# Online Appendix: Estimating the Ideology of Congressional Primary Electorates

# A Primary Electorate Extremity Model

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

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.[28]

[28]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

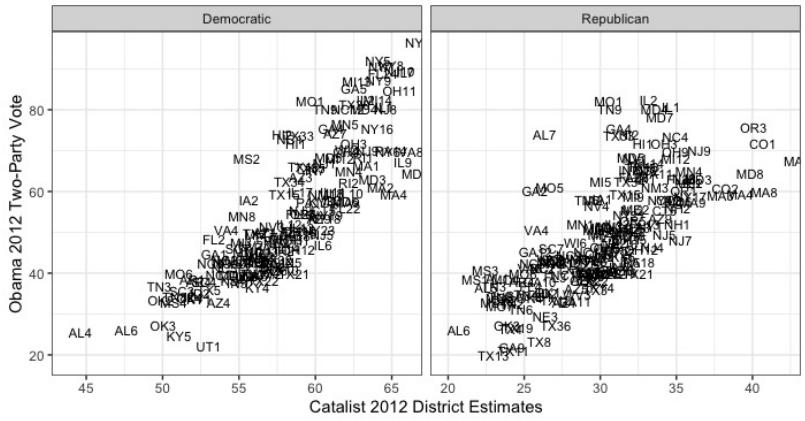


Figure 7: Face validity of Catalist Estimates for Primary Electorate Ideological Extremity
*2012 Catalist Estimates vs. 2012 Dem. Presidential Vote Share*

Plotted districts include only those where Catalist 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.

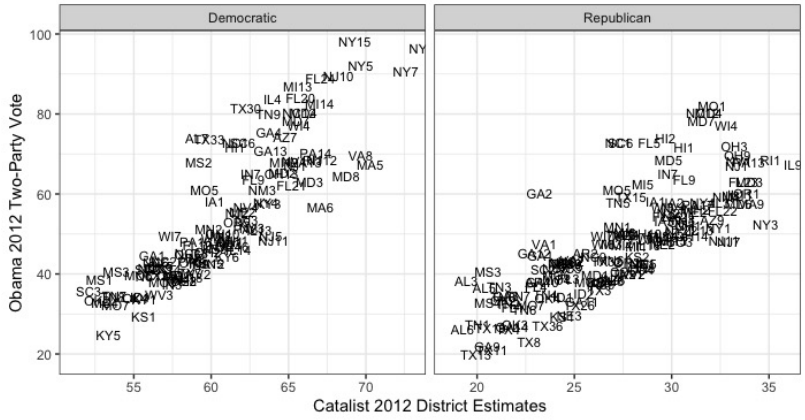


Figure 8: Face validity of Catalist Estimates for Primary Electorate Ideological Extremity
*2014 Catalist Estimates vs. 2012 Dem. Presidential Vote Share*

Plotted districts include only those where Catalist 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.

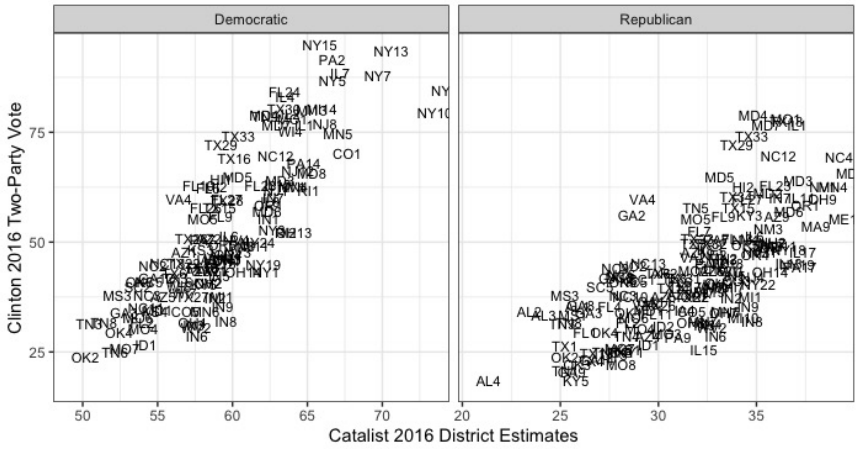


Figure 9: Face validity of Catalist Estimates for Primary Electorate Ideological Extremity
*2016 Catalist Estimates vs. 2016 Dem. Presidential Vote Share*

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.

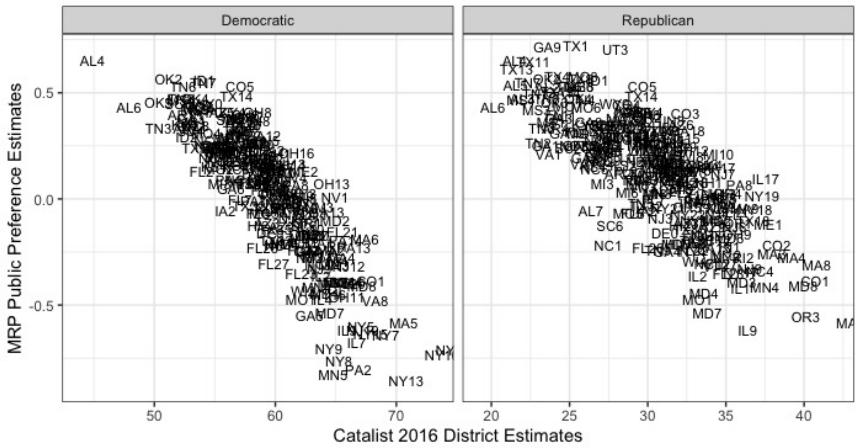


Figure 10: Face validity of Catalist Estimates for Primary Electorate Ideological Extremity
*Average Electorate Estimate vs. MrP Public Preference Estimates*

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.