Final Data Essay
Spatial Voting in US Presidential Elections

Denis Cohen | Marcel Neunhoeffer | Oliver Rittmann

Due: 22 December 2020 by 23h59

Purpose

The purpose of this data essay is to test a theory about vote choice in US presidential elections. Specifically, you will have to test two arguments about how individuals' choose to vote for one presidential candidate over another.

Include a write-up of no more than 2000 words ($\pm$ 10%), font size 12pt, double-spaced with at least 2.5cm margins on all sides. Focus on a professional presentation of your results (for example, meaningful and neatly formatted tables and publication-quality graphs). Please only include one table or figure per page, appropriately named and referenced in the text. Figure/table-generating code must not be printed in the essay. Essays not adhering to these formatting guidelines will be marked down, as will sloppy presentations of tables and figures.

Apply in this essay what you have learned in this class about causal reasoning, statistical techniques, the interpretation of your estimates, and the presentation of statistical results. Do not include a theoretical discussion of the hypotheses. Instead, focus on the specification of your models (i.e., why you translated a theoretical model into the respective systematic components) and the interpretation of your results. Keep in mind that there are different ways how to present results; try to choose the one that conveys as much information as possible to the reader in an intuitive way. Think about this data essay as the "methods, data, and results" section of a research paper that you would like to submit to a journal for publication. Please keep in mind that absolutely **no collaboration with others is allowed**.

Explaining Vote Choices in US Presidential Elections

The question why individuals vote the way they do is a staple in political science. Two of the most prominent theoretical approaches to explaining individual vote choices, particularly in the context of US presidential elections, are the rational choice approach of spatial theory and the Michigan School's social-psychological approach.

The proximity theory of spatial voting posits that voters' ideological views and candidates' ideological positions align along a uni-dimensional continuum (e.g., a liberal-conservative continuum). Voters determine how strongly candidate's ideological standpoint deviates from their own ideal points and vote for the candidate whose offer is most congruent with their own views. In the context of US presidential elections, this implies that voters with more liberal ideological views should on average be more likely to vote for Democratic candidates whereas voters with more conservative ideological views should on average be more likely to vote for Republican candidates. Proximity theory also implies a "tipping point": On average, voters should be *indifferent* (i.e., equally likely to vote for either of the two candidates) when positioned at the ideological midpoint between the two candidates.

This view is at odds with the social-psychological model, which posits that voters, under the formative impression of their socio-structural identity and their socio-economic position, develop a stable affective orientation toward parties: *party identification.* As an 'unmoved mover', party identification then determines both candidate evaluations and policy views which ultimately result in an individual's vote choice.

A lot of debate surrounds the question whether party identification is truly a stable social identity that determines ideological views and candidate evaluations, or whether it is in turn the outcome of voter ideology. Here, we adopt the framework of Jessee (2009), who argues that party identification and voter ideology are independent but systematically interact in determining individuals' vote choices. Specifically, party identification moderates how stringently voters rely on ideological views: Whereas pure spatial voting predicts that voters will on average be indifferent between two candidates at the ideological midpoint between the two candidates, the hybrid model (see the Figure below) predicts that this will only apply to independents. Republican identifiers, on the other hand, will need to be much closer to the Democratic candidate while Democrats will need to be much closer to the Republican candidate to reach their indifference point.

Additionally, Jessee argues that spatial voting is cognitively more demanding than other mechanisms of voting because it requires that voters acquire a solid understanding of where candidates stand ideologically. Voters who do not trust their perception of candidates' positions may therefore be forced to rely more strongly on other heuristics, e.g., their party identification or the national political mood. Pure spatial voting should thus be more pronounced among politically informed or knowledgeable voters, and most pronounced among politically informed independents.

From these arguments, we derive two hypotheses:

Hypothesis 1 *Party identification moderates the effect of ideological self-placement: The expected relationship that the more conservative (more liberal) voters' ideological views, the higher their probability of voting for the Republican (Democratic) candidate is stronger for independents than it is for either Democrats or Republicans.*
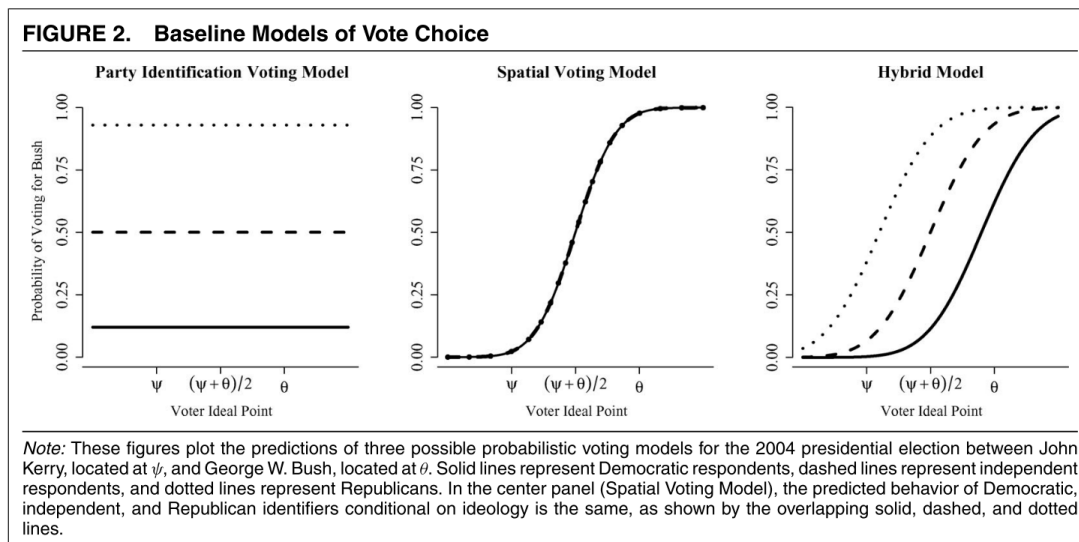
FIGURE 2. Baseline Models of Vote Choice

Note: These figures plot the predictions of three possible probabilistic voting models for the 2004 presidential election between John Kerry, located at ψ, and George W. Bush, located at θ. Solid lines represent Democratic respondents, dashed lines represent independent respondents, and dotted lines represent Republicans. In the center panel (Spatial Voting Model), the predicted behavior of Democratic, independent, and Republican identifiers conditional on ideology is the same, as shown by the overlapping solid, dashed, and dotted lines.

Figure 1: A schematic illustration of Hypothesis 1 in the context of the 2004 US Presidential Election, taken from Jessee (2009).

Hypothesis 2 *The effect of ideological self-placement has a "pure" spatial effect among highly informed independents: On average, members of this group are indifferent between the two presidential candidates when positioned exactly at the midpoint of the candidates' average perceived positions.*

The Case: The 2008 US Presidential Election

We ask you to test these arguments in the context of the 2008 US presidential election, in which Democratic presidential candidate Barack Obama ran against Republican presidential candidate John McCain. Toward this end, we provide you with pre-processed data from the post-election wave of the *American National Election Study 2008* (`anes08_data_essay.dta`).

The central variables are:

- `obama`: Coded 1 if the respondent voted for Obama and 0 if the respondent voted for McCain (non-voters and voters for third-party candidates have been removed).

- `lib_con_self`: Respondent's self-placement on the liberal (1) to conservative (7) continuum. *Treat this as a continuous variable.*

- `party_id`: Respondents's party identification, distinguishing Republicans, Democrats, and Independents.

Additional variables included in the data are described in the codebook below. Please also note:

- Across the sample, the average perceived ideological positions of Obama, $\hat{\psi}$, and McCain, $\hat{\theta}$, are $\hat{\psi} = 2.83$ $[2.74, 2.92]$; $\hat{\theta} = 5.28$ $[5.19, 5.37]$ on the liberal (1) to conservative (7) continuum. The average perceived midpoint is $\frac{\hat{\psi}+\hat{\theta}}{2} = 4.05$ $[4.01, 4.09]$.

- Among highly informed independents, the average perceived ideological positions of Obama, $\hat{\psi}^*$, and McCain, $\hat{\theta}^*$, are $\hat{\psi}^* = 2.29$ $[2.07, 2.51]$; $\hat{\theta}^* = 5.73$ $[5.53, 5.94]$ on the liberal (1) to conservative (7) continuum. The average perceived midpoint is $\frac{\hat{\psi}^* + \hat{\theta}^*}{2} = 4.00$ $[3.85, 4.15]$.

Your Assignment

This section details what we expect you to do in your essay.

- Use descriptive statistics and variable descriptions to familiarize the reader with the data that you use. Think hard about which variables you select and which descriptives are most relevant and informative in the context of the theory.

- Use appropriate models to empirically test the hypotheses formulated above. Note that your test of **Hypothesis 1** may require a different model specification and/or sample of observations from those required for your test of **Hypothesis 2**.

- Present the results of your main models in one or two nicely formatted output tables.

- In addition to the estimation results, make sure to present interesting quantities of interest and to include the uncertainty surrounding your quantities. Your interpretation should focus on the direction and substantive magnitude of your quantities.

  - For **Hypothesis 1**: Based on your main model, plot predicted probabilities of voting for Obama for the three groups (Democrats, Republicans, and Independents) along the liberal-conservative continuum (akin to the schematic presentation in the third plot of the Figure above). Make sure to discuss at which values of the liberal-conservative continuum each group, on average, reaches its indifference point. Then, for each group, calculate the first differences in the predicted probabilities of voting for Obama that result from changing voters' ideal points from $\hat{\theta}$ to $\hat{\psi}$. Compare these quantities to establish for which groups the effect is strongest and weakest.

  - For **Hypothesis 2**, stipulate a precise numerical hypothesis and present appropriate numerical and/or visual evidence that allows you test it. Make sure to transparently communicate your strategy for testing this hypothesis.

- Remember to carefully specify your models and gradually build more inclusive models to test the robustness of your findings. Defend your model choices carefully. The selection of control variables should follow concise causal reasoning. You may (but you do not have to) present causal graphs.

- Remember to think before you run models. Do not just dive in without reflection. Stay focused. Have models in mind. Step back when things do not work out. Give yourself time to write and re-write.

- You will have to **upload** both your PDF and your `.R` script or `.Rmd` documents to ILIAS by 22 December 2020, 23h59. The files should be personalized with your university login name / GitHub usename (e.g., `dcohen.pdf` and `dcohen.R` or `dcohen.Rmd`). Make sure that your code is well documented and runs through smoothly. Late submissions of either the paper or the code will not not accepted. Late submissions will be marked as "failed (5.0)".

Codebook

The codebooks describes all variables from the data set and explains what they measure.

Variable	Description
obama	Coded 1 if the respondent voted for Obama and 0 if the respondent voted for McCain (non-voters and voters for third-party candidates have been removed).
vote_house	Vote choice in the 2008 election to the US House of Representatives.
lib_con_self	Respondent's self-placement on the liberal (1) to conservative (7) continuum.
lib_con_obama	Respondent's placement of Barack Obama on the liberal (1) to conservative (7) continuum.
lib_con_mccain	Respondent's placement of John McCain on the liberal (1) to conservative (7) continuum.
like_obama	Respondent's feeling thermometer score (0: cold, 100: warm) for Barack Obama.
like_mccain	Respondent's feeling thermometer score (0: cold, 100: warm) for John McCain.
party_id	Respondents's party identification, distinguishing Republicans, Democrats, and Independents.
pol_knowledge	Coded 1 if the respondent correctly identified the majority parties in the House and Senate prior to the election; coded 0 otherwise.
race	Respondent's race.
male	Coded 1 if respondent is male; coded 0 if respondent is female.
age	Respondent's age in years.
relig_denomination	Respondent's religious denomination.
above_median_inc	Coded 1 if respondent's family income is above the national median household income; coded 0 otherwise.
college_degree	Coded 1 if respondent has a four-year college degree or higher educational degree; coded 0 otherwise.
state	Respondent's state of residence.
weights	Post-stratification weights for post-election wave.

References

Jessee, S. A. (2009). *Spatial Voting in the 2004 Presidential Election.* American Political Science Review, 103(1), 59–81.