ESTIMATING THE IDEOLOGY OF PRIMARY ELECTORATES

Rachel A. Surminsky

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Political Science, concentration in American Politics.

Chapel Hill
2018

Approved by:

Sarah A. Treul

Jason M. Roberts

Santiago Olivella

© 2018
Rachel A. Surminsky
All rights reserved

ABSTRACT

RACHEL A. SURMINSKY: Estimating the Ideology of Primary Electorates
(Under the direction of Sarah A. Treul.)

Studies of racial profiling typically focus on a White/Black or White/minority dichotomy. In this project, I extend that analysis to multiple racial, gender, and class groups. I use data from every traffic stop that occurred in six states over multiple years, amounting to more than 15 million traffic stops. Using this original and unique dataset, I am able to draw conclusions about the outcomes that individual drivers face as a result of their intersectional racial, gender, and class-based perceived identities. I attribute this phenomenon to widely held stereotypes about social groups, rather than to individually racist police officers. Overall, I find that social groups that are stereotyped as more suspicious receive the harshest treatment from police, while those who are not considered suspicious receive lighter treatment, in the aggregate.

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vi

INTRODUCTION 1

REFERENCES 21

LIST OF FIGURES

Figure

1 Republican Party Primary Electorate Ideology by Congressional District . . 16

2 Democrat Party Primary Electorate Ideology by Congressional District . . 16

3 Democratic Party Primary Electorate vs. General Electorate Ideological Estimates 18

4 Republican Party Primary Electorate vs. General Electorate Ideological Estimates 18

INTRODUCTION

More so today than in the past, congressional districts favor one party over the other (Carson, Crespin, Eaves and Wanless 2012). With increased frequency, voters are casting partisan ballots (Jacobson 2009). In the 2016 general election, just thirteen House seats switched party control and 96% of general election voters selected a presidential and congressional candidate from the same party (Carson, Roberts and Surminsky 2017). General elections have become increasingly consistent, producing predictably partisan outcomes. Primary elections, however, have become less predictable. Since 2010, the number of unopposed primary elections has dropped dramatically, demonstrating a shift towards greater intra-party competition (Porter and Treul 2018). Incumbents from safely partisan districts who would presumably win in the general election now fear losing to an in-party challenger in the primary. For some members of Congress it seems the threat of losing reelection has shifted from the general election to the primary election.

Competitive primaries can put an incumbent in a tough position because she must appeal to two different voting electorates. She must please voters in the primary election to secure her party's nomination before running in the general election where she must please a different set of voters. In contemporary elections, the average primary voter is thought to be more ideologically extreme[1] than the average general election voter. Primary constituents are more likely to reward or punish an incumbent for her voting record than are general election constituents (Sides, Tausanovitch, Vavreck and Warshaw Forthcoming). Ideologically extreme voters encourage more ideological challengers to enter at the primary stage, elevating the likelihood of intra-party competition (Butler 2009; Maestas, Fulton, Maisel and Stone 2006). Ideological challengers may better fit the preferences of the primary elec-

[1] My characterization of primary electorate ideology follows current characterizations in the primary electorate literature. I make no underlying assumptions about an individual's placement on a scale from liberal to conservative utilizing policy positions. I am interested in the extremity of an individual's ideology; this could just as well be labeled "partisan extremity"

torate, swaying voters away from the incumbent. Faced with divergent constituencies, an incumbent may choose to appeal more heavily to her ideologically extreme primary electorate instead of the moderate general electorate. To the extent that many members of Congress are elected from safe districts, an uncertain or vulnerable primary election may make pleasing primary voters a higher priority. In turn, if an incumbent is trying to win over ideologically extreme primary voters, she will engage in more ideologically extreme behavior.

This theory for an ideologically extreme primary electorate pulling incumbents away from the median voter is intuitive. Current research on ideological candidate emergence and polarized member behavior fit with this conception of primary elections (Thomsen 2014; Hall 2015; Jewitt and Treul 2018). Anecdotally, examples like the primary election defeat of Richard Lugar (R-IN) fall in line with this argument as well. However, our ability to directly investigate the influence of ideologically extreme primary voters is limited by the lack of better measures for each district's primary election constituency (Brady, Han and Pope 2007).

Current measures of primary constituency ideology may not accurately capture district-by-district heterogeneity in the ideological distribution of constituents. The limitations of these measures may account for mixed findings on the influence, or even existence, of ideologically extreme primary election voters (for a review see Hirano, Jr., Ansolabehere and Hansen 2010; Brady, Han and Pope 2007; Boatright 2014; Clinton 2006; Sides et al. Forthcoming; Abramowitz 2008). If we want to be able to definitively demonstrate that primary election voters are more ideological and test claims about their impact on candidate behavior, we need an new, more direct measure for constituency ideology.

I create a measure of the ideology of primary and general electorates for both parties at the congressional district level. My approach is a multi-stage process. Using voter files aggregated by Catalist, LLC, I model constituent ideological extremity as a function of demographic and geographic predictors. I use multilevel regression with synthetic post stratification (MrsP) to create ideological point estimates for the primary electorate and general electorate for each party in a district. This allows me to compare primary and

general electorate ideology within and across congressional districts.

Examining primary and general electorates for U.S. House districts in the 2012 election, I show that primary voters are more ideologically extreme than general election voters. I also find evidence that there is variation in the extent to which primary and general election voters ideologically diverge. These findings support my theory that the ideological distribution of primary voters is heterogeneous across districts. While primary voters may be more extreme on average, my estimates suggest that some districts have electorates that are very ideologically similar. Comparing between parties, I additionally demonstrate that Republican primary voters are often more ideologically extreme than Democratic primary voters.

My measure demonstrates the necessity for a district-level examination of electorate ideology. Candidate behavior is conditional on district context (Fenno 1978; Maestas et al. 2006). Our theories about the influence of primary voters on candidate behavior must be conditional as well, accounting for ideological heterogeneity across districts. Future work should tailor expectations for candidate behavior to the specific ideological distribution of voters in each candidate's district.

A Need for Electorate-Level Estimates of Ideology

A variety of approaches have been taken to demonstrate that primary voters are ideological extreme. Brady, Han and Pope (2007) use vote shares in elections to demonstrate that primary voters favor more ideologically extreme candidates. Measuring candidate ideological positions using FEC data, Hall and Snyder (2015) find similar results. This literature assumes a spatial model of voting where voter i chooses the candidate j that is the most ideologically proximate to himself. If voter i prefers the ideologically extreme candidate, then he also should be ideologically extreme. Additional research looks at ideological candidate emergence and nomination; it suggests that the emergence of more extreme candidates is motivated by the preferences of ideologically extreme primary voters (Thomsen 2014; Thomsen and Hall 2018).

Several studies use survey data to directly measure the ideological predispositions of primary voters. Clinton (2006) and Butler (2009) use data from the Congressional Cooperative Elections Survey (CCES) to show that same-party constituency preferences motivate candidates to move their positions towards the extreme. Jacobson (2009) looks at the ideological predispositions of primary voters using data from the American National Election Survey (ANES), finding them more extreme than general election voters. These findings point to the presence of ideologically extreme primary *voters* but do not make conclusions about in which districts we should expect to find ideologically extreme primary *electorates*.[2]

A sizable literature argues that incumbent behavior is conditional on her constituency. Fenno's (1978) seminal work finds that the representative-constituent relationship is district-specific. Further work suggest an incumbent will use updated information about her constituency to shift her behavior in the ideological direction of constituents (Rabinowitz and Macdonald 1989; Kousser, Lewis and Masket 2007; Fleisher and Bond 2004; Sulkin 2005; Jewitt and Treul 2018; Clinton 2006). Political economy models of two-stage electoral competition in primary elections predict that an incumbent will diverge from her party conditional on the ideological clustering of her primary constituency (Aranson and Ordeshook 1972; Owen and Grofman 2006; Grofman 2004).[3] In theory, the electoral clout of ideological primary voters should increase with the quantity of ideologically extreme primary voters in a given district. To understand the electoral influence of primary voters, it is not enough to know that primary voters can be extreme. We need to know in which districts these voters tend to be more extreme, forming expectations about incumbent behavior accordingly.

A similar argument can be made for candidate emergence. Strategic, politically experienced candidates are more likely to run when national and local conditions are favorable,

[2] Several analyses have looked at the ideological extremity of primary voters at the national or state level, finding that on aggregate primary voters are not that much more extreme than general election voters (Sides et al. Forthcoming). These explorations provide interesting insight but mask variance in the ideological distribution of voters across districts, which could solicit a polarizing effect on incumbent behavior

[3] Roemer (2001) also formalizes the role of primary constituents in the context of the median voter theorem, showing ideological primary voters as a cause of divergent candidate behavior.

acutely aware of the costs and benefits to running (Jacobson 1989; Hetherington, Larson and Globetti 2003; Maestas et al. 2006). Theories of candidate emergence in the general election have been applied at the primary level, finding similar results: candidates run in primary elections when district-level conditions are the most favorable (Thomsen 2014; Porter and Treul 2018). Without a way to distinguish one primary electorate from another, we cannot pinpoint those districts where we would expect an extreme primary electorate to motivate ideological candidate emergence. If candidate behavior is conditional on a particular district's ideological composition and the ideological extremity of the primary electorate varies across districts, a direct estimate for electorate ideology is necessary to test for a primary constituency effect on candidate emergence.

Current Conceptions of Primary Voter Ideology

Conventional methods for estimating public opinion use data on individual-level voter preferences from national surveys. Survey respondents can be disaggregated into smaller subsamples, for instance by county or congressional district, to estimate voter ideology at the subnational level (Miller and Stokes 1963; Gelman and Little 1997; Leemann and Wasserfallen 2017). However, there is very little survey data on primary elections. National surveys such as the American National Election Survey (ANES) and Cooperative Congressional Elections Survey (CCES) infrequently ask questions regarding primary election vote choice or participation. Additionally, sampling for these large, national-level surveys is not representative of each primary constituency for each party at the congressional district level. Once survey respondents are disaggregated into subnational units, for example the primary constituency for the Democratic party in a given district, there is a significant small-N problem. Limited survey data restricts our ability to use traditional approaches to estimate primary electorate ideology.

In lieu of a direct measure for electorate ideology, scholars have used other identification strategies to pinpoint those districts where we may expect to find an ideological primary electorate. These approaches infer district ideology by looking at the characteristics of primary elections that may correlate with the presence of ideologically extreme

primary electorate. For example, to determine if primary voters influence candidate behavior, McGhee, Masket, Shor, Rogers and McCarty (2014) compare districts with open primary institutions to more closed systems. They assume that states with exclusionary closed primaries will have more partisan, ideological primary electorates; representatives from these states should be more polarized. The authors find no evidence that incumbent behavior is more polarized in districts with closed institutions. Hill (2015) tests McGhee et al.'s (2014) assumption finding that the distribution of voter ideology within primary and general electorates does not correlate with a state's type of primary institution. His work demonstrates that proxy measures for district ideology can mischaracterize the extremity of a district's primary electorate.

In a similar vein, Boatright (2014) and Lawless and Pearson (2008) find little difference in member behavior when comparing across different levels of primary competition, assuming that highly competitive elections should produce an incumbent behavior shift to accommodate primary voter preferences. Jewitt and Treul (2018) argue that close elections may not lead to changes in incumbent behavior, instead pointing to divisive races — races that are ideological in nature — as challenges with behavior-altering consequences. Hirano et al. (2010) find no evidence of incumbent behavior change before and after the introduction of primary elections in the late 1950's to early 1960's. Hill and Tausanovitch (2017) demonstrate that primary electorates today are more ideological than they were in the past. No change in incumbent behavior should have been expected after the introduction of primary elections because primary constituencies in the 1960's were not yet ideologically extreme. These types of investigations into the influence of ideologically extreme primary voters make a strong assumption, inferring that there is ideological consistency in primary electorates across most districts.[4]

Restricted by data availability, our capacity to investigate the electoral impacts of primary voters has previously been limited to these types of approaches. Building on the groundwork laid by Leemann and Wasserfallen (2017), I employ a new method to develop

[4] For additional examinations of indirect measures of primary constituency ideology see Hill and Tausanovitch (2017), Levendusky and Fiorina (2008) and Burden (2001)

a measure for primary and general electorate ideology in both parties at the congressional district-level. I demonstrate that the ideological composition of primary and general election voters does, indeed, vary across districts. My estimates serve to further our understanding of subnational public opinion in the United States and allow for more thorough investigations of primary constituency influence on candidate behavior.

Subnational Public Opinion Estimation: A Methodological Overview

Significant strides in the study of subnational public opinion have been made recently using multi-level regression with postratification (MrP). The utility of MrP comes from its ability to produce more precise estimates of subnational public opinion than traditional disaggregation (Lax and Phillips 2009*a*). MrP up-weights specific geographies that may be undersampled in survey data and down-weights oversampled subpopulations to make estimates more representative. This approach developed by Gelman and Little (1997) has been applied in a variety of contexts to measure public opinion in the United States at the state and local level (Park, Gelman and Bafumi 2004; Lax and Phillips 2009*b*; Tausanovitch and Warshaw 2013; Warshaw and Rodden 2012).

To illustrate how MrP works, consider the example of measuring public opinion on gun control for each state. In traditional MrP, individual-level opinion provided by survey responses is modeled as a function of demographic and geographic information. For example, what does a 20-25-year-old male with a college education in a given state think about gun control? MrP employs hierarchical modeling to exploit the fact that individuals are nested within the geographic subunit of interest (in this case their state) and are nested within other demographic groupings. Hierarchical modeling allows for the borrowing of information across these groups by leveraging demographic and geographic correlations, thus compensating for small sample sizes in national surveys. Using the model, a prediction is made about gun control opinion for every voter type. In other words, a prediction is made for all combinations of demographic and geographic predictors. To make these predictions representative at the state level, they are weighted with the percentage of that voter type within the actual population of the state. Aggregating the weighted predictions by group

within a state produces an estimate of public opinion at the state level.

To weight — or poststratify — predictions, fine-grain data from the U.S. census is often employed. U.S. census data is excellent for postratification because: (1) it serves as a reliable picture of the population, (2) information is broken down by various subnational units of interest for instance county or congressional district, and (3) data is provided in the form of joint distributions. This means that the census not only provides a count of the number of men in a given state but also the number of men who are 20-25 and have a college education in a given state.

The ability of MrP to precisely measure public opinion at very localized levels makes it a promising solution for estimating electorate ideology. However, joint distributions are necessary for MrP; without them public opinion predictions cannot be weighted to reflect the distribution of voter types in a given population. This stringent data requirement has made it impossible to use MrP to estimate constituency ideology. The U.S. census does not release any information on individual-level electoral participation or partisan affiliation. In other words, there are no available joint distributions for the number of men in a given congressional district who are 20-25, have a college education, *and* belong to the Republican party primary electorate.[5]

A recent development by Leemann and Wasserfallen (2017) relaxes the necessity of joint distributions in estimating subnational public opinion.[6] This variation of MrP – known as multilevel regression with synthetic poststratification or MrsP — allows for marginal distributions to be used to impute unknown joint distributions. For example, with a known joint distribution of college-educated men in a specific district and a known

[5] This limitation is not exclusive to examinations of voter participation or partisanship, Warshaw and Rodden (2012) could not use age as a predictor in the MrP model for district level public opinion on individual issue areas.

[6] Before MrsP, scholars attempted to circumvent the limitations of MrP to measure district-level ideology using creative approaches. Tausanovitch and Warshaw (2013) used classic MrP to develop ideological estimates of constituents at the congressional district level. Their estimations include all individuals within a district – voters and non-voters – because the census does not include information about electoral participation. Hill (2015) additionally used MrP to create primary and general electorate estimates for each party for each congressional district. Without census-level data to poststratify his predictions, Hill used survey weights. This approach, however, could bias his estimates if survey weights do not accurately capture population characteristics at the requisite level of disaggregation.

marginal distribution of registered Republicans in a specific district, MrsP can be used to impute how many college-educated men in that district are registered Republicans. This imputed joint distribution is what Leemann and Wasserfallen (2017) call a "synthetic joint distribution". Correlations amongst empirical joint distributions of individual-level variables can be used to help inform the creation of synthetic joint distributions. For example, if college educated men tend to be more Republican, this information is used to help guide our estimations. These "adjusted synthetic joint distributions" take into account dependencies across covariates.

Data, Measurement, & Analysis of District Ideology

I use MrsP to create a direct measure for primary and general electorate ideology. MrsP performs just as well, and in some instances better than, traditional MrP. It does not require census-level joint distributions that include a person's electoral participation, which has prevented its application for this purpose in the past (Leemann and Wasserfallen 2017). MrsP overcomes the methodological hurdle in producing of a district-level measure for constituency ideology, but data availability is still an obstacle.

To nest survey respondents within the primary or general electorate for each district requires knowing if they voted in the primary election, the general election, or both. As previously noted, questions about primary voter participation are highly infrequent in national-level surveys.[7] Further, self-reported participation measures often over-report election turnout (Butler 2009; Sides et al. Forthcoming; Vavreck 2007). Using validated voter turnout in the CCES resolves this problem, but drastically reduces already small subsamples of respondents nested within each party's primary and general electorate for each district.[8] The average number of CCES respondents who voted in the primary election

[7] For example, the American National Elections Survey only asked about primary election turnout in 1958, 1964, 1966, and 1978. The ANES asked about presidential primary participation in 1992, not congressional primary participation.

[8] The CCES does not directly ask questions about primary voter participation, but instead validates voter turnout in the primary and general election using voter files. The CCES posed a question about primary turnout in 2008 exclusively. Voter validation with voter files for the CCES is completed via Catalist data.

for a given party in each congressional district is approximately 15; some district primary constituencies have no respondents, a few have as many as 100.[9]

Instead of national surveys, I use Catalist LLC's Validated Voter Database as a data source for MrsP estimation. Catalist aggregates voter files for all 50 states, also drawing on external data sources to build profiles of individuals. The academic subscription provides a 1% randomly drawn sample of their database. My sample drawn in 2014 for the 2012 election includes 3.1 million cases, approximately 890,000 primary voters. The Catalist Voter Database includes demographic information, validated election participation, and party registration when available. This data from Catalist, LLC acts as a good substitute to survey data, including similar information with a much larger sample size of primary election voters. However, using the data involves making some assumptions, discussed in the following section.

Estimating Voter Ideology

To build estimates of constituency ideology by party for each congressional district, I employ a hierarchical linear model. It is similar to those adopted in previous studies examining voter ideological extremity (Hill and Tausanovitch 2017; Sides et al. Forthcoming). I regress an individual's ideological extremity on a standard set of demographic characteristics including age, gender, education level, and race. I also include indicators variables for state and congressional district along with dummy variables for 2012 Democratic primary participation and 2012 Republican primary participation.

[9] It is worth noting that pooling surveys creating what is called a mega-poll across CCES years, similar to the approach used by Warshaw and Tausanovitch (2013), would not work for my purposes. My postratification involves using voter turnout for a given election year as the marginal distribution to produce synthetic joint distributions. This marginal distribution, the number of voters in a given district's primary and general election, would change from year to year and make this approach unuseable.

$$
\begin{aligned}
Ideology_i &= \beta_0 + \alpha_{race_{[i]}} + \alpha_{gender_{[i]}} + \alpha_{edu[i]} + \beta_{demprimary_{[i]}} + \beta_{repprimary_{[i]}} \\
&\quad + \alpha_{age[i]} + \alpha_{state[i]} + \alpha_{district[i]} + \varepsilon_{[i]} \\
\alpha_{\cdot} &\sim \mathcal{N}(0, \sigma_{\cdot}^2) \\
\alpha_{district} &\sim \mathcal{N}(\gamma_{1\text{PresVote}}, \sigma_{district}^2) \\
\varepsilon_{[i]} &\sim \mathcal{N}(0, \sigma_y)
\end{aligned}
$$

All predictors are modeled using random effects except party primary participation modeled using fixed effects.[10] I let my model intercept vary by congressional district and state. Random effects are drawn from a zero mean normal distribution, though the district level covariate is drawn from a distribution centered on the the Democratic presidential vote share for that district.[11] A discussion of the measurement for each variable in the model is as follows.

Voter Ideological Extremity

To help clients analyze the electorate and target constituents, Catalist, LLC builds predictive scores using their Voter Validated Data for a respondent's ideological extremity. This synthetic score is scaled from 0-100 with 0 being the most conservative and 100 being the most liberal. The variable is constructed using more than 150 covariates and is generally accepted as a reliable measure for relative ideological extremity between individuals (Hersh 2015).

[10] I use fixed effects because I assume Democratic and Republic primary voter ideological extremity is not drawn from a common distribution. There is no borrowing of information across groups to inform an individual's level of ideological extremity. The ideology of Democratic and Republican primary voters will be fundamentally different.

[11] Presidential vote share was calculated as the percent of the two-party vote in a district that went to the Democratic presidential nominee in the previous election year.

Age

The age variable in the Catalist database is respondent's age at the time of the next general election; subtracting two from this number gives a respondent's age at the time of the 2012 election. Only individuals in the voting age population are included in the analysis. Data on respondent age provided by Catalist is drawn from state voter files.

Gender

This dummy variable equals 1 if the respondent is female, 0 otherwise. Data on respondent gender provided by Catalist is also drawn from state voter files.

Education

State voter files do not include information about an individual's education level. Based on geographic information, consumer information, and other covariates, Catalist, LLC creates a propensity score for a respondent's likelihood to have a certain level of educational attainment. Catalist includes six potential bins for respondent education, I collapse this into three categories: Not High School Graduate, High School Graduate, and Bachelor's Degree/Postgraduate.[12] Comparing across these three categories, the respondent is coded for having the educational attainment of whichever propensity score is the highest.

Race

The race variable provided by Catalist includes race and ethnicity categories different from those included in U.S. census data. In order to weight model predictions in the postratification stage of MrsP, these categories are binned to match those in the census. Race categories include Caucasian, Black, Hispanic, and Other. Data is drawn from state voter

[12] The categories originally provided by Catalist for the imputed education variable are the following: Not High School Graduate, High School Graduate, Some College, Bachelor's Degree, Bachelor's or Postgraduate Degree, Postgraduate degree. These categories proved to be very messy, with Some College, Bachelor's, Bachelor's or Postgraduate Degree, Postgraduate degree having similar propensity scores. I drop the Some College category, Bachelor's and Postgraduate are binned together to improve the binning accuracy.

files and commercial sources.[13]

MrsP Poststratification

I use this model to calculate predicted values of ideological extremity for all voter types in the Catalist Validated Voter Database. To make these predictions representative at the district-level, they are weighted with the percentage of that voter type within the actual population of each congressional district. The U.S. census provides the joint distributions for all but two individual-level predictors in the model. The census factfinder does not include age or voter participation in its joint distributions. For these demographic characteristics I impute joint distributions using marginal distributions. To account for correlations amongst individual-level predictors, I use adjusted synthetic joint distributions instead of simple synthetic joint distributions.

To produce an adjusted synthetic joint distributions for age, I use the marginal distribution for age provided by the U.S. census. However, the U.S. census provides no information about voter participation in the primary or general election. In lieu of census data, I characterize the marginal distribution for the Republican (Democratic) primary electorate as the total number of voters who participated in the Republican (Democratic) primary. For the general electorate, I use the total number of voters who voted for the Republican candidate and the total number of voters who voted for the Democratic candidate as my marginal distribution.

Using voter turnout as my marginal distribution could be problematic for several reasons. First, if a race is unopposed, there is no recorded vote total in that party's primary. Therefore, no marginal distribution exists for voter turnout and no ideological estimate can be produced in that district for the party's primary constituency. On one hand, this could indicate that a representative matches her constituency well; on the other, it may simply be that no challenger decided to run. Regardless, this limits the explanatory power of my estimates.

[13] This other categories matches the "Other" race category in the census; it includes Asian Native-American / Pacific Islander.

Second, voter turnout in elections fluctuates year-to-year, therefore the marginal distribution for primary and general election voters fluctuates year-to-year. This could introduce bias into my estimates. Some may argue that if turnout in the primary election vacillates, it is not a good measure of a district's primary constituency. I would argue otherwise. Hill (2015) and Sides et al. (Forthcoming) demonstrate that the demographic characteristics and ideological predispositions of voters participating in primary elections do not vary widely across years. Per Fenno (1978), primary voters should be the most dedicated individuals within a constituency. Therefore, while we may see variability in turnout, we should not see too much variability in the types of voters participating.

Third, defining primary and general election constituencies as only those people who voted in an election requires making an assumption about who candidates pay attention to when campaigning. However, incumbent attention allocation and responsiveness has been widely tested in the literature finding that representatives are more responsive to partisan voters than non-voters (Fenno 1978; Clinton 2006; Bafumi and Herron 2010; Bartels 2016).

Disaggregation by Primary and General Electorate

Including an indicator variable for a subunit of interest in the preceding response model allows for estimates to be disaggregated after the postratification step of MrsP by that subunit. For instance, including an indicator variable for state and congressional district allows for estimates to be disaggregated at the congressional district level. To disaggregate by primary electorate for each party at the congressional district level requires an indicator variable for a respondent's state and congressional district, as well as a dummy variable for the partisan primary he voted in.

As stated previously, I use party primary participation as a proxy for whether an individual is part of a party's primary electorate. Catalist LLC's Validated Voter Database provides more thorough information about voter participation than traditional survey data through the aggregation of state voter files. The dataset provides complete information for whether a respondent is a voter or non-voter.

The Validated Voter Dataset does not have complete information on *which* primary a

voter voted in; data availability varies with a state's type of primary institution. States with closed and semi-closed primary institutions require voters to register with a party to participate in the primary election. For these states I place voters in the primary constituency corresponding to their party registration. For independents in semi-closed systems and all voters in open systems, I do not know for certain in which party's primary a voter participated. Additionally, several states with semi-closed systems do not disclose party registration information in their voter files. In these instances, I assume a voter participates in the party primary matching their party registration. In the absence of party registration, I use the Catalist partisanship propensity score as a substitute. Much like the ideological extremity score, the partisanship propensity score uses covariates in the Catalist, LLC database to predict an individual's partisan affiliation. I assume a voter participates in the party primary most closely matching their party propensity score.

To estimate the ideological extremity of general electorates, I use general election vote choice as a proxy for whether an individual is part of a party's general election constituency. Unlike primary electorate affiliation, I cannot make strong inferences for a voter's general electorate affiliation because it is impossible to know for certain which candidate a person voted for. I classify voters as Democrat or Republican following the same procedure explained above, using an individual's party registration whenever available or the Catalist party propensity score to infer vote choice.

Results

Figures 1 and 2 summarize my initial findings. I have completed estimates for all districts that had a contested race in the 2012 primary election. Recall, marginal distributions for voter turnout are only available for contested primaries. I cannot create primary electorate ideological estimates for districts with an uncontested party primary. The gray districts denote those primaries elections that were uncontested or did not have any candidates run. Therefore, these gray districts have no estimates produced. Further, I only include states that have partisan primaries in my anlysis, excluding those top-two primary states of California, Washington, and Louisiana.

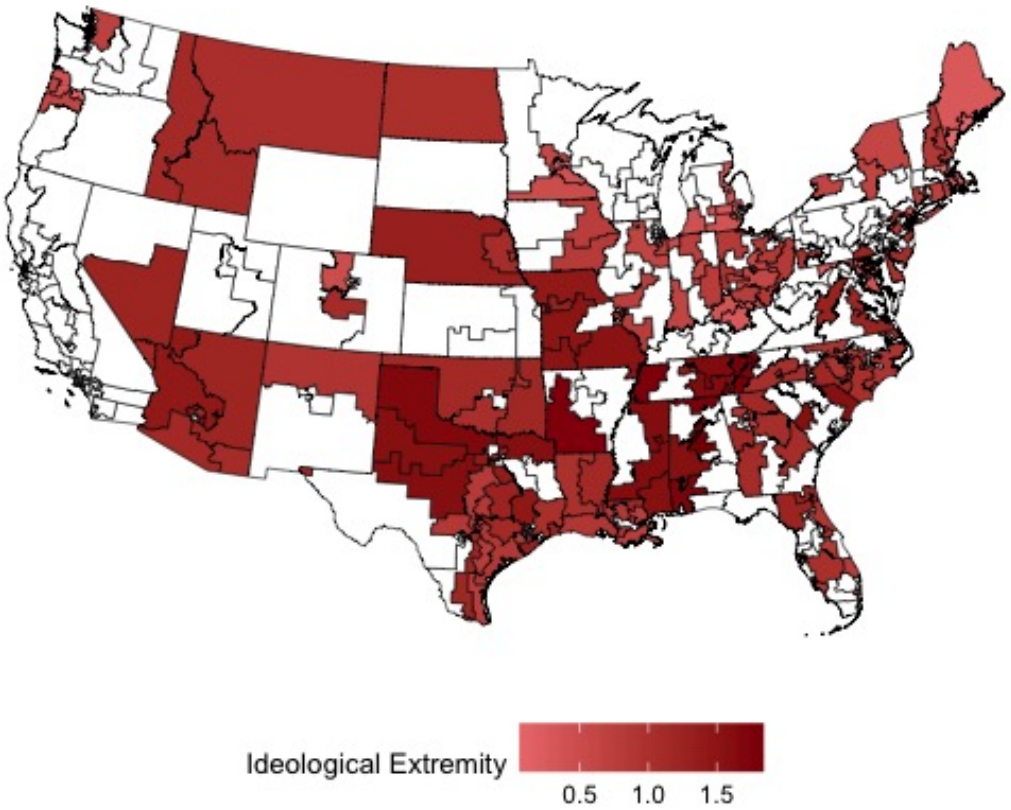

Fig. 1: Republican Party Primary Electorate Ideology by Congressional District

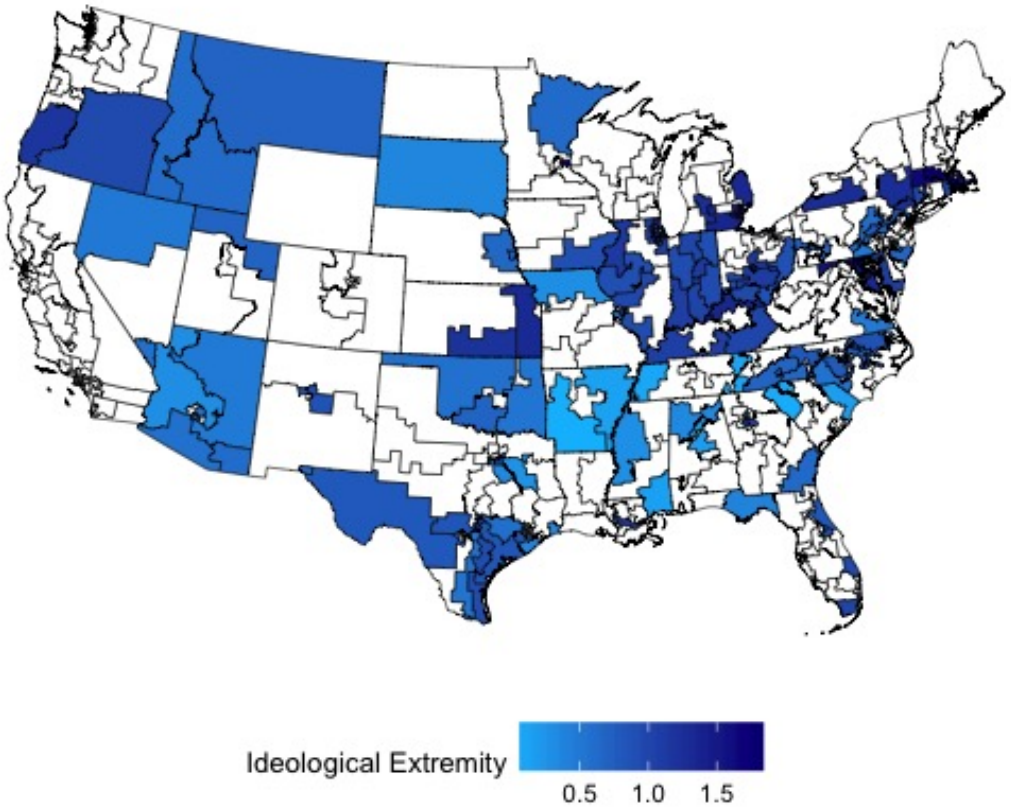

Fig. 2: Democrat Party Primary Electorate Ideology by Congressional District

Figure 1 depicts ideological estimates for Republican primary electorates, Figure 2 depicts ideological estimates for Democratic primary electorates. The scale for ideological extremity in Figure 1 ranges from 0-50, with 0 being the most conservative. The scale for ideological extremity in Figure 2 ranges from 50-100, with 100 being the most liberal. In Figure 1, the districts with the most conservative primary electorates for the Republican party are shaded lighter. In Figure 2, the districts with the most liberal primary electorates for the Democratic party are shaded darker. The most notable findings in these figures are (1) the ideological variation between states, and (2) the ideological variation within states. Also worth noting is the variation in ideological extremity between Republican and Democratic district estimates. Republican districts seem to exhibit a higher level of variation in extremity, while Democratic districts seem to be more consistently moderate. If ideological extremity in primary constituencies varies by district as well as within and across states, evaluating primary voters on aggregate will not capture ideological heterogeneity. These estimates demonstrate the necessity for a direct measure of primary constituency ideology in evaluations of primary voter influence on candidate behavior.

Comparing primary electorate estimates across districts with similar electoral characteristics demonstrates how indirect measures for district ideology gloss over this ideological heterogeneity. Finding similar results to Hill (2015), there is no consistency in the ideological extremity of districts in states with the same type of primary institution. Comparing Nebraska and Georgia, both states with open systems, the Republican primary electorates in these states clearly have different ideological distributions. States like Florida — which has a closed system — and Georgia – which has an open system — have comparable levels of ideological extremity in their primary constituencies despite having different types of primary election laws.

Moving to Figure 3 and 4, I examine differences in ideology between primary and general electorates to investigate if primary voters are more ideologically extreme than general election voters. For districts in six randomly selected states, I produce ideological estimates for both party's primary and general election constituencies. I plot these estimates to explore their relative ideological extremity. Each point represents a single district. Primary

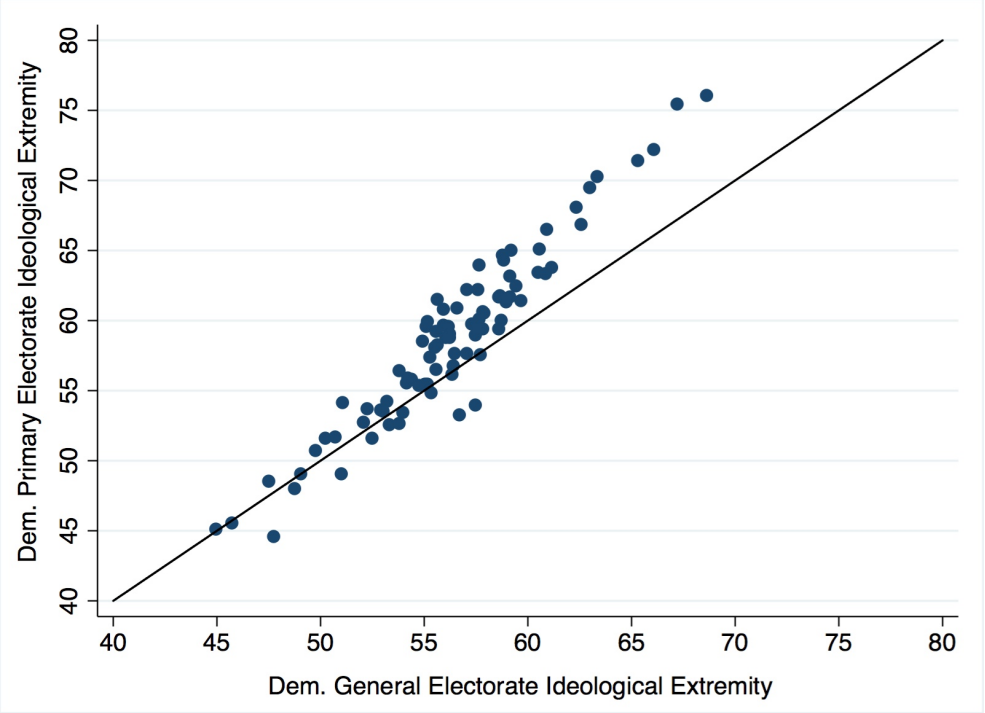

Fig. 3: Democratic Party Primary Electorate vs. General Electorate Ideological Estimates

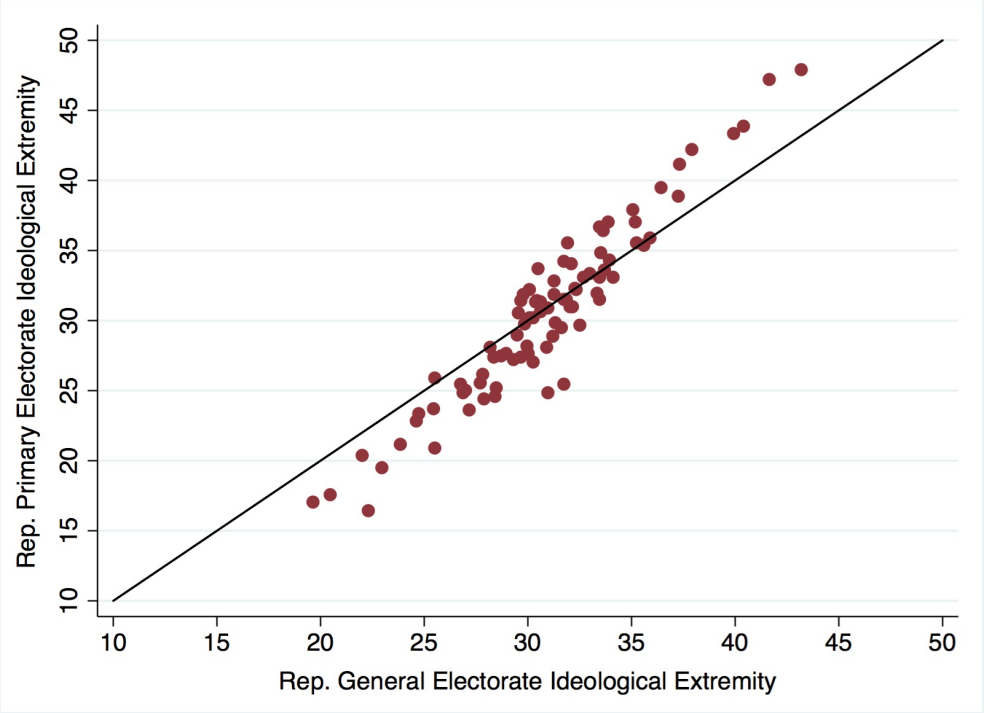

Fig. 4: Republican Party Primary Electorate vs. General Electorate Ideological Estimates

constituency ideology is plotted on the y-axis, general electorate ideology is plotted on the x-axis. Similar to Figures 1 and 2, ideological extremity is on a 0-100 scale with 0 being the most conservative and 100 being the most liberal.

Figure 3 plots Democratic primary and general electorates; districts falling above the reference line have partisan primary electorates that are more ideologically extreme than the general electorate. Figure 4 plots Republican primary and general electorates; districts falling below the reference line have primary electorates that are more ideologically extreme than the general electorate. In most cases for Democratic constituencies, the primary electorate is more ideologically extreme than the general electorate. This relationship is less prevalent when examining Republican constituencies. These mixed findings could be motivated by several different factors.

First, in this analysis I am looking at a small sample of cases, I may observe greater variation when I estimate the remaining districts for the 2012 congressional election. Second, to estimate relative primary and general electorate extremity I require marginal distributions for *both* party's primary electorates *and both* party's general electorates. In other words, if any party's primary race is uncontested in a district I cannot create an estimate for the general electorate's ideological extremity. This means that those districts for which I can produce estimates may be innately more competitive; both parties ran contested primary elections in these districts. In turn, I may expect constituencies in this subset of districts to be generally more moderate. Third, estimating primary constituency ideology in the year of a presidential election could moderate my estimates. Creating estimates for the 2014 primary and general election would remove the influence of presidential primary participation.

Discussion & Next Steps

This paper serves as a proof-of-concept: using MrsP, direct estimates of primary and general electorate ideological extremity can be produced at the congressional district level. My new measure should encourage scholars to revisit existing findings regarding the influence of primary voters on candidate behavior. Future theories should be conditioned on the ideological extremity of each district's primary electorate, allowing expectations for

incumbent behavior and candidate emergence to vary across contexts.

Continuing to work on this project, my next step is to conduct robustness checks for my estimates. Hill (2015) notes, "Without clear benchmarks, it is hard to evaluate the procedure outside of the statistical theory that demonstrates that both hierarchical models and post-stratification improve the validity of...estimates to corresponding population statistics." (Appendix, 6). Using CCES survey data I will estimate an IRT model to predict individual-level ideology as a function of survey policy questions as an attempt to validate my measure. This IRT model will allow for the evaluation of Catalist LLC's synthetic ideology score in capturing actual voter predispositions. Even if certain districts do not have enough data to produce estimates, I can compare my Catalist-based estimates to those districts that could be estimated with CCES data.

Expanding my dataset to include years beyond 2012 will allow me to average estimates of primary constituency ideology over multiple years. Thus may allow me to correct for any bias that may stem from postratification using a single election's voter turnout. Ideally in future work on this project when more data becomes available, I will be able to compare across years to see how widely ideological estimate vary based on election turnout.

REFERENCES

Abramowitz, Alan. 2008. "Don't Blame Primary Voters for Polarization." *The Forum* 5(4).

Aranson, Peter H. and Peter C. Ordeshook. 1972. Spatial strategies for sequential elections. In *Probability Models of Collective Decision Making*, ed. RG Niemi and HF Weisberg. Columbus, Ohio: Charles E. Merrill Publishing Company.

Bafumi, Joseph and Michael C. Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104(3):519–542.

Bartels, Larry M. 2016. *Unequal democracy: The political economy of the new gilded age*. Princeton University Press.

Boatright, Robert G. 2014. *Getting Primaried: The Changing Politics of Congressional Primary Chal- lenges*. University of Michigan Press.

Brady, David W., Hahrie Han and Jeremy C. Pope. 2007. "Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32(1):79–105.

Butler, Daniel M. 2009. "The Effect of the Size of Voting Blocs on Incumbents' Roll-Call Voting and the Asymmetric Polarization of Congress." *Legislative Studies Quarterly* 34(3):297–318.

Carson, Jamie L., Jason M. Roberts and Rachel Surminsky. 2017. "The Nationalization of Congressional The Nationalization of Congressional Elections." Presented at the Annual Meeting of the Midwest Political Science Assocatiation.

Carson, Jamie, Michael H. Crespin, Carrie P. Eaves and Emily O. Wanless. 2012. "Constituency Congruency and Candidate Competition in Primary Elections for the U.S. House." *State Politics and Policy Quarterly* 12(2).

Clinton, Joshua. 2006. "Representation in Congress: constituents and roll calls in the 106th House." *The Journal of Politics* 68(2).

Fenno, Richard F. 1978. *Home Style: House Members in Their Districts*. New York: Longman.

Fleisher, Richard and John R. Bond. 2004. "The shrinking middle in the US Congress." *British Journal of Political Science* 34(3):429–451.

Gelman, Andrew and Thomas C. Little. 1997. "Poststratification Postratification into Many Categories Using Hierarchical Logistic Regression." *Survey Research* 23(2):127–35.

Grofman, Bernard. 2004. "Downs and Two-Party Covergence." *Annual Review of Political Science* 7:25–46.

Hall, Andrew. 2015. "What Happens When Extremists Win Primaries?" *American Political Science Review* 109(1):18–42.

Hall, Andrew and James M. Snyder. 2015. "Candidate Ideology and Electoral Success.".

Hersh, Eitan D. 2015. *Hacking the Electorate: How Campaigns Perceive Voters*. Cambridge University Press.

Hetherington, Marc J., Bruce Larson and Suzanne Globetti. 2003. "The Redistricting Cycle and Strategic Candidate Decisions in U.S. House Races." *Journal of Politics* 65(4):1221–1234.

Hill, Seth J. 2015. "Institution of nomination and the policy ideology of primary electorates." *Quarterly Journal of Political Science* 10(4).

Hill, Seth J. and Chris Tausanovitch. 2017. "Southern realignment, party sorting, and the polarization of American primary electorates, 1958-2012." *Public Choice* (1-26).

Hirano, Shigeo, James M. Snyder Jr., Stephen Ansolabehere and John Mark Hansen. 2010. "Primary Elections and Partisan Polarization in the U.S. Congress." *Quarterly Journal of Political Science* 5.

Jacobson, Gary C. 1989. "Strategic Politicians and the Dynamics of U.S. House Elections, 1946-1986." *American Political Science Review* 83(3):773–93.

Jacobson, Gary C. 2009. *The Politics of Congressional Elections*. 7th edition ed. New York: Longman.

Jewitt, Caitlin E. and Sarah A. Treul. 2018. *Routledge Handbook on Primary Elections*. Routledge chapter Ideological Primaries and their Influence in Congress.

Kousser, Thad, Jeffrey B. Lewis and Seth E. Masket. 2007. "Ideological adaptation? The survival instinct of threatened legislators." *Journal of Politics* 69(3):828–843.

Lawless, Jennifer L. and Kathryn Pearson. 2008. "The Primary Reason for Women's Underrepresentation? Reevaluating the Conventional Wisdom." *Journal of Politics* 70(1):67–82.

Lax, Jeffrey R. and Justin H. Phillips. 2009*a*. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Journal of Political Science* 103(3):367–86.

Lax, Jeffrey R. and Justin H. Phillips. 2009*b*. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1):107–21.

Leemann, Lucas and Fabio Wasserfallen. 2017. "Extending the Use and Prediction Precision of Subnational Public Opinion Estimation." *American Journal of Political Science* 61(4):1003–1022.

Maestas, Cherie D., Sarah A. Fulton, L. Sandy Maisel and Walter J. Stone. 2006. "When to risk It? Institutions, ambitions, and the decision to run for the U.S. House." *American Political Science Review* 100(2):195–208.

McGhee, Eric, Seth Masket, Boris Shor, Steven Rogers and Nolan McCarty. 2014. "A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology." *American Journal of Political Science* 58(2).

Miller, Warren E. and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 1(57):45–56.

Owen, Guillermo and Bernard Grofman. 2006. "Two-stage electoral competition in two-party contests: persistent divergence of party positions." *Social Choice and Welfare* 26(3):547–569.

Park, David K., Andrew Gelman and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12(4):375–85.

Porter, Rachel and Sarah A. Treul. 2018. "The Primary Decision: Candidate Emergence and Success in Congressional Primaries." Presented at the Annual Meeting of the Midwest Political Science Assocatiation.

Rabinowitz, George and Stuart Elaine Macdonald. 1989. "A directional theory of issue

voting." *American Political Science Review* 83(1):93–121.

Roemer, J.E. 2001. *Political Competition: Theory and Applications*. Harvard University Press.

Sides, John, Chris Tausanovitch, Lynn Vavreck and Christopher Warshaw. Forthcoming. "On the Representativeness of Primary Electorates." *British Journal of Political Science* .

Sulkin, Tracy. 2005. *Issue Politics in Congress*. Cambridge University Press.

Tausanovitch, Chris and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2):330–342.

Thomsen, Danielle M. 2014. "Ideological Moderates Won't Run: How Party Fit Matters for Partisan Polarization in Congress." *Journal of Politics* 76(3).

Thomsen, Danielle M and Andrew Hall. 2018. "Who Punishes Extremist Nominees? Candidate Ideology and Turning Out the Base in US Elections." *American Political Science Review* .

Vavreck, Lynn. 2007. "The exaggerated eects of advertising on turnout: The dangers of self-reports." *Quarterly Journal of Political Science* 2(4):325–343.

Warshaw, Christopher and Jontahan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues." *Journal of Politics* 74(1):203–19.