

UC San Diego

UC San Diego Previously Published Works

Title

Institution of Nomination and the Policy Ideology of Primary Electorates

Permalink

<https://escholarship.org/uc/item/6zp95036>

Journal

QUARTERLY JOURNAL OF POLITICAL SCIENCE, 10(4)

ISSN

1554-0626

Author

Hill, Seth J

Publication Date

2015

DOI

10.1561/100.00015023

Peer reviewed

Institution of Nomination and the Policy Ideology of Primary Electorates*

Seth J. Hill[†]

September 29, 2015

Abstract

Many hypothesize that the divergence between Democratic and Republican members of Congress is partly attributable to partisan primary elections. Yet most empirical evidence on the influence of primary elections finds small to no effect on member behavior. I argue that existing designs that compare members elected out of nomination systems with more open rules of access to members elected out of more closed systems rest on the crucial and untested assumption that more closed institutions lead to more polarized primary electorates. With survey opinions, turnout validated to voter files, and an IRT model of ideology, I characterize the preferences of Democratic and Republican primary electorates and general electorates in each House district in 2010 and 2012. To the extent that there is a relationship between primary ideology and closed primary institution, it is in the direction opposite that hypothesized. I then show that the primary electorate diverges from the general electorate in every House district and even from supporters of the party in the general election in almost every district, which is consistent with a centrifugal influence of primary voters. These results suggest that institution of nomination may not have a large influence on the type of voters who turn out, and that some other feature of nominating contests must be implicated in polarized primary voters.

*I have benefited from comments by seminar participants at Princeton, Stanford, UCLA, and Washington University, Andy Hall, Greg Huber, Gary Jacobson, Thad Kousser, Jeff Lewis, Sam Popkin, Chris Tausanovitch, Lynn Vavreck, the editors, and anonymous reviewers.

[†]Department of Political Science, University of California, San Diego, 9500 Gilman Drive #0521, La Jolla, CA 92093-0521; sjhill@ucsd.edu, <http://www.sethjhill.com>.

Although primary elections were introduced by reformers around the turn of the twentieth century to democratize American politics, most worry today that primaries are part of a disconnect between what representatives do in Congress and what most citizens would prefer (e.g., Fiorina and Abrams, 2009). Democrats in Congress seem to be more liberal than their districts, Republicans seem to be more conservative than theirs, and compromise is less common than voters desire. The theory connecting primary elections to representative divergence from the general electorate is straightforward. Candidates for office need to win the votes of those who turn out in the elections for which they stand. If a candidate need only win a general election to gain a seat in Congress, the candidate needs to please those citizens who vote in the general election. If the candidate has to first win a partisan primary prior to winning a general election, however, the candidate must please two different voting electorates. Thus, the candidate may have to appeal to the divergent preferences of the primary electorate to the detriment of the general electorate when the primary electorate is filled with citizens of more partisan or extreme preferences than the general electorate. To the extent that many members of the U.S. House are elected out of districts safe for one party or the other, it may be that primary electorates are of highest electoral concern, and thus dominate many of the choices representatives must make.

Because the arguments connecting primary electorates to representative divergence from the general electorate are both reasonable and intuitive, many believe primary elections to be a key component of the large differences between Democrats and Republicans in the contemporary Congress. One of the preeminent scholars on congressional elections summarizes: “the most common explanation for the failure to observe Downsian convergence [is] the extremism of primary electorates (Jacobson, 2012, p. 1615).”

Despite the straightforward appeal of primaries as the basis of partisan divergence in Congress, an impressive array of empirical evidence comes up with mostly null evidence on the relationship between primaries and representative behavior. These studies often compare the behavior of representatives elected out of more open primary systems, where most or all registrants are eligible to participate, to more closed primary systems, where participation is limited to partisan registrants

or in some other way. Studies such as these make the assumption that more open primary institutions, which allow more citizens to participate at the nomination stage, should have more moderate primary electorates and thus nominate more centrist candidates.¹ For example, McGhee, Masket, Shor, Rogers, and McCarty (2014) measure the behavior of tens of thousands of state legislators over more than a decade of elections and find little relationship between the type of primary out of which representatives are nominated and the ideology in their roll call votes. Hirano, Snyder, Jr., Ansolabehere, and Hansen (2010) analyze more than sixteen thousand House elections from 1932 to 2006 and find little evidence that members of Congress subject to primary elections vote more extreme than members not subject to primaries, and Bullock and Clinton (2011) find only limited effects of California's blanket primary in the 1990s.

One interpretation of the null results referenced above is that primary elections do not influence member voting behavior. An alternative interpretation is that the assumption that more open primary institutions lead to more centrist primary voters does not hold, and thus existing null results do not refute an influence of primary elections. Although reasonable to think that variation in nominating institution would lead to variation in the divergence between the preferences of primary and general electorates, this has not been empirically established. I argue that if the important influence on member behavior is the preferences of the voters in primary elections, then the institution of nomination relevant when it materially changes these preferences. Statutory rules on participation may change the rate of turnout in primary elections or the composition of the primary electorate. These changes, however, may not materially change the distribution of preferences of primary voters. This could be because institution of nomination has marginal influence on turnout or composition, or because regardless of rate of turnout, composition, or institution, the same types of voters are motivated to participate in primary contests.²

There are two important implications of the argument that institution of nomination has little

¹ “[F]ew doubt that opening nomination procedures to previously excluded nonpartisans will increase mass participation in the nomination process (Gerber and Morton, 1998, p. 305)” or “[A] more onerous system should produce less moderation (McGhee et al., 2014, p. 339).”

² As noted in some of the first scholarship on primary elections, statutes “are not the end but the beginning (Merriam and Overacker, 1928, p. 196).”

influence on the ideology of primary voters. First is that, even if members represent their primary constituencies, variation in institution of nomination is unlikely to have a measurable relationship to member behavior. Second, it suggests that there must be other incentives, separate from rules regulating who may participate in nominations, that (a) influence who makes the effort to participate in primary elections, and that (b) generate the divergence in behavior of Democratic and Republican members of Congress. More broadly, the result has implications for reformers who aim to moderate national politics by changing the rules and institutions of primary elections. If the results I present here generalize to other contexts, liberalizing access to the primary ballot may not broaden or increase the representativeness of the primary electorate.³

In this essay, I characterize the preferences of Republican and Democratic primary electorates and the general electorate in each House district with an item-response theory (IRT) model of policy ideology. The data source is opinion surveys with turnout validated to voter files. This allows me to compare the policy views of voters who actually turned out in primary and general elections in both closed and less-closed primaries. I then estimate district-level preferences both through simple aggregation and through multi-level regression with post-stratification (MRP, Gelman and Little, 1997). Looking at the distributions of individual preferences or using either estimator for congressional district preferences, I find no evidence that in 2010 and 2012 closed and semi-closed primary states had more ideological primary voters than states with more open primary systems. I then show that the primary electorate has divergent preferences from the general electorate in every House district. I make this comparison in each district, which is an improvement over previous comparisons of primary voters to general voters nation or statewide. I also show that primary electorates diverge not only from the general electorate in every district, but also from the party's voters in the general electorate in almost every district (the *party following*, e.g. Geer, 1988). Finally, I consider change in California from 2010 to 2012 when the primary system moved from closed to non-partisan, again finding little evidence of an influence on primary ideology.

This article contributes to our understanding of regulations on the franchise, the importance

³ See Hill and Kousser (2015) for further evidence on this point.

of who turns out in low salience elections, and representation, and proceeds as follows. I first present the theory of the influence of primary elections, summarize the current mixed evidence on nominating elections, primary electorates, and representative behavior, then summarize the data sources, research design, and results.

Primary elections, electorates, and representatives

Why should members of Congress be responsive to primary electorates? I assume that members represent their election constituencies as ambitious office-seeking politicians. Canonical studies of representation have considered the relationship of member behavior to the preferences of general electorates, from Downs (1957) to Miller and Stokes (1963) to dozens subsequent. The logic of representation through election is that candidates who want to win office must win the votes of the electorates for which they stand. Most candidates for Congress today must first win nomination through a primary election, and second win the general election. If the policy preferences of the two electorates diverge, candidates must determine how to present themselves so as to most effectively navigate the electoral process.⁴

If members are concerned about primary elections, they must make tradeoffs between their primary and general electorates when they cast roll call votes. Averaged across issues, member roll call behavior might be something like $y_i = \alpha x_i + \beta z_i + \epsilon_i$, where y_i is an ideological summary of the member's behavior, x_i is the ideology of the general electorate, z_i is the ideology of the primary electorate, α and β are coefficients that capture the average influence of each of the two electorates on member vote choices, and ϵ_i is an error term that encompasses other factors of member voting other than electorate ideology. This highlights not only that measurement of the preferences of primary electorates is important to evaluate their potential influence, but also that the primary electorate should also be contrasted to the general electorate. Correlation in the preferences of the two groups across space or time could lead to omitted variables bias if both are not measured

⁴ This does depend upon how strategic the primary voters are: In anticipation of a more moderate general electorate, primary voters may demand less ideological behavior from their candidate than they would otherwise prefer so that they have a better chance of winning the general election. See Aranson and Ordeshook (1972) for a formal derivation of this tradeoff and the implications for candidate strategies. Alvarez and Nagler (2002), however, provide evidence that primary voters in California are much more sincere than strategic.

and considered concurrently.

Of course, x_i and z_i are not exactly observed, either by the analyst or the member. I follow the logic of Fenno (1978) and Arnold (1990) that members work hard to anticipate the preferences of their electorates through constant interaction with constituents and district interests. Through these efforts, along with their aptitude as professional politicians, they gather a sense of the wants of their two electorates. Note that this constant search means that members need not necessarily even be subjected to a competitive primary or general election to be responsive to the interests of the two electorates. If they sufficiently anticipate and respond to those interests, no challenger may want to waste their time.⁵ My empirical efforts acknowledge this reality by measuring the preferences of members' primary electorates even in places where incumbents run unopposed. Because primary voters vote for multiple offices, I am able to observe the set of citizens who turn out in primaries in most House districts, and use this as a measure of z_i , even if the member is not challenged.⁶

It is this logic of an electoral penalty for members who vote against their primary electorates that has motivated research to determine how primaries are implicated in partisan differences in roll call voting. One challenge is measuring electorate preferences in each primary electorate in each district. Existing research has often used proxy measures rather than direct measures of z_i . Most research on the effects of congressional primaries considers the relationship between the institution out of which the elected representative was nominated and summaries of their roll call votes. The basis for these designs is the assumption that more open primary institutions allow less partisan and more heterogeneous electors to participate in primaries, thus lessening the extremity of the primary electorate and its divergence from the general electorate. If this assumption holds, legislators elected under more open primary systems should feel on average less pressure to diverge from the general electorate than legislators elected under closed primary systems. These studies assume that the institution of nomination v_i influences the preferences of the primary electorate z_i ,

⁵ In the words of Hirano et al. (2010), "Since strategic candidates are likely to adjust their position to minimize electoral threats, whether MCs face primary competition is unlikely to be an accurate measure of the actual underlying primary threat they face (172)."

⁶ Of course, many members of the House are unopposed at the general election, as well, and the same logic applies to my measurement and theory about the influence of the general electorate.

and estimate reduced form regressions of y_i on v_i instead of on z_i .

For example, McGhee et al. (2014) compare thousands of state legislators in the United States from 1992 to 2010 elected out of more and less open primary systems. The authors find no evidence that openness of primary system influences either roll call voting behavior or candidate responses to policy surveys. If anything, they find suggestive evidence that more open primaries lead to *more* divergent roll call voting. Likewise, Hirano et al. (2010) consider the relationship between three characteristics of primaries and the partisan voting of elected representatives, none of which show much relationship with party polarization measured by DW-NOMINATE score using methods such as differences-in-differences.⁷ Others who analyze variation in institution (Bullock and Clinton, 2011; Gerber and Morton, 1998) or variation in competition (Boatright, 2013; Burden, 2001) find little or no influence of primary elections on representative behavior.

These results are consistent with a literature on the representativeness of *presidential* primary voters. Much of the work on presidential primary voters finds only limited evidence of divergence from general election voters (e.g. Ranney and Epstein, 1966; Geer, 1988; Norrander, 1989; Kaufmann, Gimpel, and Hoffman, 2003; Abramowitz, 2008; Sides and Vavreck, 2013). These two sets of evidence together have led many to conclude that primary elections may *not* influence the behavior of representatives.⁸ The evidence on congressional primaries, however, does rest on the assumption that institution of nomination v_i is a reasonable proxy measure for primary preferences z_i , and it is unclear what to make of presidential primary voters given the sequential and national nature of those contests.⁹

In contrast, at least three papers do find suggestive evidence that preferences of primary voters are related to divergent member behavior. Brady, Han, and Pope (2007) use primary and general

⁷ DW-NOMINATE is a summary measure of the liberal and conservative preferences of members of Congress based upon their roll call voting behavior (see Lewis and Poole, 2004; Carroll et al., 2009). Hirano et al. (2010) also look for an effect of moving from a system without primary elections to one with primary elections. This assumes that the nominating electors prior to the implementation of primary elections, for example party bosses or party caucuses, had preferences divergent from those who participated in the primary election system, which also may not hold.

⁸ “[T]he polarized state of American politics today reflects the polarized state of the overall American electorate rather than any peculiar characteristics of primary voters (Abramowitz, 2008).”

⁹ Recent evidence comparing congressional primary voters to congressional general voters finds primary voters to have more divergent policy views, but again the difference with general election voters is not dramatic (Jacobson, 2012, Table 2).

election vote shares by congressional district from 1956 to 1998 to argue that primary electorates cause divergence in representative behavior. As the authors note, however, “Our ability to make broader claims is limited by the lack of better measures of each district’s primary election constituency (p. 99).”¹⁰ Butler (2009) and Clinton (2006) use survey data to find an influence of primary voters or partisan subconstituencies on representative voting.¹¹ Gerber and Lewis (2004), however, do not find an effect of legislator partisan subconstituency on roll call voting when measuring preferences using proposition votes.

Research design

The empirical goals of this article are to test the assumption that institution of nomination is related to the ideology of who votes in primary elections, and to describe and evaluate the relationship between the preferences of primary electorates and general electorates by institution. To address these questions, I require concurrent measures of ideology and primary voting across institutions. I use opinion surveys to measure policy preferences and voter files to measure the validated turnout of these same individuals. The voter files mitigate the common problem of over-reporting of turnout. Because the survey policy preferences are likely to be measured with error, I use an item-response theory (IRT) model to collapse responses to multiple issues into one summary value. I then compare the distribution of these values for primary voters who reside in states with closed and not-closed primary institutions.

The data sources are the 2010 and 2012 CCES (Ansolabehere, 2010, 2012), both nationally representative samples of around 55,000 Americans with interviews before and after the 2010 midterm and 2012 presidential elections, stratified by state. The surveys asked standard sets of political questions about attitudes, preferences, and beliefs, and also validated turnout records by matching respondents to voter files. The surveys are large enough to include some validated pri-

¹⁰ Multiple pieces of evidence are presented in support of the effect of primary elections. For example, members of Congress whose NOMINATE score is closer to district presidential vote have more challengers and do worse in primaries, all else equal; turnout in primary elections is lower and more stable than turnout in general elections; and primary losses for incumbents, though rare, more often happen to incumbents with moderate voting records in Congress.

¹¹ The Clinton (2006) party subconstituency is defined as respondents sharing the incumbent’s partisanship, though his discussion (p. 398) suggests primary constituencies are part of what motivate this choice.

mary voters in almost every district.¹² The study here thus includes both midterm and presidential year primary voters.

To summarize the policy conservatism (I code ideology in the conservative direction) of each respondent to each of the two surveys, I estimate the grouped IRT model proposed by Lewis (2001) on expressed preferences over a set of policy issues. Each CCES asked respondents how they would vote on a set of roll calls actually considered in the House and Senate, as well as other policy preferences not specific to any roll call vote. I identified 17 questions from each survey that serve as the items in the model, the full list of which are available in Appendix Section B.2. I group respondent ideal points by the intersection of three characteristics: their state of residence, their partisanship (coded three ways, with leaners collapsed as partisans), and their primary turnout. With model estimates, I calculate the expected *a posteriori* ideal point for each respondent, conditional on their responses and group membership (see Lewis, 2001, p. 279), and post-process the ideal points to have mean zero and unit variance in each year. The ideology estimates correlate with partisanship and self-reported ideology, and they are superior predictors of vote choice than self-reported ideology, which I document in Appendix Section B.2.¹³

With individual-level estimates of policy conservatism, I can first compare distributions and summary statistics for primary voters in states with closed and less-closed primary institutions, the classification of which I borrow from McGhee et al. (2014). I also consider the question of divergence of primary voters relative to general voters and party supporters in the general electorate by institution. I estimate electorate preferences in two ways. First, I simply aggregate the CCES respondents up to the congressional district using the CCES post-stratification weights.¹⁴ Second, I implement a hierarchical model to ameliorate sampling error. The hierarchical model smooths across geographies, turnout, party, and respondents to provide best estimates for each electorate

¹² See Appendix Table A1 for counts of the number of primary and general validated voters by state and year. The validation has no coverage of Virginia in 2010.

¹³ The 2010 estimates correlate with the IRT estimates of Tausanovitch and Warshaw (2013) for 2010 CCES respondents at $r = 0.961$.

¹⁴ Formally, if y_i is the conservatism for respondent i , then my estimate of the conservatism for electorate e in congressional district c is $\sum_{i \in e, c} w_i^{-1} \times \sum_{i \in e, c} y_i w_i$, where w_i is the survey weight for respondent i and $i \in e, c$ evaluates to the set of respondents i validated to have voted in primary or general election e and residing in district c .

in each district. This in general shrinks estimates towards the grand mean across individuals and districts, reducing the influence of outlying values in small sample primary electorates. I present full details of the hierarchical model and construction of MRP (Gelman and Little, 1997; Park, Gelman, and Bafumi, 2004) district estimates in Appendix Section A.

For states with open or not-fully-closed primary elections, I am uncertain in which primary each validated primary voter voted. To make estimates somewhat consistent across states, I use respondents' self-reported party of registration to construct estimates. That is, no matter the primary institution in place, Democratic primary electorate estimates are the weighted average of voters validated to have voted in the primary and who report being registered Democrats, and Republican primary electorate estimates are the weighted average of voter validated to have voted in the primary and who report being registered Republican. Clearly this choice induces measurement error in the district estimates, though not the individual estimates.¹⁵

Institution of nomination and primary voter policy ideology

In this section, I present results testing for a relationship between institution of nomination and the ideology of primary voters. I first plot the individual distributions of conservatism in each class of institution, finding little evidence of difference. I then calculate differences of means and variances for individual- and district-level distributions of conservatism. Statistical tests also find no evidence of difference.

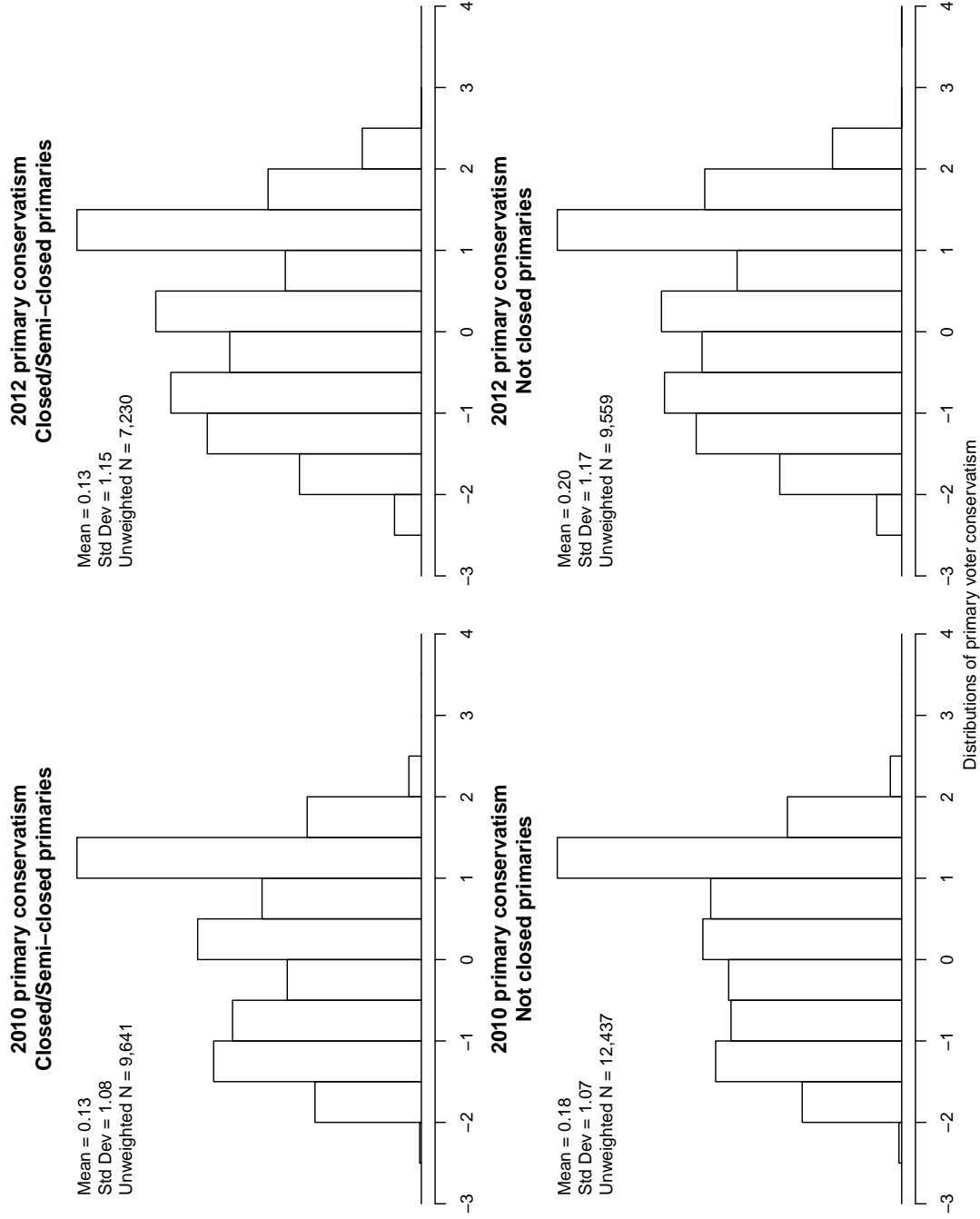
In Figure 1, I construct histograms of conservatism using the IRT policy ideology scores for each respondent validated to have voted in congressional primary elections in the 2010 and 2012 CCES.¹⁶ I plot separate distributions for respondents who reside in states with closed primary systems (states classified as “closed” or “semi-closed” by McGhee et al., 2014) and respondents who reside in remaining states. One implication of closed primary states leading to more ideological primary voters is more variance in the distributions in closed states than in non-closed states.

Figure 1 presents little difference by institution of nomination. For 2010 (frames in first col-

¹⁵ In the two surveys, 18.8 percent (2010) and 19.7 percent (2012) of validated congressional primary voters reported being registered Decline-To-State or third party.

¹⁶ Histograms are weighted with CCES post-stratification weights.

Figure 1: Primary voter ideology by year and type of primary institution



Note: Each frame is the distribution of conservatism for voters validated to have voted in the congressional primary in that year. Distributions are separated by states where both party primaries are classified as “closed” or “semi-closed” by McGhee et al. (2014) and the remainder. Alaska is excluded due to mixed primary type across parties. Distributions include all validated primary voters, even those who report a non-Democrat/non-Republican party of registration. The figure shows little difference in conservatism for primary voters participating in states with closed vs open primaries.

Table 1: Ideology of validated voters, closed versus not closed primary institution

Year	Aggregation	Estimator	Difference of Means	Difference of Variances
2010	Individual	Raw	-0.04 [-0.09, -0.00]	+0.03 [-0.01, +0.07]
2012	Individual	Raw	-0.08 [-0.13, -0.03]	-0.06 [-0.12, +0.00]
2010	Congressional district	Raw	-0.19 [-0.25, -0.13]	+0.02 [-0.01, +0.05]
2012	Congressional district	Raw	-0.09 [-0.15, -0.02]	-0.01 [-0.04, +0.02]
2010	Congressional district	MRP	-0.15 [-0.20, -0.09]	+0.00 [-0.02, +0.03]
2012	Congressional district	MRP	-0.04 [-0.09, +0.01]	-0.01 [-0.03, +0.01]

Note: Cells present the difference of mean ideology and difference of variance of ideology in states with closed minus states with not closed institutions of nomination, separately by year, level of aggregation, and estimator. Confidence intervals derived from 1,000 nonparametric bootstrap samples. The results show little difference in means or variance of voter ideology by institution of nomination.

umn), shape, spread, mean, and standard deviation are very close to equivalent between the two types of institution.¹⁷ Likewise for 2012, there is little difference in the distributions of ideology for voters in closed and non-closed states.

As a statistical test for these differences, I present in Table 1 differences of means and variances for both years at the individual level in rows one and two. For each year, I calculate the difference in the mean and variance of conservatism between closed and not closed states along with 95 percent confidence intervals constructed from a nonparametric bootstrap.¹⁸ While the confidence intervals on differences of means are statistically significant, the magnitudes are small, and the more relevant differences of variances have confidence intervals of small magnitude that cross zero. The largest difference in variance within a confidence interval is -0.12 for 2012, and this is a difference in the direction opposite that assumed (-0.12 means closed primary states had smaller variance).

Along with differences in the individual distributions, I aggregate preferences up to congressional districts by party of registration for each year, with both MRP and raw estimates. In rows

¹⁷ Note that the identifying restriction for the Lewis (2001) IRT model is mean zero and variance one for one of the groups in the estimation, so there is no requirement the distribution across groups be unimodal or normal.

¹⁸ I use the bootstrap percentile method, which resamples from the data at random 1,000 times, and on each resample calculates the difference of means and difference of variances. Confidence intervals extend from the 2.5th to the 97.5th percentile of the bootstrap statistics.

three through six, I again find statistically significant but materially small differences in means, and differences of variances that are small with confidence intervals that always cross zero.

In sum, these observations suggest that in 2010 and 2012, institution of nomination would be a poor proxy for the policy ideology of the voters who participated in primary elections. Both at the individual and the district level, I find no relationship between the statutory rules for which individuals may participate in primary elections and the conservatism of the individuals validated to have voted in those elections.

This result is very interesting. It either means that the citizens who want to participate in primary elections do so regardless of institutions in place, that institution of nomination is seriously confounded with ideological features of the states, or that different regulations on the franchise influence the composition of voters who participate but not their preferences. I turn next to further interrogate whether primary elections produce different electoral signals for candidates for office with respect to voter conservatism.

Comparison of general and primary voters

In the previous section, I showed that variation in institution of nomination is not related to variation in the ideology of primary voters or electorates. In this section, I evaluate whether primary voters actually do diverge in their preferences from general election voters, and whether the level of divergence varies by the institution under which access to the primary ballot is regulated. I find that primary electorates are less centrist than general electorates in every House district, and as before little relationship to institution of nomination.

I plot in Figure 2 the conservatism of each partisan primary electorate in each congressional district (y-axis) against the conservatism of the general electorate in that same district (x-axis). Each district is represented by a text code, and I compare the general electorate to the primary electorate of the party of the member who represents that district. The dashed lines in each plot are 45 degree lines; points falling on that line would indicate the district primary electorate has the same conservatism as the district general electorate. The top set of plots use the MRP estimates of

primary and general electorate conservatism, the bottom set of plots the raw survey aggregates. I make the plots separately for Democrats and Republicans, by election year, and by institution of nomination. Each plot also includes a loess smooth through the points to highlight the trend.

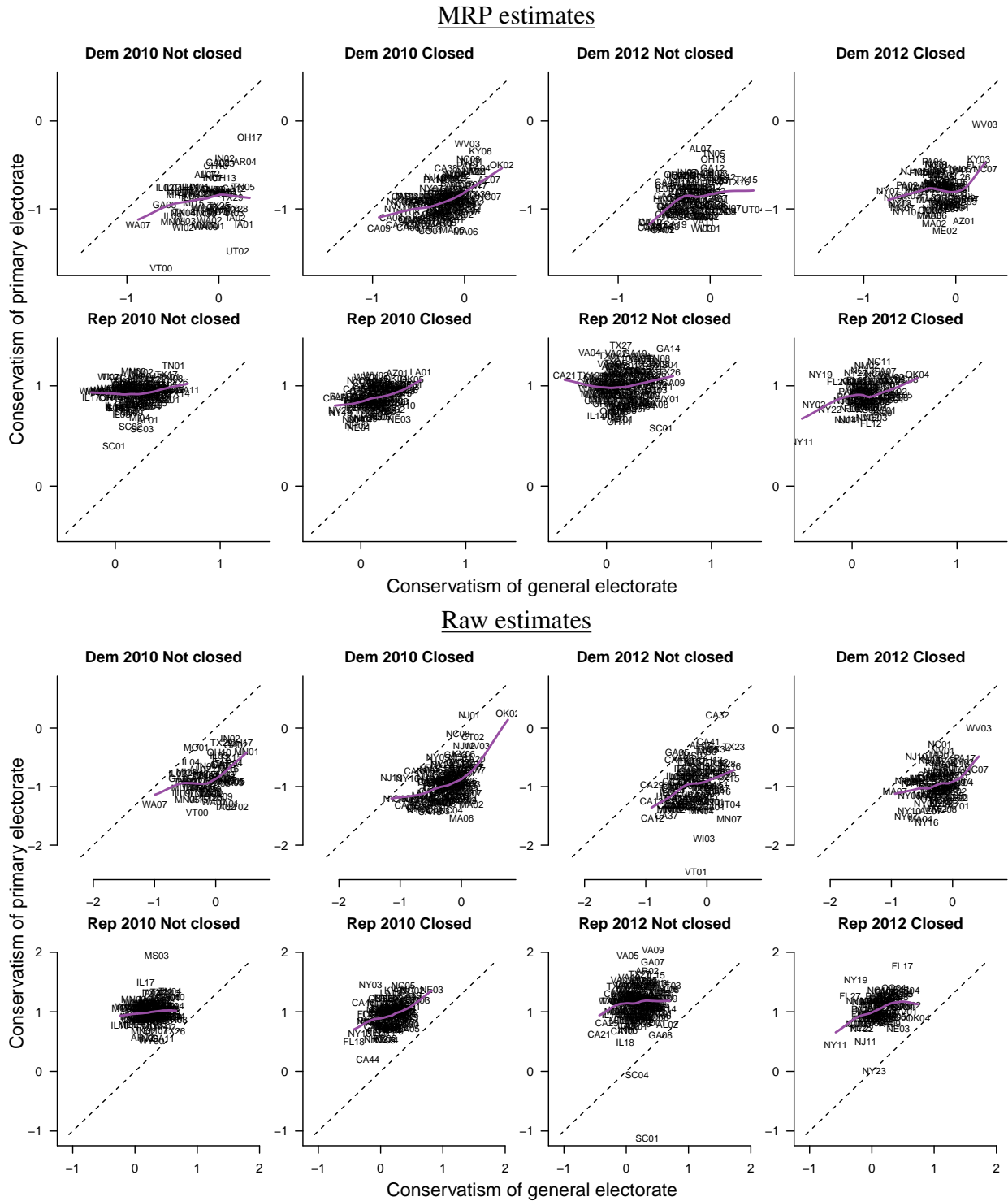
The first thing to note from Figure 2 is that not a single district in either party, election year, or type of primary has a primary electorate more centrist than the general electorate with the MRP estimates; all points in the top Democratic plots fall below the 45 degree line (primary electorate more liberal than general electorate), and all points in the bottom Republican plots fall above the 45 degree line (primary electorate more conservative than general electorate). In the lower frames using the raw survey aggregates estimator, a few districts fall above the 45 degree line, but the pattern is of consistent divergence between primary and general electorates. Although there are many House districts dominated by voters from one party, the general electorates in even the safest districts are not as ideological as the primary electorates in most districts around the nation. Looking at variation by institution of nomination, there is suggestive evidence of a stronger relationship between general electorate and primary electorate conservatism in states with closed institutions.

Also of note is the relatively modest relationship between general electorate conservatism and primary electorate conservatism, especially for districts represented by Republicans in Congress. Comparing loess smooths to the angle of the 45 degree line shows that the relationship is not just an intercept shift down, but rather that the conservatism of primary voters is only modestly related to the conservatism of general election voters. While the most liberal general electorates have some of the most liberal Democratic primary electorates (e.g., in 2010 Democrats Barbara Lee [CA-09] in Oakland and Jim McDermott [WA-07] in Seattle), primary electorate conservatism increases only moderately in relation to general electorate conservatism for Democrats. For Republicans, the relationship between primary and general election conservatism is even more attenuated.

Primary voters and party voters

Previous research (e.g., Geer, 1988) evaluates the divergence of primary voters by comparing primary voters to general election supporters of that party. I show in Figure 3 a similar pattern to that

Figure 2: Ideology of primary and general electorates, closed versus not closed primary institution



Note: Each frame plots the ideology of the primary electorate for that member's party in the district on the ideology of the general electorate in that district. Member party, election year, and state institution of nomination by McGhee et al. (2014) indicated in each title. The first eight frames use the MRP estimates, the second eight are the raw survey aggregates. Lines are loess smooths.

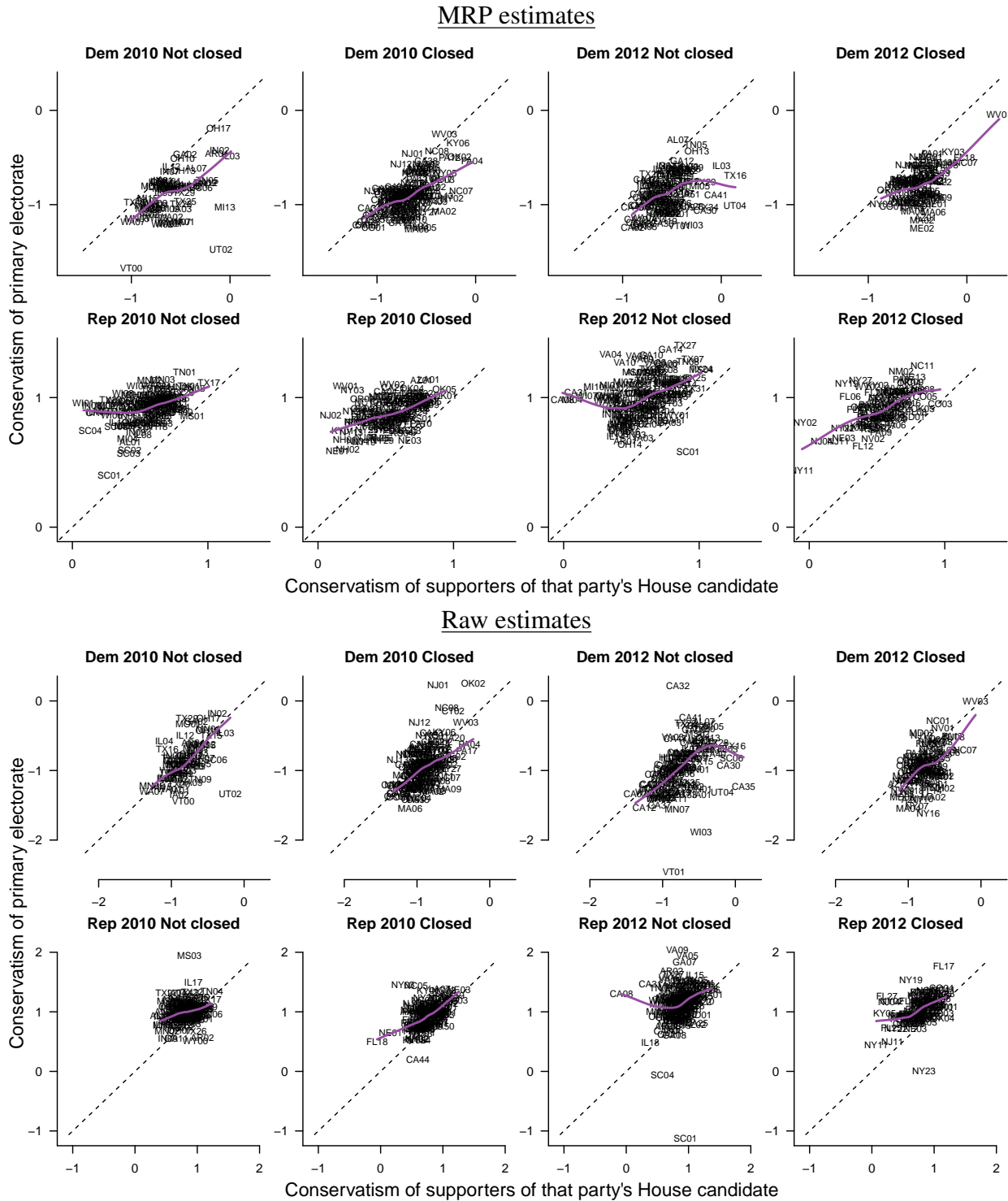
in Figure 2, with primary voters notably divergent even from the party's voters in the general election. I calculate the conservatism of the validated general election voters who report voting for that party's House candidate and compare that conservatism to the conservatism of validated primary voters from that House candidate's party. Plots are partitioned as before by party, year, institution of nomination, and MRP versus raw estimator of electorate conservatism. Here, we find a closer relationship to primary preferences – not surprising because this is the subset of the general electorate who preferred the candidate of that primary – but continued divergence in the preferences of primary voters. Although the clouds are closer to the 45 degree line, and in some cases I estimate primary electorates more centrist than party voters, the overall story remains divergent primary electorates. These results contrast existing findings on presidential primary voters.

With respect to variation by institution of nomination, there appears to be a stronger relationship between primary conservatism and party voter conservatism in states with closed primary systems, at least for Republicans, but the difference is not dramatic. Institution of nomination does not seem to be strongly related to the ideology of primary electorates, general electorates, or party voters.

As a statistical test of the relationship of institution of nomination to partisan primary ideology in each congressional district, I present OLS regressions in Table 2. The dependent variable is the conservatism of the primary electorate, with separate specifications for Democratic primaries (limited to districts represented by Democrats) and Republican primaries (limited to districts represented by Republicans). The coefficient of interest is the effect of a closed primary institution, with the assumption in the literature being a negative coefficient for Democrats (primary more liberal in closed), and a positive coefficient for Republicans (primary more conservative in closed). I present models where the indicator for closed primary is also interacted with the preferences of general electorates and party voters in the general electorate to account for potential spurious correlation between the types of places with closed institutions and the types of voters who reside in those places.¹⁹

¹⁹ Excluding states using the top-two primary system, Washington in 2010 and 2012 and California in 2012, yields similar results.

Figure 3: Ideology of primary and party voters, closed versus not closed primary institution



Note: Each frame plots the ideology of the primary electorate for that member's party in the district on the ideology of the general election voters who voted for that party's candidate in that district. Member party, election year, and state institution of nomination by McGhee et al. (2014) indicated in each title. The first eight frames use the MRP estimates, the second eight are the raw survey aggregates.. Lines are loess smooths.

When controlling for the ideology of the general electorate or party general electorate voters, the direct effect of a closed primary is in the wrong direction for both parties and for both estimators, MRP and raw. For Democrats, a closed primary is associated with a more *conservative* primary electorate all else equal, though the relationship is not statistically significant. For Republicans, a closed primary is associated with a more *liberal* primary electorate all else equal, a relationship that is statistically significant at $p < .05$ in three of four specifications. The evidence is contrary to the presumed effects of more accessible primary elections, which assume that broader participation should lead to more centrist primary voters. I also find consistent though not always statistically significant evidence that the correspondence between primary ideology and general ideology is stronger in places with closed primary systems. This is again inconsistent with the presumed effect of closed primaries, which are thought to lessen the connection between primary voters and general voters. The magnitude, also, is small relative to differences between the parties.

In sum, I find consistent support for the argument that primary and general electorates diverge in their policy ideology. I also present evidence that primary ideology is less centrist than the ideology of the party's voters in the general election. Yet this measured divergence appears mostly unrelated to the system of nomination. If there is a relationship with institution, it is in the opposite direction of what is usually assumed, with closed primaries having more centrist primary electorates more closely aligned with general electorates. While I cannot attribute this as a causal relationship, it does support the argument that previous findings of little influence of primary institutions on roll call behavior is not necessarily evidence that primaries don't matter.

Changing institutions: California 2010 to 2012

One aspect of my analysis that merits consideration is the cross-sectional nature of the comparison. I observe variation on institution of nomination across states within these two years, thus leaving open the possibility that some other feature of the states that is correlated with institution of nomination is masking an effect.²⁰ One state did change its primary system in this time period.

²⁰ Of course, analysis of over time change in institutions must also assume that the change in institution is not confounded with some other ideological feature of the state in time.

Table 2: Tests of relationship of primary institution to electorate ideology

	MRP estimates				Raw survey estimates			
	Dem	Dem	Rep	Rep	Dem	Dem	Rep	Rep
(Intercept)	-0.83*	-0.65*	0.95*	0.79*	-0.87*	-0.51*	1.05*	0.77*
	(0.02)	(0.04)	(0.01)	(0.03)	(0.03)	(0.07)	(0.02)	(0.07)
Closed primary	0.05*	0.12*	-0.09*	-0.09*	0.03	0.18	-0.12*	-0.16
	(0.03)	(0.05)	(0.02)	(0.04)	(0.04)	(0.10)	(0.03)	(0.10)
Ideology general electorate	0.28*		0.09*		0.38*		0.09	
	(0.06)		(0.05)		(0.08)		(0.07)	
General electorate*Closed	0.09		0.17*		0.06		0.29*	
	(0.09)		(0.07)		(0.11)		(0.11)	
House Democrat voters		0.42*				0.56*		
		(0.06)				(0.08)		
House Dem voters*Closed		0.13				0.17		
		(0.09)				(0.12)		
House Republican voters				0.30*				0.37*
				(0.04)				(0.08)
House Rep voters*Closed				0.06				0.14
				(0.07)				(0.12)
N	378	378	453	453	378	378	453	453
R ²	0.14	0.27	0.10	0.23	0.14	0.22	0.07	0.13
adj. R ²	0.13	0.26	0.10	0.22	0.13	0.22	0.06	0.13
Resid. sd	0.20	0.18	0.14	0.13	0.30	0.29	0.26	0.25

Standard errors in parentheses

* indicates significance at $p < 0.05$

Note: OLS regression coefficients with standard errors in parentheses. Dependent variable is ideology of the primary electorate for party of member representing that district.

California moved from a semi-closed system in 2010 to a non-partisan top-two system in 2012. Because California is a large state, the CCES records include thousands of validated primary voters from each election.

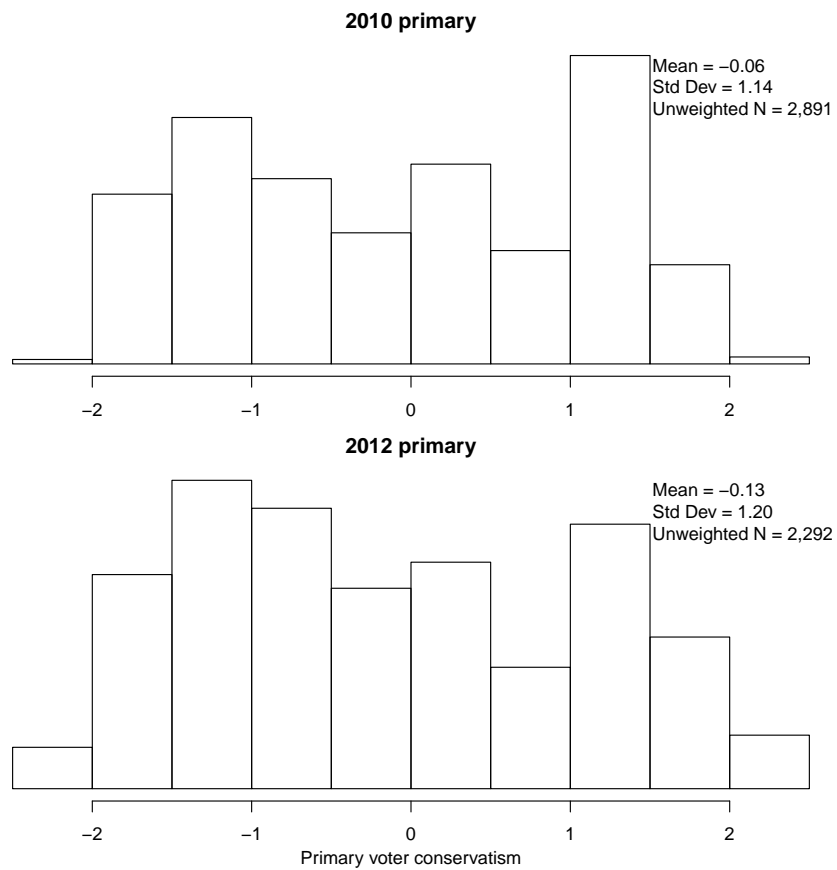
Figure 4 evaluates whether there are notable differences in the conservatism of those who turned out in California in 2010 under a semi-closed party primary compared to 2012 under a non-partisan top-two primary. The distributions look roughly similar, and their statistics of spread are off by one twentieth of a standard deviation. Because the IRT models from 2010 and 2012 are constructed from different survey observations, the estimated conservatism is only comparable by assumption of similar meanings of the latent variable. To more specifically benchmark the change in spread between the two years, I calculate the ratio of the standard deviation of California primary voter ideology in each year to the standard deviation of primary voters from all states with closed or semi-closed primaries. This ratio asks how much more or less variable California primary voters are in their ideology relative to primary voters in states with closed systems, while allowing for changes in the nature of contests from 2010 to 2012 and in the IRT model and inputs from 2010 to 2012. The ratio of standard deviations between California and other closed primary states in 2010 is 1.05, and the same ratio of California to the closed primary states in 2012 is also 1.05. Neither graphically nor with this statistical test did the change in institution change primary voter ideology in California from 2010 to 2012.

In sum, the comparison of voters in California from 2010 to 2012 across a change of nominating institution corresponds to little change in the ideology of primary voters. Institution of nomination does not seem to be a driving factor in the set of preferences of the voters who turn out in primary elections.

Discussion

I began this essay presenting the quandary of theoretical reasons why primary elections should polarize members of Congress coinciding with empirical evidence that is mixed and often null. I argued that one reason that current designs find limited effects may be that variation in institution

Figure 4: Ideology of primary voters in California, 2010 and 2012



Note: California moved from a semi-closed party primary in 2010 to a non-partisan top-two primary in 2012.

of nomination does not correspond to much variation in the preferences of primary electorates. I find evidence consistent with this argument: My measures of the conservatism of the primary electorate from each district are unrelated to the type of nominating institution under which the primary is held. While the institution of nomination is unrelated, I do find that primary voters and primary electorates are less centrist than general voters and general electorates around the nation, and are even consistently less centrist than party voters in the general electorate. These two results together suggest that primary electorates remain a candidate for a polarizing influence on Congress.

These results confirm some stylized facts that to date have lacked strong evidence. First, that the preferences of congressional primary electorates do diverge importantly from the preferences of congressional general electorates district by district. Second, that the primary electorate diverges from the voters who support that party at the general election. My results also present a new stylized fact to keep in mind as we explore the relationship of voters to their representatives and the influence of primary elections. I find little cross-sectional relationship between the institution of nomination and the ideology of primary voters in 2010 and 2012. This holds when analyzing all primary voters as individuals, or when analyzing aggregated electorate preferences. This finding should be replicated in other years and with other methods, but it does suggest some caution to assuming that institutions generate the individual voter behavior that they are designed to create. That said, given the level of elite conflict and polarization in the two years of this study, if we should ever expect to find polarization in primary ideology, it might be in the present time. That I do not suggests even more strongly that institution may not be an important influence on primary composition.

These results add to lines of inquiry on primary elections. First, with respect to debates about open versus closed primary institutions, my results here suggest the importance of focus on the sets of voters who actually turn out in each primary election, not just the institution in place. The findings do not necessarily contradict the results of institutional analysis that find no effect of primaries. Instead, what may matter is not so much the institution that is in place, but the set of voters who show up at primary elections in each district. While the institution may be important in

setting the stage for who turns out, it may not have large influence on the preferences of the voters who participate.

Second, with respect to questions on the representativeness of primary voters, my results suggest that primary voters are more divergent from even the party's supporters at the general election than has previous research. Three features of my analysis are distinct from most previous comparisons. First, I examine congressional primary voters in each district, rather than presidential or congressional voters nationwide. Second, I use validated as opposed to self-reported primary turnout. And third, I use a scaled measure of ideology across multiple items, which may be a more accurate measure of preferences subject to less measurement error than individual survey responses. Future work could more carefully consider the most accurate way to measure the distinctiveness of primary voters.

Finally, these results suggest that incentives beyond the range of primary institutions currently in place generate the set of voters who turn out in primary elections. More empirical and theoretical work should consider the act of turnout in nominating elections, and the influence of these choices on candidate and incumbent behavior. For example, does the same subset of the citizenry vote in primary elections regardless of institution of nomination? How much is the composition of this subset due to individual characteristics versus entrepreneurial candidates constructing their own primary coalitions? What motivates some citizens to participate in primary elections and others to stay home? Because nomination contests generate the candidates who eventually run in general elections and who win seats in Congress, these questions are of crucial importance to the functioning of American representation.

More broadly, these results suggest the importance of considering both institutions and individuals in evaluating large questions such as the effect of primary elections on representation and polarization. While primary elections as an institution may be the experimental "treatment" in such research questions, this treatment operates through mediators importantly including the behavior of individuals. Members of Congress do not necessarily respond to institutions per se, but rather to the signal received from their constituents as filtered through those institutions. The potential of

primary elections to influence member behavior likely depends upon the set of citizens who decide to participate.

References

- Abramowitz, Alan. 2008. "Don't Blame Primary Voters for Polarization." *The Forum* 5(4).
- Alvarez, R. Michael, and Jonathan Nagler. 2002. "Should I Stay or Should I Go? Sincere and Strategic Crossover Voting in California Assembly Races." In *Voting At The Political Fault Line: California's Experiment with the Blanket Primary*, ed. Bruce E. Cain, and Elisabeth R. Gerber. Berkeley and Los Angeles: University of California Press.
- Ansolabehere, Stephen. 2010. *Cooperative Congressional Election Study, 2010: Common Content. [Computer File]*. Cambridge, MA: Harvard University.
- Ansolabehere, Stephen. 2012. *Cooperative Congressional Election Study, 2012: Common Content. [Computer File]*. Cambridge, MA: Harvard University.
- Aranson, Peter H., and Peter C. Ordeshook. 1972. "Spatial Strategies for Sequential Elections." In *Probability Models of Collective Decision Making*, ed. Richard G. Niemi, and Herbert F. Weisberg. Columbus: Charles E. Merrill Publishing Co.
- Arnold, R. Douglas. 1990. *The Logic of Congressional Action*. New Haven: Yale University Press.
- Boatright, Robert G. 2013. *Getting Primaried: The Changing Politics of Congressional Primary Challenges*. Ann Arbor: University of Michigan Press.
- Brady, David W., Hahrie Han, and Jeremy C. Pope. 2007. "Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32(1): 79–105.
- Bullock, Will, and Joshua D. Clinton. 2011. "More a Molehill than a Mountain: The Effects of the Blanket Primary on Elected Officials' Behavior from California." *Journal of Politics* 73(3): 915–930.
- Burden, Barry C. 2001. "The Polarizing Effects of Congressional Primaries." In *Congressional Primaries and the Politics of Representation*, ed. Peter F. Galderisi, Marni Ezra, and Michael Lyons. Oxford: Rowman & Littlefield.
- 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.
- Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. 2009. "Measuring Bias and Uncertainty in DW-NOMINATE Ideal Point Estimates via the Parametric Bootstrap." *Political Analysis* 17(3): 261–275.
- Clinton, Joshua D. 2006. "Representation in Congress: Constituents and Roll Calls in the 106th House." *Journal of Politics* 68(2): 397–409.

- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper Collins.
- Fenno, Richard F. 1978. *Home Style: House Members in Their Districts*. Boston: Little, Brown.
- Fiorina, Morris, and Samuel J. Abrams. 2009. *Disconnect: The Breakdown of Representation in American Politics*. Norman, OK: University of Oklahoma Press.
- Geer, John G. 1988. "Assessing the Representativeness of Electorates in Presidential Primaries." *American Journal of Political Science* 32(4): 929–945.
- Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23(2): 127–135.
- Gerber, Elisabeth R., and Jeffrey B. Lewis. 2004. "Beyond the Median: Voter Preferences, District Heterogeneity, and Political Representation." *Journal of Political Economy* 112(6): 1364–1383.
- Gerber, Elisabeth R., and Rebecca B. Morton. 1998. "Primary Election Systems and Representation." *Journal of Law, Economics, & Organization* 14(2): 304–324.
- Hill, Seth J., and Thad Kousser. 2015. "Turning Out Unlikely Voters? A Field Experiment In The Top-Two Primary." Working paper, University of California, San Diego.
- Hirano, Shigeo, James M. Snyder, Jr., Stephen Ansolabehere, and John Mark Hansen. 2010. "Primary Elections and Partisan Polarization in the U.S. Congress." *Quarterly Journal of Political Science* 5: 169–191.
- 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.
- Kaufmann, Karen M., James G. Gimpel, and Adam H. Hoffman. 2003. "A Promise Fulfilled? Open Primaries and Representation." *Journal of Politics* 65(2): 457–476.
- Lewis, Jeffrey B. 2001. "Estimating Voter Preference Distributions from Individual-Level Voting Data." *Political Analysis* 9(3): 275–297.
- Lewis, Jeffrey B., and Keith T. Poole. 2004. "Measuring Bias and Uncertainty in Ideal Point Estimates via the Parametric Bootstrap." *Political Analysis* 12(2): 105–127.
- McGhee, Eric, Seth Masket, Boris Shor, Steven Rogers, and Nolan McCarty. 2014. "A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology." *American Journal of Political Science* 58(2): 337–51.
- Merriam, Charles Edward, and Louise Overacker. 1928. *Primary Elections*. Chicago: University of Chicago Press.
- Miller, Warren E., and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57(1): pp. 45–56.
- Norrander, Barbara. 1989. "Ideological Representativeness of Presidential Primary Voters." *American Journal of Political Science* 33(3): 570–587.

- Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12(4): 375–385.
- Ranney, Austin, and Leon D. Epstein. 1966. "The Two Electorates: Voters and Non-Voters in a Wisconsin Primary." *Journal of Politics* 28: 598–616.
- Sides, John, and Lynn Vavreck. 2013. "On the Representativeness of Primary Electorates." Working paper, George Washington University.
- Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2): 330–342.

On-line Appendix
Institution of Nomination and the Policy Ideology of
Primary Electorates
Quarterly Journal of Political Science

Seth J. Hill
Department of Political Science
University of California, San Diego
9500 Gilman Drive #0521
La Jolla, CA 92093-0521
sjhill@ucsd.edu
<http://www.sethjhill.com>

Contents

A	Measurement approach	3
A.1	Purpose of the hierarchical model and weights	3
A.2	Connection to Multi-level Regression with Post-Stratification	5
B	Implementation	6
B.1	Data sources	6
B.2	IRT model of conservatism	6
B.3	Procedure for district estimates	7
B.4	How the model changes district estimates	9
C	Validation of measure and improvement over raw survey margins	9
C.1	Testing the two estimates against each other	14

Table A1: Number of validated voters by year, state, and election

State	2010		2012	
	General	Primary	General	Primary
AK	83	74	90	54
AL	361	289	444	230
AR	322	176	329	13
AZ	1,035	751	1,025	674
CA	3,629	2,891	3,243	2,292
CO	705	490	756	295
CT	518	272	370	131
DC	56	52	71	23
DE	148	70	162	55
FL	2,784	1,923	2,554	1,285
GA	1,056	640	1,131	192
HI	118	108	124	98
IA	417	214	470	104
ID	171	84	256	119
IL	1,573	1,048	1,339	701
IN	544	449	704	391
KS	380	246	490	242
KY	465	352	537	182
LA	390	292	444	NA
MA	693	399	683	250
MD	655	406	727	314
ME	245	180	294	96
MI	1,190	812	1,237	623
MN	653	347	757	245
MO	766	544	848	477
MS	212	42	276	135
MT	97	75	179	126
NC	897	448	1,132	758
ND	86	56	75	53
NE	89	51	409	243
NH	244	164	250	131
NJ	884	321	856	283
NM	294	150	300	153
NV	430	297	445	201
NY	1,733	633	1,702	180
OH	1,697	1,143	1,432	755
OK	353	247	415	162
OR	593	464	864	599
PA	1,735	1,240	1,472	740
RI	127	64	171	60
SC	387	271	581	5
SD	105	50	112	46
TN	537	405	708	325
TX	2,150	1,362	2,205	1,046
UT	211	107	353	141
VA	0	0	1,064	228
VT	56	29	107	9
WA	965	811	1,059	784
WI	742	488	850	412
WV	210	131	203	150
WY	63	46	97	55

Note: Cell counts are the number of validated general and primary voters from each year's CCES by state.

A Measurement approach

In this section, I describe my measurement approach for the MRP district estimates and in the section following present the details of its implementation. The basic intuition is to use opinion surveys to describe the political conservatism of primary election voters and general election voters in each district. Because sample sizes are often small, I use a hierarchical model to pool observations, allowing the data to indicate if the preferences of, say, Democratic primary voters in New York congressional districts are more similar to each other than they are to Democratic primary voters in West Virginia. The model allows observed correlation within New York to influence my estimate of the preferences of the New York Democratic primary electorates relative to the preferences of the West Virginia electorates. I summarize the political preferences of each electorate in each district to a single value on a liberal-conservative dimension so that I may characterize the relationship between the conservatism of congressional voting and the conservatism of primary and general electorates.

While we have opinion survey samples, moving from the sample to a population estimate has many challenges. The main problem in my case is small samples of primary voters in each district, but additional problems include measurement error in sampled preferences, differential survey non-response, and uncertainty about turnout. I apply the current best practices in response to each of these problems to best characterize the preferences of primary and general election voters in each district, and then provide tests of fit and validation.

A.1 Purpose of the hierarchical model and weights

I use the 2010 and 2012 Cooperative Congressional Election Studies (CCES) below. Even with sample sizes of around 55,000 respondents, dividing them among 435 congressional districts yields on average about 125 respondents per district per year. If primary turnout is 20 percent of the voting age population, this means around 25 total primary voters, which must then be divided into two party primaries.

Hierarchical models are a natural solution to the problem of small samples when covariates are available and/or observations nest in some natural way. They have been shown to have the lowest mean-square error for this kind of estimation (see, e.g., Jackman, 2009, ch. 7 for citations). In this case, my hierarchical approach models preferences as a function of individual covariates and nested geographies. I include the individual demographic characteristics of each respondent along with their turnout history in the primary and general election and their geographic location. Turnout is validated to administrative records with statewide voter files. The hierarchical model smooths across geographies, turnout, and respondents to provide best estimates for each electorate in each district. This in general shrinks estimates towards the grand mean across individuals and districts, reducing the influence of outlying values in small sample primary electorates. I estimate the model separately for each of three types of party of registration, Democrat, Republican, and Decline to State/other party. Formally, my model of individual preferences y_i within each party of registration with respondents nested within districts within states within Census regions,

$$\begin{aligned}
y_i &\sim N(\mu_i, \sigma_y^2) \\
\mu_i &= \beta' \mathbf{x}_i + \boldsymbol{\alpha}'_{c(i)} \mathbf{t}_i + \delta_{c(i)} \\
\boldsymbol{\alpha}_{c(i)} &\sim N(\boldsymbol{\alpha}_{s(c)}, \Sigma_c) \\
\boldsymbol{\alpha}_{s(c)} &\sim N(\bar{\boldsymbol{\alpha}}, \Sigma_s) \\
\delta_{c(i)} &\sim N(\delta_{s(c)}, \sigma_c^2) \\
\delta_{s(c)} &\sim N(\delta_{r(s)}, \sigma_s^2) \\
\delta_{r(s)} &\sim N(0, \sigma_r^2)
\end{aligned} \tag{A1}$$

where y_i is a summary of the preferences of respondent i , μ_i is the expected value of y_i and σ_y^2 is the variance of the distribution of y_i . β is a vector of coefficients mapping individual characteristics \mathbf{x}_i to the expected value of y_i , $\boldsymbol{\alpha}_{c(i)}$ is a vector of coefficients for congressional district $c(i)$ mapping a vector of turnout history \mathbf{t}_i (e.g., primary and general turnout) to the expected value of y_i , $c(i)$ is a function that returns the congressional district of respondent i , and $\delta_{c(i)}$ is an intercept shift for the expected value of y_i in congressional district $c(i)$. The hierarchical structure of the model proceeds in the following lines, with the congressional district turnout coefficients $\boldsymbol{\alpha}_{c(i)}$ distributed multivariate normal with means specific to the state of that district through the function $s(c)$ and covariance matrix Σ_c . The state means for the turnout coefficients, $\boldsymbol{\alpha}_{s(c)}$, are distributed multivariate normal with mean vector $\bar{\boldsymbol{\alpha}}$ and covariance matrix Σ_s . Finally, each congressional district has an intercept shift $\delta_{c(i)}$ nested within states via a normal distribution with mean $\delta_{s(c)}$ and variance σ_c^2 , which are nested within regions with mean $\delta_{r(s)}$ and variance σ_s^2 . The region means are distributed normal with mean 0 and variance σ_r^2 .

In words, this hierarchical model allows the preferences of each respondent to vary with their party, individual characteristics such as age or race, their turnout history in primary and general elections, and the geography in which they reside. Their congressional district is related to their political preferences not only through an intercept shift, but also interacted with their turnout history, allowing primary and general election voters in each congressional district and of each party to have different coefficients relating turnout to their preferences. Each Democratic, Republican, and independent primary and general electorate in each congressional district has a different intercept for conservatism, holding constant individual demographics. I specify the effects in this way because turnout in congressional primaries may vary with the candidates on offer as well as the primary institution and local party dynamics, which vary by state and district.

To map the hierarchical model of the individual preferences of voters from each electorate in each congressional district, I apply the following procedure:

Note that in states with more open primary rules, I do not know in which primary the decline-to-state/other party registrants with a validated primary turnout record voted. Unlike turnout, party of registration is self-reported in the 2010 CCES such that, even in states without party of registration, I partition respondents by their self-report party of registration. I replicate this choices in the 2012 CCES. Thus in closed states I am likely better able to characterize the full primary electorate because of the required party registration.

Table A2: Procedure to estimate district-electorate preferences

1. Apply the estimated individual-level coefficients ($\hat{\beta}$ and $\hat{\alpha}_{c(i)}$) from the hierarchical models by party of registration along with the posterior mixed (random) effects ($\hat{\delta}_{c(i)}$) to each respondent. This creates the modeled conservatism for each respondent. For example, a Democratic registrant who voted only in the 2010 general election from New York's 22nd Congressional district would have predicted conservatism given her set of demographics and turnout history and the coefficients on age, gender, marital status, etc., plus the geography shift for Democratic general election voters in New York state, plus the geography shift for Democratic general election voters in New York's 22nd district.
2. Average the predicted conservatism of validated voters in the target election to each congressional district using the CCES post-stratification weights for those validated voters. For example, if the target electorate is general election voters in NY-22, take all of the survey respondents with validated vote in that election, Democrat, Republican, and other party, and calculate the weighted average of their predicted conservatism with CCES survey weights. This weighted average is my estimate of the conservatism of NY-22 2010 general election voters. To calculate the conservatism of NY-22 Democratic primary voters, calculate the weighted average only for those respondents with Democratic party of registration and with validated vote in the 2010 primary.

A.2 Connection to Multi-level Regression with Post-Stratification

The procedure in Table A2 above, using hierarchical models and post-stratification weights to aggregate to subnational geographies, is related to a line of recent work on estimating the policy opinions of subnational electorates using hierarchical models (e.g. Gelman and Little, 1997; Park, Gelman, and Bafumi, 2004; Lax and Phillips, 2009; Warshaw and Rodden, 2012). These methods, called multi-level regression with post-stratification (MRP), use post-stratification along with the hierarchical model to improve estimation. The hierarchical model helps improve the estimation of small-area quantities, essentially by shrinking small-area estimates towards larger areas in which they are nested. Post-stratification is a standard statistical adjustment for nonresponse bias in sample surveys. The combination of the two improves estimates of small-area opinion (e.g., Gelman and Little, 1997; Park, Gelman, and Bafumi, 2004; Lax and Phillips, 2009; Warshaw and Rodden, 2012).

However, I do not exactly follow MRP because I do not have Census targets for turnout in primary elections. Other research designs applying MRP create estimates based on known joint distributions of Census demographics at subnational levels.¹ The Census, however, does not provide joint distributions of demographics for primary electorates in each House district. What I have instead is the CCES data with post-stratification weights. The CCES weights post-stratify to the Census distributions (along with other known targets) in each state. In addition, I have the validated primary turnout for each respondent. I use the joint distribution of the CCES post-stratification

¹ Though, some applications also use external targets, such as religion, that are not included in the Census distribution, and others use the Census Current Population Survey product, which is subject to sampling error.

weights with the validated primary turnout as the target joint distribution for primary voters, just as MRP uses the Census joint distribution. My joint distribution is based upon the CCES sample, and so if it is not as high in quality as the Census estimates, my targets may be less effective at the post-stratification stage. 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 survey estimates to corresponding population statistics. I present efforts below to validate the approach.

B Implementation

In this section I describe the data to which I apply the procedure described in Table A2 above, the details of the implementation, and provide efforts to validate the procedure.

B.1 Data sources

The data source to which I apply this model is the 2010 and 2012 CCES (Ansolabehere, 2010, 2012). Both surveys are a nationally representative sample of around 55,000 Americans with interviews before and after the 2010 midterm and 2012 presidential elections, stratified by state. The surveys asked standard sets of political questions about attitudes, preferences, and beliefs, and behaviors such as vote choice and turnout. The survey also validated turnout records by matching respondents to administrative records instead of relying only on self-reported turnout.

B.2 IRT model of conservatism

I describe in Section A above how to take individual-level measures of preferences and map them into best district estimates through a hierarchical model and stratification weights. In this section, I describe how I measure individual-level preferences. I aggregate responses to multiple policy questions into a single summary value of conservatism through an item-response theory (IRT) model. Aggregating across multiple responses mitigates measurement error and mimics the aggregation across roll call votes used to summarize roll call voting behavior in congress with NOMINATE.

To characterize the policy conservatism for each respondent to the 2010 and 2012 CCES, I estimate the grouped IRT model proposed by Lewis (2001) on respondent expressed preferences over a set of policy issues. Each CCES asked respondents how they would vote on a set of roll calls actually considered in the House and Senate, as well as other policy preferences not specific to any roll call vote. I identified 17 questions from each survey that serve as the items in the model. For 2012, the questions query preferences about abortion, the war in Iraq, environment vs jobs, the Ryan budget, the Middle Class Tax Cut, the Tax Hike Prevention Act, birth control exemption, Keystone pipeline, the Affordable Care Act, ending Don't Ask Don't Tell, the war in Afghanistan, gun control, climate change, immigration, gay marriage, affirmative action, and balancing the budget. For 2010, the questions query preferences about gun control, climate change, immigration, abortion, environment vs jobs, gay marriage, affirmative action, balancing the budget, the Stimulus, SCHIP, Carbon Tax, Affordable Care Act, Kagan nomination to the Supreme Court, Dodd Frank Act, ending Don't Ask Don't Tell, funding stem cell research, and the Troubled Asset Relief Program.

The Lewis (2001) model allows categorical, rather than binary responses, so I use all response categories available in the CCES on these questions. The model estimates group-specific intercept shifts and variances for the distributions of respondent ideal points. I group respondent ideal points by the intersection of three characteristics: their state of residence, their partisanship (coded three

ways, with leaners collapsed as partisans), and their primary turnout. Thus, for example, the model can estimate a different intercept and variance for the ideal points for respondents from New York who identify as Democrats and voted in the 2010 congressional primary relative to respondents from New York who identify as Democrats and did not vote in the primary.

The model estimates the item parameters and the group distributions. To calculate the ideology of each individual, I calculate the expected a posteriori ideal point for each respondent, conditional on their responses, the estimated item parameters, and their group membership (see Lewis, 2001, p. 279 for details). As with all ideal point models, the latent scale of ideology is only identified up to an affine transformation. I post-process the ideal points to have mean zero and unit variance for each survey.² My 2010 estimates correlate with the IRT estimates of Tausanovitch and Warshaw (2013) for 2010 CCES respondents at $r = 0.961$.

The ideology estimates correspond well to what one would think ideology should relate to. They correlate with partisanship and self-reported ideology, and they are superior predictors of vote choice than self-reported ideology: The R^2 of a linear model predicting Democratic House vote in 2010 with my estimate of ideology, partisanship, and state fixed effects is 0.87, compared to 0.85 with self-reported ideology (don't know respondents set to moderate). When both self-reported ideology and my measure are included in this model, the coefficient on self-reported ideology is 28 percent of its size without my measure of conservatism in the model, while the coefficient on my measure of conservatism is 88 percent of its original size. These same numbers for 2012 are R^2 of 0.78 versus 0.76, and coefficient ratios of 18 percent self-reported versus 96 percent my estimate.³ This suggests my estimate of conservatism is more closely related to vote choices than self-reported ideology, and so is more closely related to the preferences that motivate member behavior.

B.3 Procedure for district estimates

To estimate the conservatism of each primary and general electorate in each district, I implement the procedure in Table A2. I first estimate the hierarchical model of individual conservatism on the IRT results. In the hierarchical model, I use as individual characteristics in the model a standard set of variables used in political models: age, gender, race, marital status, church attendance, income ordinal and income missing, vote in the 2010 (2012) general, vote in the 2010 (2012) congressional) primary, and voted in both.⁴ I let the coefficients on the three turnout indicators vary by congressional district and state through random effects, and let the intercept term vary by congressional district, state, and region through random effects. I estimate the model separately for each of three types of party of registration, Democrat, Republican, and Decline to State/other party.⁵

In Table A3, I present the model R^2 from the individual hierarchical models. The models ex-

² Other estimation choices required for the ? EM implementation are number of quadrature points (for the approximation of group normal distributions), and convergence criteria. I use five quadrature points and iterate the EM algorithm until the maximum parameter change across all item and group parameters is less than $1e-5$.

³ Regression results available from the author on request.

⁴ I chose demographic variables based on the variables used by the CCES to construct its post-stratification survey weights. These are also a standard set of demographics used in studies of voting and turnout.

⁵ I estimate the hierarchical model using the R package `lme4` (R Development Core Team, 2015; Bates, Maechler, and Bolker, 2013). Coefficient estimates are consistent with other models of conservatism: males, the married, and the less secular are more conservative. I estimate primary voters are more conservative (liberal) in the Republican (Democratic) models.

plain a reasonable amount of variance in the individual scores within-party. Across-party variance explained is higher, at 0.60 in 2010 and 0.58 in 2012.⁶

Table A3: Hierarchical model fit

Year	Party	Model R ²
2010	Dem	0.19
2010	Rep	0.19
2010	DTS/Oth	0.21
2010	All	0.60
2012	Dem	0.22
2012	Rep	0.21
2012	DTS/Oth	0.21
2012	All	0.58

Note: Cell entries present the coefficients of determination for each mixed model by year and party of registration.

I next use the model estimates to calculate predicted values for each respondent in the survey. This replaces the respondent’s IRT-estimated conservatism with the hierarchical model predicted conservatism. Then, for each electorate, I take the weighted average of the predicted preferences using the survey weights.⁷ This is similar to the final step of MRP, where opinion is imputed from the hierarchical model to the known Census targets. For example, to calculate the conservatism of the Republican primary electorate in the 7th district of Illinois, I average the predicted conservatism of all the respondents surveyed in IL-07 who are validated to have voted in the Republican primary, weighted by each survey weight. Note that the set of Republican primary voters includes both primary-only voters and primary voters who also voted in the general election. Likewise, my estimate of the general electorate includes the conservatism of general-only voters and general voters who also participated in the primary. This procedure provides an estimate of the preferences for each electorate that pools and smooths observations through the hierarchical model and stratifies to targets through the survey weights and validated turnout.

For states with open or not-fully-closed primary elections, I am uncertain in which primary each validated primary voter voted. To make estimates somewhat consistent across states, I use respondents’ self-reported party of registration to construct estimates. That is, no matter the primary institution in place, Democratic primary electorate estimates are the weighted average of voters validated to have voted in the primary and who report being registered Democrats, and Republican primary electorate estimates are the weighted average of voter validated to have voted in the primary and who report being registered Republican. Clearly this choice induces measurement error in my estimates. To the extent this error might bias my estimate of the relationship, it seems more

⁶ Full coefficient estimates and estimates of variance of the mixed effect distributions are available from the author on request or from the replication archive.

⁷ Formally, if \hat{y}_i is the predicted conservatism for respondent i given i ’s party, demographics, turnout history, and geography, then my estimate of the conservatism for electorate e in congressional district c is $\sum_{i \in e, c} w_i^{-1} \times \sum_{i \in e, c} \hat{y}_i w_i$, where w_i is the survey weight for respondent i and $i \in e, c$ evaluates to the set of respondents i validated to have voted in primary or general election e and residing in district c .

likely to do so in states with non-closed primaries, because I may be missing many non-Democrat and non-Republican registrants who actually participated in these primaries.⁸

B.4 How the model changes district estimates

To present intuition about how the hierarchical model estimates differ from the raw survey aggregates, I plot the district-electorate estimates against each other in Figures A1 and A2. Each figure compares the modeled versus raw district estimates for general electorates, Democratic primary electorates, and Republican primary electorates, separately for 2010 and 2012. The x-axis in each frame is the simple weighted mean respondent conservatism aggregated to each electorate in each congressional district.⁹ On the y-axis is the electorate conservatism estimated by the hierarchical model and the aggregation procedure of Table A2. Each circle is a district, and point size is proportional to the number of CCES respondents in that district matching that electorate characteristic (general voters or primary voters of that party of registration).

The figures highlight three key features of my procedure. First, when there is enough data, as in most of the districts in the two general election frames, the modeled estimates do not differ much from the raw estimates. In each frame, the dashed line is the 45 degree line indicating perfect correspondence, and in the general election frames the circles cluster around this line. Also apparent in the general election frames, but more so in the primary frames, is the slope being less than 1: hierarchical models shrink estimates towards the overall mean when there is not enough data. Because the primary electorates have smaller samples, it is apparent that the shrinkage (the attenuation of the slope) is larger for the primary district electorates.

The third key feature is the re-scaling of district estimates for outlying cases with small sample sizes. In each of the primary electorate frames, outlying cases along the x-axis, those with raw district estimates far away from most of the other primary electorates of that party, are brought towards the center of the distribution by the hierarchical procedure. For example, in the primary electorate frame for Republicans in 2010, one district electorate far to the left has a raw aggregated ideology of -0.8, or one standard deviation more liberal than average. This district is NY-07, and the CCES has one validated Republican primary voter in that district. The hierarchical model shrinks the district estimate to about 0.6, certainly on the liberal side of Republican primary electorates, but nowhere near as outlying as the raw estimate. Similar shrinkage is applied for other outlying raw estimates for both Democrat and Republican validated primary voters. This is exactly what the hierarchical model is supposed to do. This shrinkage lessens the leverage outlying cases have on parameter estimates and improves my statistical power to estimate the relationship between district primary preferences and member roll call behavior.

C Validation of measure and improvement over raw survey margins

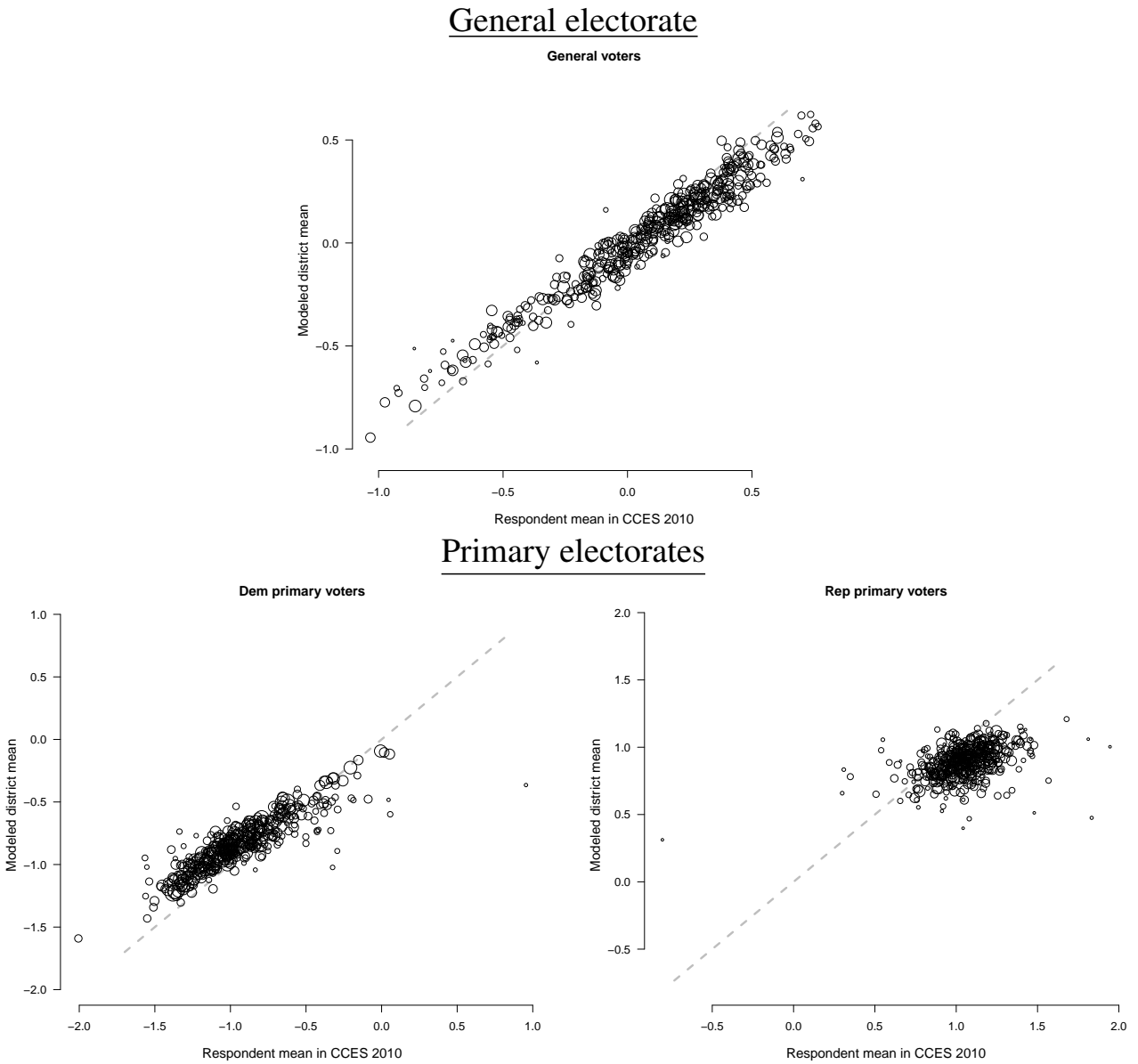
In this section, I present evidence on the validity of my estimates and then present some evidence of improvement through the hierarchical model over raw survey aggregates.

To demonstrate that my estimates of electorate conservatism in each district have reasonable face validity, I compare the estimated conservatism of general electorates to standard measures

⁸ In the two surveys, 18.8 percent (2010) and 19.7 percent (2012) of validated congressional primary voters reported being registered Decline-To-State or third party.

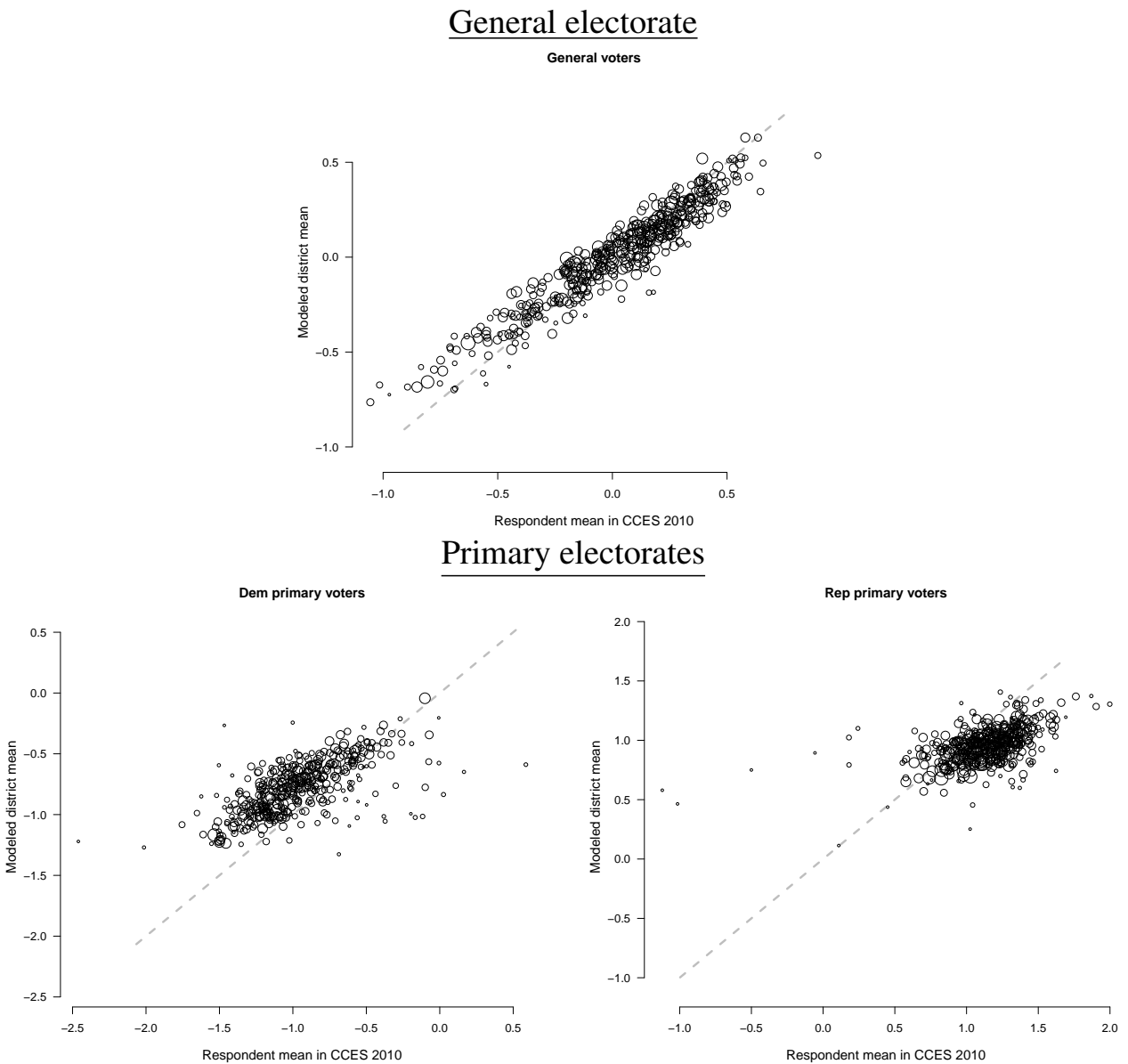
⁹ Formally, if y_i is the conservatism for respondent i , then my estimate of the conservatism for electorate e in congressional district c is $\sum_{i \in e, c} w_i^{-1} \times \sum_{i \in e, c} y_i w_i$, where w_i is the survey weight for respondent i and $i \in e, c$ evaluates to the set of respondents i validated to have voted in primary or general election e and residing in district c .

Figure A1: Hierarchical versus raw district estimates of electorates, 2010



Note: Each point is the estimated conservatism of a set of voters in one congressional district. The x-axis is the survey-weighted aggregate, the y-axis the result of the hierarchical estimation procedure. Point size proportional to the number of CCES respondents in that district-electorate.

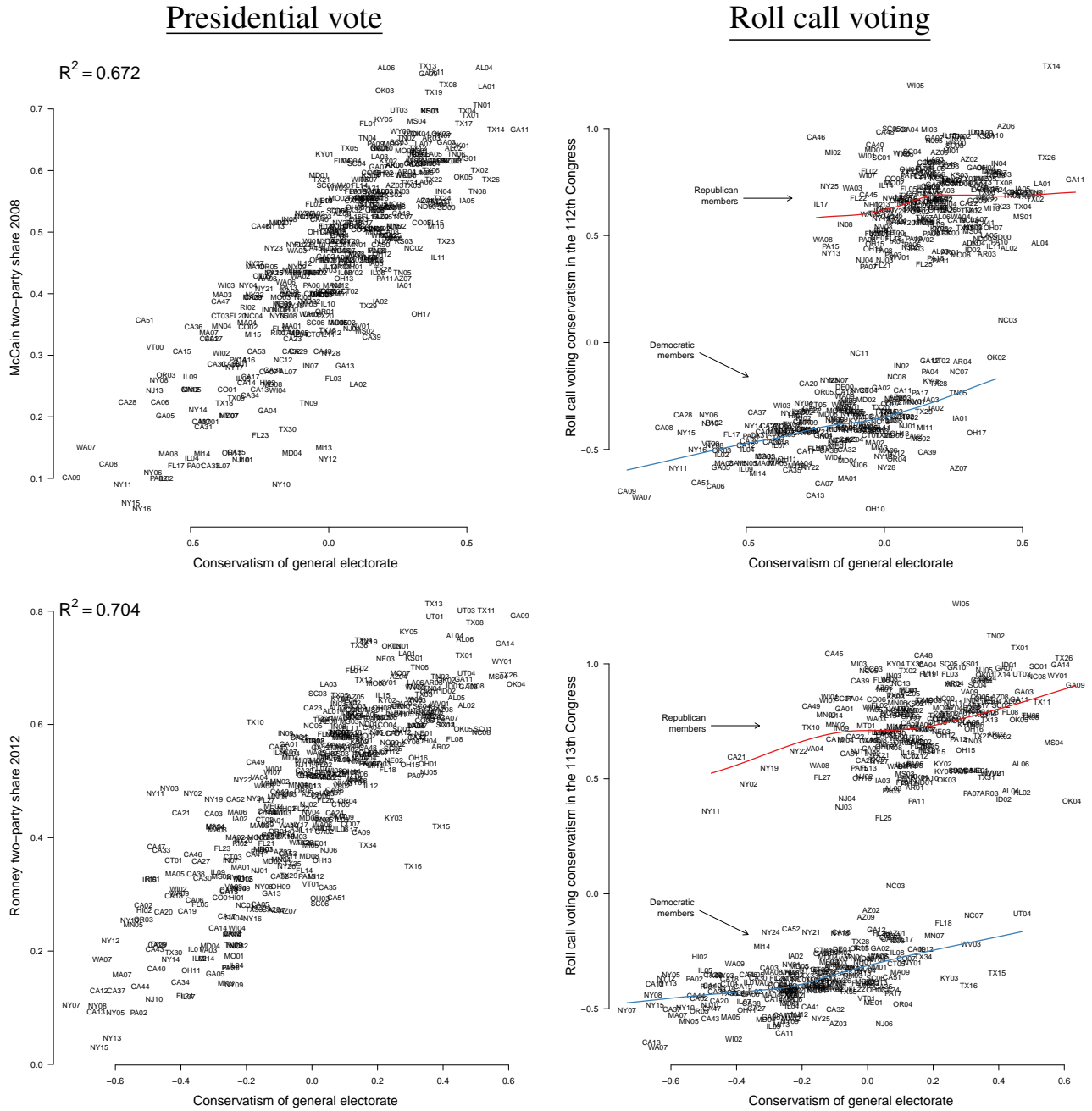
Figure A2: Hierarchical versus raw district estimates of electorates, 2012



Note: Each point is the estimated conservatism of a set of voters in one congressional district. The x-axis is the survey-weighted aggregate, the y-axis the result of the hierarchical estimation procedure. Point size proportional to the number of CCES respondents in that district-electorate.

of political preferences and representative behavior. Presidential vote is often used as a summary of the political preferences of a district. In the first column of Figure A3, I compare presidential vote share in 2008 and 2012 in the congressional district to my estimate of general election voter conservatism in that district and election. In the right column, I compare DW-NOMINATE score of the member of congress elected out of that district to the 112th (top frame) or 113th (bottom frame) House to the estimates of electorate conservatism in that election. Both frames suggest good validity to my estimates. Presidential vote share tracks electorate conservatism in a linear fashion with R^2 of 0.67 and 0.70, suggesting a good connection between my estimates of district conservatism and voting behavior in that district. The plot of member NOMINATE scores is common for modern congresses, with some responsiveness of Democrats to electorate conservatism, smaller responsiveness of Republicans to electorate conservatism, and a large intercept shift between Democratic and Republican members of the House. These plots suggest that my estimation procedure is capturing similar relationships as other measures of preferences used to measure the responsiveness of the House of Representatives (e.g., Ansolabehere, Snyder, Jr., and Stewart III, 2001; Clinton, 2006).

Figure A3: Face validity of hierarchical model estimates of district conservatism



Note: The x-axis in each frame is the estimated conservatism of the general electorate in each congressional district. The left column compares the estimates to 2008 and 2012 district presidential vote for the Republican. The right column compares the estimates to the roll call voting behavior of the House member in the following 112th and 113th Congresses summarized by DW-NOMINATE score.

C.1 Testing the two estimates against each other

Figure A3 presents face validity to the hierarchical estimates of district conservatism. To evaluate improvement in fit using the modeled estimates over the raw survey aggregates, I present six regression models in Table A4. The dependent variable is two-party Republican presidential vote share in the district (actual votes cast, not estimated from the CCES), with 2008 McCain vote matched to 2010 CCES districts (as there was no presidential election as a benchmark in 2010), and with 2012 Romney vote matched to 2012 CCES districts. For each year, I present three specifications, one with only modeled conservatism, one with only raw survey aggregated conservatism, and one with both terms. All models include state fixed effects to mimic my main analysis of member roll call voting behavior, which include state fixed effects.

Table A4: Relationship of raw and modeled district conservatism to actual votes

	2008	2008	2008	2012	2012	2012
Intercept	0.57*	0.51*	0.53*	0.54*	0.49*	0.50*
	(0.09)	(0.09)	(0.08)	(0.10)	(0.09)	(0.09)
Survey conservatism	0.35*		0.16*	0.39*		0.11*
	(0.02)		(0.03)	(0.02)		(0.03)
Modeled conservatism		0.43*	0.26*		0.51*	0.40*
		(0.02)	(0.04)		(0.02)	(0.04)
N	424	424	424	435	435	435
R ²	0.70	0.72	0.74	0.67	0.73	0.74
adj. R ²	0.66	0.68	0.70	0.63	0.70	0.71
Resid. sd	0.09	0.08	0.08	0.10	0.09	0.09

Standard errors in parentheses

* indicates significance at $p < 0.05$

Note: Dependent variable is Republican two-party presidential vote share in the district, with OLS coefficients and standard errors in parentheses. All models include state fixed effects. The 2008 Republican vote share is modeled with district conservatism estimates from the 2010 CCES, which may be one reason the coefficients are of smaller magnitude relative to 2012.

The results of Table A4 suggest the modeled estimates are closer to actual voting behavior than the raw survey aggregates. While the raw survey estimates do predict presidential vote share, with coefficients of 0.35 and 0.39 and adjusted R² of 0.66 and 0.63 (columns one and four), the modeled estimates do better – coefficients of 0.43 and 0.51, adjusted R² of 0.68 and 0.70 (columns two and five). When both estimates are included in the specification, the coefficient on the raw score is attenuated much more than the coefficient on the modeled score. The raw coefficients attenuate from 0.35 to 0.16 and 0.39 to 0.11 while the modeled coefficients attenuate from 0.43 to 0.26 and 0.51 to 0.40, respectively. Overall, the data suggest that modeled conservatism is a better predictor of actual presidential votes cast.

Note that the improvement of fit suggested in Table A4 is for the estimates of the preferences of the general electorates, which I noted in Figures A1 and A2 are less changed than estimates of primary electorates. Unfortunately, I do not possess an obvious national benchmark for primary

electorates along the lines of presidential vote share to validate. However, if we assume that the hierarchical procedure is doing more to alleviate measurement error for the primary electorates than for the general electorates, as is consistent with the smaller sample sizes, then the improvement of fit for general electorates presented in Table A4 would be a lower bound on the improvement of fit for primary electorates.

References

- Ansolabehere, Stephen. 2010. *Cooperative Congressional Election Study, 2010: Common Content*. [Computer File]. Cambridge, MA: Harvard University.
- Ansolabehere, Stephen. 2012. *Cooperative Congressional Election Study, 2012: Common Content*. [Computer File]. Cambridge, MA: Harvard University.
- Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45(1): 136–159.
- Bates, Douglas, Martin Maechler, and Ben Bolker. 2013. *lme4: Linear mixed-effects models using Eigen and syntax*.
- Clinton, Joshua D. 2006. "Representation in Congress: Constituents and Roll Calls in the 106th House." *Journal of Politics* 68(2): 397–409.
- Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23(2): 127–135.
- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. West Sussex: John Wiley & Sons.
- Lax, Jeffrey R., and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 107–121.
- Lewis, Jeffrey B. 2001. "Estimating Voter Preference Distributions from Individual-Level Voting Data." *Political Analysis* 9(3): 275–297.
- Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12(4): 375–385.
- R Development Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2): 330–342.
- Warshaw, Christopher, and Jonathan Rodden. 2012. "How Should We Measure District-Level Opinion on Individual Issues?" *Journal of Politics* 74(1): 203–219.