

# Estimating the Ideology of Congressional Primary Electorates

Rachel Porter\*

## Abstract

Contemporary accounts of elections characterize voters who participate in primaries as the extreme subset of their party. However, previous methodological limitations have impeded efforts to directly measure the ideological extremity of primary voters at the congressional district level. Using multilevel regression with synthetic poststratification (MrsP), I create a direct measure of primary constituency ideology for both parties in each congressional district. My analysis demonstrates that there is significant variation in constituency extremity across districts and that this variation is dependent on district characteristics. Examining the effect of primary constituency extremity on polarization, I show that (1) ideological candidates are more likely to emerge in districts with extreme primary electorates, and (2) incumbents are less representative of primary voters when the general election is competitive for both parties. The estimates developed here offer numerous opportunities for future work to explore patterns in candidate emergence and success at the primary election level.

---

\*The University of North Carolina at Chapel Hill, Department of Political Science

Primary elections are often blamed for furthering partisan polarization in Congress. Recent work suggests that primary voters hold more ideologically extreme policy positions and prefer more ideologically extreme primary candidates (Brady et al. 2007; Hall and Snyder 2015). Some argue that it is the preferences of these voters that impact candidate behavior in a way that facilitates polarization. For instance, at the candidate emergence stage, Thomsen (2014) finds that moderate potential candidates increasingly opt out of running for Congress, perceiving their chances of winning as low because they do not appeal to primary voters. In turn, retiring moderate incumbents have slowly been replaced by more ideologically extreme members thus driving the parties apart (Bafumi and Herron, 2010; Theriault, 2012). Others, however, find little evidence for a connection between the primary electorate and partisan polarization (e.g. Hirano et al. 2010; Sides et al. 2018).

These contradictory findings may in part be explained by the lack of a district-level measure of primary electorate extremity. Presumably, the relative extremity of primary voters varies across districts, just like the ideological distribution of general election voters. To reflect this variation, the relationship between primary voters and polarization should be defined at the district level, with more extreme electorates fostering greater congressional polarization. However, without a direct measure for primary constituency extremity, accounting for this heterogeneity has been challenging. Some scholars have relied on noisy proxy indicators, which may muddle the connection between primary voters and polarization.<sup>1</sup> Others have focused their research on aggregate-level questions, assuming that ideological extremity is relatively homogeneous across districts.

To more directly measure primary electorate extremity, I produce estimates for the ideology of each party’s primary electorate in congressional districts for the most recent redistricting cycle. Using voter files aggregated by Catalist, LLC, I model constituent ideological extremity as a function of demographic and geographic predictors. Improving on current

---

<sup>1</sup>In lieu of a direct measure for primary electorate extremism, the average DIME campaign contribution score within a district, the average self-reported ideology of survey respondents, and presidential two-party vote share have all been used to capture this important quantity. One noteworthy exception, Hill (2015) produces a more direct measure for primary constituency extremity.

measures, these estimates are then corrected for sampling bias using an approach introduced by Leemann and Wasserfallen (2017). The stringent data requirements associated with existing methods for sampling bias correction—such as multi-level regression with poststratification (MrP)—have hindered their broader application, particularly in models that incorporate partisanship or ideology among their predictors. This is because the kinds of high-quality, census-level data that are required for MrP rarely include information on individuals’ political affiliation. Leemann and Wasserfallen’s (2017) extension of MrP called multilevel regression with *synthetic* poststratification (MrsP), however, relaxes these data requirements, allowing for strong predictors like partisanship to be incorporated into poststratification and, thus, included in model estimation.

My estimates demonstrate that primary constituency extremity varies substantially across districts and between parties. This finding demonstrates the importance of conditioning on the district to study the relationship between extremism and other outcomes. Without properly accounting for the connection between candidate behavior and ideologically extreme voters, current work may underestimate the influence of primary electorates on polarization. Using my estimates, I further examine the effect of primary electorate extremity on congressional polarization. I show that ideological challengers are more likely to emerge in districts with more extreme primary electorates. I also find that the relationship between member ideology and primary electorate extremity is weaker when the general election is competitive for both parties. This finding suggests that general election dynamics produce conditions that make members of Congress more responsive to primary voters. Finally, in producing these estimates, I provide a road-map for other scholars studying partisan polarization who may wish to implement this approach in their own work.

## A Need for Electorate-Level Estimates of Ideology

Having a measure of primary electorate ideology is essential to study the effects of constituency extremity on candidate behavior in elections and, further, incumbent behavior in

Congress. Such a measure is especially important given ongoing shifts in the dynamics of congressional elections. More so today than in the past, congressional districts favor one party over the other and, with increased frequency, voters are casting partisan ballots (Jacobson and Carson, 2016). 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. General elections have become more and more 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, 2019). 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.

Increasingly competitive primaries put incumbents in a tough position because, in today's elections, the average primary voter is thought to be more ideologically extreme<sup>2</sup> than the average voter in the general election. A sizable literature argues that incumbent behavior is conditional on the preferences and demographics of her constituency (e.g. Fenno 1978). Other work more specifically suggests that an incumbent will use updated information about her constituency to shift her behavior in the ideological direction of her constituents (Rabinowitz and Macdonald, 1989; Kousser et al., 2007; Fleisher and Bond, 2004; Sulkin, 2005; Clinton, 2006). However, when the nominating electorate and general electorate are ideologically distinct, tailoring legislative and electoral behavior to the constituency becomes a far more difficult task for incumbents.

Beyond ideological distinctiveness, primary voters present other challenges to an incumbent seeking reelection. For example, primary constituents are more likely to reward or punish an incumbent for her voting record than are general election constituents (Sides et al.,

---

<sup>2</sup>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."

2018). If primary voters are more attuned to congressional behavior then, perhaps, it is more strategic for an incumbent to align with the primary electorate’s ideology—especially in an era where general election voters cast predictably partisan ballots. Ideologically extreme voters also encourage more ideological challengers to enter at the primary stage, elevating the likelihood of intra-party competition (Butler, 2009; Maestas et al., 2006; Thomsen, 2014). These ideological challengers may better fit the preferences of the primary electorate, swaying voters away from the incumbent and putting her reelection in jeopardy.

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 makes 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.

A similar argument about strategic behavior can be made for challenger emergence in primary elections. Strategic, politically experienced candidates are more likely to run when national and local conditions are favorable, acutely aware of the costs and benefits to running (Jacobson, 1989; Hetherington et al., 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, 2019). If this is the case, then extreme candidates should emerge and succeed more often in districts where they align well with an ideologically extreme primary constituency.

Shifting electoral competition has made understanding the ideological composition of primary electorates an increasingly interesting and important topic of study. Characterizing the preferences of primary voters as more ideologically extreme has provided some insight into incumbent behavior and candidate emergence. However, to measure the electoral influence of the primary electorate it is not enough to know that these voters can be extreme.

If incumbent 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 incumbent behavior. Further, without a way to distinguish one primary electorate from another, we cannot pinpoint those districts where one would expect an extreme primary electorate to motivate ideological 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 the 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 on the participation and preferences of primary election voters 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 strategies to identify those districts where one 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 an ideologically extreme primary electorate. Proxy measures for extremity, however, have proven to be inconsistent or erroneous. For example, to determine if primary voters influence candidate behavior, McGhee et al. (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. 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. In a similar vein, 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. Studies such as these demonstrate that proxy measures for district ideology can mischaracterize the extremity of primary electorates.

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 measure primary electorate ideology for both parties at the congressional district-level. I demonstrate that the ideological composition of primary 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.

## Data & Methodological Approach

Significant strides in the study of subnational public opinion have been made recently using multi-level regression with poststratification, also known as MrP (Gelman and Little, 1997).<sup>3</sup>

---

<sup>3</sup>For a selection of articles employing MrP see: Park et al. (2004); Lax and Phillips (2009); Tausanovitch and Warshaw (2013); Warshaw and Rodden (2012).

The utility of MrP comes in part from its ability to produce more precise estimates of subnational public opinion by up-weighting specific demographic groups that may be under-sampled in survey data and down-weighting over-sampled subpopulations. In order to weigh subpopulations, census-level data is used. To be compatible with MrP, census data on demographic characteristics of interest must be available in the form of joint distributions. For example, if gender, education, and age are predictors of interest, census data must provide the proportion of 25 year old men that are college educated who live in a given subnational unit. Providing the *marginal* proportion of residents who are male, the proportion who are 25 years old, and the proportion who are college educated will not suffice. This is problematic for studying primary electorates given that the U.S. census does not release joint distributions that include electoral participation or party affiliation. In other words, the proportion of 25 year old men that are college educated who live in a given subnational unit *who also* voted in the Republican Party primary election is not provided.<sup>4</sup> This limitation makes it impossible to use traditional MrP models to estimate partisan electorate ideology.

A recent development by Leemann and Wasserfallen (2017) relaxes the necessity of joint distributions in estimating subnational public opinion.<sup>5</sup> 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. To compute these missing joint distributions based on marginal information, Leemann and Wasserfallen (2017) propose two approaches. The first simply assumes independence between poststratification variables and imputes joint values by multiplying their marginal frequencies. The second approach, which

---

<sup>4</sup>This limitation is not exclusive to examinations of voter participation or partisanship. For example, Warshaw and Rodden (2012) could not use age as a predictor in their MrP model for district level public opinion on individual issue areas.

<sup>5</sup>Before MrsP, scholars attempted to circumvent the limitations of MrP to measure district-level ideology using creative approaches. Tausanovitch and Warshaw (2013) use 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.



I adopt here, relaxes the unrealistic assumption of independence, and attempts to capture dependence between poststratification variables using observed covariation between them in the available survey data. More specifically, a multi-step procedure corrects the observed joint distribution of poststratification variables in the survey data to match the target population’s known joints and marginal distributions, and then uses these corrected sample-based joints to impute the population-level distribution of interest.<sup>6</sup> For my own purposes in employing MrsP, I can use the marginal distribution of voters who belong to a district’s partisan primary constituency to impute census-level joint distributions that include primary electorate membership. To define each district’s marginal distribution of Democratic (Republican) primary electorate members, I use total voter turnout for that district’s Democratic (Republican) primary election.<sup>7</sup>

Beyond methodological limitations, a simple lack of survey data on primary election voters also impedes the estimation of primary electorate extremity. As previously noted, questions about primary voter participation are highly infrequent in national-level surveys. For example, the American National Elections Survey only asked about primary election turnout in 1958, 1964, 1966, and 1978. For those few surveys that *do* ask about primary voting, self-reported participation measures often over-report election turnout (Butler, 2009; Sides et al., 2018; Vavreck, 2007). Using validated voter turnout in the CCES resolves this problem, but drastically reduces the already small sub-samples of respondents nested within each party’s primary electorate for each congressional district.<sup>8</sup> Only about twenty percent of those surveyed by the CCES are validated primary voters. The median number of CCES respondents who voted in a party’s primary by congressional district is just over 10.<sup>9</sup> And

---

<sup>6</sup>Interested readers can refer to Leemann and Wasserfallen (2017) for a more detailed discussion of the two approaches, and their performance vis-a-vis alternative strategies.

<sup>7</sup>This approach requires making several assumptions about the “primary constituency.” First, that the types of voters who participate in the primary are consistent from year-to-year and, second, that the primary constituency is defined as only those people who voted in the primary. I address these concerns in greater detail in section A of the online appendix.

<sup>8</sup>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.

<sup>9</sup>For many primary constituencies represented in CCES data there are only a handful of respondents, a

although combining multiple surveys across elections allows for the estimation of primary electorate ideology at the national level (Sides et al., 2018), the small sample problem persists when disaggregating to partisan subconstituencies in each congressional district.<sup>10</sup>

Despite small sample sizes, estimating a model of ideology is still possible with survey data. Limited data at the constituency-level can be overcome by employing a hierarchical model with a random intercept for the partisan primary constituency in each congressional district (Park et al., 2004). This hierarchical intercept parameter allows for the partial pooling of information across partisan primary constituencies; meaning that data from those districts with an adequate number of respondents can be used to make predictions about districts for which there is an inadequate number of respondents. The very shrinkage that makes such predictions possible, however, induces bias that must then be corrected at the poststratification stage. Having a larger probability sample with more observations within constituencies would allow me to take advantage of the regularization afforded by such partial-pooling models without paying too high a price in terms of bias at the lowest levels of data aggregation.

Therefore, in lieu of survey data, I instead use Catalist LLC’s Validated Voter Database as a source of primary voter data. Catalist aggregates voter files for all 50 states and draws on external data sources to build individual profiles of voters and non-voters. Using separate samples for the 2012, 2014, and 2016 elections,<sup>11</sup> my sample contains over 9 million cases and approximately 2.7 million primary voters.<sup>12</sup> The Catalist voter file provides some clear advantages: primary election participation is verified, the sample size is large, and all

---

few have as many as 50, and some districts have no respondents.

<sup>10</sup>The pooling of surveys to create what is called a mega-poll across CCES years—similar to the approach used by Warshaw and Rodden (2012)—would not work for my purposes. My poststratification 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 election, would change from year to year and make this approach infeasible.

<sup>11</sup>In MrP, it is common practice to aggregate multiple data sources in order to produce a single point estimate. I produce separate estimates for each election because voter turnout varies across year.

<sup>12</sup>To ensure that Catalist voter profiles are complete, samples for a given election cycle should be taken in the year following the election. Yearly samples represent a random 1% of all Catalist records. In this paper the sample for the 2012 election was drawn in 2013, the sample for the 2014 election was drawn in 2015, and the sample for the 2016 election was drawn in 2017.

demographic variables common in survey data are present. Using Catalist data allows me to continue taking advantage of partial pooling through hierarchical modeling while also reducing the bias of primary constituency-level estimates substantially.

## Modeling Voter Ideology

Employing MrsP and Catalist data, I generate point estimates for the ideological extremity of each party’s primary electorate in congressional districts for the U.S. House of Representatives. To build these estimates I employ a hierarchical linear model. My model specification is similar to those adopted in previous studies examining voter ideological extremity (Hill and Tausanovitch, 2017; Sides et al., 2018). I regress an individual’s ideological extremity on a standard set of demographic characteristics including age, gender, education level, and race, which are provided by the aforementioned Catalist data.<sup>13</sup> My model specification departs from existing models of constituency extremity by including a variable for partisan primary election participation. This allows me to produce separate weighted estimates for Democratic and Republican Party primary electorates—a key innovation of this paper. I produce separate estimates for 2012, 2014, and 2016, along with an average estimate across years.<sup>14</sup> All predictors are modeled using random effects except party primary participation modeled using fixed effects.<sup>15</sup> I let my model intercept vary by congressional district and state. Random intercepts are drawn from a zero mean normal distribution, though the district level random intercept is drawn from a distribution centered on Democratic presidential vote share for that district.<sup>16</sup> The full model specification is outlined below.

---

<sup>13</sup>A full discussion of each independent variable is available in Section A of the Online Appendix.

<sup>14</sup>Because my estimation approach relies on voter turnout, if a race is uncontested I cannot produce an estimate for that primary electorate in that year. Combining estimates and taking an average across years produces a more complete set of partisan electorate estimates.

<sup>15</sup>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. The omitted category is no participation.

<sup>16</sup>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.

$$\begin{aligned}
Ideology_i &= \beta_0 + \alpha_{race[i]} + \alpha_{gender[i]} + \alpha_{edu[i]} + \beta_{demprimary[i]} + \beta_{repprimary[i]} \\
&\quad + \alpha_{age[i]} + \alpha_{state[i]} + \alpha_{district[i]} + \varepsilon[i] \\
\alpha. &\sim \mathcal{N}(0, \sigma^2) \\
\alpha_{district} &\sim \mathcal{N}(\gamma_{1PresVote}, \sigma_{district}^2) \\
\varepsilon[i] &\sim \mathcal{N}(0, \sigma_y)
\end{aligned}$$

To define voter extremity, I rely on Catalist’s predictive scores for individual ideology. This variable is constructed using more than 150 covariates and is generally accepted as a reliable measure for relative ideological extremity between individuals (Hersh, 2015). This synthetic score is scaled from 0-100 with 0 being the most conservative and 100 being the most liberal. To postratify my estimates, I rely principally on joint distributions of voter types (e.g. men who are college education that live in a given subnational unit) provided by the U.S. census. Data from census.gov provides joint distributions that include all but two of the individual-level predictors I specify in the model. For the two predictors not included—age and primary participation—I impute joint distributions using Leemann and Wasserfallen’s (2017) multilevel regression with synthetic postratification (MrsP) methodology. Marginal distributions for age are provided by the U.S. census. To define each district’s marginal distribution of Democratic (Republican) primary electorate members, I use total voter turnout for that district’s Democratic (Republican) primary election.<sup>17</sup> These voter turnout totals—which I employ as marginal distributions for primary participation—are provided by the American Votes book series.<sup>18</sup> Because I define the primary electorate as those who participated in the primary, estimates are only produced for those constituencies that had a contested partisan primary in 2012, 2014 and/or 2016.

---

<sup>17</sup>As previously noted, this approach requires making several assumptions about the “primary constituency.” I address these concerns in greater detail in section A of the online appendix.

<sup>18</sup>A complete discussion of independent variables for which joint distributions were imputed can be found in section A of the online appendix.

## Evaluating Estimate Validity

While I produce estimates of ideological extremity for all contested partisan primary elections from 2012 to 2016, determining the validity of these estimates presents a challenge. As Hill (2017) 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.” To demonstrate that my measure of partisan primary electorate extremity aligns with other measures for constituency ideology, I compare my estimates to a variety of other measures.

First, I replicate my estimation approach using data from the 2012 CCES and compare this measure with my Catalist estimates for primary electorate extremity. I produce these CCES estimates using the same set demographic covariates—age, gender, education, and race—which are self-reported. To determine whether or not respondents participated in their district’s primary election, the CCES validates voter turnout.<sup>19</sup> However, like Catalist, the CCES does not report in which primary a voter participated. I rely on self-reported party affiliation to determine partisan primary participation, placing respondents in the primary that matches their self-reported party.<sup>20</sup>

In the CCES, ideology is self-reported on a 7 point Likert scale from very conservative to very liberal. Because these ideological scores are likely reported with error, I employ estimates produced by Hill’s (2017) item-response theory (IRT) model for ideological conservatism to measure CCES respondents’ ideological leanings. The CCES regularly asked respondents how they would vote on a set of roll call votes that were considered by the U.S. House and Senate. Hill (2017) uses these expressed policy preferences to create a measure of conservatism, where a score of 0 is the ideological center, 4 is most conservative, and -3

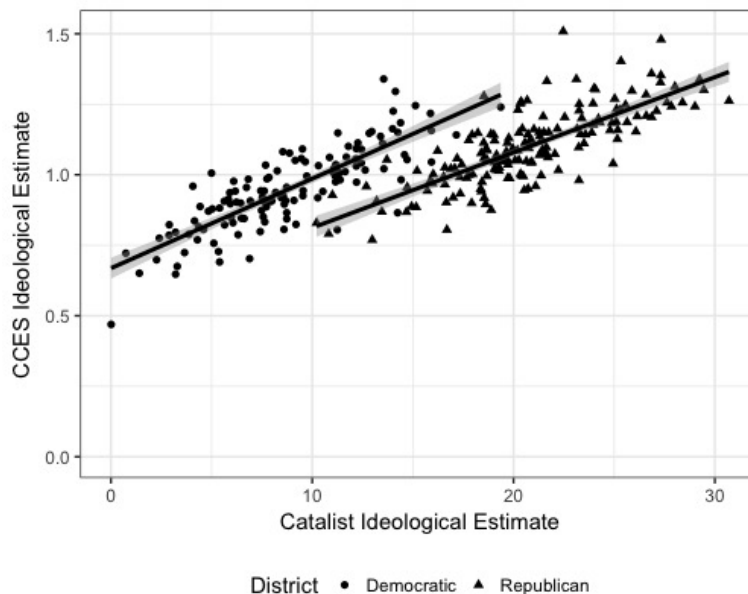
---

<sup>19</sup>This voter validation is performed using data from Catalist, LLC—the same data I to produce my estimates of voter ideological extremity.

<sup>20</sup>While this categorization decision could introduce bias into my estimates, Hill (2015) and Sides et al. (2018) 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 there may be variability in turnout, the types of voters participating in the primary should remain relatively consistent.

is most liberal. The clear advantage of these IRT estimates of ideology is that the summary value produced is constructed using identical schema for each respondent. In other words, Hill’s (2017) IRT model for ideology does not succumb to the same measurement error frequent in self-reported measures, where, for instance, a Likert score of “4” may not be conceptualized the same way across respondents. The same post-stratification technique (MrsP) outlined above is applied to the produced CCES estimates.

Figure 1: Catalist vs. CCES Estimates of Ideological Extremity



Plotted districts include only those where estimates were produced using both the CCES and Catalist data. Using CCES data produced 32% fewer estimates than Catalist data. For estimates produced using CCES data, only those districts with a sample size of ten or more were included. California, Washington, and Louisiana are excluded from this analysis. For Democratic districts, the correlation is .8262. For Republican districts, the correlation is .7787.

To plot ideological extremity rather than partisan extremity, both sets of estimates are transformed such that 0 is the most moderate and increasingly positive integers indicate greater ideological extremity. For estimates produced using CCES data, only those districts with a sample size of ten or more were included. Top-two and jungle primary states—California, Washington, and Louisiana—are excluded from this analysis because primary electorate composition is atypical. The CCES estimates are plotted against Catalist esti-

mates for primary electorate ideology in Figure 1. For Democratic districts, the correlation between Catalist and CCEs estimates is .8262. For Republican districts, the correlation between estimates is .7787. Figure 1 clearly demonstrates that there is a strong relationship between Catalist and CCES scores—where both sets of estimates are produced using identical variables and the same poststratification technique but different conceptualizations of ideological extremity. The preceding analysis provides some reassurance that Catalist’s predictive scores for individual ideology are a sound measure of ideological extremity, correlating well with estimates produced using a more traditional measure for ideology.

To further demonstrate the robustness of my estimates, I plot my yearly estimates for partisan district ideology (individual estimates for 2012, 2014, and 2016) against Democratic two-party presidential vote share for each election year. These plots are presented in section B of the online appendix. Correlation coefficients between Catalist district estimates and presidential vote share are above 0.70 across all election years. I further compare my average estimates of district ideological extremity with Warshaw and Tausanovitch’s district-level public preference estimates from the American Ideology Project. These non-partisan estimates of district extremity are based on responses to survey questions by 275,000 Americans who participated in the Annenberg National Election Study and the Cooperative Congressional Election Study. Similar to Hill (2017), Warshaw and Tausanovitch’s estimates are produced using an item-response theory (IRT) model and are weighted using U.S. census data. The correlation coefficients between my Catalist district estimates and Warshaw and Tausanovitch’s estimates—presented in appendix Figure 10—are, once again, above 0.70. This robustness exercise demonstrates that—at the very least—the face validity of my estimates for partisan primary constituency ideological extremity is strong.

## Descriptive Results

Much of our knowledge about the composition of the primary electorate comes from data on election outcomes (see Brady et al. 2007). While these discussions are valuable, they are

limited. The estimates I present in this paper provide new insights into primary constituencies by examining the ideologies of voters themselves. For that reason, I present descriptive features of my estimates in this section, demonstrating significant variation in primary electorate extremity within states and across parties. I further show that the overall distribution of primary electorate extremity remains relatively consistent across elections.

Aligning with my expectation, I find that the ideological extremity of primary electorates varies substantially between parties and across districts. To illustrate this heterogeneity, Figure 2 compares Democratic and Republican primary constituency extremity for congressional districts in North Carolina (2a, 2b), Indiana (2c, 2d), and Arizona (2e, 2f). For clearer comparison, average point estimates of ideological extremity are converted to percentiles, where districts are shaded to reflect extremity relative to the overall distribution of Democratic (Republican) primary electorates.<sup>21</sup> For both parties, more moderate districts are denoted by the lighter gradient and more extreme districts are denoted by the darker gradient. For example, North Carolina’s 12th district and Indiana’s 1st district—the darkest shaded districts in Figure 2 (a) and (c)—are in the 70th percentile of Democratic districts, making them some of the more extreme liberal primary constituencies in the country. This evaluation seems appropriate given that the NC-12 is a gerrymandered Democratic strong hold predominantly comprised of African Americans. Similarly, the IN-01 has not sent a Republican to Congress in 90 years, electing Obama in 2012 by a twenty-four point margin.

Figure 2 notably demonstrates instances of within state variation in same-party electorate extremity. Turning to Republican primary electorates in North Carolina, depicted in Figure 2 (b), there is little difference in electorate extremity. All Republican-controlled districts fall between the 70th and 80th percentiles of the overall distribution of Republican primary electorates. Conversely, the extremity of electorates in Arizona, presented in Figures 2 (e) and (f), are far less consistent. While Arizona’s 1st congressional district is represented by

---

<sup>21</sup>Percentiles are generated with an Empirical Cumulative Distribution Function (ECDF), where a distribution is produced for each party. Expressed percentiles are proportion of district ideological scores that are less than or equal to that district’s score on the partisan ECDF.



Figure 2: Primary Electorate Ideological Extremity in Partisan Electorates Across States



Ideological Extremity  
0.00 0.25 0.50 0.75 1.00

(a) NC - Democratic Primary Electorates



Ideological Extremity  
0.00 0.25 0.50 0.75 1.00

(b) NC - Republican Primary Electorates



Ideological Extremity  
0.00 0.25 0.50 0.75 1.00

(c) IN - Democratic Primary Electorates



Ideological Extremity  
0.00 0.25 0.50 0.75 1.00

(d) IN - Republican Primary Electorates



Ideological Extremity  
0.00 0.25 0.50 0.75 1.00

(e) AZ - Democratic Primary Electorates

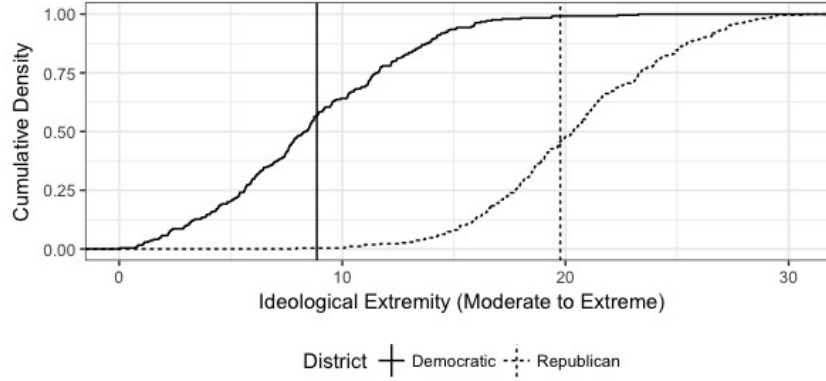


Ideological Extremity  
0.00 0.25 0.50 0.75 1.00

(f) AZ - Republican Primary Electorates

Gradient references extremity, with lighter districts being more moderate within the party and darker districts more extreme. White districts have no estimate due to an uncontested primary in 2012, 2014, and 2016. Estimates reflect partisan constituency ideology are pooled across all estimated years.

Figure 3: Empirical Cumulative Density Function of Electorate Ideological Extremity



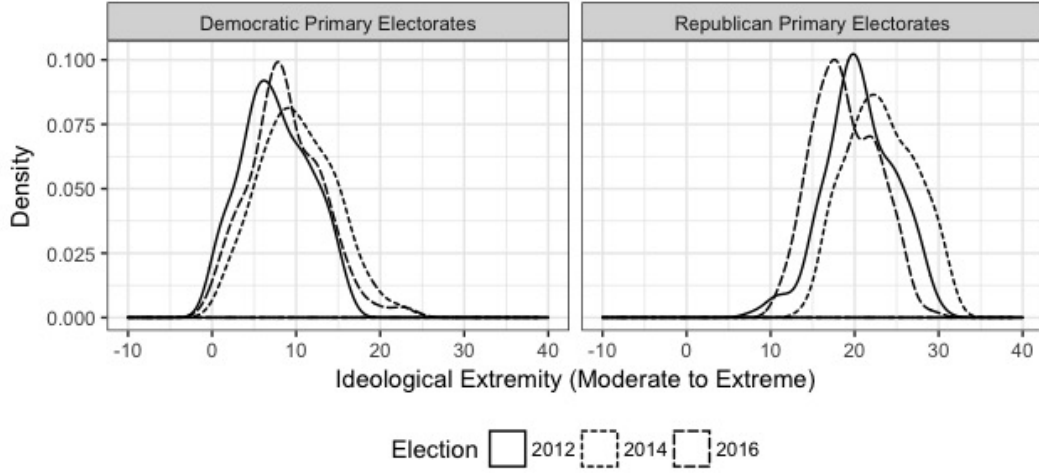
Catalist ideology scores are scaled from 0-100 with 0 being the most conservative, 100 being the most liberal, and 50 being exactly moderate. For comparison, estimates have been transformed such that moderate for *both* parties is 0 and most extreme is 50. Estimates reflect partisan constituency ideology are pooled across all estimated years (2012, 2014, and 2016).

a Democrat, this district narrowly favored the Republican presidential candidate in both the 2012 and 2016 general elections. Accordingly, the Democratic and Republican primary constituencies in AZ-01 skew moderate.<sup>22</sup> On the other hand, the AZ-04 and AZ-07 have far more extreme constituencies, both falling in the 70th and 77th percentile of the Republican and Democratic primary electorates, respectively.

Variation in the extremity between parties is explored nationwide in Figure 3, depicting an empirical cumulative distribution function of partisan primary electorate ideological extremity. Recall that Catalist scores are scaled from 0-100 with 0 being the most conservative and 100 being the most liberal. In Figure 3, these scores have been transformed such that moderate Democrat and Republican constituencies are centered at zero to allow for direct comparisons. Indicated by the vertical reference lines, the median Republican primary constituency is significantly more extreme than the median Democratic primary constituency. Divergence between the ECDFs indicates asymmetric polarization among primary voters—while both parties have a distribution of moderate and extreme electorates, Republican constituencies are skewed more extreme. Figures 2 and 3 clearly demonstrate that primary constituency extremity varies within and across states, as well as across parties.

<sup>22</sup>The Democratic primary constituency for AZ-01 is the 25th percentile and the Republican primary constituency is in the 55th percentile.

Figure 4: Primary Electorate Ideological Extremity Disaggregated by Year



Exploring the ideological extremity of primary electorates further, several noteworthy patterns emerge. In Figure 4, point estimates for primary electorate extremity are disaggregated by year to evaluate shifts in primary electorate ideology between presidential and midterm election years. Fenno (1978) and many other scholars suggest that primary electorate voters are among the most dedicated partisans. For that reason, the ideological distribution of partisan primary electorates should—in expectation—remain relatively consistent across time, with the same types of voters showing up year in and year out. However, it could also be the case that primary constituencies are generally more moderate in presidential election years due to higher voter turnout. This question has not yet been addressed at the congressional district level.<sup>23</sup> To conduct this comparison, Catalist ideological scores have once again been transformed such that a score of 0 is most moderate and a score of 50 is most ideologically extreme for both parties. Comparing the extent of overlap across all three distributions, I find that the ideological distribution of primary electorates remains relatively consistent across elections, with electorates in the 2014 midterm being slightly more extreme than those in presidential election years.

<sup>23</sup> Sides et al. (2018) investigate the ideological consistency of the national primary electorate and find consistency in the types of voters who participate in these primaries.

## Polarization & Primary Electorates

To demonstrate the utility of my estimates for partisan primary electorate ideology, I assess the relationship between primary electorate extremity and ideological challenger emergence as well as electorate extremity and incumbent legislative behavior. Turning first to candidate emergence, existing theories highlight the polarizing impacts of ideologically extreme primary electorates on the types of candidates who choose to run for Congress. Thomsen (2014, 2017) demonstrates that moderates increasingly opt-out of running for Congress because they see themselves as “out of step” with ideologically extreme primary electorates and, therefore, perceive their chances of winning as low. This trend may help to explain why so few Republican women choose to run, given that these candidates have been historically to the left of their male counterparts (Thomsen, 2015). Bolstering this finding, Hall and Snyder (2015) find that extreme primary candidates tend to receive more votes and are more likely to win their primary than are moderate candidates. If and when these extremist nominees reach the chamber, they contribute to an ongoing cycle of ideologically extreme member replacement, driving the parties farther apart (Bafumi and Herron, 2010; Theriault, 2012).

These analyses are predicated on the idea that ideologically extreme challengers are strategic and, therefore, tend to emerge in districts where they will fit the preferences of an ideologically extreme primary constituency. Such an assumption is less tenuous if partisan primary electorates are quite homogeneous in their extremity. My preceding analyses, however, demonstrate that partisan constituency extremity does indeed vary greatly across districts. Employing my novel measure for partisan constituency extremity, I more directly assess the strategic emergence decisions of extremist candidates in the following analysis.

To assess ideological primary candidate emergence in congressional elections, I fit a Poisson model where the dependent variable is a count of the ideologically extreme primary challengers who emerged in a given partisan primary election from 2012-2016. Independent variables include race-level characteristics that may influence candidate emergence such as primary type, redistricting, partisan seat-safety, and, of course, primary constituency ide-

ological extremity. Recall, Catalist’s ideological extremity measure varies from 0 (extreme conservative) to 100 (extreme liberal) with 50 indicating the ideological center. Following my descriptive analyses above, this measure has been transformed such that 0 indicates the most moderate partisan primary constituency for both parties and increasingly positive integers indicate greater ideological extremity.<sup>24</sup> Per work on strategic candidate emergence, politically sophisticated candidates are most likely to emerge when their chances of success are maximized (e.g. Jacobson and Kernell 1983; Cox and Katz 1996; Maestas et al. 2006; Jacobson and Carson 2016). The following analysis, therefore, examines candidate emergence in open races (i.e. district where the incumbent is not seeking reelection) because strategic and, moreover, successful challenges to sitting incumbents are rare.<sup>25</sup>

To measure the ideological extremity of primary election candidates, I employ Bonica’s (2014) CFscores. This measure uses millions of political contributions to estimate the ideology (liberal vs. conservative) of congressional candidates while also allowing for direct comparisons across actors. Bonica (2019) validates the predictive accuracy of his CFscores, demonstrating that they perform similarly to scaling roll call votes in legislative settings as a means to intuit ideology. This measure is centered around zero with negative integers indicating a liberal candidate and positive integers indicating a conservative candidate. To allow for comparison across parties, CFscore scores have been transformed such that 0 indicates a moderate candidate and increasingly positive integers indicate greater ideological extremity among both Democrats and Republicans. A CFscore was generated for 80% of all primary election candidates from 2012-2016 who garnered more than 5% of the vote-share in their party’s primary election. Furthermore, over 90% of primary election winners had a CFscore.

The results are presented in Table 1 with Democratic and Republican races in the left and right columns, respectively. The number of observations indicates that nearly 200 primary races for open seats occurred between 2012 and 2016. For Republican contests, fewer ideolog-

---

<sup>24</sup>To allow for the maximum number of races to be evaluated—similar to Figure 2—a simple average is taken across estimates created for the 2012, 2014, 2016 congressional primary elections.

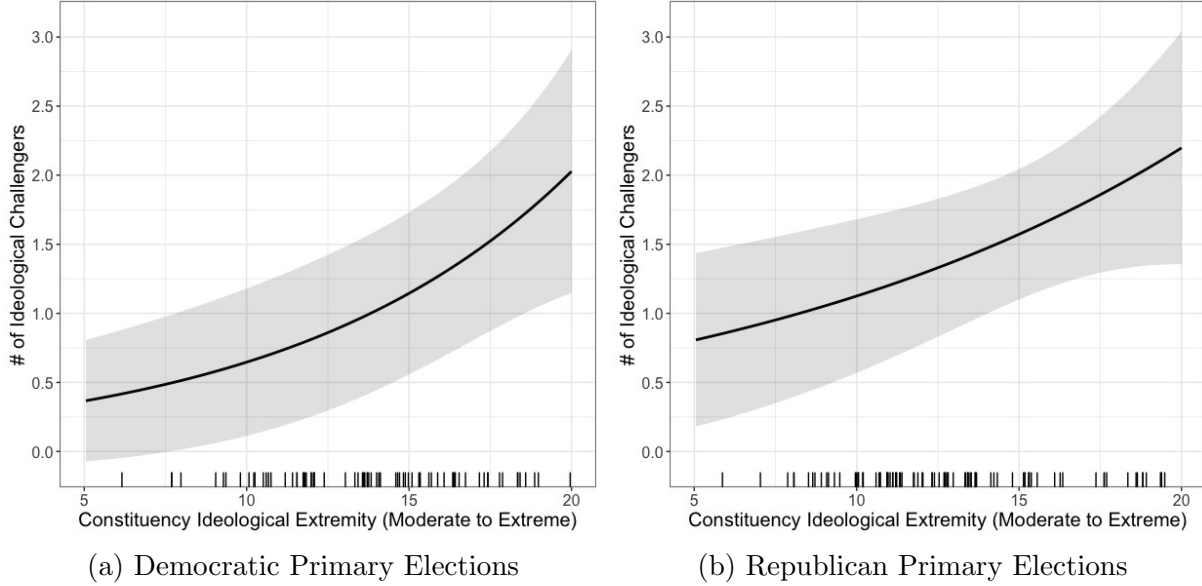
<sup>25</sup>Replicating this analysis with all races (against an incumbent or otherwise) produces substantively identical results. See Figure 11 and Table 2 in the online appendix.

Table 1: Ideologically Extreme Challenger Emergence in Primary Elections, 2012-2016

	<i>DV: Count of Ideological Challengers</i>	
	Democratic Races	Republican Races
Primary Type: Closed	−0.318 (0.118)	−0.890* (0.127)
Primary Type: Semi-Closed	−0.088 (0.318)	0.007 (0.262)
Redistricting	−0.368 (0.234)	0.190 (0.205)
Race Type: Partisan Safe-Seat	−0.341 (0.371)	0.309 (0.340)
Race Type: Two-Party Competitive	−0.169 (0.317)	−0.040 (0.389)
Primary Electorate Ideological Extremity	0.114* (0.042)	0.067* (0.033)
Constant	−1.237* (0.548)	−0.861* (0.480)
Observations	85	95
Log Likelihood	−105.864	−125.596
Akaike Inf. Crit.	225.729	265.192

*Note:* The dependent variable is a count of the number of ideological challengers running in a given partisan primary election from 2012-2016. Independent variables are race-level characteristics that could impact challenger emergence. Candidates are considered “ideologically extreme” if their CFscore is above the average ideological score for co-partisans across the time period of interest (1.266 for Republicans and 1.106 for Democrats). For my analysis, 106 out of 369 Republicans were labeled ideologically extreme and 91 out of 230 Democrats. Coefficient estimates are generated with 95% confidence intervals.

Figure 5: Predicted Count of Ideologically Extreme Challenger Emergence, 2012-2016



Predicted probabilities are generated using simulated data where all predictors are held constant at 0 and primary electorate ideological extremity is varied from the minimum to maximum observed value for each party. The x-axis is a normalized average of constituency ideological extremity where a score of 0 indicates “moderate” and increasingly positive integers indicate greater ideological extremity. The y-axis is a predicted count of the number of ideological challengers in simulated races. Predicted probabilities are generated with 95% confidence intervals.

ical challengers emerge in closed primaries—where only registered partisans can participate in primary election voting—than in contests with open institutions. Among all included variables, primary constituency ideological extremity is the only positive, statistically significant predictor for greater ideological challenger emergence.

Using predicted probabilities, I explore the association between of primary electorate extremity and ideological candidate emergence in more detail. Data is simulated such that all predictors are held at 0 and district extremity is varied in a sequence from the minimum to the maximum observed value for each party. Separate probabilities are generated for Democratic and Republican primary elections. Predicted counts are presented in Figure 5. For both parties, as district ideological extremity increases so does the count of the expected number of ideological challengers; moving from less than 1 to 2 or more for Democratic and Republican races. This is especially noteworthy given the distribution of constituency ideological extremity, displayed in the rug plots in Figure 5. Primary constituencies for

both parties skew ideological, indicating that more electorates than not have at least one ideological primary challenger. This expectation is bore out in the data, over 60% of primary elections have at least one ideological challenger.

Next, I assess if the relationship between members of Congress and primary constituents is conditional on electorate extremity, where more extreme primary electorates should elicit more extreme representation. Since 2010, the number of unopposed primary elections has dropped dramatically, demonstrating a shift towards greater intra-party competition. While incumbent defeats in the primary are still few and far between, noteworthy losses—like Eric Cantor (VA-07) in 2014 and Joe Crowley (NY-17) in 2018—signal to incumbents that winning their party’s nomination is not a guarantee. Theories on representation suggest that an incumbent will be highly responsive to the opinions of her district if she believes that it will win her votes (Erikson, 1978). By this logic, members of Congress have an incentive to better represent those constituents who can help them to get reelected (Fenno, 1978). Griffin and Newman (2005) and Clinton (2006) find evidence of this, demonstrating that member ideology aligns more closely with electorally valuable constituents—like voters and copartisans—than with other constituents in their district. If members perceive the primary electorate as increasingly important to their bid for reelection, then their legislative behavior may more closely align with the ideology of these voters.

Similar to Hall (2015), however, I anticipate that this connection may be mitigated by district safety. In districts that are safe for one party, members may be more representative of their primary voters because—with minimal competition in the general election—winning the nomination may be their only major obstacle to attaining reelection. In districts competitive for both parties, on the other hand, the general election is not a foregone conclusion. In these kinds of districts, incumbents have a greater incentive to represent the broader constituency in order to shore up their chances of defeating an out-partisan in the general. For this reason, I expect to observe a stronger, positive relationship between electorate extremity and member behavior in safe districts.



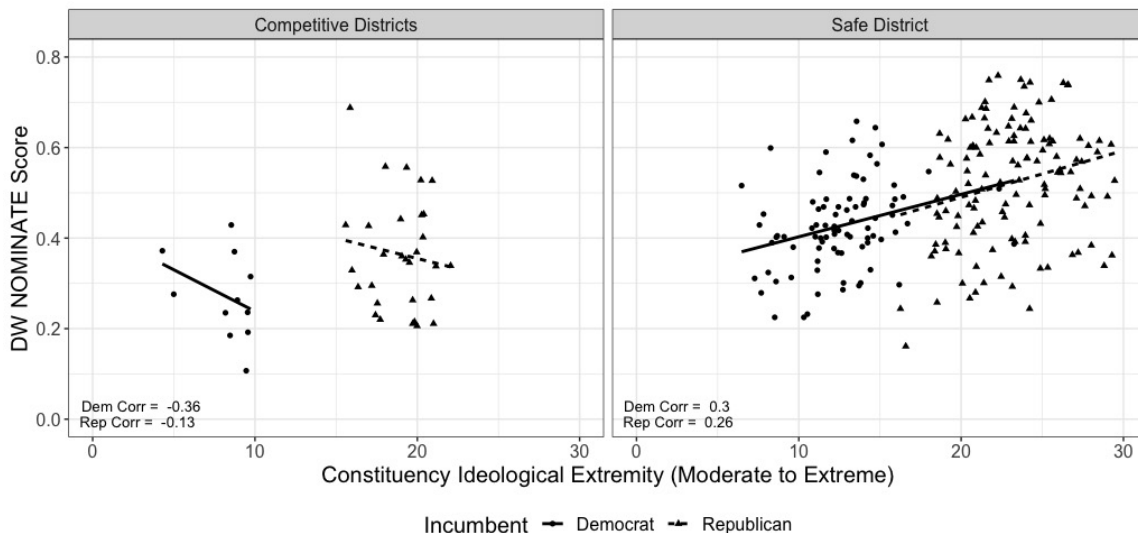


Figure 6: Primary Electorate Representation by Seat Safety in the 114th Congress

Plotted points are individual members of Congress. The y-axis is the absolute value of a member’s DW-NOMINATE score in the 114th Congress. The x-axis is the extremity of the incumbent’s same-party primary constituency. To allow for the maximum number of races to be evaluated a simple average is taken across estimates created for the 2012, 2014, 2016 congressional primary elections.

Figure 6 examines this connection between legislator voting and primary electorate ideology conditional on seat safety.<sup>26</sup> The right panel assesses this relationship in safe districts, which I define as congressional districts with a 55% or above same-party presidential vote share from 2012-2016, and the left panel assess this relationship in competitive districts, those district with a same-party presidential vote share below 55% and above 45% from 2012-2016.<sup>27</sup> I characterize member legislative behavior using the absolute value of NOMINATE scores. Confirming my expectations, the relationship between the primary electorate and member extremity is conditional on electorate extremity and district seat safety. For safe districts, there is a positive relationship between legislator NOMINATE scores and primary electorate extremity. The inverse is true in competitive district, where there is a negative relationship between legislator NOMINATE scores and primary electorate extremity.<sup>28</sup>

<sup>26</sup>I examine the relationship between members and the primary electorate that shares their partisanship (ie. the Republican primary electorate for a Republican member).

<sup>27</sup>Characterizing safe and competitive districts using Cook’s Political Report — a similarly restrictive measure — produces substantively similar results.

<sup>28</sup>The difference in correlation between safe and competitive districts is statistically significant.

## Discussion & Conclusion

Methodological and data-driven obstacles have previously impeded the development of ideological estimates for partisan subconstituencies. Without a clear measure for electorate extremity, scholars have been limited in their ability to examine the connection between primaries and polarization. Relying on validated voter data and an extension of MrP, this article presents a new measure for the ideological extremity of primary electorates in congressional districts. My measure improves on existing methods for estimating the ideology of primary electorates by (1) accounting for respondent partisanship in estimations of extremity, and (2) correcting for sampling bias in data on validated voter preferences.

The estimates presented here are not a perfect representation of primary electorate extremity but are rather a more direct, fine-grained measure. When employed by researchers, these estimates could serve as an independent or control variable in analyses.<sup>29</sup> Furthermore, this analysis serves as a template for other researchers who wish to pair postratification methods for sampling bias correction with data on partisan subpopulations.

These estimates open up the range of questions that Congress scholars can explore regarding the connection between primaries and polarization. To demonstrate the utility of my estimates, I show that primary electorate extremity is highly variable between parties and across districts. These findings suggest that future work should assess the polarizing influence of primary voters conditional on district-level extremity. Second, I illustrate one of the numerous potential applications of this new measure, demonstrating a connection between extreme primary voters and extreme representation conditional on district seat safety. This finding suggests that general election dynamics may, in part, contribute to congressional polarization by producing conditions that make members of Congress more responsive to voters who participate in primary elections.

---

<sup>29</sup>For election-specific research questions, disaggregated yearly estimates should be used rather than the aggregate point estimates. While estimates have been produced for top-two primary states, caution should be taken in when using the estimates in these applications. Because top-two primaries are not partisan, these estimates draw on a data generating process that is fundamentally different than a traditional primary.

## References

- Bafumi, J. and Herron, M. C. (2010). Leapfrog representation and extremism: A study of american voters and their members in congress. *American Political Science Review*, 104(3):519–542.
- Bartels, L. M. (2016). *Unequal democracy: The political economy of the new gilded age*. Princeton University Press.
- Brady, D. W., Han, H., and Pope, J. C. (2007). Primary elections and candidate ideology: Out of step with the primary electorate? *Legislative Studies Quarterly*, 32(1):79–105.
- Butler, D. M. (2009). The effect of the size of voting blocs on incumbents’ rollcall voting and the asymmetric polarization of congress. *Legislative Studies Quarterly*, 34(3):297–318.
- Clinton, J. (2006). Representation in congress: constituents and roll calls in the 106th house. *The Journal of Politics*, 68(2).
- Cox, G. W. and Katz, J. N. (1996). Why did the incumbency advantage in u.s. house elections grow? *American Journal of Political Science*, 40(2):478–497.
- Erikson, R. S. (1978). Constituency opinion and congressional behavior: A reexamination of the miller-stokes representation data. *American Journal of Political Science*, pages 511–535.
- Fenno, R. F. (1978). *Home Style: House Members in Their Districts*. Longman, New York.
- Fleisher, R. and Bond, J. R. (2004). The shrinking middle in the us congress. *British Journal of Political Science*, 34(3):429–451.
- Gelman, A. and Little, T. C. (1997). Poststratification postratification into many categories using hierarchical logistic regression. *Survey Research*, 23(2):127–35.
- Griffin, J. D. and Newman, B. (2005). Are voters better represented? *Journal of Politics*, 67(4):1206–1227.

- Hall, A. B. (2015). What happens when extremists win primaries? *American Political Science Review*, 109(1):18–42.
- Hall, A. B. and Snyder, J. M. (2015). Candidate ideology and electoral success.
- Hersh, E. D. (2015). *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press.
- Hetherington, M. J., Larson, B., and Globetti, S. (2003). The redistricting cycle and strategic candidate decisions in u.s. house races. *Journal of Politics*, 65(4):1221–1234.
- Hill, S. J. (2015). Institution of nomination and the policy ideology of primary electorates. *Quarterly Journal of Political Science*, 10(4).
- Hill, S. J. and Tausanovitch, C. (2017). Southern realignment, party sorting, and the polarization of american primary electorates, 1958-2012. *Public Choice*, (1-26).
- Hirano, S., Jr., J. M. S., Ansolabehere, S., and Hansen, J. M. (2010). Primary elections and partisan polarization in the u.s. congress. *Quarterly Journal of Political Science*, 5.
- Jacobson, G. C. (1989). Strategic politicians and the dynamics of u.s. house elections, 1946-1986. *American Political Science Review*, 83(3):773–93.
- Jacobson, G. C. and Carson, J. L. (2016). *The Politics of Congressional Elections*. Longman, New York, 7th edition edition.
- Jacobson, G. C. and Kernell, S. (1983). *Strategy and Choice in Congressional Elections*. Yale University Press, New Haven, 2nd edition edition.
- Jewitt, C. E. and Treul, S. A. (2018). *Routledge Handbook on Primary Elections*, chapter Ideological Primaries and their Influence in Congress. Routledge.
- Kousser, T., Lewis, J. B., and Masket, S. E. (2007). Ideological adaptation? the survival instinct of threatened legislators. *Journal of Politics*, 69(3):828–843.

- Lawless, J. L. and Pearson, K. (2008). The primary reason for women’s underrepresentation? reevaluating the conventional wisdom. *Journal of Politics*, 70(1):67–82.
- Lax, J. R. and Phillips, J. H. (2009). How should we estimate public opinion in the states? *American Journal of Political Science*, 53(1):107–21.
- Leemann, L. and Wasserfallen, F. (2017). Extending the use and prediction precision of subnational public opinion estimation. *American Journal of Political Science*, 61(4):1003–1022.
- Maestas, C. D., Fulton, S. A., Maisel, L. S., and Stone, W. J. (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, E., Masket, S., Shor, B., Rogers, S., and McCarty, N. (2014). A primary cause of partisanship? nomination systems and legislator ideology. *American Journal of Political Science*, 58(2).
- Miller, W. E. and Stokes, D. E. (1963). Constituency influence in congress. *American Political Science Review*, 1(57):45–56.
- Park, D. K., Gelman, A., and Bafumi, J. (2004). Bayesian multilevel estimation with post-stratification: State-level estimates from national polls. *Political Analysis*, 12(4):375–85.
- Porter, R. and Treul, S. A. (2019). Inexperience and success in congressional primaries. <https://rachelporter.org/wp-content/uploads/2019/09/PorterTreul1919.pdf>.
- Rabinowitz, G. and Macdonald, S. E. (1989). A directional theory of issue voting. *American Political Science Review*, 83(1):93–121.
- Sides, J., Tausanovitch, C., Vavreck, L., and Warshaw, C. (2018). On the representativeness of primary electorates. *British Journal of Political Science*.
- Sulkin, T. (2005). *Issue Politics in Congress*. Cambridge University Press.

- Tausanovitch, C. and Warshaw, C. (2013). Measuring constituent policy preferences in congress, state legislatures, and cities. *Journal of Politics*, 75(2):330–342.
- Theriault, S. (2012). *Party Polarization in Congress*. Cambridge University Press.
- Thomsen, D. M. (2014). Ideological moderates won’t run: How party fit matters for partisan polarization in congress. *Journal of Politics*, 76(3).
- Thomsen, D. M. (2015). Why so few (republican) women? explaining the partisan imbalance of women in the u.s. congress. *Legislative Studies Quarterly*, 40(2):295–323.
- Thomsen, D. M. (2017). *Opting Out of Congress: Partisan Polarization and the Decline of Moderate Candidates*. Cambridge University Press.
- Vavreck, L. (2007). The exaggerated effects of advertising on turnout: The dangers of self-reports. *Quarterly Journal of Political Science*, 2(4):325–343.
- Warshaw, C. and Rodden, J. (2012). How should we measure district-level public opinion on individual issues. *Journal of Politics*, 74(1):203–19.

Online Appendix:  
Estimating the Ideology of Congressional Primary  
Electoralates

## A Primary Electorate Extremity Model

$$\begin{aligned} Ideology_i &= \beta_0 + \alpha_{race[i]} + \alpha_{gender[i]} + \alpha_{edu[i]} + \beta_{demprimary[i]} + \beta_{repprimary[i]} \\ &\quad + \alpha_{age[i]} + \alpha_{state[i]} + \alpha_{district[i]} + \varepsilon[i] \\ \alpha. &\sim \mathcal{N}(0, \sigma^2) \\ \alpha_{district} &\sim \mathcal{N}(\gamma_{1PresVote}, \sigma_{district}^2) \\ \varepsilon_{[i]} &\sim \mathcal{N}(0, \sigma_y) \end{aligned}$$

Predictors of interest include race, gender, education, primary participation, age, voter state, and voter district. Variables are indexed by individual ( $i$ ) and congressional district ( $district$ ). All predictors are modeled using random effects ( $\alpha$ ) except party primary participation modeled using fixed effects ( $\beta$ ). 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. 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 Democratic presidential vote share for that district. 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.

### Independent Variables with Known Joint Distributions

#### *Gender*

This dummy variable equals 1 if the respondent is female, 0 otherwise. Data on respondent gender provided by Catalist is 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 Bachelor's Degree. Individuals are coded for having a bachelor's degree if their propensity score is greater or equal to 50.

#### *Race*

The race variable provided by Catalist, LLC—which is drawn from state voter files—includes more detailed race and ethnicity categories than those provided in U.S. census data. In order to weight model predictions in the poststratification stage of MrsP, the race and ethnicity Catalist categories are binned to match those in the census data. Race categories include Caucasian, Black, Hispanic, and Other.<sup>30</sup>

---

<sup>30</sup>This other categories matches the Other race category in the census; it includes Asian Native-American / Pacific Islander.



## Independent Variables with Imputed Joint Distributions

### *Age*

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. However, as previously stated, the U.S. census factfinder does not include age in reported joint distributions. To produce an adjusted synthetic joint distributions for age, I use the marginal distribution for age provided by the U.S. census.

### *Primary Election Participation*

Catalist provides verified information on individual turnout from state voter files. These voter files, however, do not always specify in *which* party’s primary a voter participated. Data availability on partisan primary participation varies with each state’s type of primary electoral 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 can assign voters to the partisan primary constituency matching their party registration.

For independents in semi-closed systems and all voters in open systems, it is impossible to 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.

The census provides no partisan or voter participation information in open-access data files, therefore I cannot use data from the U.S. census to weight my estimates. 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. 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. Second, voter turnout in elections fluctuates year-to-year, therefore the marginal distribution for primary election voters fluctuates year-to-year. While this could introduce bias into my estimates, Hill (2015) and Sides et al. (2018) 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 there may be variability in turnout, the types of voters participating in the primary should remain relatively consistent. Finally, defining primary election constituencies as only those people who voted requires making an assumption about who candidates pay attention to when campaigning. While a representative’s electorate includes all individuals in the district—voters and non-

voters alike—incumbent attention allocation and responsiveness has been widely explored in the literature. A bevy of finding show that representatives are more responsive to partisan voters than non-voters (Fenno, 1978; Clinton, 2006; Bafumi and Herron, 2010; Bartels, 2016).

## B Model Robustness

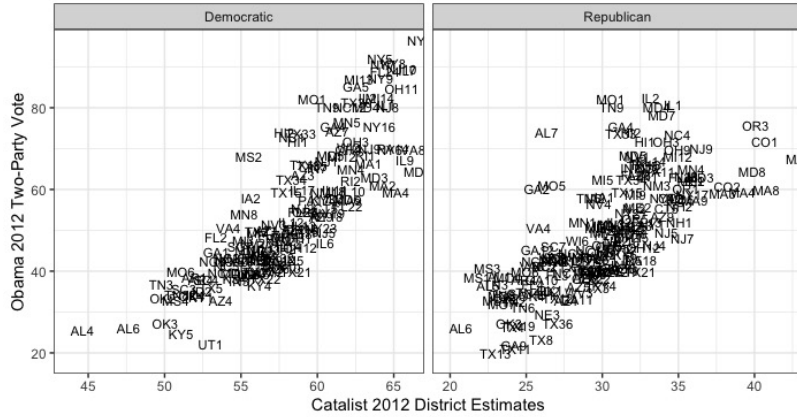


Figure 7: Face validity of Catalyst Estimates for Primary Electorate Ideological Extremity  
*2012 Catalyst Estimates vs. 2012 Dem. Presidential Vote Share*

Plotted districts include only those where Catalyst estimates were produced for the 2012 election year. The x-axis is the post-stratified 2012 electorate ideology, the y-axis is Democratic presidential two-party vote share. For Democratic districts, the correlation is 0.845. For Republican districts, the correlation is 0.715.

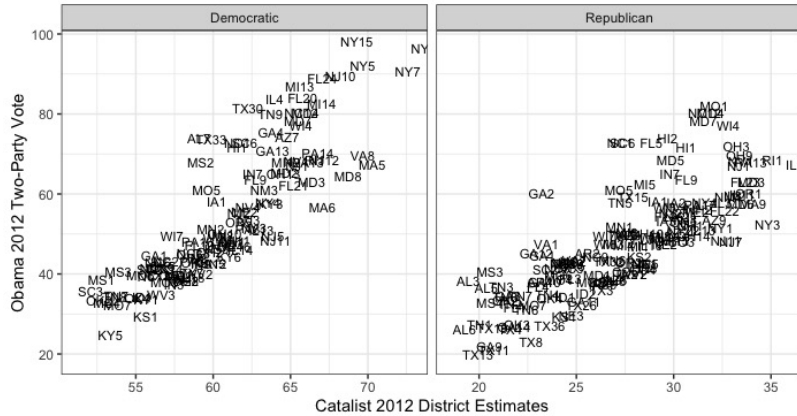


Figure 8: Face validity of Catalyst Estimates for Primary Electorate Ideological Extremity  
*2014 Catalyst Estimates vs. 2012 Dem. Presidential Vote Share*

Plotted districts include only those where Catalyst estimates were produced for the 2014 election year. The x-axis is the post-stratified 2012 electorate ideology, the y-axis is Democratic presidential two-party vote share. For Democratic districts, the correlation is 0.849. For Republican districts, the correlation is 0.789.

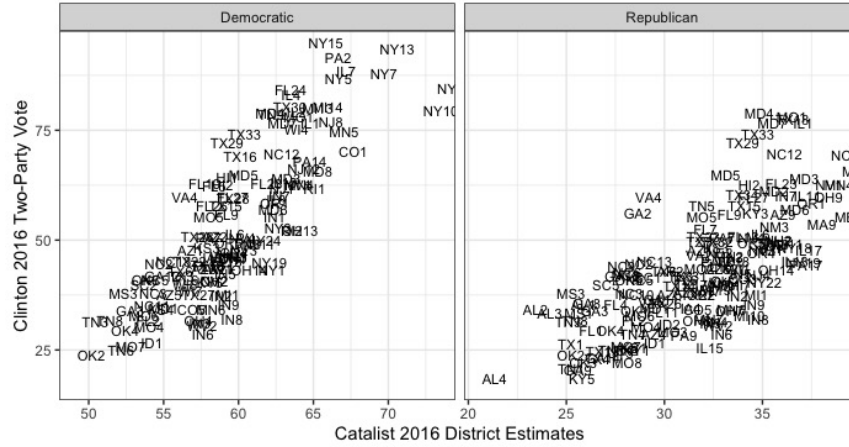


Figure 9: Face validity of Catalist Estimates for Primary Electorate Ideological Extremity  
*2016 Catalist Estimates vs. 2016 Dem. Presidential Vote Share*

Plotted districts include only those where Catalist estimates were produced for the 2016 election year. The x-axis is the post-stratified 2012 electorate ideology, the y-axis is Democratic presidential two-party vote share. For Democratic districts, the correlation is 0.830. For Republican districts, the correlation is 0.707.

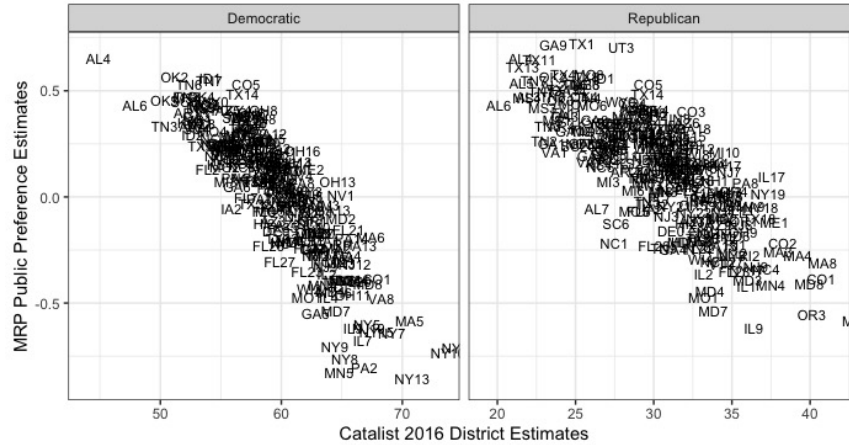
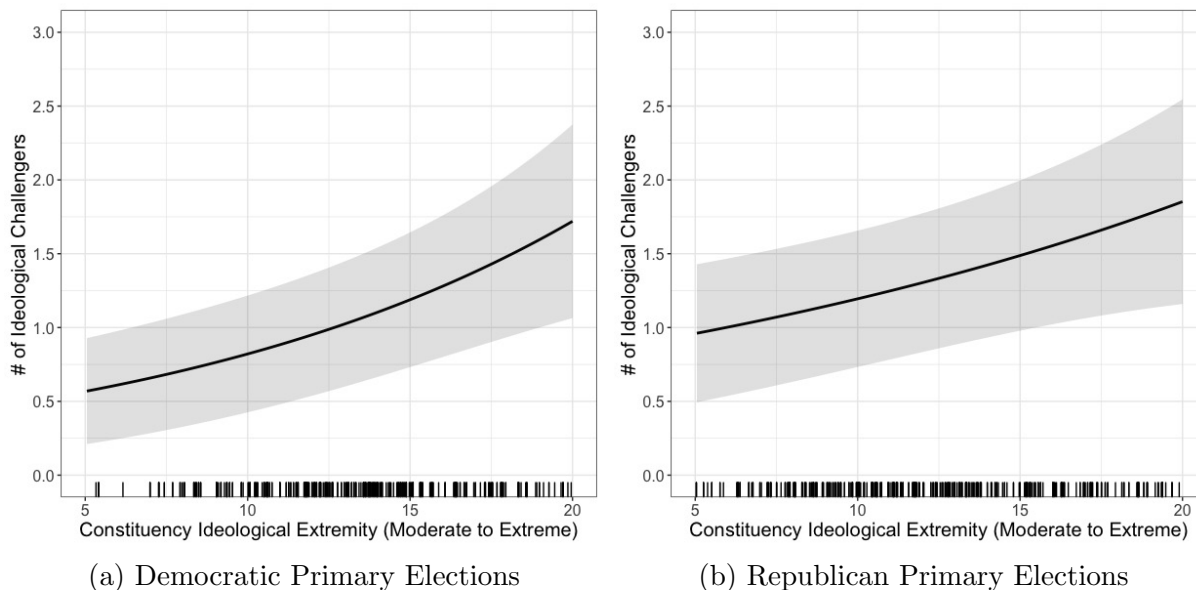


Figure 10: Face validity of Catalist Estimates for Primary Electorate Ideological Extremity  
*Average Electorate Estimate vs. MrP Public Preference Estimates*

Plotted districts estimates are the average ideological extremity of partisan primary elections across the 2012, 2014, and 2016 elections. The x-axis is the post-stratified average electorate ideology, the y-axis is MrP public preference estimates produced by Warshaw and Tausanovitch for the *American Ideology Project*. For Democratic districts, the correlation is -0.830. For Republican districts, the correlation is -0.707.

## C Alternative Specification: Candidate Emergence

Figure 11: Predicted Count of Extreme Challenger Emergence: All Races, 2012-2016



Predicted probabilities are generated using simulated data where all predictors are held constant at 0 and primary electorate ideological extremity is varied from the minimum to maximum observed value for each party. The x-axis is a normalized average of constituency ideological extremity where a score of 0 indicates “moderate” and increasingly positive integers indicate greater ideological extremity. The y-axis is a predicted count of the number of ideological challengers in simulated races. Predicted probabilities are generated with 95% confidence intervals.

Table 2: Ideologically Extreme Challenger Emergence: All Races, 2012-2016

	<i>DV: Count of Ideological Challengers</i>	
	Democratic Races	Republican Races
Primary Type: Closed	−0.088 (0.121)	−0.529* (0.129)
Primary Type: Semi-Closed	−0.075 (0.153)	−0.200 (0.146)
Redistricting	−0.066 (0.109)	0.066 (0.102)
Race Type: Partisan Safe-Seat	−0.069 (0.217)	0.284 (0.193)
Race Type: Two-Party Competitive	−0.121 (0.140)	0.134 (0.171)
Race-Type: Incumbent in Same Primary	−1.166* (0.238)	−0.405* (0.173)
Race-Type: Open Seat	−0.061 (0.141)	0.016 (0.163)
Primary Electorate Ideological Extremity	0.074* (0.021)	0.044* (0.017)
Constant	−0.868* (0.258)	−0.545* (0.208)
Observations	440	475
Log Likelihood	−489.115	−544.054
Akaike Inf. Crit.	996.230	1,106.108

*Note:* The dependent variable is a count of the number of ideological challengers running in a given partisan primary election from 2012-2016. Incumbents are not included in challenger counts. Independent variables are race-level characteristics that could impact challenger emergence. Candidates are considered “ideologically extreme” if their CFscore is above the average ideological score for co-partisans across the time period of interest (1.266 for Republicans and 1.106 for Democrats). For my analysis, 413 out of 955 Republicans were labeled ideologically extreme and 91 out of 230 Democrats. Coefficient estimates are generated with 95% confidence intervals.