Supplementary Appendix: On the Representativeness of Primary Electorates

Monday 24th July, 2017

Appendix A: Ideal Point Scale

In order to measure citizens' ideal points, we need surveys with information on voters' policy preferences. To this end, we pool the 2008, 2010, 2012, and 2014 Cooperative Congressional Election Surveys.[1] Each of these surveys asked between 14 and 32 policy questions. This enables us to jointly scale each respondent's ideal point using an approach similar to that of Tausanovitch and Warshaw (2013). We measure the ideal points of the 2008 CCAP respondents separately due to the lack of overlap in their policy questions with the CCES.

To estimate voters' ideological positions, we assume that all survey respondents have a quadratic utility function with normal errors (Clinton, Jackman, and Rivers, 2004). Each item presents individuals with a choice between a "Yes" position and a "No" position.[2] We use the two-parameter IRT model introduced to political science by Clinton, Jackman, and Rivers (2004), which characterizes each response $y_{ij} \in \{0, 1\}$ as a function of subject i's latent *ability* (x_i), the *difficulty* (α_j) and *discrimination* (β_j) of item j, and an error term (e_{ij}), where

$$\Pr[y_{ij} = 1] = \Phi(\beta_j x_i - \alpha_j) \tag{1}$$

where Φ is the standard normal CDF. β_j is referred to as the "discrimination" parameter because it captures the degree to which the latent trait affects the probability of a yes answer. The "cut point" is the value of α_j / β_j at which the probabilities of answering yes or no to a question are 50-50. We assume a one-dimensional policy space because a two-dimensional model shows little improvement in terms of model fit. The ideal point, x, for individual i signifies the "liberalness" or "conservativeness" of that individual. We orient our values so that lower values are associated with more liberal preferences and higher values with more conservative preferences. We approximate the joint posterior density of the model

[1] The 2008 CCES has 32,800 respondents, the 2010 CCES has 55,400 respondents, the 2012 CCES has 54,535 respondents, and the 2014 CCES has 56,200 respondents.

[2] We dichotomize each of the survey questions for our scaling model. For example, if a question asks whether respondents "agree" or "strongly agree" to a statement, both responses would be coded simply as "Yes."

Table 1: Symbolic Ideology and Citizen Ideal Points

Symbolic Ideology	Mean Ideal Point
Very Liberal	-1.30
Liberal	-1.03
Moderate	-0.31
Conserative	.83
Very Conservative	1.34

parameters using a Markov chain Monte Carlo (MCMC) method. Non-response is handled straightforwardly in MCMC: if a question is not answered, then that question is effectively dropped since it does not inform the respondent's ideal point. To validate our estimates, Table 1 shows the strong relationship between symbolic ideology and our scaled measure of citizens' ideal points.

Appendix B: Voter File Matching Process

We use data from five large-scale surveys of the American public: the 2008 Cooperative Campaign Analysis Project (CCAP) (Jackman and Vavreck, 2009) and the 2008-2014 CCES (Vavreck and Rivers, 2008; Ansolabehere and Rivers, 2013). Once sample weights are applied, CCAP respondents are representative of registered voters and CCES respondents are representative of the American public.

The matching of our surveys to voter files was conducted by the survey provider YouGov in conjunction with the voter file firm Catalist (see Ansolabehere and Hersh, 2012, for a detailed description of this process). Table 2 illustrates the results of the matching process, focusing on the 2008 CCAP data. Catalist was able to match 16,792 CCAP respondents (84%) to a state voter file.[3] Six hundred of the respondents (3%) were confirmed as unregistered. The remainder, 2,608 (13%), could not be matched by Catalist to any record on a voter file nor a record on a consumer file. These respondents may or may not be registered to vote. Catalist's inability to match them does not necessarily mean they are unregistered, only unverified.[4] The validated turnout rate in the general election of the CCAP registered voter sample was 68%. The primary turnout rate was 48%. (These percentages are calculated by leaving the unmatched respondents in the denominator and classifying them as not having voted in the 2008 election.)[5]

[3] State parties that nominate candidates using a caucus instead of a primary may not report participation data for the caucus to Secretary of State's offices. The matching process between YouGov and Catalist did return some validated data in some caucus states, but we believe that this data generally does not represent turnout in presidential caucuses. Rather, it may represent turnout in congressional primaries in those states.

[4] None of the CCAP respondents living in Nevada were matched to the voter file because Catalist did not have access to a Nevada voter file at that time. Virginia also does not make its voter file available except for a fee, and thus respondents living in Virginia were not matched to the file. Overall, the rate of successfully matching respondents to the voter file varied across states. In Mississippi, known to have one of the least advanced voter files, 67 of 100 CCAP panelists were found on the file. Similar rates obtained in Wyoming (21 out of 31, or 67.7%), the District of Columbia (29 out of 40, 72.5%) and Alaska (38 out of 52, or 73.1%). States with high rates of matching include the Dakotas (SD: 46 out of 51 respondents, or 90.2%; ND, 40 out of 43 respondents, 93%) and Montana (53 out of 55 respondents, 96.4%). That said, most states' rates clustered around the 84% average. We thus do not believe that differences in match rates across states affect our inferences about the characteristics of primary and general electorates.

[5] Similar results obtained when matching CCES respondents to the voter file. In 2008, Catalist was able to match 25,381 (77%) CCES respondents to one of their state voter files (Note that similarly to the CCAP,

Most states indicate which primary a person voted in, but some do not. Among the 68% that voted in both the primary and general election are 1,714 respondents (19% of this group) for whom we have no way of knowing in which party primary they voted. In cases where the state voter files require party registration, we allocated people in states with no party primary indicator on the file to the primary for the party in which they were registered. Even after this allocation, we are left with 1,714 respondents who we cannot classify into one or the other of the primaries.

	General Election			
Primary Election	Registered and voted	Registered and did not vote	Not matched to a voter file	Verified unregistered
Registered and voted	9,119 (68%)	488 (17%)	0	0
Registered and did not vote	4,389 (32%)	2,433 (83%)	0	0
Not matched to a voter file	0	0	2,833 (100%)	0
Verified unregistered				738 (100%)
TOTAL	13,508	2,921	2,833	738

Cell entries are the weighted number of respondents with column percentages in parentheses.

Table 2: Validated Turnout in 2008 Primary and General Election, CCAP Registered Voters

Unsurprisingly, the validated turnout rates for both the primary and general election in the CCAP survey are much lower than the corresponding self-reported turnout rates.[6] For instance, 57% of the respondents in the 2008 CCAP reported that they voted in a primary (either when asked just after their state's primary or when asked in September to recall whether they had voted in the primary). The validated vote data, however, indicate that of those who reported turning out in a primary, only 60% or so actually did, according to state

Virginia does not make its voter file available except for a fee, and thus respondents living in Virginia were not matched to the CCES file.) The validated turnout rate in the general election of the 2008 CCES sample, according to Catalist, was 62%. The primary turnout rate was 40%. The slightly higher turnout rates in the 2008 CCAP compared to the 2008 CCES likely reflects the fact that the CCAP had a registered voter sample.

[6] Self-reported turnout in primary is only inconsistently available for the CCES. For instance, self-reports on primary turnout are not available in 2012 or 2014. Therefore, the comparisons of self-reported and validated turnout in this section focus on the CCAP.

records. This raises an important question about previous literature. The earlier studies of the 1976 and 1980 elections (Geer, 1988; Norrander, 1989) rely on data about verified voters—via exit polls and validated vote measures in the ANES, respectively. Jacobson (2012) relies on self-reported turnout, as do some other studies (Butler, 2009; Peress, 2013). Jacobson acknowledges that, as in many surveys in which turnout is assessed via self-reports, respondents to the 2010 CCES over-reported turnout (Vavreck, 2007). But he argues that "comparisons across participation categories remain informative" (1615). However, introducing validated voting data can indeed alter comparisons across categories of participation. As Ansolabehere and Hersh (2011) show in their study of general election voters and non-voters, using validated turnout data reduces the (already small) ideological gaps between these groups.

We can also show the importance of using validated turnout in making these comparisons. In Table 3, we focus on the differences between the views of primary voters and general election-only voters on the six policy questions in the 2008 CCAP—calculated in percentage points. We present the differences separately for each party and using both validated and self-reported turnout in the primary and general.

	Democrats		Republicans	
	Validated	Self-reported	Validated	Self-reported
Arrest, deport illegal immigrants	-1	8	4	20
Support gov. health insurance	4	13	-7	-2
Withdraw from Iraq immediately	-1	10	-3	0
Raise taxes on incomes $200K+	0	19	-11	-1
Abortion always legal	6	5	-1	-3
Abortion legal in special cases	-4	-1	3	5
Support gay marriage	-2	-2	2	-4

Table 3: Differences in Support between Primary and General Election-Only Voters (Primary-General), Comparing Validated and Self-Reported Vote (2008 CCAP).

Using self-reported turnout usually exaggerates the difference between these two groups, much as it exaggerates the difference between general election voters and non-voters. For example, among Republicans validated primary voters are only 4 points more supportive

of deporting illegal immigrants, compared to those who voted only in the general election. Among self-reported voters, that difference balloons to 20 points. The same pattern obtains among Democrats on several issues, including support for government health insurance, withdrawal from Iraq, and raising taxes on the wealthy.

These differences between validated and self-reported turnout may be one reason for the differences in the findings of Geer (1988) and Norrander (1989) on the one hand, and Jacobson (2012) on the other. More generally, these differences suggest that self-reported turnout data are problematic for comparing primary and general electorates and artificially inflate the level of polarization.

References

Ansolabehere, Stephen, and Douglas Rivers. 2013. "Cooperative survey research." Annual Review of Political Science 16: 307–329.

Ansolabehere, Stephen, and Eitan Hersh. 2011. "Who Really Votes?" Facing the Challenge of Democracy: Explorations in the Analysis of Public Opinion and Political Participation pp. 267–291.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." Political Analysis 20(4): 437–459.

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.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." American Political Science Review 98(2): 355–370.

Geer, John G. 1988. "Assessing the Representativeness of Electorates in Presidential Primaries." American Journal of Political Science 32(4): 929–945.

Jackman, Simon, and Lynn Vavreck. 2009. "The 2008 Cooperative Campaign Analysis Project, Release 2.1. Palo Alto, CA USA: YouGov/Polimetrix.".

Jacobson, Gary C. 2012. "The Electoral Origins of Polarized Politics: Evidence From the 2010 Cooperative Congressional Election Study." American Behavioral Scientist 56(12): 1612–1630.

Norrander, Barbara. 1989. "Ideological Representativeness of Presidential Primary Voters." American Journal of Political Science 33(3): 570–587.

Peress, Michael. 2013. "Candidate Positioning and Responsiveness to Constituent Opinion in the US House of Representatives." Public Choice 156(1-2): 77–94.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." The Journal of Politics 75(02): 330–342.

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

Vavreck, Lynn, and Douglas Rivers. 2008. "The 2006 cooperative congressional election study." Journal of Elections, Public Opinion and Parties 18(4): 355–366.