI. The Cook Partisan Voting Index (PVI) was first developed in 1997 and “measures how each district performs at the presidential level compared to the nation as a whole.”\(^5\) This measure identifies which districts are reliably Republican, reliably Democratic or “toss-up.” A district that is categorized

as “R+4 means the district performed [in the previous two presidential elections] four points more Republican than the national average.” The PVI cutoffs provided above can otherwise be interpreted (from left to right) as representing districts that are “safe Democrat,” “lean Democrat,” “toss-up,” “lean Republican” and “safe Republican.”

As Table 4.1 demonstrates, both Democratic and Republican primary voters are more likely to be extreme relative to nonvoting co-partisans in safe Democratic and safe Republican districts, respectively. For example, 100% of Republican primary voters in districts that are R+20 or greater are more ideologically extreme than nonvoters in those electorates and nearly 98% of Democratic primary voters in districts that are D+20 or greater are more liberal. However, it is worth noting that the relationship of percent primary voter extremism is not linear across the PVI bins provided. Further, this table does not explore the magnitude of these ideological distances. Analysis 1 below explores this relationship more explicitly and considers alternative explanations for extremism.

**Hypotheses**

Are primary voters in seats that are safe in the general election more likely to be ideologically extreme? When districts are more reliably Republican(/Democrat) are primary voters more likely to be conservative(/liberal) compared to non-voting co-partisans? As Gary Jacobson explains in his *The Politics of Congressional Elections:*

“The primary election system also complicates the pursuit of a congressional career. Candidates must be prepared to face two distinct, if overlapping, electorates. Primary electorates are much more partisan and prone to ideological extremity, and the need to please them is one force behind party polarization in Congress” (Jacobson, 2009, p. 20)
This chapter explicitly explores the implied empirical question of primary electorates “prone[ness] to ideological extremity” as it relates to electoral context.

Following from this assumption, this chapter tests the following hypothesis:

\[ H_1 \textbf{Extremist Voters in Extreme Districts.} \] Primary voters in more partisan-favorable districts are relatively more ideologically extreme compared to primary non-voting co-partisans, \textit{ceteris paribus}.

Considering the preponderance of evidence that the two major political parties are not equally polarized (i.e. with evidence suggesting Republicans have moved rightward more than Democrats have moved leftward) (see, e.g., Mann and Ornstein 2012; Grossmann and Hopkins 2016; Kamarck, Podkul and Zeppos 2017), a second hypothesis aimed toward understanding any partisan asymmetries is also explored:

\[ H_2 \textbf{Asymmetric Primary Voter Polarization.} \] Relative to co-partisan non-voters, Republican primary voters are more ideologically extreme in reliably Republican districts but Democratic primary voters are not more ideologically extreme in reliably Democratic districts, \textit{ceteris paribus}.

The following section explains the data and methods for the two analyses seeking to test these hypotheses.

4.2 \textbf{Data and Methods}

Similar to the format presented in the previous chapter, Chapter 4 relies upon a pair of related analysis to test its hypotheses. The first analysis explores primary voter relative ideological extremism across different “swing” and “safe” districts. The second analysis explores individual proclivities to vote for ideologically extreme candidates across these types of districts using the 2016 and 2018 congressional primaries exit polls.
4.2.1 Analysis 1: District-Level Ideological Estimates

To explore the relationship between district-level ideological estimates, this analysis surveys the district-level ideological estimates developed in Chapter 2. These estimates provide the unique perspective for looking at relative primary voter extremism for each party’s constituency in each district across four different election cycles.

The dependent variable for this analysis is the measure of relative primary voter extremism. Looking at the ideological score of primary voters compared to primary nonvoters allows us to explore not just the conditions of primary voter extremism across districts but rather the conditions of primary voter extremism within a particular district. Focusing on this theme of relative ideological representativeness this chapter the following measure is again utilized as the dependent variable:

\[ Extr_{c,d,p} = \theta_{c,d,p,\text{voter}} - \theta_{c,d,p,\text{nonvoter}} \]  

where \( Extr \) measures the relative ideological extremism for a constituency in district \( d \) during cycle \( c \) for political party \( p \) and \( \theta \) represents the subpopulation’s MRP estimated ideological score.

To measure “swing” and “safe” seats, this section relies upon various measures of district partisanship. The first is the Cook PVI rating of a district in a particular cycle. As stated earlier, this metric measures the relative “partisanship” of a district by drawing on how the district performed in the previous presidential elections relative to other congressional districts.\(^6\) As an alternative metric of district partisanship, this analysis also studies the district’s Democratic candidate’s top-two vote share.

\(^{6}\)Cook PVI is usually presented in a format indicating the party and relative strength (e.g. R+5). For the analyses below, this variable was recoded numerically from Republican to Democratic. (e.g. R+5 is measured as -5, D+2 is measured as 2, and EVEN is measured as 0.)
in the most recent presidential election as a secondary measure. The statistical and substantive findings across these two metrics are equivalent.

This analysis also considers potential alternative explanations for variations in the dependent variable by controlling for a variety of political, economic, demographic, and electoral factors. (For more details on the specific measures included, see page 56.)

In this analysis, the following two-way fixed effects linear model is estimated:

\[
Extr_{c,d} = \beta_0 + \beta_1 Partisanship_{c,d} + \beta_2 Partisanship^2_{c,d} + \beta_3 Type_t[c,d] \\
+ \beta_5 Unemploy_{c,d} + \beta_6 Nonwhite_{c,d} + \beta_7 ENC_{c,d} + \beta_8 Turnout_{c,d} \\
+ \alpha_d + \tau_c + \epsilon_{c,d}
\] (4.2)

for each district \(d\) in cycle \(c\) where \(Extr\) is a primary constituency’s relative ideological extremity and \(Partisanship\) measures the relative partisanship for a district in a cycle using the two metrics noted above. (For interpretations of the control variables, see the explanation following Equation 3.1 in the previous chapter.) The model is estimated separately for Republicans and Democrats.

Equation 4.2 notably includes a polynomial term for the measure of district partisanship \((Partisanship^2)\). This term is included to account for nonlinearities in the data. As demonstrated above in Table 4.1 (which measures the percentage of constituencies with extreme primary electorates by Cook PVI), relative primary voter extremism may not necessarily follow a pattern of (party safe seat – swing seat – opposite party safe seat). The squared partisanship term attempts to estimate this nonlinear relationship.\(^7\)

\(^7\)Additional polynomial models are estimated in the Appendix (see Table A8).
4.2.2 Analysis 2: Congressional Primaries Exit Polls

In the next analysis, we turn again to the congressional primaries exit polls from 2016 and 2018. While Analysis 1 focuses on relative primary vote ideological extremism as its key dependent variable, Analysis 2 studies ideological vote choices made by primary voters.

In this analysis, the dependent variable is whether an exit poll respondent voted for the ideologically extreme candidate. Among sampled Republican primary voters, this would include voters who self-reported voting for the relatively more conservative candidate and, among Democratic primary voters, this would include voters who opted for the more liberal candidate in situations where the top two candidates included a discernibly moderate candidate facing off against an extreme candidate. Following Analysis 1, this analysis also explores district partisanship – whether a district is a safe or swing seat – using Cook PVI. Figure 4.1 displays the distribution of sampled races where there was an extreme candidate versus a moderate candidate across Cook PVI ratings. As the figure demonstrates, there are races across the district partisanship spectrum for both of the party primaries. (For robustness, additional model specification use the Democratic candidate’s most recent top-two presidential vote share as another measure of district partisanship.)

Again controlling for demographic, political, and electoral factors, this analysis controls for explanations caused by respondent-level and race-level descriptors. (For more details, see the controls explanation provided on page 65.)

Following the structure outlined above, the following logistic regression was estimated:
\[ Pr(Extr_i = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \text{Partisanship}_i + \beta_2 \text{Republican}_i + \beta_3 \text{Partisanship}_i \times \text{Republican}_i + \beta_4 \text{PID}_i + \beta_5 \text{IID}_i + \beta_6 \text{Nonwhite}_i + \beta_7 \text{College}_i + \beta_{8:9} \text{Type}_{t[i]} ) \] (4.3)

where \( Pr(Extr_i) \) explores the probability of primary voter \( i \) voting for the extremist primary candidate and Partisanship measures a district partisanship for the district in which voter \( i \) is participating. Like in the last chapter, this partisanship term is interacted with whether the voter was participating in the Republican primary election. The models estimating Equation 4.3 below are weighted according to the
survey weights described above and estimated with robust standard errors clustered by congressional district.

4.3 Republican Extremism and Democratic Representativeness in Safe Seats

The results from Analyses 1 and 2 are presented below in order.

4.3.1 Analysis 1

The coefficient estimates from Equation 4.2 are presented in Table 4.2. Each of the four columns display estimates of different model specifications for testing the relationship of district partisanship on primary electorates’ relative extremism. The first two columns isolate Democratic primary electorates and the last two columns exclusively study Republican primary electorates. Each of the four columns explores all partisan constituencies from 2010 to 2016 where a primary election was actually held. The first model specification for each of the parties (columns 1 and 3) measure district partisanship via Cook PVI ($CookPVI$) while the second model specification (columns 2 and 4) measures it using the Democratic candidate’s top-two vote share in the most recent presidential election ($Dem.PresVoteShare$). For each of this model, the partisanship measure was included along with partisanship-squared ($CookPVI^2$ and $DemPresVoteShare^2$, respectively).\(^8\)

As Table 4.2 finds, there is a statistically significant relationship between district partisanship and relative primary vote extremism. Looking first at the models measuring partisanship with PVI (observed from -20 to 40), the coefficients for the

\(^8\)Additional models featuring higher order polynomials estimate similar results. To see these model estimates, see Table A8.
Table 4.2: District Partisanship Linear Model: District Ideology Estimates

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th></th>
<th>Republican</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cook PVI</td>
<td>0.00***</td>
<td></td>
<td>−0.00***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Cook PVI²</td>
<td>−0.00***</td>
<td></td>
<td>−0.00*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Dem. Pres Vote Share</td>
<td>0.47***</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td></td>
<td>(0.09)</td>
<td></td>
</tr>
<tr>
<td>Dem. Pres Vote Share²</td>
<td>−0.32***</td>
<td>−0.20*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>Type: Challenger</td>
<td>0.00</td>
<td>0.01</td>
<td>−0.02**</td>
<td>−0.02**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Type: Open Seat</td>
<td>0.01</td>
<td>0.01*</td>
<td>−0.01</td>
<td>−0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>District Unemployment (%)</td>
<td>0.18</td>
<td>0.18</td>
<td>−0.37***</td>
<td>−0.41***</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.09)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>District Nonwhite (%)</td>
<td>0.06**</td>
<td>0.05*</td>
<td>0.15***</td>
<td>0.14***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Effective Num. of Candidates</td>
<td>−0.01</td>
<td>−0.00</td>
<td>0.00</td>
<td>−0.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Primary Turnout (%)</td>
<td>0.08</td>
<td>0.08</td>
<td>−0.04</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.18***</td>
<td>0.02</td>
<td>0.26***</td>
<td>0.31***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>State and Cycle F.E.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>954</td>
<td>968</td>
<td>1042</td>
<td>1058</td>
</tr>
<tr>
<td>R²</td>
<td>0.83</td>
<td>0.83</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.82</td>
<td>0.82</td>
<td>0.76</td>
<td>0.76</td>
</tr>
</tbody>
</table>

***p < 0.001, **p < 0.01, *p < 0.05
first order term are statistically significant in the positive direction and in the negative direction for the Democratic and Republican models, respectively. The second order terms for both are negative. These results are mostly consistent across partisanship measures. In the models measuring partisanship via Democratic Presidential Vote Share (measured from 0 to 1), the first order variables are positively statistically significant and not statistically distinguishable from zero for the Democratic and Republican models, respectively. The second order estimates, however, are statistically significant in the negative direction.

For ease of interpretability of the polynomial coefficients, Figure 4.2 demonstrates the predicted primary voter relative extremism across the full range of observable Cook PVI ratings (arranged from most Republican to most Democratic, left to right) using the models presented in the first and third columns of Table 4.2. Each point along the lines (with the solid red line representing Republican electorates and the dashed blue line representing Democratic electorates) represents the predicted value of relative primary vote extremism at a given value of Cook PVI while holding the additional control variables at their averages.9 Values above the solid horizontal line (where the y-axis = 0) show primary voters in a district are ideologically extreme compared to primary nonvoters while values below the line find primary voters to be more moderate. Shaded bars represent two standard errors around the predicted values from the estimates.

The findings from Figure 4.2 demonstrate a great deal of heterogeneity of the influence of district partisanship on primary voter extremism across voters belonging to the two major political parties. On the one hand, among Democratic primary electorates, relative primary vote extremism is not affected by district partisanship. For districts that are R+10 to D+40, Democratic primary voters are generally represent-

---

9Numeric variables were held to their median and categorical variables at their mode. 88
tive of primary non-voters, *ceteris paribus*. However, these predictions due reveal that in safe Republican districts Democratic primary voters are slightly more moderate than Democratic primary non-voters.

On the other hand, among Republican electorates, Republican primary voters are expected to be more ideologically extreme than primary non-voters across all levels of district partisanship, especially in safer seats. In Republican safe seats (e.g. near R+20), primary voters are expected to be significantly more extreme than non-voters – at a rate much higher than the average value of primary voter extremism (0.2). Even in swing seats (e.g. PVI: EVEN or 0), these voters as still expected to be more
ideologically extreme. Going even further into more Democratic districts, Republicans do not “moderate” like Democratic electorates do in Republican districts but rather remain extreme compared to non-voters (albeit less so than those Republican primary voters in Republican districts).

Analysis 1 seems to confirm the Asymmetric Primary Voter Polarization hypothesis listed above. Primary voters are extremist in extreme districts – but, according to these data, only those voters belonging to the Republican Party.

4.3.2 Analysis 2

While Analysis 1 identifies the asymmetric relationship of district partisanship and primary voters’ relative ideology, Analysis 2 focuses on ideological vote choice. Table ?? presents the model estimates from the exit polls model presented in Equation 4.3. The first column uses Cook PVI as the measure of district partisanship and the second uses Democratic Presidential Vote Share. In both models, this district partisanship measure is interacted with a dummy variable identifying whether the voter is participating in a Republican primary.

For a simpler interpretation, Figure 4.3 presents the predicted probability of voting for the extreme candidate across observed Cook PVI (using the first column in Table ??) with variables set to their median and modal values. Solid lines represent the predicted probability score for Republican (dark red) and Democratic (navy blue) primary voters while dashed lines represent boot-strapped 95% confidence intervals.\(^{10}\)

According to Figure 4.3, there is an interesting finding: both Democratic and Republican primary voters are much more likely to vote for extremist candidates over moderate ones in districts’ that are considered safe by the other party. For example, from this data, Republicans are much more likely to vote for a more conservative

\(^{10}\)Using a percentile-based bootstrap confidence interval with 5,000 bootstrap samples.
Table 4.3: District Partisanship Logistic Model: Exit Poll Vote Choices

<table>
<thead>
<tr>
<th></th>
<th>Voting for Extremest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Cook PVI</td>
<td>-0.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Dem. Pres Vote Share</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Republican</td>
<td>-0.57</td>
<td>-0.74</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>Cook PVI*Republican</td>
<td>0.07**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Dem. Pres Vote Share*Republican</td>
<td>0.03*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Respondent-Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party-ID</td>
<td>-0.07</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Ideological-ID</td>
<td>0.13*</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>College Grad</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Race-Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type: Challenger</td>
<td>-1.03</td>
<td>-1.01</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
<td>(0.94)</td>
</tr>
<tr>
<td>Type: Open Seat</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.63)</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.62</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>(0.54)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>7128</td>
<td>7128</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>L.R.</td>
<td>492.06</td>
<td>504.85</td>
</tr>
</tbody>
</table>

---

***p < 0.001, **p < 0.01, *p < 0.05.

Robust standard errors clustered by congressional district.

candidate in a safe Democratic district and much less so in a safe Republican district. Across the aisle, Democratic primary voters act similarly: primary voters in very safe Democratic districts are less likely to pull the lever for an extreme candidate but are much more willing to do so in very red districts. Further, it is also noteworthy, according to this metric, that in even districts (e.g. where the x-axis value is = 0) Democratic primary voters are actually more likely to vote for the extreme candidate when compared to Republican primary voters. In these even districts, Republican
voters’ likelihood of voting for the extreme candidate is statistically indistinguishable from a coin flip but Democratic voters’ are likely to vote for the extreme candidate nearly 65% of the time, holding all else equal.

According to these findings, primary vote choices between extreme and moderate candidates do not follow the expected direction. Instead, voters in “safe seat” districts have a greater tendency to support the more moderate candidate. While this finding may be a function of the sampling procedure of the congressional primaries exit polls (viz. that they sample “competitive primaries” from a restricted sampling frame) or the non-randomness of cases in that sample where an extreme candidate is facing a moderate candidate, this analysis nevertheless presents an interesting study that requires further exploration,
4.4 Discussion

The findings from this chapter reveal a bit of a paradox. When it comes to the relative ideological extremism of primary voters compared to nonvoters in the same district, it appears that broader electoral context, viz. a district’s overall partisanship, is associated with ideological extremism among adherents to the Republican Party but not those of the Democratic Party. Specifically, for Republicans, Republicans voters are more likely to be relatively ideologically extreme the more Republican-lean ing a district is but, for Democrats, a similar relationship was not found.

Yet, when it comes to ideological voting measured by primary vote choices between moderates and extremists, it appears that voters in primaries for both major political parties tend to reveal somewhat of an opposite tendency. Instead, voters in both parties tend to support relatively moderate candidates over more extreme ones in safer seats. That is, Republicans(/Democrats) are more likely to support moderate candidates the more Republican(/Democratic)-leaning a district is.

Why might this be the case? Assuming this finding is not driven by any of the methodological concerns written about above, there might be a handful of explanations. First, one’s ideology does not necessarily dictate which candidate they may support at the polls. There are a number of other considerations that voters may weigh at the ballot box, such as electability, identity, aptitude, and other strategic concerns. Although voters in more extreme districts might be further to the right ideologically-speaking (as is the case for Republicans), it does not necessarily follow that these ideological positions translate to voting patterns. For example, a voter may explicitly vote against her ideological preference in a primary to support a candidate of her party who she perceives has a better chance against the opposite party in the November election. Even in these “safe seats,” voters may have to weigh such concerns.
The findings from this chapter demonstrate that asymmetry in ideological representativeness among primary electorates across the two major parties when it comes to district partisanship. However, it also finds that voters within both parties who cast ballots in safer seat districts tend to elect the more moderate candidate, even controlling for other factors. In the next chapter, I examine how these patterns may relate to greater trends in polarization among candidates and elected officials produced by congressional primaries.
Before the 2018 congressional midterm elections, conservative Senator Jeff Flake (R-AZ) found himself facing a precarious electoral fortune. Despite his reliably conservative track record in Congress – he was once described as being “the Tea Party before the Tea Party was cool” (Wilson, 2017) – Flake was often the target of right-leaning activists due to his vocal criticisms of the sitting Republican president with respect to a number of political and practical issues. As he approached the election season of his first senatorial reelection campaign, Flake needed to make a decision: run for re-election but risk an ideological primary challenge or retire from the U.S. Senate altogether.

Ultimately, Flake decided not to run for reelection. In sharing this decision with the American public, the Senator made a retirement speech on the floor of the U.S. Senate. In the message, Flake provided a candid glimpse into the thought process that guides legislators’ decision-making. Lamenting the polarized state of U.S. politics, Flake clarified:

“When we remain silent and fail to act when we know that that silence and inaction is the wrong thing to do because of political considerations, because we might make enemies, because we might alienate the base, because we might provoke a primary challenge, because ad infinitum, ad nauseam, when we succumb to those considerations in spite of what should
be greater considerations and imperatives in defense of our institutions and our liberty, we dishonor our principles and forsake our obligations” (emphasis added) (Flake, 2017)

Confirming a dynamic (viz. the significance of primary threats) that qualitative political science research has uncovered via legislator interviews (Mayhew, 1974; Fenno, 1978; Kamarck and Wallner, 2018), Flake connects contemporary polarization to primaries via both candidates (i.e. “provoking a primary challenge”) and voters (i.e. “alienating the base”). Following from this motivation, the present chapter examines the relationships of primary voter ideology with both the ideology of political actors and the presence of ideological primary challenges. Summarized more explicitly, this chapter tests two closely related research questions on the topic of primary voter ideology and representation (or polarization):

1. Do relatively more extreme primary electorates facilitate the nomination or election of more ideologically extreme political actors?

2. Do more extreme primary electorates bring about a greater number of ideological primary challenges to sitting party incumbents?

The remainder of this chapter seeks to shed light on each of these two research questions in turn. The first section analyzes explicitly how relative primary voter extremism might be connected to more ideologically extreme political actors. Relying upon the constituency-level estimates described above, this section explores a variety of extremism measures and their potential influence on existing ideological measures for actors such as party nominees and elected officials. The second section explores the relationship of relative primary voter extremism on the dynamic of getting primaried,
that is the explicit launch of a primary campaign to challenge a sitting incumbent on expressly ideological grounds. Finally, the chapter concludes with a brief discussion of the findings and suggested avenues for future research.

5.1 Voter Extremism and Polarization in Representation

This present section aims to explore the relationship between primary voter ideological extremism and the ideological positions of political actors – viz. elected politicians and political candidates – across congressional elections. This section begins with a brief outline of the political science motivations behind the research question and approach and is then followed by an explanation of the data and methods. The section concludes by describing the statistical and substantive results of the analyses.

5.1.1 Exploring Polarization in Primaries

As explored in previous chapters, party activists and primary electorates are often pointed to as a potential cause of political polarization. Standard accounts explain that as primary electorates become more extreme, politicians must placate them and subsequently become just as extreme (King, Orlando and Sparks, 2016). Whether this mechanism happens via adaptation or replacement (see, e.g., Theriault 2006, 2008) or at the stage of candidate emergence (Thomsen, 2014), the account explains that primary electorates have further to the left and right and subsequently representation in Congress has also moved further to the left and right, respectively.

Previous research indicates that politician ideological positioning is nearer primary electorate preferences than general elections. Brady, Han and Pope (2007) finds that
“congressional candidates who face a strategic dilemma between primary constituencies and general-election constituencies will work to build up a primary-election constituency” thereby moving these candidates further left or right (summarizing Fenno 1978 on p. 99). However, the authors admit greater attention to “primary constituencies” should be paid in future studies. Further research confirms this general finding in presidential elections (Jessee, 2010).\footnote{Jessee (2010) finds that in the 2008 presidential election, “Obama and McCain are estimated to have taken positions near both their party and primary constituency medians” rather than positions near the median general election voter.}

Responding to the charge cast in Brady, Han and Pope (2007) to study primary constituencies, this dissertation chapter provides a unique view into the question surrounding politician ideological positioning. While this project cannot go as far as Jessee (2009) which places candidates and voters on the same scale (for presidential candidates), it does often greater information surrounding the variation among primary voter extremism. By estimating the ideological positions of primary electorates, this project is able to explore whether more extreme primary electorates are associated with more extreme candidates or representatives.

5.1.2 Data and Methods

In order to examine the research question of this section, the data and methods will first be explained. First, a brief discussion of the three dependent variable measures used in this analysis will begin. Next, an explanation of two different measures of primary voter extremism will be developed. Finally, after a summary of control variables precedes an explanation of the models estimated.
**Dependent Variables: The Ideological Positions of Political Actors**

Measuring the policy preferences of political actors in an ideological space has been a significant agenda in political science research for decades. Different estimation strategies have been used to examine legislators (Poole and Rosenthal, 1985; Clinton, Jackman and Rivers, 2004; Shor and McCarty, 2011), candidates for political office (Bonica, 2014), cross-institutional actors (Bailey and Chang, 2001; Bailey, 2007), policy demanders and coalition merchants (Noel, 2013), constituents (Bafumi and Herron, 2010), and voters (Jessee, 2012) using roll-call votes, campaign donations, op-ed pages, and surveys.

While this present dissertation is unable to place candidates and voters on the same ideological scale (since survey responses to the scores of policy preferences are not available for congressional candidates), this chapter instead relies upon a number of existing ideological estimates for both legislators as well as candidates measured from governing and electoral behaviors.

The first dependent variable is the first dimension *NOMINATE* score for elected officials. This measure, estimated from roll call behavior, measures the “ideology” of members of Congress and allows these measures to be compared across years (Poole and Rosenthal, 1985, 2006; McCarty, Poole and Rosenthal, 2006). One benefit of using this measure is that it correlates very highly with other measures of ideology. Measured from observable political behavior, it allows researchers to place legislators in an ideological space to compare them to other legislators from different Congresses. One shortcoming of this measure is that it is only provided for members who serve in the legislature (that is, we do not have scores for losing candidates). Nevertheless, this measure will allow us to explore if primary voter extremism is associated
with legislative behaviors that are further left or right for Democrats or Republicans, respectively. (Data source: Lewis et al. 2017.)

The next dependent variable is the *cfScore* for candidates.\(^2\) This metric, estimated from donor behavior and campaign finance patterns, estimates a relative ideology for all candidates (i.e. not just elected politicians) (Bonica, 2014). Using this metric, we can explore if the ideological measure of winning primary candidates (i.e. candidates who win their primary but can win or lose in the general election) is associated with primary electorate extremism. (Data source: Bonica 2018.)

Lastly, echoing a strategy utilized in Chapter 3 above, this third metric seeks to capture more electoral considerations by estimating the percentage of the top-two vote share allocated to extremist candidates. To identify extremist candidates, this project relies upon a similar strategy of Hall (2015). Using the full population of congressional candidate *cfScores* (Bonica, 2018) from 1980 to 2016,\(^3\) a primary race is determined to be between an extremist and a moderate if the distance between the first and second place primary candidates’ *cfScores* was greater than the median distance between all top-two primary candidates.\(^4\) For Democrats, the left (negative) most candidate was marked the extremist while, for Republicans, the right (positive) most candidate was marked the extremist. From this strategy, the dependent variable is the top two primary vote percentage received by the extremist candidate. One shortcoming to using this approach is that not all primary races feature an extremist and a moderate. However, paired with the earlier two metrics, a rather full perspective of measuring ideology among political actors is considered in this chapter.

\(^2\)More specifically, the *recipient cfScore*.

\(^3\)With election results supplemented from Boatright and Moscardelli (2017).

\(^4\)For a review of this estimation procedure, see page 138.
INDEPENDENT VARIABLES: MEASURES OF PRIMARY VOTER EXTREMISM

Two separate measures of primary voter extremism were developed for this chapter. The first measure echoes that from earlier chapters and looks at the distance between primary voters and their nonvoting co-partisans in a certain district during a cycle. Borrowing from equation 2.9 above, this first extremism measure is measured with:

\[ Extremism_{c,d,p} = \theta_{c,d,p,voter} - \theta_{c,d,p,nonvoter} \]  

(5.1)

where \( \theta \) measures the estimated ideological ideal point for the constituency in cycle \( c \), district \( d \), part \( p \) across primary voters and nonvoters. For interpretability, when \( p = \) Democratic the measure is multiplied by (-1) so higher numbers above zero on this metric correspond to higher levels of extremism for both parties. This measure simply looks at the distance between voters and nonvoters.

The second measure of primary voter extremism looks instead at the distance between primary voters and that district’s median voter. Following previous literature that provides empirical measures of district median positions (e.g., Brady, Han and Pope 2007; Bafumi and Herron 2010; Tausanovitch and Warshaw 2013), this metric leverages the constituency-level estimates in relation to that constituency’s median voter. Similar to the previous metric, this estimated extremism is measured by:

\[ Extremism_{c,d} = \theta_{c,d,voter} - \theta_{c,d,medianvoter} \]  

(5.2)

following the above definitions for \( c, d, p \), and \( \theta \) and again multiplying Democratic constituency estimates by (-1). In order to estimate, \( \theta_{c,d,medianvoter} \), we slightly adjust the post-stratification formula presented by Equation 2.8 to the following:

\[ \theta_{c,d,medianvoter} = \frac{\sum_{p \in g} N_p \cdot y_p}{\sum_{p \in g} N_p} \]  

(5.3)
where \( \theta_{c,d,\text{medianvoter}} \) represents the estimated ideological position for the median voter in district \( d \) for cycle \( c \) where \( p \) represents all rows of the aggregated voter file belong to group \( g \) (district by cycle), \( y \) serves as the predicted ideological score for each row and \( N \) represents the number of voters represented by each row.\(^5\) The estimate from Equation 5.3 is plugged into Equation 5.2, which then reveals constituency-level estimates for primary voter relative distance from the district’s median voter.

Figure 5.1 provides a brief schematic of Virginia’s 8th congressional district during the 2014 midterm. The plot shows the estimated positions of different groups on the ideological line (arranged from liberal to conservative, left to right). The first extremism score measures the distance between Primary Voter (dark blue) and the Primary Nonvoter (light blue). The second extremism score measures the distance between Primary Voter (dark blue) and the Median Voter (gray). In this example, the values are as follows: Democratic primary voters in VA-8 were estimated at the ideological position of -0.93 in 2014, Democratic primary nonvoters were estimated

\(^5\)The median voter score estimates are highly correlated with district-level measures for Democratic presidential vote share. See Figure A7 in the Appendix.
at -0.77, and the median voter score was estimated near -0.26. Following the formulas above, the first extremism score is 0.16 and the second is 0.67.

**Figure 5.2** plots the features of the second measure of extremism. The top panel shows the district-level estimates from 2010 to 2016 across different groups. From left to right, these density plots represent the relative distributions of ideological scores for Democratic primary voters (dark blue), Democratic primary nonvoters (light blue), median voters (grey), Republican primary nonvoters (light red), and Republican primary voters (dark red). The panels beneath show the distribution of
the extremism measure (i.e. the distance between primary voters and median voter) for Democrats (left) and Republicans (right).

Although this pair of extremism estimates measure slightly different concepts, their substantive meanings are very similar. The larger the extremism score, the more out-of-step primary voters are from their districts. By exploiting the variation within these measures, we can better explore if extremism among primary electorates is associated with more ideological extremism among political actors.

**Additional Control Variables**

Additionally, following the structure of analyses from previous chapters, the following covariates are also considered in the models below:

- **district PVI** – a measure for the Cook Partisan Voting Index (see chapter 4) arranged across values from -33 (the most Republican-leaning district observation) to +44 (the most Democratic-leaning district observation) with 0 serving as evenly partisan

- **primary type** – a series of binary variables indicating the type of the primary (incumbent, challenger, or open seat) with incumbent serving as the reference category (for this measure’s significance see, e.g., Podkul N.d.)

- **district unemployment (%)** – a measure of the district’s annual unemployment (measured as a proportion from 0 to 1)

- **district nonwhite (%)** – a measure of the proportion of a district’s nonwhite population (from 0 to 1)

---

6Despite the differences across the metrics, the two measures are moderately correlated. For the district-level estimates, the extremism measures are correlated at 0.60 among Democratic constituencies and 0.39 for Republican constituencies.
• **effective number of primary candidates (ENC)** – the effective number of primary candidates in the primary election (see Boatright 2013, 2014)

• **primary turnout (%)** – the number of a party’s primary voters divided by the population older than 18 in a district (measured from 0 to 1)

• **candidate incumbent status** – a dichotomous measure indicating whether the legislator or candidate is a current incumbent during the primary (see, e.g., Jacobson 2009)

The data source for these additional covariates is the Primaries Timing Project (using their district-party-year observations from 2010 to 2016) (Boatright and Moscardelli, 2017).

***

In the sections that follow, a series of linear models are estimated regressing the dependent variables above across a number of model specifications for the two independent variables of interest while controlling for the covariates listed in the controls as well as fixed effects for state and election cycle. Additionally, models are estimated separately for Republicans and Democrats to account for any heterogeneous effects of the independent and control variables.

### 5.1.3 Results

This section demonstrates the results produced from the data and methods outlined above. First, descriptive data connecting primary voter ideology to candidate ideology is explored. Second, model estimates looking at the influence of relative primary voter extremism on legislator and candidate ideology is examined. Finally, model estimates...
analyzing the possible connections between primary voter extremism and extremist primary voting patterns are reported.

Figure 5.3 plots the estimated average primary voter ideological score for constituency-level for both the state- (left) and district-level (right) estimates described above against that constituency-level primary winner’s cfScore for both Republican (red x points) and Democratic (blue circular points) constituencies from 2010 to 2016. Each of the plots is presented with the within-party linear best fit (with 95% confidence interval) for this bivariate relationship. Across both panels of the plot, there is not a clear relationship between primary voter ideology and
the relative ideology of primary winners (i.e. eventual party nominees). None of the bivariate relationships uncover a statistically significant relationship (i.e. for each of the bivariate regressions \( p > 0.05 \)) and almost all of the estimate slopes are near zero. Although this is meaningful descriptive evidence of a lacking relationship between primary voter and candidate ideology, the following sections probe this relationship more closely by accounting for additional factors and leveraging relative primary voter extremism (within-district extremism) rather than just estimated primary voter ideology (across-district) using the 2010 to 2016 district-level estimates.\(^7\)

Table 5.1 and Table 5.2 present the model estimates for the regression analyses introduced above. Table 5.1 outlines eight separate linear models exploring the potential effects of primary voter extremism (measured via primary voter distance from nonvoting co-partisans and primary voter distance from median voters) on the ideology of legislators (measured via NOMINATE scores from legislator) and candidates (measured via \( cfScore \)) broken out by Democratic and Republican district-level constituencies. The first group of models regresses the primary-winning candidates’ first dimension NOMINATE score on the predictors, which by virtue of the NOMINATE measure only includes cases where the primary-winning candidate is an incumbent (since NOMINATE scores are only available for elected legislators who have cast roll votes in Congress). The second group of models regresses the primary-winning candidates’ estimated \( cfScore \) on the predictors (which is available for most candidates so long as they raise funds during the campaign). Due to these data constraints produced by these variables, there is a bit of variation among the within-party, across-model numbers of observations. Taken together, these models provide a robust view of the

\(^7\)State-level estimates are excluded from this analysis due to the significantly lower number of observations and non-random, biased inclusion of states into the sample due to the 6-year term of U.S. Senators. Using the district-level estimates, there is an observation per district for each year studied.
The influence of primary voter relative extremism on extremism among political actors, *ceteris paribus*.

The statistical estimates for the coefficients of the independent variables of interest (Extremism from Nonvoter and Extremism from Median voter) indicate a lack of statistical significance for each of the models, except one. In addition to the lack of statistical significance, the estimates on these coefficients also alternate direction, further suggesting the lack of a consistent finding of primary voter ideology on political actor ideology. The only model that finds a statistically significant finding is the influence of extremism from nonvoter on the estimated primary-winning candi-
date ideology, controlling for other factors. While there may not be an overwhelming amount of evidence of voter ideology effects on political candidates, this might suggest that as primary voters become more ideologically extreme compared to nonvoting co-partisans primary winners are – on average – more conservative. Especially considering the results from column 3, this may suggest that extreme Republican primary electorates nominate more extreme candidates but that these candidates do not make it to Congress (cf. Hall 2015).

Table 5.2 presents the results of a similar model where the dependent variable is an extremist candidate’s top-two vote share (%) in all moderate versus extremist
primary match-ups for both Democratic and Republican constituencies. Considering
the procedure for identifying moderate versus extremist primaries (see above), many
primary races are not including (which accounts for the severe drop in the number of
observations, especially among Democratic constituencies). The covariate predictors
estimated in this series of models are identical to those presented in Table 5.1.

The findings from Table 5.2 explain that extremism of primary electorates (using
both measures) are unrelated to increased extremist candidate vote shares for both
Democrats and Republicans controlling for other factors. Among Republican elec-
torates, greater levels of extremism are associated with lower extremist vote shares,
ceteris paribus; however, this association is not statistically significant according to
conventional levels. For Democratic electorates, the relative magnitude and direc-
tion of voter extremism’s influence of extremist candidate vote share differs across
measures for voter extremism but both are similarly not statistically significant.

From the findings presented in this section, it appears that there is a lack of
connection between primary voter relative extremism and the extremism of winning
primary candidates. In the next section of this chapter, we explore whether this voter
extremism is associated with greater numbers of ideological primary challenges.

5.2 Voter Extremism and Ideological Primary Challenges

In this second section of this present chapter on primary electorate extremism and
polarization among political actors, we explore the influence on how extremism may
or may not influence ideological primary challenges in House races. This section first
outlines the context of ideological primary challenges, it then identifies the data and
methods for seeking associations between relative primary voter extremism and ide-
ological challenges, and finally presents the results from these analyses.
5.2.1 When Primary Challenges Are Ideological

Primary challenges to incumbent members of Congress have been on the rise in the past decade (Kamarck and Podkul, 2017). Despite this increase in electoral competition, however, incumbents very rarely lose these challenges. From the five cycles between 2000 and 2008, fewer than twenty incumbents lost their primaries – many of whom lost due to congressional redistricting caused by the decennial reapportionment (Cook, 2010). Further, incumbents seeking renomination to be their party’s candidate in the general election generally win their primaries by over margins of victory greater than 50% (Kamarck and Podkul, 2018).

Even though members of Congress rarely face serious primary challenges, they report being greatly influenced by these challenges when it comes to decision-making in the legislature (Fenno, 1978), especially when it comes to potential ideological challenges (Kamarck and Wallner, 2018). Although incumbents do a good job deterring potential challengers (see, e.g., Maisel and Stone 2001), ideological challenges nevertheless occur (albeit not as often as pundits and scholars might expect) (Boatright, 2013). According to Boatright (2013), the number of ideological primaries is on the rise with fourteen incumbents facing ideological challenges in 2010 – a four-decade high.8

Understanding primary challenges – i.e. the electoral dynamic of a sitting incumbent facing a challenge in her primary – is fairly simple. But what is an ideological primary challenge? Defined by Boatright, these cases are those when an “Incumbent [is] criticized for being too moderate or insufficiently partisan.” For example, in the 2010 Republican primary, incumbent Alaska Senator Lisa Murkowski faced an ideological challenge from attorney Joe Miller. Miller – who enjoyed support from the

---

8See Figure 2.1 in Boatright (2013) for more detailed information.
Tea Part Express in the form of six figure ad buys – attacked his fellow Republican for being insufficiently conservative on the issues of taxes, abortion, the environment, and many other policies (McCormack, 2010). In the media, Miller boasted support from former Governor and Vice Presidential Republican candidate Sarah Palin while Murkowski had to fight off being labeled a RINO – Republican in Name Only (Rudin, 2010). Ultimately, Miller narrowly won the primary demonstrating that ideological challenges can be successful.

Even when ideological challenges are unsuccessful they can still influence legislator behavior. Evidence from roll call votes in Congress finds that ideological primary challenges are related to stricter party voting among minority party members (Jewitt and Treul, 2018). Additional scholarship indicates that primary challenges are also associated with members having a greater propensity to respond to their ideological “bases” (Caldeira and Zorn, 2004). Even further, the very threat of a primary challenge may have polarizing effects, echoing the charge from Sen. Flake above.

While previous research explores the relationship of how ideological primary challenges might be influenced by the ideology of members (Boatright, 2013; Pyeatt, 2015), this present section explores how these primary challenges may be related to the ideology of voters. Are extremist primary electorates associated with more ideological primary challenges? The following section outlines the data and methods used to explore this research question.

5.2.2 Data and Methods

To test the possible influence of primary voter relative ideological extremism on the likelihood of ideological primary challenges, this project conducts a number of quantitative tests. This section first introduces the data and next explains the modeling strategies used to estimate this relationship below.
Measuring Ideological Primary Challenges

Determining whether a primary challenge is ideological is not a straightforward task. Considering candidate-level ideological measures discussed above (for example, $cfScores$ developed by Bonica 2014), this measurement strategy may seem simple: if a primary challenger is more extreme than an incumbent than that may conceivably constitute an ideological challenge. However, although this relative approach might be helpful for identifying extremists and moderates (see, e.g., Table 3.5 and Table ??) it fails to identify whether the nature of the primary was necessarily ideological.

Although ideological challenges seem to capture a lot of public attention from media and political actors (Boatright, 2013), there are many additional circumstances that may cause a candidate to get into the race and challenge an incumbent. For example, New Jersey’s ninth congressional district featured Reps. Bill Pascrell (D) and Steve Rothman (D) facing off in a hotly contested primary, drawing endorsements from Former President Bill Clinton (for Pascrell) and political advisor David Axelrod (for Rothman). However, the race was not prompted by an ideological challenge but rather the decennial redistricting (Zernike, 2012). In another example, during the 2016 primary cycle in Pennsylvania’s 2nd congressional, incumbent Rep. Chaka Fattah was challenged by state legislator Dwight Evans. Although the Evans campaign was not averse to some ideological language, the main cause of the challenge was the Democratic incumbent’s indictment in a series of federal corruption charges (Brennan, 2015). While ideological challenges are common in congressional primaries, they are not the only reason why incumbents face intra-party competition.

To determine whether a primary challenge is ideological in nature, this dissertation relies upon the coding used in Getting Primaried (Boatright, 2013) and provided by the Primaries Timing Project (Boatright and Moscardelli, 2017). This dataset pro-
vides a variable “reason for challenge” (for all primary races from 1970 to 2012) coded from *The Almanac of American Politics* and *Politics in America*. This measure codes whether a primary challenge is due to the following: scandal, competence/age, local issues, national issues, centrist challenge, ideological challenge, race, machine/party faction, ambitious challenger, redistricting, other, or none.\(^9\) Using this measure, I seek to explore the predictors for ideological challenges.

Figure 5.4 displays the number of ideological and national issue-based U.S. House of Representatives primary challenges from 1970 to 2012. As the chart demonstrates, there has been a slight increase in the number of ideological primary challenges over the last decade. In the models below both ideological and national issues challenges

---

\(^9\)Further details on the methodological decisions behind this coding practice can be found on pg. 66 in *Boatright* (2013).
are considered “ideological challenges” (cf. Figure 2.1 and Table 2.4 from Boatright 2013).

**Primary Extremism, Additional Covariates, and Models**

Following previous chapters, this study explores primary extremism relative to non-voting co-partisans. For robustness, this study also looks at primary extremism by just measuring extremism as the mean primary voter ideological score.

To account for potentially confounding factors, the tests in this study also include the previously utilized controls of district partisan vote index (PVI), district unemployment (measured as %), and district nonwhite population (measured as %). Further, to consider candidate effects related to the primaried incumbent, the tests below also control for the incumbent’s cfScore (Bonica, 2018) and the incumbents’ seniority measured by the years served in Congress (logged).

The models estimated below follow the familiar form of:

\[
Pr(IdeoChall_i = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \text{Extremism}_i + \beta_2 \text{IncIdeo}_i + \beta_3 \text{IncSeniority}_i + \beta_4 \text{PVI}_i + \beta_5 \text{Unemploy}_i + \beta_6 \text{Nonwhite}_i)
\]

(5.4)

where for each constituency-cycle \(i\): Extremism is the primary voter extremism scores described above and IdeoChall is a dichotomous variable of whether a primary challenge was ideological in nature. These logistic regressions were estimated separately by party for all incumbent House primaries from 2010 to 2012.

**5.2.3 Handling Imbalanced Data**

As evidenced from Figure 5.4 above, there are not many ideological challenges (\(n < 20\) per cycle) relative to the number of total primary challenges each cycle. Among
Republican incumbent primaries, only 7% are confidently considered as “ideological challenges” and, among Democratic incumbent primaries, the amount is less than 2%. These relative percentages represent a severe imbalance in the data, which may lead to errors in attempting to estimate the factors related to studying the probability of an ideological challenge (see, King and Zeng 2001).

To attempt to correct for the biased estimates that might be produced this imbalanced data, this project leverages an algorithmic over-sampling technique. Using Synthetic Minority Over-Sampling Technique (SMOTE) to produce a “new” dataset that over-samples from the ideological challenges in the data (Chawla et al., 2002), this section produces a new dataset that better balances the dependent variable. Overall, to create SMOTE’d data, new “instances” of the minority class (in the dependent variable) are created (synthesized) using K-nearest neighbors (using the independent variables). The application of this algorithm to political science is not particularly common but there is precedence among work seeking to join machine learning approaches to political research questions (e.g. Muchlinski et al. 2016; Kaplow and Gartzke 2016).

In practice, SMOTE was applied to the data separately by party.\textsuperscript{10} In the Democratic data, ideological challenges were slightly more oversampled than in the Republican case due to fewer cases in the original data.\textsuperscript{11} In the results below, models are estimated both with the original data and the SMOTE’d data. Although the SMOTE’d data significantly improves each of the models a great deal (demonstrated below), both datasets produce similar substantive results.

\textsuperscript{10}The SMOTE algorithm was performed using the DMwR package in R.

\textsuperscript{11}Specifically, the Democratic data was undersampled from an additional 200 cases that were generated and oversampled and oversampled from 1,500 cases. The Republican data as undersampled from 200 cases and oversampled from 400 cases. Both data used the number of nearest neighbors at k = 5.
The findings from both the balanced and imbalanced data are presented below.

5.2.4 Results

The results from the models introduced above are estimated in Table 5.3. The first two columns show models estimated from a logistic regression on the dependent variable of an ideological challenge from the Democratic incumbent primaries from 2010 to 2012 and the second two columns show the estimates from Republican incumbent primaries. In the first and third columns the main independent variable of interest is the extremism score measured as the ideological distance between primary voters and primary nonvoting copartisans for district-cycle units while the second and fourth columns estimate the raw primary voter score for district-cycle units. As described in the previous section, each of the columns represents data that was produced using synthetic minority over-sampling SMOTE. (Appendix Table A9 displays the results using the original raw data and finds no substantive differences from the Table 5.3.)

The estimates from Table 5.3 reveal unexpected cross-party differences of the influence of primary voter extremism on the propensity of their being an ideological challenge, *ceteris paribus*. Among Republican electorates, there does not appear to be a statistically meaningful relationship of electorate ideology on the likelihood of an ideological challenge using either of the ideology measures when controlling for other factors.

Among the Democratic electorates, however, there is a consistent statistical finding that relates more moderate behavior with higher likelihoods of an ideological challenge. In the first model (column 1), we find that as primary voters move further to

\[ \text{Further, Figure A8 and Figure A9 plot Receiver Operating Characteristic (ROC) curves and demonstrate an improvement in classification in the models using the SMOTE’d data, especially among the Democratic models. For example, the area under the curve for the first column in the original data is 0.89 and improves to 0.958 using the SMOTE’d data.} \]
Table 5.3: Ideological Primary Challenge Logistic Model by Party, 2010–2012

<table>
<thead>
<tr>
<th>Ideological Primary Challenge</th>
<th>Democratic</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremism from Nonvoter</td>
<td>−18.08*** (2.66)</td>
<td>−4.21 (2.23)</td>
</tr>
<tr>
<td>Primary Voter Score</td>
<td>8.21*** (1.44)</td>
<td>1.27 (0.93)</td>
</tr>
<tr>
<td>Incumbent cfScore</td>
<td>2.60* (1.15)</td>
<td>−1.15 (0.61)</td>
</tr>
<tr>
<td></td>
<td>1.63* (0.78)</td>
<td>−0.07 (0.62)</td>
</tr>
<tr>
<td>Incumbent Seniority (years logged)</td>
<td>1.03* (0.51)</td>
<td>0.15 (0.36)</td>
</tr>
<tr>
<td></td>
<td>−0.07 (0.36)</td>
<td>−0.25 (0.17)</td>
</tr>
<tr>
<td>District PVI</td>
<td>−0.36*** (0.06)</td>
<td>−0.21*** (0.03)</td>
</tr>
<tr>
<td></td>
<td>0.02 (0.02)</td>
<td>−0.00 (0.02)</td>
</tr>
<tr>
<td>District Unemployment (%)</td>
<td>−19.65 (12.50)</td>
<td>−10.77 (10.92)</td>
</tr>
<tr>
<td></td>
<td>11.05* (4.92)</td>
<td>11.92* (5.09)</td>
</tr>
<tr>
<td>District Nonwhite (%)</td>
<td>11.10*** (3.13)</td>
<td>7.50*** (2.22)</td>
</tr>
<tr>
<td></td>
<td>0.65 (1.53)</td>
<td>−0.77 (1.43)</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>−0.05 (1.86)</td>
<td>5.32** (1.77)</td>
</tr>
<tr>
<td></td>
<td>0.44 (0.98)</td>
<td>−2.19* (1.08)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>−65.84</td>
<td>−88.67</td>
</tr>
<tr>
<td>Smote’d Data</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>274</td>
<td>271</td>
</tr>
</tbody>
</table>

Logistic model. ***p < 0.001, **p < 0.01, *p < 0.05.

the left (i.e. extreme; a one-unit change in the positive direction represents greater liberalism among primary voters) compared to their co-partisan primary nonvoters that the likelihood of there being an ideological challenge in that district-cycle falls. In the second model (column 2), we find that as primary voters move further center (i.e. moderate; a one-unit change in the positive direction represents greater conservatism among primary voters) the likelihood of a primary challenge increases. In other words, both models find that as primary voters move further to the left, the likelihood of an ideological challenge decreases, controlling for the other listed factors.
This finding appears consistent across whether we leverage the variation in relative extremism within- or across-units.

To aid in the interpretability of the logistic regression models from Table 5.3, the predicted probabilities from columns 1 (Democratic constituencies, dashed blue line) and 3 (Republican constituencies, solid red line) are plotted in Figure 5.5 with 95% confidence intervals (shadings) across observed relative primary voter extremism scores with control variables held to their party medians. For both Democratic and Republican constituencies, the likelihood of an ideological primary challenge decreases as primary voters tend to be more extreme. For Republicans, the difference between no extremism (at 0.0) and the upper quartile of primary voter extremism is not statistically significant, echoing the substantive findings presented in the estimates from Table 5.3. For Democrats, the predicted probability chart demonstrates that the statistically significant finding is driven by cases where primary voters are more
moderate and the difference in the predicted probability between a very moderate primary electorate (-0.1) and one that is representative (0) is also not statistically significant. When presented as a predicted probability, this statistically significant finding demonstrates a minor substantively meaningful result.

Before concluding, there are a number of control variables that remain interesting in their relationship to ideological primary challenges. Among Democratic electorates, district nonwhite (%), district PVI, and incumbent cfScore are consistently statistically significant. While districts increase in their nonwhite population, the likelihood of an ideological primary challenge increases. Regarding the ideological and partisan measures, ideological primary challenges are associated with more conservative incumbents (measured via cfScore) and more Republican districts (via PVI). Among Republican primary electorates, the only consistently statistically significant control coefficient is district unemployment (%), where we see in areas of greater unemployment there is a higher likelihood of their being a primary challenge. Further research might explore this relationship and how it might be relate to conservative or Tea Party related primary challenges.

5.3 Conclusion and Discussion

This chapter seeks to demonstrate that despite broad claims suggesting that polarized primary electorates may be responsible for polarization among elected representatives that there is little empirical evidence that primary voters are moving Democratic politicians further left and Republican politicians further right. Leveraging the heterogeneity of primary electorate extremism (using the ideological scores developed in Chapter 2), this present chapter finds that ideological electorates are neither associated with more extreme representatives in Congress nor more extreme candidates on
general election ballots. Further, this chapter fails to find substantive evidence of voter extremism influencing the likelihood of an incumbent facing an ideological challenge. Primary electorates differ greatly in their relative ideological positioning across districts and cycles but these differences fail to explain polarization in elections or representation (cf. Brady, Han and Pope 2007; Hirano et al. 2010).

While this chapter demonstrates a number of tests to bolster its null claims, there are a few areas of suggested future research. First, as stated earlier, this research agenda would benefit from estimation strategies that place primary voters and candidates on the same ideological scale. Considering the sparseness of information available on primary elections (especially those who raise little amounts of money), this task has been set aside for future projects.

Second, regarding the association of extremist electorates with ideological primary challenges, the tests in this chapter would benefit from additional data. In this section, the findings draw upon challenges in 2010 and 2012. While this approach expands some prior research, further data collection – before 2010 with respect to ideological scores of voters and after 2010 with respect to the nature of primary challenges – may help to identify additional trends. Additionally, interviews or surveys of primary challengers may help to suss out whether perceptions of primary electorate ideology contribute to the calculus behind candidate emergence.

Lastly, further considerations in future work may focus on the secular trends of polarization and their relationship to ideological primary challenges. As Boatright’s data demonstrates, ideological challenges are increasingly more common now than they were forty years ago. Data beyond just two primary election cycles may be able to leverage differences-in-differences, viz. whether the leftward or rightward movement of primary electorates across time is associated with further polarization by elected officials or candidates or by increasing primary challenges.
Nevertheless, as I wrote elsewhere with my co-author Elaine Kamarck during the 2016 election cycle, it may be that the possibility of primary *threats* are responsible for polarization regardless of whether extremist challengers win or even run:

“Even if these candidates lose congressional primaries, they can send a message to members of Congress to always be looking over his or her shoulder for the next challenge from the left or right – especially if the challenger does better than expected.”

Regardless, the evidence does not find that primary voters’ relative ideological representativeness has much to do with influencing this dynamic.

---

Chapter 6

Conclusions and Discussion

This brief chapter provides a conclusion along with related discussion for this dissertation project. This chapter is divided into three sections: first, the research presented above will be summarized; second, areas for future research that both directly and indirectly build on this project will be suggested; and, finally, the project will conclude.

6.1 Dissertation Conclusions

This section will provide a brief conclusion of the results and findings from the four empirical chapters above (Chapters 2 through 5).

Chapter 2 introduces the novel measure of constituency-level ideological estimates for primary voters and copartisan nonvoters within states and congressional districts resulting in relative ideological measures for constituencies across both major political parties from 2010 to 2016. Using the Cooperative Congressional Election Study to uncover ideological scores from survey respondent answers to a host of a policy preference questions ranging from social issues like gay marriage and abortion to economic policies like budget priorities, survey ideal points for over 300,000 respondents were subsequently re-weighted using multiple regression and poststratification using novel voter file data to understand the ideological positions of primary voters and nonvoters within districts.
The descriptive results from this chapter demonstrate the significance of constituency-level results. When looking at state-level responses, primary voters are generally ideologically to the left and right of nonvoting Democrats and Republicans, respectively. However, the extent of this extremism varies a great deal. In some constituencies – like for Massachusetts Republicans (2010) – primary voters are actually more moderate than nonvoters while in other constituencies – like Vermont Democrats (2010) – primary voters are significantly more extreme. The district-level responses reveal a similar pattern where most primary electorates are extreme but this finding does not hold for all constituencies across these four cycles. The relative ideological representativeness of primary voters is heterogeneous across cycles and constituencies. These estimates featured prominently in the following chapters by leveraging this variation in relative extremism.

Chapter 3 began to explore whether more closed primary systems are associated with more polarized voters in primary elections. While some states observe open primary participation rules that allow any registered voter to participate in a primary election, others are more “closed” and restrict participation based upon your party registration. In this chapter, a pair of analyses were conducted to explore relative ideological representativeness of primary voters as well as their voting behavior. In the first analysis, we reported a lack of a clear substantive effect between primary openness and primary voter extremism. Although there were moderate statistically significant findings, these results – especially those examining trends across election cycles – failed to find any clear patterns of openness being related to polarization among the electorate.

The second analysis in Chapter 3 introduced the congressional primaries exit polls, a series of district-level polls conducted in 2016 and 2018. Using these surveys, this analysis examined whether primary rules were associated with more extremist voting
patterns. In this section, the chapter isolated the behavior of primary voters (that is, respondents to the exit poll surveys) and explored any possible effects of rules on vote choice. After identifying the extremist and moderate candidates in the sampled data, results found little effects between primary system and their reported ideological extremism at the ballot when controlling for other respondent and electoral-level factors. Despite the claims from many commentators, reform groups, and politicians, there does not appear to be a clear connection between primary participation rules and extremism among primary voters.

Chapter 4 investigated how district partisanship exacerbates primary voter ideological unrepresentativeness. District partisanship, explored using Cook PVI and how the district voted in the most immediately previous presidential election, is often used as a heuristic for distinguishing “safe” seats from “toss-up” seats by elections prognosticators. From this perspective, primary voters tend to have a disproportionate significance in safe seats for their party, considering their co-partisan representatives has a greater likelihood of winning the general election. Following a similar format of Chapter 3, this chapter explored district partisanship as it relates to primary voters’ relative ideology and their voting behavior. In the first analysis, the relationship of district partisanship and primary electorate extremism was explored, \textit{ceteris paribus}. Results found that while for Democrats, there was not overwhelming evidence of a relationship between partisanship and electorate extremism there \textit{was} evidence of extremism among Republican electorates in more Republican congressional districts. Among Republicans, voters in safe Republican districts tend to be more extreme than those in more moderate and Democratic districts.

Turning to vote choice, however, the findings produced a different conclusion. For both Democratic and Republican primary voters, they were much less likely to vote for the extremist primary candidate in their safe seats. Even controlling for primary
type (incumbent, challenger, and open seat) and other respondent-level factors, the chapter discovered that primary voters who did not need to worry about strategic concerns (that is, primary voters in districts leaning in the direction of the opposite party) are more likely to vote for more extreme candidates.

In the final empirical chapter, this dissertation examined primary voter extremism and the ideology of political actors who are involved in these elections. Although this project was unable to place voters and candidates on the same ideological scale, it sought to examine associations between the survey-based constituent ideological measures developed above and other ideological measures of political actors. The first section of this chapter isolated primary voters’ influence on the ideological positions of candidates and elected officials and the second section looked at primary campaign dynamics. In this first section, it was found that regardless of how relative ideology of primary voters is measured (that is, relative to co-partisan nonvoters or the district’s median voter), there is no clear relationship between primary voter extremism and the candidates it produces nor the legislators it ultimately elects.

In the second section, this chapter explored whether more extremist primary electorates were associated with more ideological primary challenges to incumbent members of Congress. The substantive results from this analysis neglects to find a connection between extremism and legislators “getting primaried” on ideological grounds.

The next sections place these findings in the broader contexts of conversations of political polarization.

6.2 SUGGESTIONS FOR FUTURE RESEARCH

Although this present dissertation explored many of the research questions related to how primary voters might be related to political polarization, there are a number of
areas suggested for future research. First, the most obvious area for future research is to improve upon the data collection of this project to better explore temporal patterns of primary voter ideological representation as they relate to secular polarization (cf. McGhee et al. 2014). In studying the “national primary voter,” this project is only able to explore from 2008 to 2016 using present data sources; in studying primary voters at the constituency-level, it only studies voters from 2010 to 2016 due to further constraints produced by the scope of the voter file data provided to the author; and, finally, the congressional primaries exit polls were only fielded beginning in 2016. Each piece of this project would greatly benefit from further data collection. Regarding “national primary voter” and constituency-level primary voter ideology, historical data utilizing additional sources of respondent policy preferences (such as the American National Elections Studies) and additional administrative data on voter turnout might help the project. Nevertheless, going further back in history would require additional assumptions that this dissertation did not have to make (e.g. regarding the connection of administrative data with policy preference surveys, etc.).

Second, campaign effects related to primary voter ideological representativeness are another avenue for future research. Do primary campaigns actively influence the turnout behavior in congressional primaries? For example, voter filer firms and political campaigns often “score” voters’ perceived ideology (using a variety of related inputs ranging from demographic to consumer data) (cf. Hersh 2015). It is entirely possible that more ideological candidates try to turn out more ideological voters; however, this is likely mediated by political interest, knowledge, and the relative salience of ideology in a primary. Untangling the endogeneity of campaign influences in ideological representativeness of primary electorates would be tricky but would provide a better understanding of how these sources might be influencing polarization on a smaller scale.
Third, although this present project pools all primary voters regardless of when they occur within a given cycle, future research might explore any temporal effects by election calendaring. Although primary calendars (i.e. the chronological ordering of nominating contests) have been studied with respect to presidential nominations, future research might explore whether there are any effects on the primary voter pool based on the congressional primary calendar. For example, do surprise outcomes in a state at time $t$ affect the ideological composition of a primary in a different state at time $t + 1$? Looking at the 2018 congressional primary season, only four incumbent members lost their party nominations. Are the electorates voting in contests that follow these anomalous outcomes affected by any ideological spillover? For example, did Alexandria Ocasio-Cortez’s victory inspire progressive voters in later contests to turn out? Did Mark Sanford’s surprise defeat remind moderate Republicans to support moderate candidates in their own districts’ primaries?

Finally, a further exploration of campaign finance spending and fundraising’s ideological effects in political primaries might also be explored in previous research. Although the current campaign finance landscape is much more splintered than in years past, understanding how candidate or independent expenditure committees might solicit from ideological donors (cf. Bonica 2014) may have effects at the ballot box.

6.3 Final Thought

Conventional wisdom claims about congressional primary elections overwhelmingly suggest the concerns of contemporary partisan polarization can be ameliorated by restructuring and reforming how parties nominate their candidates. This dissertation project finds that while these voters may be ideologically extreme in some cases and
under certain circumstances that there may be little room for improvement by those who wish to solve ideological issues by redesigning institutions. Instead, allowing parties to shape their rules and allowing individuals to decide whether to vote or not has produced a natural ideological diversity among voters absent clear polarizing effects.
APPENDIX

APPENDIX TO CHAPTER 2

Table A1: CCES Survey Items Used as Bridge Observations

<table>
<thead>
<tr>
<th>Topic</th>
<th>Survey Item</th>
<th>CCES File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abortion</td>
<td>Always allow a woman to obtain an abortion as a matter of choice</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Permit abortion only in case of rape, incest, or when the woman’s life is in danger</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Make abortions illegal in all circumstances</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Allow employers to decline coverage</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Prohibit after 20 weeks</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Prohibit expenditure of public funds</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affirmative Action</td>
<td>Support for affirmative action programs in employment and college admissions</td>
<td>X</td>
</tr>
<tr>
<td>Environment</td>
<td>Give EPA power to regulate Carbon Dioxide emissions</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Raise required fuel efficiency for the average automobile from 25 mpg to 35 mpg</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Require a minimum amount of renewable fuels (wind, solar, and hydroelectric) in the generation of electricity even if electricity prices increase somewhat</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Strengthen enforcement of the Clean Air Act and Clean Water Act even if it costs US jobs</td>
<td>X</td>
</tr>
<tr>
<td>Gay Marriage</td>
<td>Support for legal marriage of gays and lesbians</td>
<td>X</td>
</tr>
<tr>
<td>Gun Control</td>
<td>Background checks for all sales, including at gun shows and over the Internet</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Prohibit state and local governments from publishing the names and addresses of all gun owners</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Ban assault rifles</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Make it easier for people to obtain concealed-carry permit</td>
<td>X</td>
</tr>
<tr>
<td>Iraq Mistake</td>
<td>Invading Iraq in 2003 was a “mistake”</td>
<td>X</td>
</tr>
<tr>
<td>Immigration</td>
<td>Grant Legal Status to people who have held jobs and paid taxes for at least 3 years, and not been convicted of any felony crimes</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Increase the number of border patrols on the U.S.-Mexican border</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Allow police to question anyone they think may be in the country illegally</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Fine U.S. businesses that hire illegal immigrants</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Deport illegals</td>
<td>X</td>
</tr>
<tr>
<td>Military Use</td>
<td>Protect American allies under attack by foreign nations</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Assist the spread of democracy</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Intervene in a region where there is genocide or a civil war</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Help the United Nations uphold international law</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Ensure the supply of oil</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Destroy a terrorist camp</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>X</td>
</tr>
<tr>
<td>Health</td>
<td>Repeal Affordable Care Act</td>
<td>X</td>
</tr>
</tbody>
</table>
Table A2: Mean Ideological Score for Democratic Voters (Ideal Point and Symbolic), 2008-2016

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Voters (Ideal)</td>
<td>-0.65</td>
<td>-0.69</td>
<td>-0.71</td>
<td>-0.92</td>
<td>-1.01</td>
</tr>
<tr>
<td>General Voters (Ideal)</td>
<td>-0.61</td>
<td>-0.66</td>
<td>-0.65</td>
<td>-0.83</td>
<td>-0.84</td>
</tr>
<tr>
<td>Primary Voters (Symbolic)</td>
<td>2.33</td>
<td>2.41</td>
<td>2.42</td>
<td>2.35</td>
<td>2.38</td>
</tr>
<tr>
<td>General Voters (Symbolic)</td>
<td>2.36</td>
<td>2.42</td>
<td>2.48</td>
<td>2.42</td>
<td>2.47</td>
</tr>
</tbody>
</table>

Table A3: Mean Ideological Score for Republican Voters (Ideal Point and Symbolic), 2008-2016

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Voters (Ideal)</td>
<td>1.00</td>
<td>1.09</td>
<td>0.85</td>
<td>1.06</td>
<td>0.96</td>
</tr>
<tr>
<td>General Voters (Ideal)</td>
<td>0.92</td>
<td>1.00</td>
<td>0.74</td>
<td>0.97</td>
<td>0.81</td>
</tr>
<tr>
<td>Primary Voters (Symbolic)</td>
<td>4.18</td>
<td>4.12</td>
<td>4.08</td>
<td>4.07</td>
<td>4.03</td>
</tr>
<tr>
<td>General Voters (Symbolic)</td>
<td>4.14</td>
<td>4.07</td>
<td>4.00</td>
<td>4.02</td>
<td>3.97</td>
</tr>
</tbody>
</table>

Note: Tables present mean estimated ideological scores for Democrat (top table) and Republican (bottom table) primary election and general election voters for each election cycle by ideal point and symbolic ideology (cf. Table 3.1 from Boatright 2014 and Table 2 from Sides et al. 2018).
Figure A1: Unweighted and Weighted MRP Estimates, 2010-2016.
Table A4: Comparing Party Registration Details, State-Level Estimates

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th></th>
<th>Republican</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Non-Voter Ideal Point Difference</td>
<td>−0.61</td>
<td>−0.60</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>−0.006</td>
<td></td>
<td>[−0.02, 0.03]</td>
<td>[−0.04, 0.04]</td>
</tr>
</tbody>
</table>

Note: Tables show the results of a two sample-test (difference of means) by comparing non-voter estimated ideal points for Democrat and Republican constituencies in the state-level (top) and district-level (bottom) estimates. Primary voters are not compared because they are properly identified by the voter file; primary non-voting co-partisans, however, are harder to define, especially in states without party registration status. Using the Data Trust modeled partisanship, however, we find that – on average – primary nonvoters (broken out by party) are not systematically different across party registration rules.

Table A5: Comparing Party Registration Details, District-Level Estimates

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th></th>
<th>Republican</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Non-Voter Ideal Point Difference</td>
<td>−0.60</td>
<td>−0.62</td>
<td>0.64</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>−0.025***</td>
<td></td>
<td>[−0.04, −0.01]</td>
<td>[−0.04, −0.01]</td>
</tr>
</tbody>
</table>

Note: Tables show the results of a two sample-test (difference of means) by comparing non-voter estimated ideal points for Democrat and Republican constituencies in the state-level (top) and district-level (bottom) estimates. Primary voters are not compared because they are properly identified by the voter file; primary non-voting co-partisans, however, are harder to define, especially in states without party registration status. Using the Data Trust modeled partisanship, however, we find that – on average – primary nonvoters (broken out by party) are not systematically different across party registration rules.
Figure A2: Sample Excerpt from Exit Polls Questionnaire (2018, Iowa-3, Democratic Primary)
Table A6: Exit Poll Summary Information in Accordance with AAPOR Code of Professional Ethics & Practices Section III-A.

<table>
<thead>
<tr>
<th></th>
<th>2016 Exit Polls</th>
<th>2018 Exit Polls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poll Sponsor:</td>
<td>The Primaries Project at Brookings</td>
<td>The Primaries Project at Brookings</td>
</tr>
<tr>
<td>Conducted by:</td>
<td>Edison Research</td>
<td>Edison Research</td>
</tr>
<tr>
<td>Funded by:</td>
<td>The William and Flora Hewlett Foundation</td>
<td>Democracy Fund</td>
</tr>
<tr>
<td>Sampling frame:</td>
<td>Primary voters in competitive Republican and Democratic primaries (not including voters in states with exclusive by-mail voting or top-two primary systems)</td>
<td>Primary voters in competitive Republican, Democratic, and top-two primaries (not including voters in states with exclusive by-mail voting)</td>
</tr>
<tr>
<td>Sample design:</td>
<td>225 competitive districts entered the sampling frame. Districts were randomly selected proportional to the size of the most recent contested primary vote in each district. Sample locations within districts were sampled randomly, stratified by past presidential vote and proportional to the size of the most recent presidential election vote.</td>
<td>282 competitive districts entered the sampling frame. Districts were randomly selected proportional to the size of the most recent contested primary vote in each district. Sample locations within districts were sampled randomly, stratified by past presidential vote and proportional to the size of the most recent presidential election vote. Two of the districts were randomly sampled from the top-two primaries in California</td>
</tr>
<tr>
<td>Sample size:</td>
<td>9,201 (20 districts)</td>
<td>13,372 (20 districts)</td>
</tr>
<tr>
<td>Margin of Error:</td>
<td>±2% for both Democratic and Republican primary voters</td>
<td>±2% for Republican primary voters; ±1% for Democratic primary voters; and ±4% for Top-Two primary voters</td>
</tr>
<tr>
<td>Weighting:</td>
<td>Weighted by non-response by age, race, and gender; then by congressional vote choice final results; finally by the total number of votes each congressional district primary represented as a share of all competitive congressional districts</td>
<td>Weighted by non-response by age, race, and gender; then by congressional vote choice final results; finally by the total number of votes each congressional district primary represented as a share of all competitive congressional districts</td>
</tr>
<tr>
<td>Method:</td>
<td>In-person interviews outside polling stations.</td>
<td>In-person interviews outside polling stations.</td>
</tr>
<tr>
<td>Dates of data collection:</td>
<td>March 1, 2016 to August 30, 2016</td>
<td>March 6, 2018 to August 28, 2018</td>
</tr>
<tr>
<td>Response rate:</td>
<td>44%</td>
<td>33%</td>
</tr>
<tr>
<td>Item Wordings:</td>
<td>See Appendix to Kamarck, Podkul and Zeppos (2017)</td>
<td>See Appendix to Kamarck and Podkul (2018b)</td>
</tr>
</tbody>
</table>
Figure A3: Map of Exit Polls Sampling Frame versus Sample, 2016

Figure A4: Map of Exit Polls Sampling Frame versus Sample, 2018
Figure A5: Candidate Primary Vote Share by Unweighted Exit Polls Estimated Vote Share
Procedure for Determining Extremist and Moderate Candidates. Following Hall (2015), this project sought to identify congressional primary races when the top two candidates featured an extremist facing off against a moderate candidate. The following steps determined when primary races met these conditions:

1. A dataset was created merging estimated ideological scores (Bonica, 2018) and primary election results (Boatright and Moscardelli, 2017)\(^1\) for all major party candidates to the U.S. House of Representatives from 1980 to 2018.

2. Using the election results, all top two finishing candidates were identified for each primary race.

3. For all primary races that had at least two candidates (and estimated ideological scores for those two candidates), ideological distances were calculated by looking at the Euclidean distance between both candidates' ideological scores.

4. If the distance between the two candidates was above or equal to the median distance (0.34), then the race was identified as featuring an extremist and a moderate.

5. Candidates were assigned the extremist label if their ideological score was further negative or positive than their competitor for Democrats and Republicans, respectively; candidates were assigned the moderate labels if their ideological score was further positive or negative than their competitor for Republicans and Democrats, respectively.

Table A7 lists the identified moderate and extremist candidates from districts sampled by the Congressional Primaries Exit Polls (2016-2018). Figure A6 plots how these extremist candidates performed at the polls.

---

\(^1\) The Primaries Timing Project data was supplemented with personal data collection of election night results for candidates running in the 2018 congressional primaries.
<table>
<thead>
<tr>
<th>District</th>
<th>Party</th>
<th>Extreme Candidate</th>
<th>Moderate Candidate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016 FL-6 Dem.</td>
<td>McCullough, William (-0.3)</td>
<td>Taylor, Dwayne L. (0.3)</td>
<td></td>
</tr>
<tr>
<td>2016 AL-4 Rep.</td>
<td>Norris, Phillip Dwight (1.7)</td>
<td>Aderholt, Robert Brown (0.9)</td>
<td></td>
</tr>
<tr>
<td>2016 FL-6</td>
<td>DeSantis, Ron (1.2)</td>
<td>Costello, Fred (0.8)</td>
<td></td>
</tr>
<tr>
<td>2016 GA-9</td>
<td>Broun, Paul Collins, Jr. (1.5)</td>
<td>Collins, Doug (1.0)</td>
<td></td>
</tr>
<tr>
<td>2016 MD-1</td>
<td>Harris, Andy (0.9)</td>
<td>Smigiel, Mike (0.6)</td>
<td></td>
</tr>
<tr>
<td>2016 MN-6</td>
<td>Kern, Aliena (2.1)</td>
<td>Emmer, Tom (1.2)</td>
<td></td>
</tr>
<tr>
<td>2016 OH-14</td>
<td>Lynch, Matt (1.6)</td>
<td>Joyce, David (0.8)</td>
<td></td>
</tr>
<tr>
<td>2016 OK-2</td>
<td>Jackson, Jarrin (1.5)</td>
<td>Mullin, Markwayne (1.1)</td>
<td></td>
</tr>
<tr>
<td>2016 TX-4</td>
<td>Gliotti, Louis (1.6)</td>
<td>Ratcliffe, John (1.2)</td>
<td></td>
</tr>
<tr>
<td>2018 FL-10 Dem.</td>
<td>Demings, Val (-1.1)</td>
<td>Darius, Wade (-0.1)</td>
<td></td>
</tr>
<tr>
<td>2018 IL-3</td>
<td>Newman, Marie (-1.5)</td>
<td>Lipinski, Daniel (-0.2)</td>
<td></td>
</tr>
<tr>
<td>2018 MI-6</td>
<td>Longjohn, Matt (-1.6)</td>
<td>Franklin, George (-0.9)</td>
<td></td>
</tr>
<tr>
<td>2018 MN-1</td>
<td>Feehan, Dan (-1.5)</td>
<td>Minehart, Colin (-0.2)</td>
<td></td>
</tr>
<tr>
<td>2018 MO-1</td>
<td>Bush, Cori (-2.2)</td>
<td>Clay, Lacy (-0.6)</td>
<td></td>
</tr>
<tr>
<td>2018 NC-9</td>
<td>Cano, Christian (-1.6)</td>
<td>McCready, Dan (-1.1)</td>
<td></td>
</tr>
<tr>
<td>2018 PA-3</td>
<td>Johnson, Kevin (-1.3)</td>
<td>Evans, Dwight (-0.6)</td>
<td></td>
</tr>
<tr>
<td>2018 TX-26</td>
<td>Fagan, Linsey (-2.2)</td>
<td>Fisher, Will (-1.9)</td>
<td></td>
</tr>
<tr>
<td>2018 FL-27 Rep.</td>
<td>Salazar, Maria (0.8)</td>
<td>Barreiro, Bruno (0.3)</td>
<td></td>
</tr>
<tr>
<td>2018 ID-1</td>
<td>Fulcher, Russ (1.4)</td>
<td>Leroy, David (1.0)</td>
<td></td>
</tr>
<tr>
<td>2018 NJ-5</td>
<td>Lonegan, Steven (1.3)</td>
<td>McCann, John (0.6)</td>
<td></td>
</tr>
<tr>
<td>2018 OH-16</td>
<td>Hagan, Christina (1.4)</td>
<td>Gonzalez, Anthony (0.9)</td>
<td></td>
</tr>
<tr>
<td>2018 OK-4</td>
<td>Taylor, James (1.6)</td>
<td>Cole, Tom (0.9)</td>
<td></td>
</tr>
<tr>
<td>2018 TX-2</td>
<td>Crenshaw, Dan (1.3)</td>
<td>Roberts, Kevin (0.2)</td>
<td></td>
</tr>
<tr>
<td>2018 TX-26</td>
<td>Birkenstock, Veronica (1.5)</td>
<td>Burgess, Michael (1.1)</td>
<td></td>
</tr>
<tr>
<td>2018 VA-10</td>
<td>Hill, Shak (1.5)</td>
<td>Comstock, Barbara (1.1)</td>
<td></td>
</tr>
</tbody>
</table>
Figure A6: Extreme Candidate Top Two Vote Shares When Facing Moderate Candidate
## Appendix to Chapter 4

Table A8: Primary Context Model: District Ideology Estimates

<table>
<thead>
<tr>
<th></th>
<th>Primary Electorate Extremism</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Democrat</td>
<td>Republican</td>
<td></td>
</tr>
<tr>
<td>Cook PVI</td>
<td>0.00**</td>
<td>-0.00***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Cook PVI$^2$</td>
<td>-0.00***</td>
<td>-0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Cook PVI$^3$</td>
<td>0.00</td>
<td>-0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Dem. Pres Vote Share</td>
<td>0.85*$^*$</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.31)</td>
<td></td>
</tr>
<tr>
<td>Dem. Pres Vote Share$^2$</td>
<td>-1.03</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.61)</td>
<td>(0.61)</td>
<td></td>
</tr>
<tr>
<td>Dem. Pres Vote Share$^3$</td>
<td>0.42</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
<td>(0.38)</td>
<td></td>
</tr>
<tr>
<td>Type: Challenger</td>
<td>0.00</td>
<td>-0.02**</td>
<td>-0.02**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Type: Open Seat</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>District Unemployment (%)</td>
<td>0.18</td>
<td>-0.37***</td>
<td>-0.40***</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>District Nonwhite (%)</td>
<td>0.06*$^*$</td>
<td>0.05*$^*$</td>
<td>0.15***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Effective Num. of Candidates</td>
<td>-0.01</td>
<td>-0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Primary Turnout (%)</td>
<td>0.08</td>
<td>0.08</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.19***</td>
<td>-0.04</td>
<td>0.26***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.06)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>State and Cycle F.E.</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>954</td>
<td>968</td>
<td>1042</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.83</td>
<td>0.83</td>
<td>0.78</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.82</td>
<td>0.82</td>
<td>0.76</td>
</tr>
</tbody>
</table>

$^*$p < 0.001, $^*$p < 0.01, *p < 0.05
APPENDIX TO CHAPTER 5

Figure A7: District Median Voter Ideology by District Presidential Vote Share, 2010–2016
### Table A9: Primary Challenge Model (Raw Data)

<table>
<thead>
<tr>
<th></th>
<th>Democratic</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremism from Nonvoter</td>
<td>−6.41*</td>
<td>−6.32</td>
</tr>
<tr>
<td></td>
<td>(2.74)</td>
<td>(3.32)</td>
</tr>
<tr>
<td>Primary Voter Score</td>
<td>4.90</td>
<td>−1.73</td>
</tr>
<tr>
<td></td>
<td>(2.66)</td>
<td>(1.59)</td>
</tr>
<tr>
<td>Incumbent cfScore</td>
<td>0.07</td>
<td>−0.25</td>
</tr>
<tr>
<td></td>
<td>(1.24)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>Incumbent Seniority (logged years)</td>
<td>−0.31</td>
<td>−0.25</td>
</tr>
<tr>
<td></td>
<td>(0.58)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>District PVI</td>
<td>−0.09</td>
<td>−0.06</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>District Unemployment (%)</td>
<td>4.48</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>(21.42)</td>
<td>(21.56)</td>
</tr>
<tr>
<td>District Nonwhite (%)</td>
<td>1.23</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>(4.39)</td>
<td>(4.18)</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>−3.70</td>
<td>−0.91</td>
</tr>
<tr>
<td></td>
<td>(2.49)</td>
<td>(3.16)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>−23.91</td>
<td>−25.00</td>
</tr>
<tr>
<td></td>
<td>−80.67</td>
<td>−81.75</td>
</tr>
<tr>
<td>Smote’d Data</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>346</td>
<td>346</td>
</tr>
</tbody>
</table>

***p < 0.001, **p < 0.01, *p < 0.05
Figure A8: ROC Curve: Comparing Democratic Primary Ideological Challenge Models

Figure A9: ROC Curve: Comparing Republican Primary Ideological Challenge Models
URL: https://cran.r-project.org/web/packages/ggthemes/ggthemes.pdf

URL: https://cran.r-project.org/web/packages/gridExtra/gridExtra.pdf

URL: https://cran.r-project.org/web/packages/questionr/questionr.pdf

URL: https://cran.r-project.org/web/packages/lme4/lme4.pdf

URL: https://cran.r-project.org/web/packages/maps/maps.pdf
URL: https://cran.r-project.org/web/packages/rgeos/rgeos.pdf

URL: https://cran.r-project.org/web/packages/maptools/maptools.pdf

URL: https://cran.r-project.org/web/packages/rgdal/rgdal.pdf

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson, and Jonathan Swinton. 2018. “Export Tables to \LaTeX{} or HTML.” xtable.
URL: https://cran.r-project.org/web/packages/xtable/xtable.pdf

URL: https://cran.r-project.org/web/packages/data.table/data.table.pdf

URL: https://cran.r-project.org/web/packages/doBy/doBy.pdf

URL: https://cran.r-project.org/web/packages/rms/rms.pdf
URL: https://cran.r-project.org/web/packages/emIRT/emIRT.pdf

Leifeld, Philip. 2017. “Conversion of R Regression Output to \LaTeX{} or HTML Tables.” texreg.  
URL: https://cran.r-project.org/web/packages/texreg/texreg.pdf

URL: https://cran.r-project.org/web/packages/mapproj/mapproj.pdf

URL: https://cran.r-project.org/web/packages/showtext/showtext.pdf

URL: https://cran.r-project.org/web/packages/foreign/foreign.pdf

URL: https://cran.r-project.org/web/packages/pROC/pROC.pdf

URL: https://cran.r-project.org/web/packages/ggrepel/ggrepel.pdf
Torgo, Luis. 2015. “Functions and Data for ‘Data Mining with R.’” DMwR.

URL: https://cran.r-project.org/web/packages/DMwR/DMwR.pdf


URL: https://cran.r-project.org/web/packages/gtools/gtools.pdf


URL: https://cran.r-project.org/web/packages/stringr/stringr.pdf


URL: https://cran.r-project.org/web/packages/haven/haven.pdf


URL: https://cran.r-project.org/web/packages/tidyr/tidyr.pdf


URL: https://cran.r-project.org/web/packages/readr/readr.pdf


URL: https://cran.r-project.org/web/packages/dplyr/dplyr.pdf

**URL:** [https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf](https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf)
Bibliography


**URL**: https://www.usnews.com/opinion/articles/2017-11-24/open-primaries-are-the-answer-to-extreme-partisanship

**URL**: http://acuratings.conservative.org/

**URL**: https://doi.org/10.7910/DVN/XFXJYV

**URL**: https://doi.org/10.7910/DVN/GDF6Z0

URL: https://doi.org/10.7910/DVN/26451

URL: http://www.thearda.com/Archive/Files/Descriptions/RCMSST.asp

URL: http://www.thearda.com/Archive/Files/Descriptions/RCMSST10.asp


URL: https://wordpress.clarku.edu/primarytiming/


URL: https://web.stanford.edu/ bonica/data.html

URL: https://nyti.ms/2H2wLIu


**URL:** [http://www.cnn.com/TRANSCRIPTS/1702/17/smer.01.html](http://www.cnn.com/TRANSCRIPTS/1702/17/smer.01.html)


**URL:** [http://www.centerforpolitics.org/crystalball/articles/frc2010030401/](http://www.centerforpolitics.org/crystalball/articles/frc2010030401/)


**URL:** [https://pewrsr.ch/2Jkwbsp](https://pewrsr.ch/2Jkwbsp)


URL: http://unionstats.gsu.edu/MonthlyLaborReviewArticle.htm


URL: https://www.nolabels.org/blog/open-primaries-will-produce-moderate-office-holders/


URL: https://ivn.us/2016/02/24/independent-voter-registration-by-state/

**URL:** https://politi.co/301WTMf


**URL:** https://brook.gs/2H8wLb5


**URL:** https://brook.gs/2Y63UKe

**URL:** [https://brook.gs/2D3vOjx](https://brook.gs/2D3vOjx)


**URL:** [https://brook.gs/2q6PoTs](https://brook.gs/2q6PoTs)


**URL:** [https://brook.gs/2H3eydP](https://brook.gs/2H3eydP)


**URL:** [https://brook.gs/2GEBHV4](https://brook.gs/2GEBHV4)


**URL:** https://doi.org/10.7910/DVN/II2DB6


**URL:** https://voteview.com


**URL:** [https://nyti.ms/2MrM1iL](https://nyti.ms/2MrM1iL)


**URL:** [https://tws.io/2JoKJac](https://tws.io/2JoKJac)


**URL:** [http://www.electproject.org/home/voter-turnout/demographics](http://www.electproject.org/home/voter-turnout/demographics)


**URL:** [https://www.openprimaries.org/mission](https://www.openprimaries.org/mission)


**URL:** [https://nyti.ms/2Yd7BOo](https://nyti.ms/2Yd7BOo)


**URL:** [https://pbpo.st/2PszkdQ](https://pbpo.st/2PszkdQ)


**URL:** [https://pewrsr.ch/2TJMwJf](https://pewrsr.ch/2TJMwJf)

URL: https://bigthink.com/videos/abolish-primary-elections


URL: https://n.pr/2VmOu7e


URL: https://slate.com/news-and-politics/2014/06/eric-cantors-extreme-voters-the-house-majority-leaders-7th-district-primary-voters-were-white-old-and-conservative.html

Sandolhm, Drew. 2014. “Brat: Cantor Defeat ‘was Basically a Miracle’.” *CNBC*, June 11.

URL: https://cnb.cx/2vL039m
URL: https://nyti.ms/2KybCH6

URL: https://politi.co/2q1go6K


URL: https://www.census.gov


URL: https://www.census.gov


URL: http://historical.elections.virginia.gov/


URL: https://prospect.org/article/hardened-divide-american-politics-0


URL: https://53eig.ht/2x3bGbb


URL: https://53eig.ht/2aT6Zbv


URL: https://wapo.st/2VjasZ0


URL: https://thehill.com/homenews/senate/356952-the-original-tea-partier-exits-stage-right


URL: https://www.youtube.com/watch?v=akZZXteJWHU


URL: https://nyti.ms/2POy0z1