

# Supplementary Appendix

## Two-Step To Win: Balancing Representation of Primary and General Electorates in Congress

Seth J. Hill  
University of California, San Diego\*

February 19, 2015

### Contents

<b>A</b>	<b>List of survey questions not asked as roll call votes mapped to roll call votes</b>	<b>2</b>
<b>B</b>	<b>Specification of hierarchical model</b>	<b>4</b>
<b>C</b>	<b>Validation of hierarchical model</b>	<b>5</b>
<b>D</b>	<b>Additional figures for both types of questions</b>	<b>6</b>
<b>E</b>	<b>Survey questions asked as roll call votes</b>	<b>6</b>
<b>F</b>	<b>Survey questions of attitudes related to specific roll call votes</b>	<b>6</b>

---

\*Department of Political Science, 9500 Gilman Drive #0521, La Jolla, CA 92093-0521; [sjhill@ucsd.edu](mailto:sjhill@ucsd.edu), <http://www.sethjhill.com>.

## **A List of survey questions not asked as roll call votes mapped to roll call votes**

In Table A1, I present the CCES survey questions not specified as roll call votes that I use in this project, and link each to the roll call vote analyzed. For example, the second row shows that I use a CCES survey question on attitudes toward abortion to proxy for district preferences on House Amendment 95, which would prohibit federal funding for Planned Parenthood.

Table A1: Mapping from roll call votes to CCES survey questions that were not specified as roll calls

CCES Var no.	CCES question	House bill	Date	Roll call no.	House
CC306	All things considered, do you think it was a mistake to invade Afghanistan?	Amdt 4 to conference appropriations bill HR4899 to set a timetable for withdrawal from Afghanistan, HR 4899	7/1/2010	432	111th
CC324	Which one of the opinions on this page best agrees with your view on abortion? (1) By law, abortion should never be permitted; (2) The law should permit abortion only in case of rape, incest, or when the womans life is in danger; (3) The law should permit abortion for reasons other than rape, incest, or danger to the womans life, but only after the need for the abortion has been clearly established; (4) By law, a woman should always be able to obtain an abortion as a matter of personal choice	Prohibit federal funding for Planned Par- enthood, H Amdt 95	2/18/2011	93	112th
CC320	In general, do you feel that the laws covering the sale of firearms should be made more strict, less strict, or kept as they are?	State reciprocity for carrying concealed firearms, HR822	11/16/2011	852	112th
CC328/9	The federal budget deficit is approximately \$1 trillion this year. If the Congress were to balance the budget it would have to consider cutting defense spending, cutting domestic spending (such as Medicare and Social Security), or raising taxes to cover the deficit. What would you most prefer that Congress do - cut domestic spending, cut defense spending, or raise taxes? AND What would you least prefer that Congress do raise taxes? [Coded -1 if response to first question was raise taxes, 1 if response to second question was raise taxes, 0 otherwise.]	Fiscal cliff deal, HR8	1/1/2013	659	112th
CC325	Some people think it is important to protect the environment even if it costs some jobs or otherwise reduces our standard of living. Other people think that protecting the environment is not as important as maintaining jobs and our standard of living. Which is closer to the way you feel?	Offshore energy and jobs act, HR 2231	6/28/2013	304	113th
CC424	What is your view of the Tea Party movement – would you say it is very pos- itive, somewhat positive, neutral, somewhat negative, or very negative or don't you know enough about the Tea Party movement to say?	End shutdown, HR2775	10/16/2013	550	113th
CC328/9	[Same as CC328/9 above.]	FY14 Budget \$1.1 trillion, HR3547	1/15/2014	21	113th
CC305	All things considered, do you think it was a mistake to invade Iraq?	Prohibit combat ops in Iraq	6/19/2014	325	113th

Note: Matches compiled by the author.

## B Specification of hierarchical model

To estimate the policy attitudes or support for each roll call vote cross party of each primary and general electorate in each district, I begin with a hierarchical model of individual policy preferences. I include 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. 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 attitudes or support  $y_i$  within each party of registration with respondents nested within districts within states within Census regions,

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

where  $y_i$  is the attitude or support for a specific roll call vote of respondent  $i$ ,  $\mu_i$  is the expected value of  $y_i$  and  $\sigma_y^2$  is the variance of the distribution of  $y_i$ .  $\beta$  is a vector of coefficients mapping individual characteristics  $\mathbf{x}_i$  to the expected value of  $y_i$ ,  $\alpha_{c(i)}$  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  $\alpha_{c(i)}$  distributed multivariate normal with means specific to the state of that district through the function  $s(c)$  and covariance matrix  $\Sigma_c$ . The state means for the turnout coefficients,  $\alpha_{s(c)}$ , are distributed multivariate normal with mean vector  $\bar{\alpha}$  and covariance matrix  $\Sigma_s$ . 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  $\sigma_c^2$ , which are nested within regions with mean  $\delta_{r(s)}$  and variance  $\sigma_s^2$ . The region means are distributed normal with mean 0 and variance  $\sigma_r^2$ .

To move from the individual model to the preferences of primary and general electorates in each district, I take the predicted value for each individual given the model results, and aggregate those predicted values up to the electorate. The procedure is:

1. Apply the estimated individual-level coefficients by party of registration along with the posterior geography and turnout history effects to each respondent. This creates the predicted policy preference for each respondent, which is related to the actual preference they expressed but smoothed given their demographics, geography, and turnout history.

2. Average the predicted policy preferences of validated voters in the target election to each congressional district using the CCES survey 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 policy preferences with CCES survey weights. This weighted average is my estimate of the preferences of NY-22 2010 general election voters. To calculate the preferences 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.

This second step is very close to survey post-stratification, except that we don't know exactly the demographics of the primary or general electorates. Instead, CCES observes the validated turnout for a set of its respondents, which if it is a representative survey will approximate the demographics of the actual electorate. Because the CCES weights are post-stratification weights to Census demographics, the combination of the two creates close to, but not exactly, a post-stratification to the voting electorate. In the next section, I present a validation of the procedure.

## **C Validation of hierarchical model**

In this section, I present evidence that the hierarchical model that I use to create estimates of the preferences of each electorate in each district improves accuracy over raw survey totals. I select two issues from the 2012 CCES that are strongly related to citizen vote choice in the 2012 presidential election between Barack Obama and Mitt Romney. The Affordable Care Act (ACA) was the most salient policy item of the first term of the Obama presidency, and a large dividing issue between supporters and opponents of President Obama. Likewise, the Tea Party emerged in 2010 in strong opposition to the policies of President Obama, notably the stimulus and the ACA. Not surprisingly, these two survey responses from the pre-election survey are the most strongly related of all the issues I consider from the 2012 CCES to the post-election report by respondents of their 2012 presidential vote.

To validate the improvement of the model, I relate the CCES survey responses about support for the ACA and Tea Party favorability to presidential vote in each district – actual votes cast, not from the CCES survey responses. I compare estimates of the preferences of the general electorate (validated general election voters) in each district on each issue to Obama two-party presidential vote in the district. I estimate the preferences of the electorate in two ways. First, I use the simple survey-weighted average from the CCES data on each issue. Second, I implement my hierarchical model and procedure to calculate best estimates of the preferences of each congressional district's general electorate, the details of which I present above in Section B. If my procedure is improving accuracy of the estimation of electorate preferences, I should observe a stronger relationship between presidential vote and the modeled electorate preference than between presidential vote and the raw electorate preference.

Figure A1 indeed shows a stronger relationship between my modeled estimates of district preferences and presidential vote than the raw district preferences. The  $R^2$  from regressing vote on preferences improves from 0.55 to 0.68 for raw and modeled estimates of support for the ACA (first column), and from 0.42 to 0.65 for raw and modeled estimates of Tea Party favorability (second column). This shows that my procedure estimates preferences of the electorate in each district that are more closely related to actual votes cast than the simple tabulation of preferences from the

survey data.

## **D Additional figures for both types of questions**

In Figure A2, I compare member votes on 20 roll calls to support for those issues in the members' general electorates by issue and congress. In each frame the y-axis is a yea vote on the bill, and the x-axis is support or attitude on the bill among the general electorate. Because attitudes and support are measured on different scales, the former being some categorical outcome and the latter a proportion from 0 to 1, I standardize both measures to mean zero and standard deviation one for each issue. I plot separate smooth trends using smoothing splines for each bill. The generally positive slopes show that the more the general electorate supports the roll call, or holds attitudes consistent with that roll call, the more likely is the member to vote for that bill.

In the 111th House, all 8 bills show a positive relationship. The smoothed lines present the relationship between the district's distance from mean district support and how likely a member is to vote yea. Looking up from zero on the x-axis shows how likely a member is to vote yea in a district with average level of support for the bill – average support across house districts on that issue. This ranges from about a 20 percent likelihood of setting a timetable to withdraw from Afghanistan to more than 60 percent likelihood of voting for SCHIP.<sup>1</sup> The trends in the 112th and 113th Houses show similar positive relationships but for three roll call votes. The figure suggests a negative relationships between attitude in the general electorate and vote on the fiscal cliff deal at the end of the 112th House and vote on the 2014 budget deal in the 113th House, and a flat, non-relationship between support for Simpson-Bowles budget and support in the 112th. The two bills with a negative relationship use the same survey question asked in the 2010 and 2012 CCES that I mapped to these bills (see Appendix Table A1). While the survey question appears not to have “worked” for the general electorate, I show below a more positive relationship for primary electorates. I also include the roll call so that I am not selecting on a positive relationship and biasing in favor of finding responsiveness – this is the full set of survey questions and roll calls I selected prior to analysis.

## **E Survey questions asked as roll call votes**

I present in Figures A3 to A6 and in Table A2 replication of figures and tables in the main body and Figure A2 for survey questions that were asked specifically about the roll call votes only.

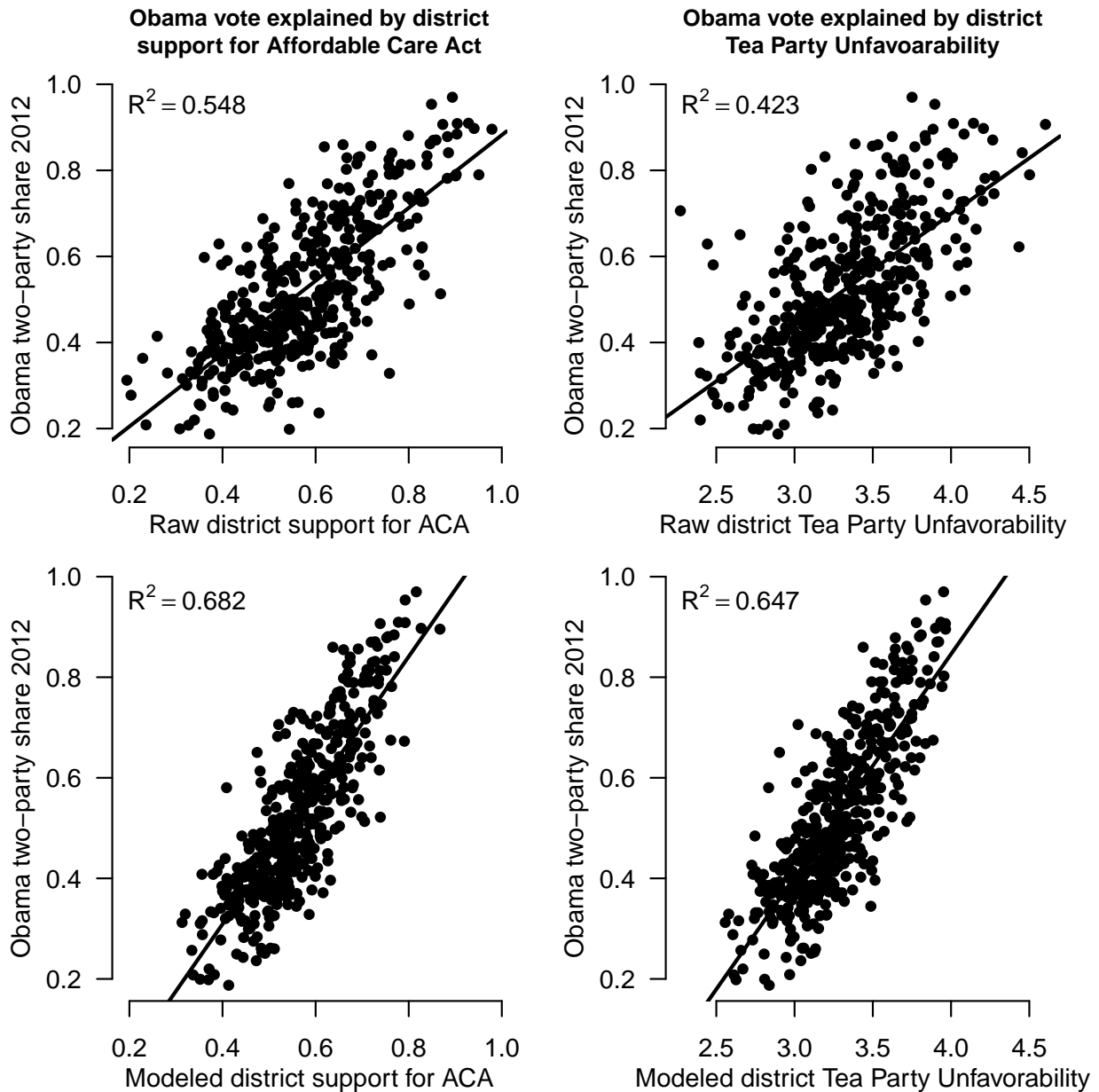
## **F Survey questions of attitudes related to specific roll call votes**

I present in Figures A7 to A9 and in Table A3 replication of figures and tables in the main body and Figure A2 for survey questions asked about respondent attitudes that I use as proxies for roll call votes only.

---

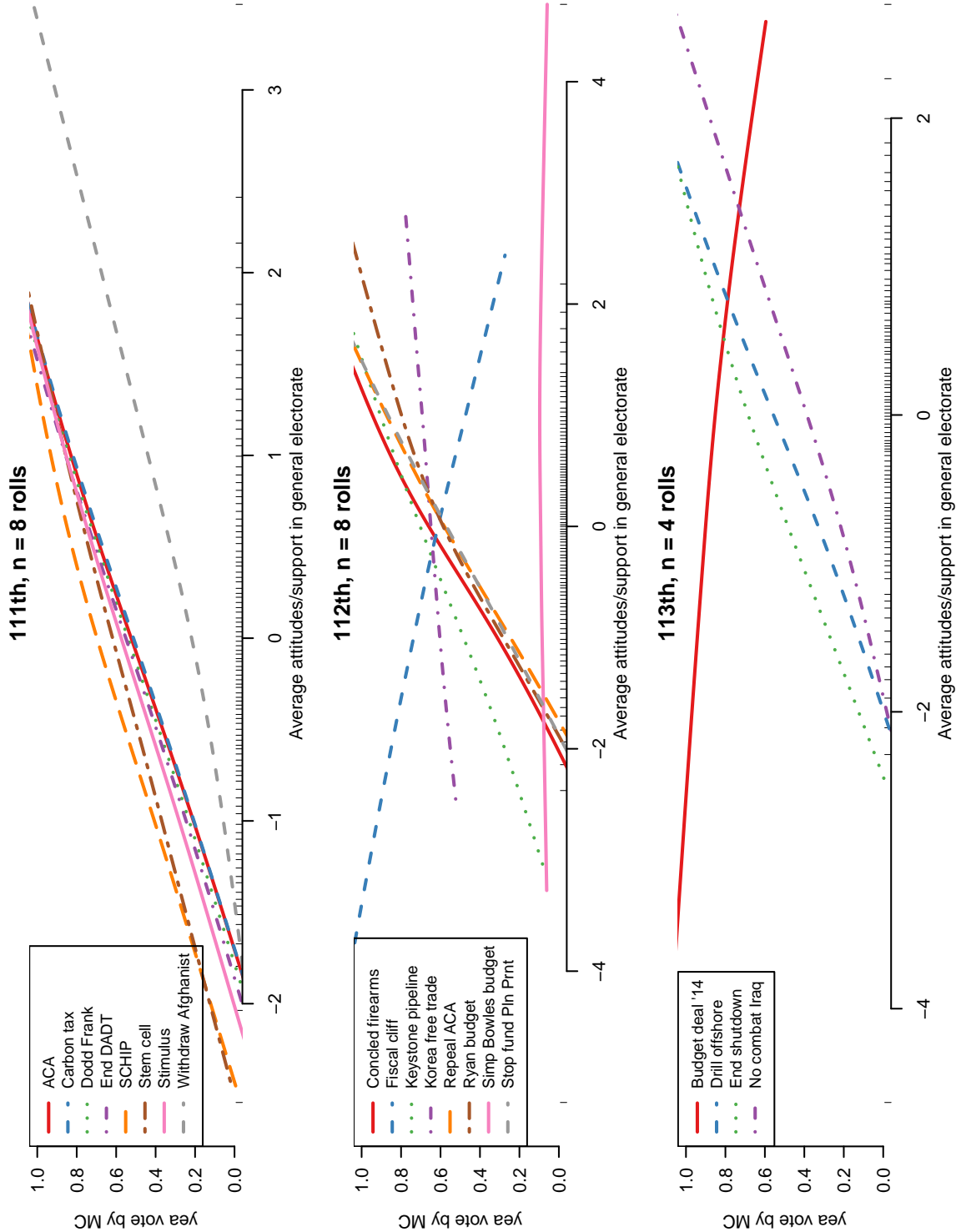
<sup>1</sup> Because I have normalized units, we cannot conclude from this figure that there is more responsiveness on one issue over the other. In Appendix Figure A3, I present this figure for only the questions asked as roll calls on the survey, where support is in comparable units.

Figure A1: Validating improvement to district estimates from the estimation procedure



Note: Each point is one congressional district from the 113th congress. The x-axis is the raw district support/attitude (first row), or the modeled district support/attitude (second row) for the general electorate on each of two issues: the Affordable Care Act roll call vote and favorability towards the Tea Party. These two survey responses are the most strongly related to Obama presidential vote at the individual level of all of the issues I consider in the 2012 CCES. The y-axis is Obama two-party vote share in the district in the 2012 election. The stronger relationship for the modeled electorate preferences suggests that these estimates are a more accurate measure of the electorate's preferences.

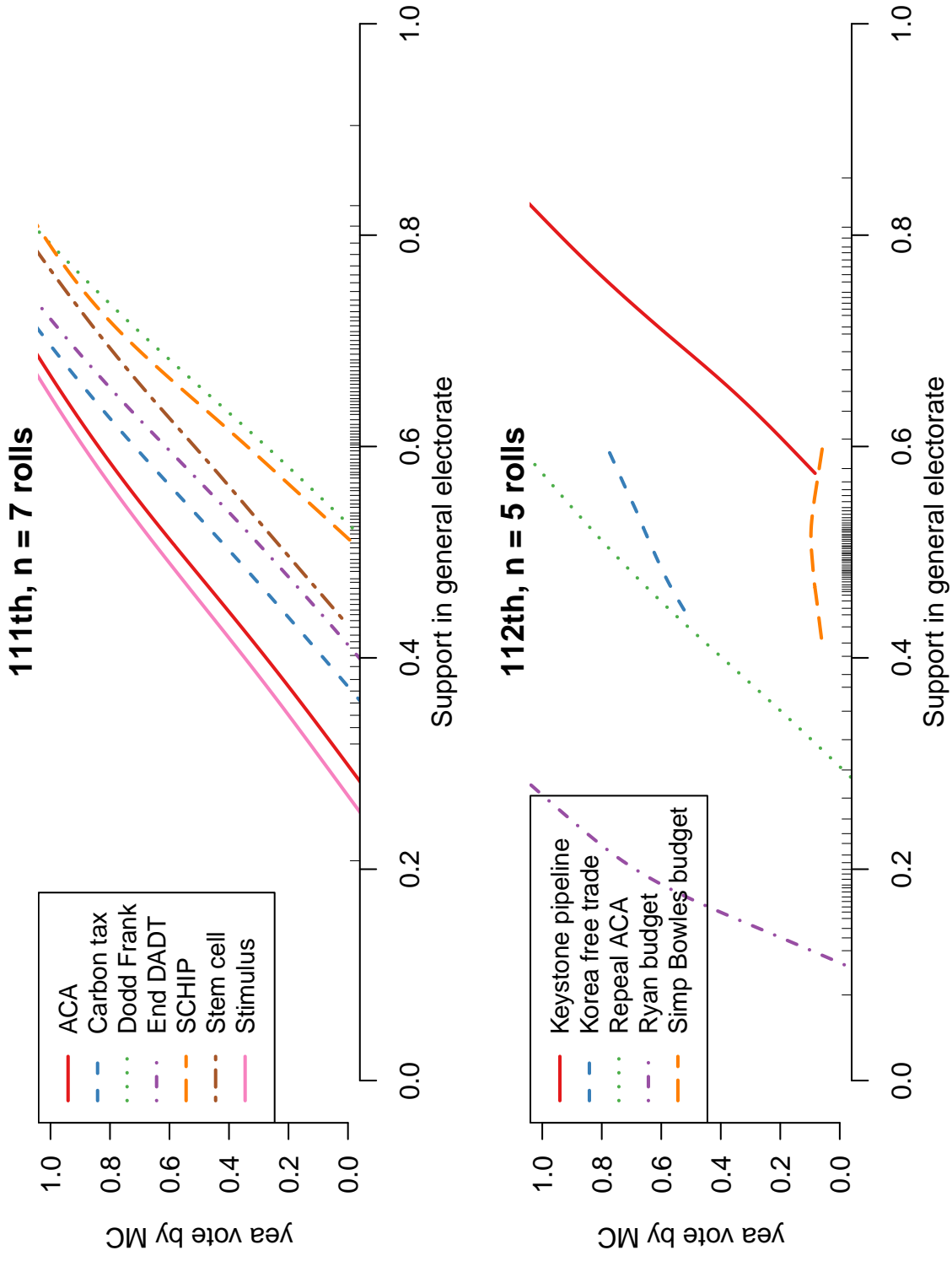
Figure A2: Relationship of representative vote to support in general electorate



Note: The x-axis in each frame is normalized support/attitudes in the house district among general election voters for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated general election voters in the 2010 and 2012 CCES.

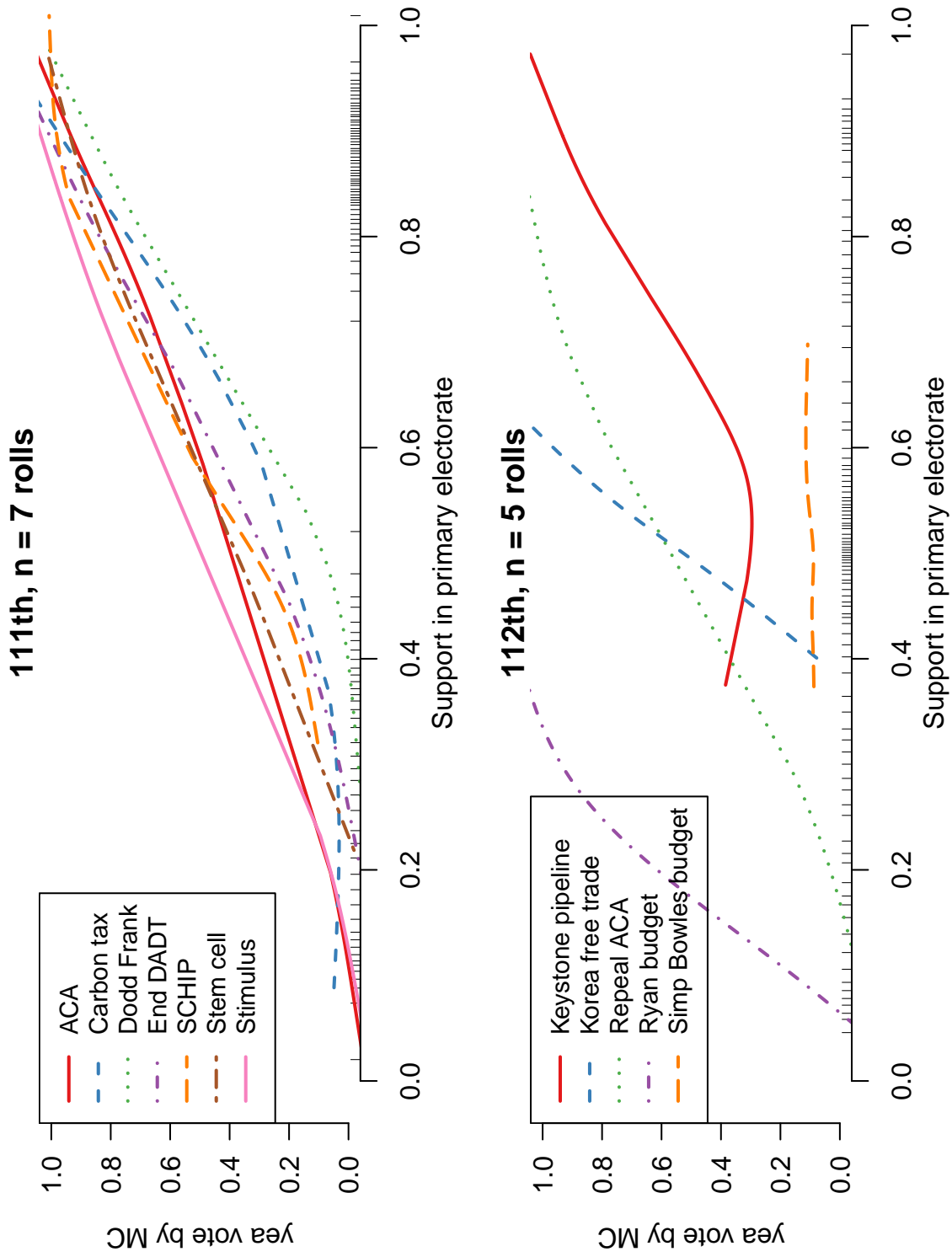


Figure A3: Relationship of representative vote to support in general electorate



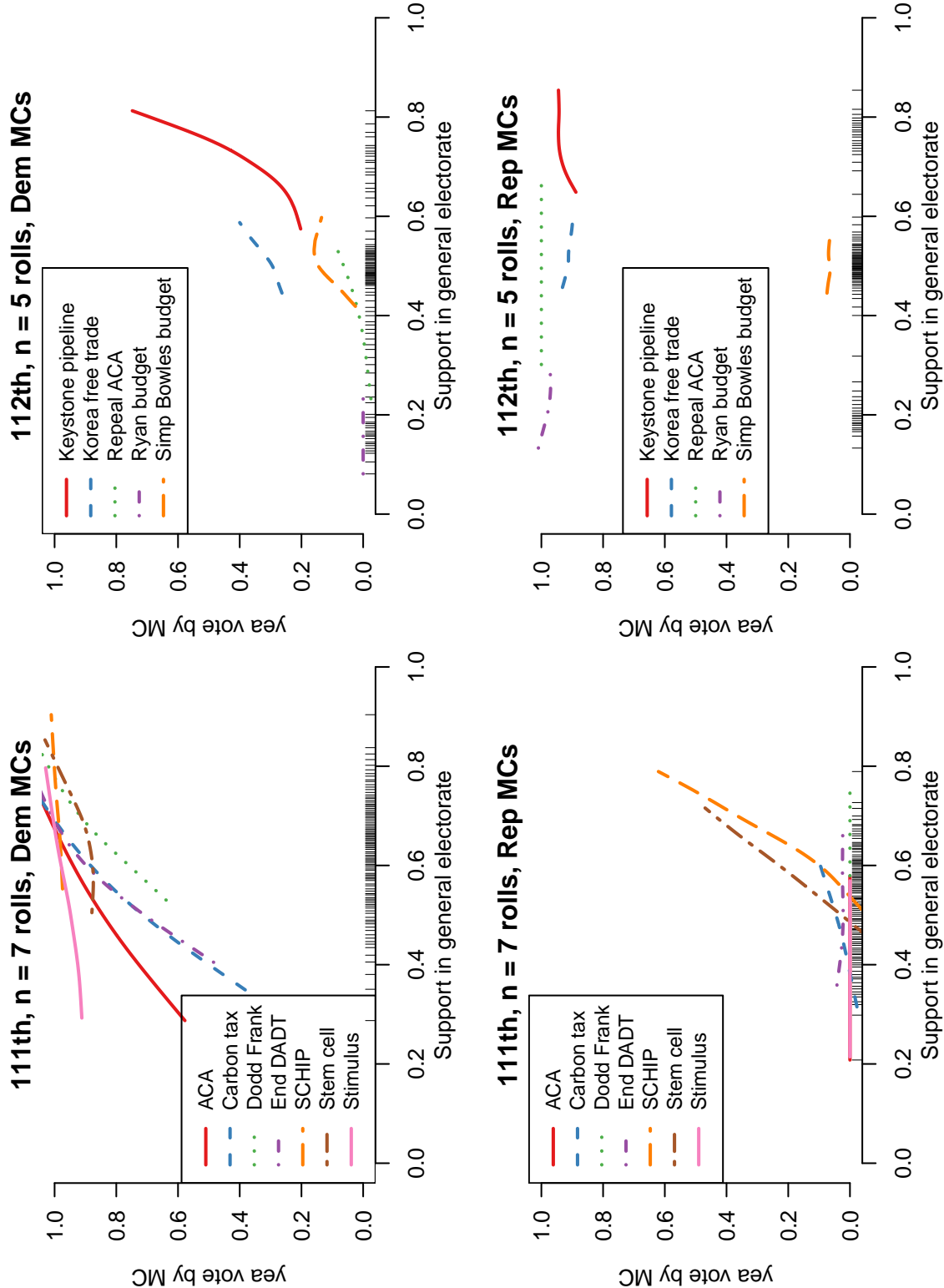
Note: The x-axis in each frame is the estimated support in the house district among general election voters for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated general election voters in the 2010 and 2012 CCES.

Figure A4: Relationship of representative vote to support in primary electorate



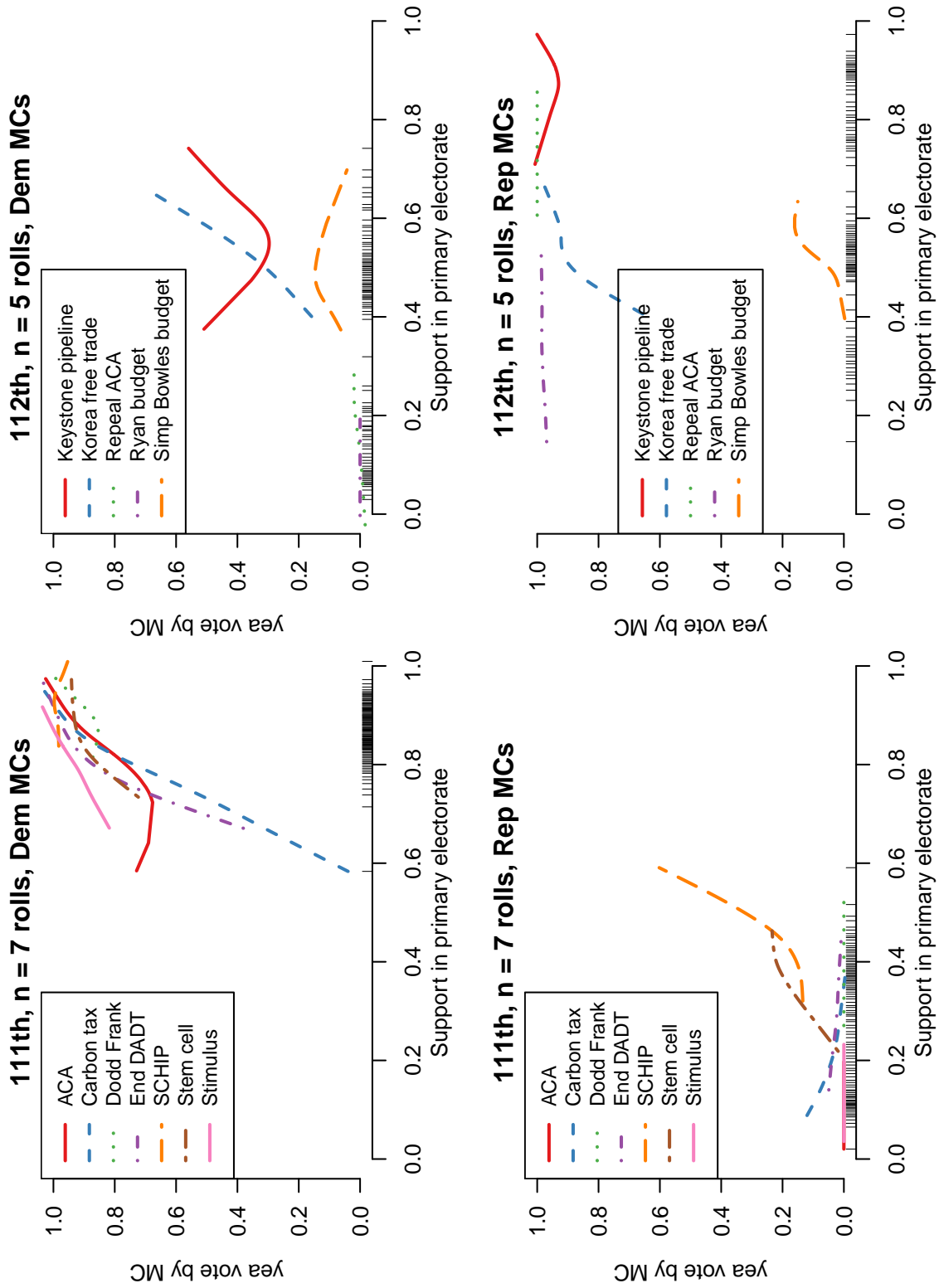
Note: The x-axis in each frame is the estimated support in the house district among primary election voters in the member's party for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated primary election voters in the 2010 and 2012 CCES.

Figure A5: Relationship of representative vote to support in general electorate, by party



Note: The x-axis in each frame is the estimated support in the house district among general election voters for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated general election voters in the 2010 and 2012 CCES.

Figure A6: Relationship of representative vote to support in primary electorate, by party



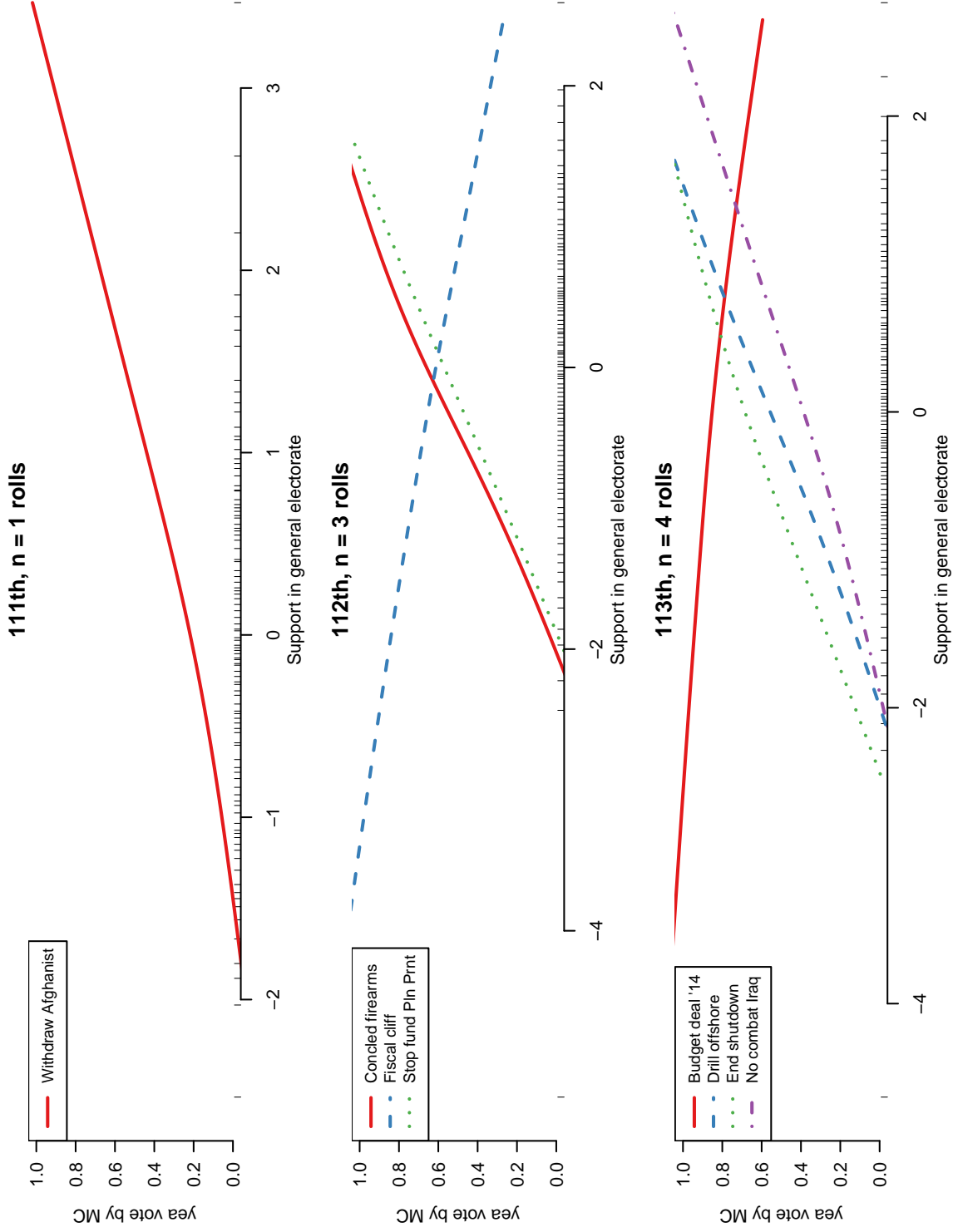
Note: The x-axis in each frame is the estimated support in the house district among primary election voters in the member's party for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated primary election voters in the 2010 and 2012 CCES.

Table A2: Preferences of general and primary electorates and roll call votes, 111th and 112th House of Representatives

	1	2	3	State*Roll FEs	GOP only	Dem only
Support in general electorate	0.74* (0.06)		0.61* (0.06)	0.39* (0.07)	0.32* (0.09)	0.72* (0.08)
Support in primary electorate		0.81* (0.09)	0.56* (0.09)	0.49* (0.10)	0.19 (0.11)	0.90* (0.14)
N	4921	4776	4776	4776	2286	2490
R <sup>2</sup>	0.87	0.87	0.88	0.90	0.89	0.87
adj. R <sup>2</sup>	0.87	0.87	0.88	0.88	0.89	0.87
Resid. sd	0.26	0.26	0.26	0.25	0.22	0.29
Standard errors in parentheses						
* indicates significance at $p < 0.05$						

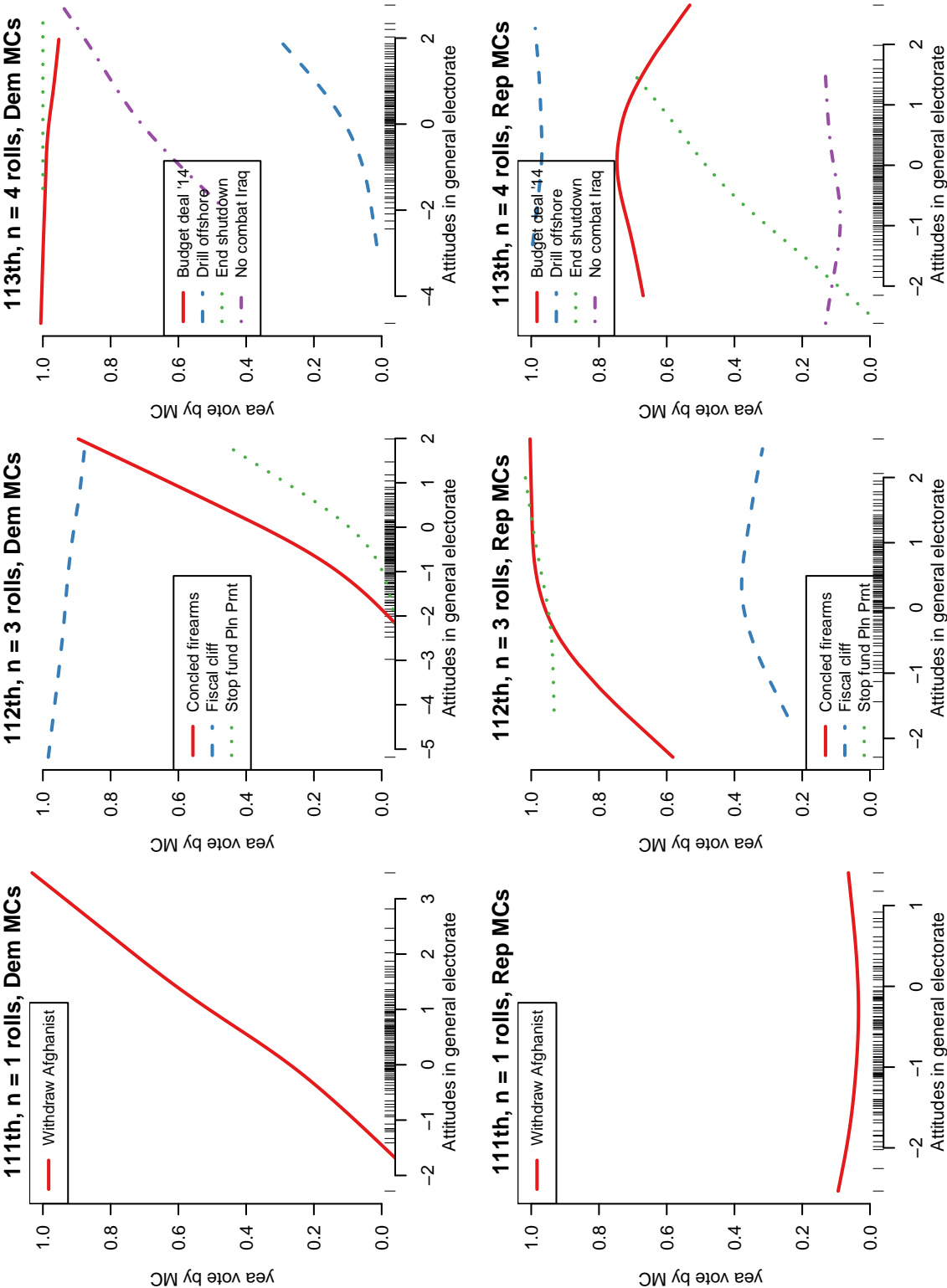
Note: Dependent variable is yea vote. OLS regression coefficients with standard errors in parentheses. Models include 7 roll call votes from the 111th House and 5 roll call votes from the 112th House. All models include separate intercepts for each roll call vote cross party. Column 4 includes intercepts for each roll call cross state.

Figure A7: Relationship of representative vote to average attitudes in general electorate



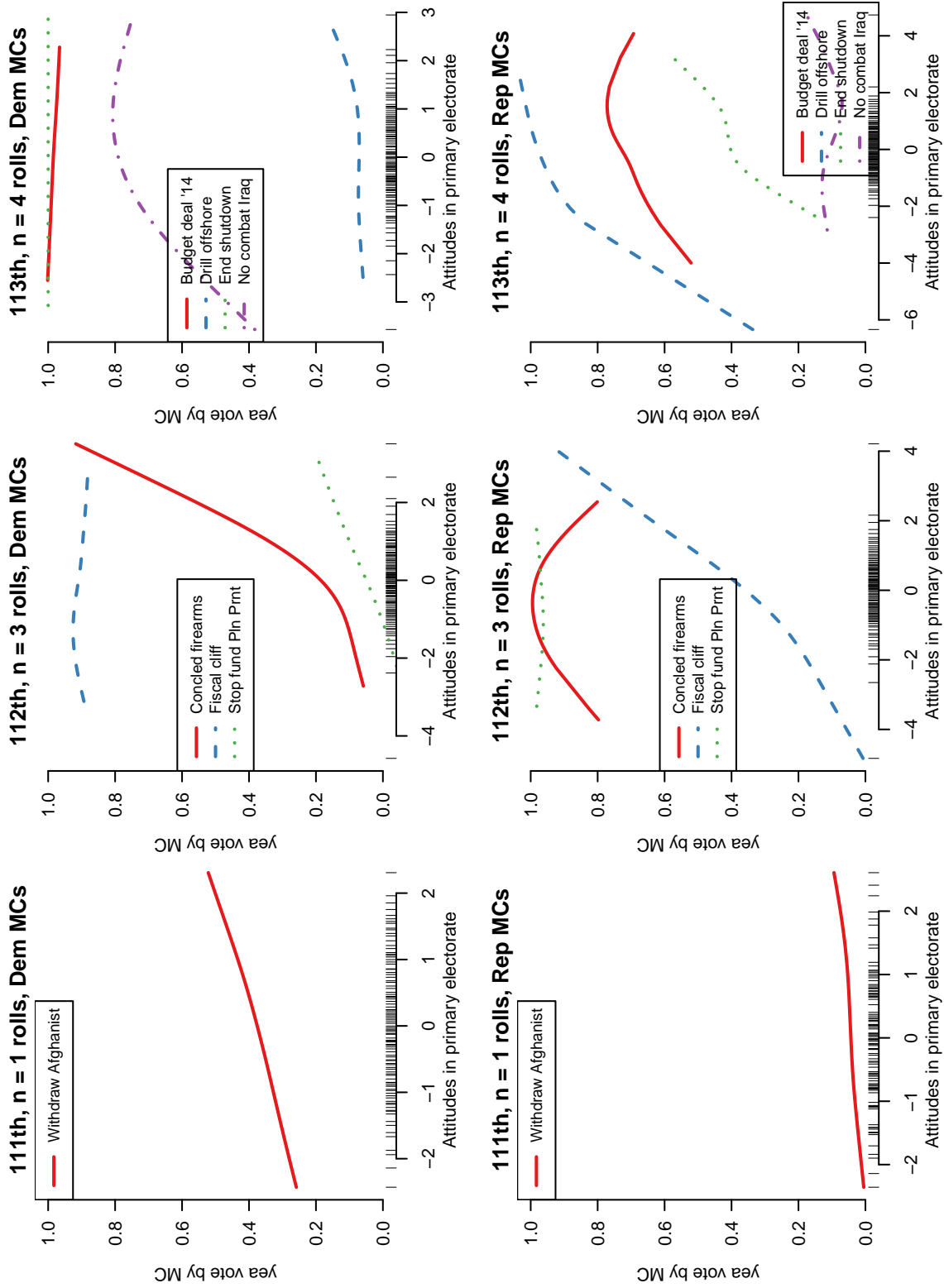
Note: The x-axis in each frame is the estimated average attitudes on the issue in the house district among general election voters for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated general election voters in the 2010 and 2012 CCES.

Figure A8: Relationship of representative vote to average attitudes in general electorate, by party



Note: The x-axis in each frame is the estimated average attitudes on the issue in the house district among general election voters for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated general election voters in the 2010 and 2012 CCES.

Figure A9: Relationship of representative vote to average attitudes in primary electorate, by party



Note: The x-axis in each frame is the estimated average attitudes on the issue in the house district among primary election voters in the member's party for each roll call. Lines are smooths of member vote on that same issue. District preferences estimated for validated primary election voters in the 2010 and 2012 CCES.



Table A3: Preferences of general and primary electorates and roll call votes, 111th, 112th, and 113th House of Representatives

	1	2	3	State*Roll FEs	GOP only	Dem only
Avg attitudes in general electorate	0.06* (0.01)		0.05* (0.01)	0.05* (0.01)	0.03* (0.01)	0.08* (0.01)
Avg attitudes in primary electorate		0.03* (0.01)	0.02* (0.01)	0.01 (0.01)	0.03* (0.01)	0.01 (0.01)
N	3370	3207	3207	3207	1708	1499
R <sup>2</sup>	0.81	0.81	0.81	0.85	0.81	0.82
adj. R <sup>2</sup>	0.81	0.81	0.81	0.83	0.81	0.82
Resid. sd	0.33	0.33	0.33	0.31	0.34	0.32

Standard errors in parentheses

\* indicates significance at  $p < 0.05$

Note: Dependent variable is yea vote. OLS regression coefficients with standard errors in parentheses. Models include 1 roll call vote from the 111th House, 3 roll call votes from the 112th House, and 4 roll call votes from the 113th House. All models include separate intercepts for each roll call vote cross party. Column 4 includes intercepts for each roll call cross state.