

Made in Congress? Testing the Electoral
Implications of Party Ideological Brand Names

Online Appendix

Jeremy C. Pope

Jonathan Woon

October 2, 2007

Data for Figure 1

The data from Figure 1 come from a survey that we designed to measure citizens' perceptions of party unity on a set of issues corresponding to key congressional votes. Such perceptions are usually not covered in traditional surveys. The data were collected as part of the 2006 Cooperative Congressional Election Study (more specifically, as part of the survey module sponsored by the Center for the Study of Elections and Democracy at Brigham Young University).¹

The survey was conducted by Polimetrix over the internet and uses a proprietary matching algorithm to produce a sample of voters as close as possible to a random sample of the American Community Study (a probability sample conducted by the U.S. Bureau of the Census). For details about the sample design, see Rivers (2006). Although some survey research analysts have expressed concerns about internet samples (e.g., Malhotra and Krosnick 2007), initial results (Hill et al 2006) suggest that even though internet surveys may contain a few minor biases, the gains from such a design are substantial.

We asked about several issues pertaining to Senate roll call votes prior to the 2006 midterm elections. The subjects of the votes included limiting “partial-birth” abortion; federally funding stem-cell research; re-deploying troops from Iraq; increasing the minimum wage; providing a path to citizenship for illegal immigrants; and extending the capital gains tax cuts beyond 2010. Respondents were presented with statements about each of the policy alternatives on a specific vote and asked to place the parties on a scale where one end indicated that the party was completely unified in favor of the proposal and the other end of the scale indicated that the party was completely unified in opposition to the proposal. A typical question went as follows:²

Earlier, we asked you about a proposal that the president begin phased redeployment of U.S. troops from Iraq starting this year and submit to Congress by the

¹General information about the project can be found at <http://web.mit.edu/polisci/port1/cces/index.html>.

²The full questionnaire is available upon request.

end of 2006: a plan with estimated dates for continued phased withdrawal.

Some politicians argue that setting out a plan to withdraw would make Iraqis take responsibility for their country and become more independent of the U.S. Others argue that it is too early to start withdrawing, and that doing so would make terrorists grow bolder.

Do you think the Democratic Party is completely unified for the proposal, the party is completely unified against the proposal, or somewhere in between?

Using the graphical interface of the internet format, respondents could then place the parties anywhere along the spectrum, yielding a measure tapping both party position and unity. We hypothesized that because parties are collections of individuals that the public would perceive variations in party unity rather than perceiving that the party takes one of two extreme positions. The results in Figure 1 bear out that assumption and suggest that voters do in fact perceive information reflected in aggregate party records. To our knowledge this is the only survey that explicitly allows voters to express their opinion of the variation or heterogeneity within political parties.

Graphical Illustration of Hypotheses

Figure A1 illustrates the three effects on the cutpoint and our hypotheses about vote share. The top panel of the figure shows a pair of party distributions that serve as the baseline for the example. Consistent with the discussion above, negative values on this continuum are more liberal, while positive values more conservative. Although in the base example the parties are not widely divergent, there is a clear difference between the liberal and conservative parties. The solid line denotes the cutpoint (v^*) between liberal party voters and conservative party voters. Although the midpoint between party means is located at

zero, the adjustment for uncertainty puts v^* less than zero due to the left party’s greater heterogeneity.

The second panel of Figure A1 illustrates Hypothesis 1 (the midpoint hypothesis). The shift in the cutpoint when both party means shift an equal distance to the right results in a shift in midpoint while keeping the GAP and DHA measures constant. (The dashed line indicates the baseline cutpoint from the top panel.) As both parties become more conservative, the Democratic candidate gains a higher vote share.

The third panel of Figure A1 illustrates Hypothesis 2 (the party heterogeneity advantage hypothesis). If a party becomes more ideologically diverse, its members will fare worse in elections. The party means and the variance of the left party are the same as in the baseline, but the right party is much more heterogeneous, resulting in a rightward shift in v^* .

The effect of party polarization on elections is stated in Hypothesis 3 (the gap hypothesis), which is illustrated in the bottom panel of Figure A1. The intuition is that as the parties become more polarized—and hence more distinct—uncertainty matters less, which benefits the more diverse party. The figure illustrates that since the left party is more heterogeneous, as the gap between the parties increases (keeping MID and DHA constant), the cutpoint shifts to the right.

Model Selection

The purpose of the model selection analysis is to determine which of three overall models best fits *all* of the data—that is, not only for subsamples of parties, incumbents, or open seats, but for the entire sample when different types of races are combined into a single statistical model, and even allowing for coefficients to vary. In the text of the paper, we initially constrain the coefficients to be the same across types of races (as in Table 1) and later allow estimates to vary across types of races by estimating the model separately for different

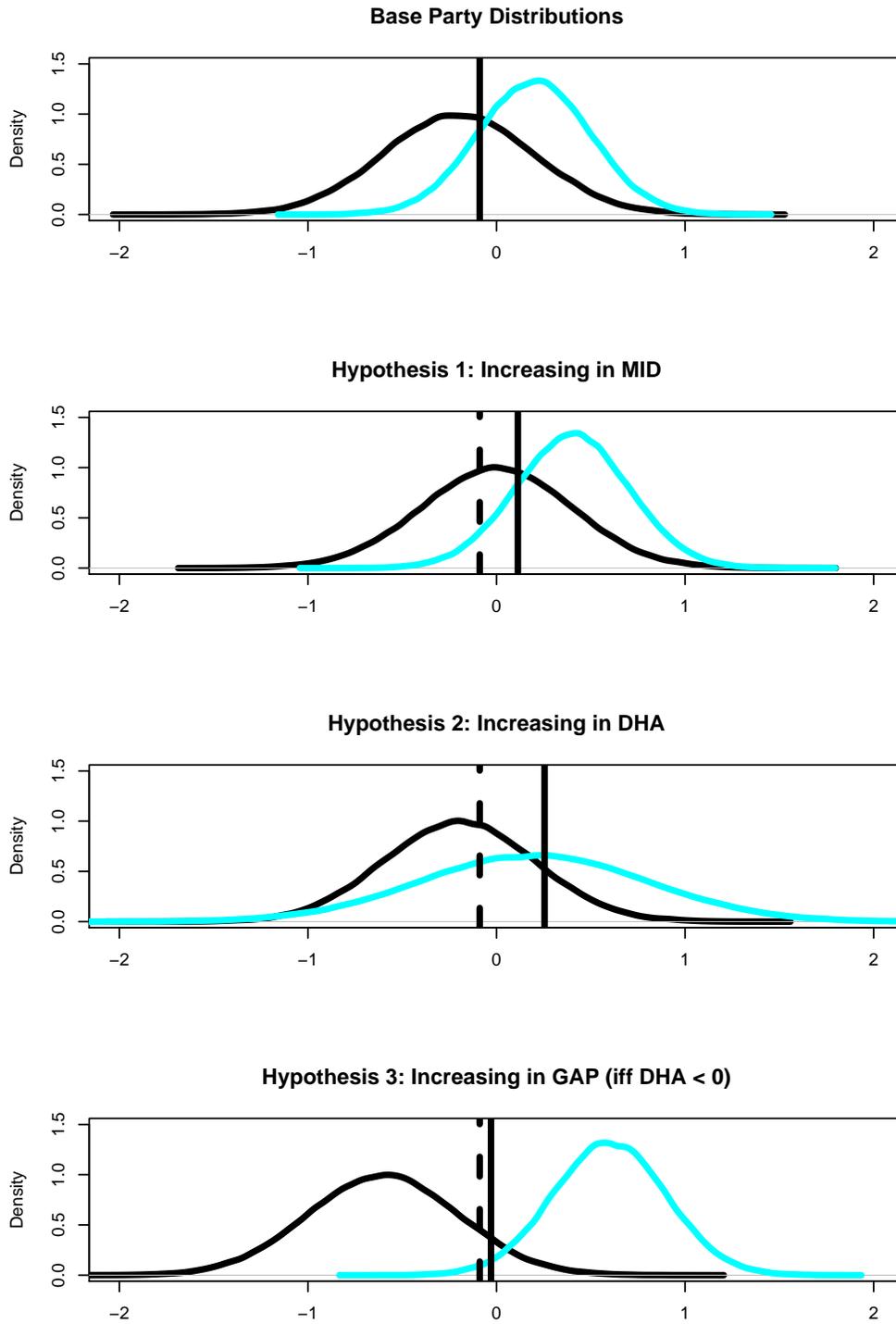


Figure A1: Illustration of Hypotheses (in the lower three panels, the dashed line indicates the original cutpoint in the base party distributions while the solid line indicates cutpoint changes with respect to each of the three hypotheses)

subsamples. Specifically, Table 2 omits open seats then estimates coefficients separately for Democratic and Republican incumbents, and columns 2 and 3 of Table 3 present separate estimates by region. We must specify a model for each version of the brand name story that accounts for this parameter variation.

Before comparing different versions of the brand name story, we first examine whether we should allow the coefficients to vary across types of races in the national brand name model. We perform a likelihood ratio test of the hypothesis that the coefficients are the same across subsamples (estimating the random effects regressions using maximum-likelihood to obtain the likelihoods for the test). The likelihood ratio test statistic is 900.95, which clearly exceeds the χ^2 critical value with 16 degrees of freedom of 10.85.

In our model selection analysis, we therefore allow the coefficients to vary across types of races in the *national brand names* model. Let y_k be the vector for the dependent variable and X_k denote the data matrix of independent variables *MID*, *DHA*, and *GAP* (excluding interactions) as in Table 1, where $k \in O, D, R$ denotes the type of race (open seats, Democratic incumbents, Republican incumbents). Let C be the data matrix of control variables, let u_k be a vector of group-specific year-level random effects, and let e_k be an error vector for each k . The model can be described in block diagonal form:

$$\begin{bmatrix} y_O \\ y_D \\ y_R \end{bmatrix} = \begin{bmatrix} X_O & 0 & 0 \\ 0 & X_D & 0 \\ 0 & 0 & X_R \end{bmatrix} \begin{bmatrix} \beta_O \\ \beta_D \\ \beta_R \end{bmatrix} + \begin{bmatrix} C_O \gamma_O \\ C_D \gamma_D \\ C_R \gamma_R \end{bmatrix} + \begin{bmatrix} u_O \\ u_D \\ u_R \end{bmatrix} + \begin{bmatrix} e_O \\ e_D \\ e_R \end{bmatrix} \quad (1)$$

To describe the model with individual incumbent positions and challenger party brand names, let Z_k denote the data matrix consisting of *incMID*, *incDHA*, and *incGAP* for $k \in D, R$. (As in Table 2, *incGAP* is interacted with the direction of *incDHA*.) The full *incumbent position and challenger party brand name* model consists of combining the national brand names for open seats with the incumbent positions and challenger party data

for incumbents:

$$\begin{bmatrix} y_O \\ y_D \\ y_R \end{bmatrix} = \begin{bmatrix} X_O & 0 & 0 \\ 0 & Z_D & 0 \\ 0 & 0 & Z_R \end{bmatrix} \begin{bmatrix} \beta_O \\ \beta_D \\ \beta_R \end{bmatrix} + \begin{bmatrix} C_O\gamma_O \\ C_D\gamma_D \\ C_R\gamma_R \end{bmatrix} + \begin{bmatrix} u_O \\ u_D \\ u_R \end{bmatrix} + \begin{bmatrix} e_O \\ e_D \\ e_R \end{bmatrix} \quad (2)$$

For the regional brand name model, we allow the coefficients to vary across regions. (We can reject the hypothesis that the coefficients are the same using the likelihood ratio test. The test statistic is 230.22 which is distributed χ^2 with 13 degrees of freedom, and the critical value is 5.89.) Let R_j be the data matrix consisting of *regMID*, *regDHA*, and *regGAP* (including appropriate interactions for *regGAP* and the sign of *regDHA* as in Table 3) for two regions $j \in N, S$. The *full regional brand names* model is:

$$\begin{bmatrix} y_N \\ y_S \end{bmatrix} = \begin{bmatrix} R_N & 0 \\ 0 & R_S \end{bmatrix} \begin{bmatrix} \beta_N \\ \beta_S \end{bmatrix} + \begin{bmatrix} C_O\gamma_O \\ C_D\gamma_D \\ C_R\gamma_R \end{bmatrix} + \begin{bmatrix} u_N \\ u_S \end{bmatrix} + \begin{bmatrix} e_O \\ e_D \\ e_R \end{bmatrix} \quad (3)$$

The likelihood dominance criterion (Pollak and Wales 1991) allows us to compare two models based on the difference in estimated likelihoods, with adjustments for differences in the number of parameters and for a given significance level. Let M_1 and M_2 be two models with n_1 and n_2 parameters, respectively, where $n_1 < n_2$ and let L_1 and L_2 be the log likelihoods. Let $C(d)$ denote a critical value of the χ^2 distribution with d degrees of freedom. (We adopt the conventional 5 percent significance level.) According to the likelihood dominance criterion:

1. Select M_1 if $L_2 - L_1 < [C(n_2 + 1) - C(n_1 + 1)]/2$.
2. Select M_2 if $L_2 - L_1 > [C(n_2 - n_1 + 1) - C(1)]/2$.
3. Otherwise, model selection is indeterminate.

M_1	n_1	M_2	n_2	Critical Values		$L_2 - L_1$
National	34	Incumb/Ch. Party	34	0	0	160.44
Regional	28	National	34	3.62	5.11	277.95
Regional	28	Incumb/Ch. Party	34	3.62	5.11	438.38

Table A1: Pairwise model comparisons using the likelihood dominance criterion

Note that when $n_1 = n_2$, the indeterminate region reduces to 0, so the criterion reduces to a simple comparison of likelihoods.

We estimate the three full models described above by maximum likelihood in order to obtain the log likelihoods. For the national brand names model (allowing for parameter variation across types of races) the log likelihood is 9,797. For the individual incumbent and challenger party brand name model, the log likelihood is 9,958. For the regional party brand names model, the log likelihood is 9,519. On the basis of the log likelihoods alone, we would select the incumbent position and challenger party brand name model.

Table A1 presents the likelihood dominance criterion critical values and differences in log likelihoods for pairwise comparisons of the models. (The number of estimated parameters includes the error variance terms σ_η^2 and σ_ϵ^2 .) Even accounting for intervals where the difference is indeterminate, the results show that the individual incumbent-challenger party brand name model clearly dominates both the national and regional brand name story.

We perform additional model selection analysis in order to compare party-only, candidate-only, and mixed incumbent-challenger party models. In order to make valid comparisons, we must restrict the sample to only races with incumbents. The party-only model is the same as the national brand name model above without the open seat races, while the mixed model is the incumbent plus challenger party model above (again, excluding open seats). The “pure” candidate-only model replaces the theoretical variables of interest (i.e. *MID*, *DHA*, and *GAP*) with only the incumbent’s DW-NOMINATE score. Similar to the analysis above, we also estimate separate coefficients for Democrats and Republicans. The results, summarized in Table A2, show that the mixed incumbent-challenger party model dominates

M_1	n_1	M_2	n_2	Critical Values		$L_2 - L_1$
Mixed	22	Parties	22	0	0	-160.44
Candidate	18	Parties	22	2.51	3.61	-152.19
Candidate	18	Mixed	22	2.51	3.61	8.25

Table A2: Pairwise model comparisons using the likelihood dominance criterion for incumbent-only models

the candidate-only model (which dominates the party-only model).

References

- Hill, Seth J., James Lo, Lynn Vavreck and John Zaller. 2007. "The Opt-in Internet Panel: Survey Mode, Sampling Methodology and the Implications for Political Research," working paper.
- Malhotra, Neil and Jon A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples" *Political Analysis* 15:286 - 323.
- Pollak, Robert A. and Terence J. Wales. 1991. "The likelihood dominance criterion." *Journal of Econometrics* 47:227-42.
- Rivers, Douglas. 2006. "Sample Matching: Representative Sampling from Internet Panels," Polimetrix White Paper Series.