alteryx

The Presidential Election Preference Application

Dan Putler, Brett Gottdener, and Richard Snow

Agenda

- The objectives of the application

- A brief demo of one part of the application and a mockup of the other part

- The underlying model used to predict area preferences for a particular candidate

- The challenges of going from the model to the local area estimates needed for the app

Objectives of the Presidential Preference Analytic App

Provide local area estimates (e.g., a 7-minute drive-time from a particular address, a zip code, a county) of the percentage of registered voters who prefer Obama, Romney, a third-party candidate, or undecided about the upcoming election

Caveat: The application looks at the preferences of *registered* voters, not likely voters, making things skew a bit Democratic

Analytic App Demonstration

The Presidential Candidate Preference Model 1

- The data: Respondent level data to Wave 7 of the USA Today / Gallup Presidential Election preference poll
  - 1,446 contacted adults age 18 and older
  - 1,301 registered voters
  - 1,026 usable (complete) responses
  - The survey was conducted from September 24 to 27, 2012
  - The data is made available through the Roper Center at the University of Connecticut
- Target variable: The candidate the respondent would vote for or leans towards
  - Respondents that support no candidate or do not know who they would vote for are classified as undecided

The Presidential Candidate Preference Model 2

- The important predictors
  - Party affiliation/identification (the dominant predictor)
  - Age
  - The region of the country in which the respondent resides
  - Income
  - Educational attainment
  - Race
  - The "red" to "blue" hue of the respondent's state of residence (based on a cluster analysis of 2004 and 2008 state level election returns)
- Models that use these variables are able to correctly classify preferences in over 75% of cases in a validation sample

You can't always get what you want, but if you try sometimes, you just might find, you get what you need.

Mick Jagger and Keith Richards

From the Preference Model to the Local Area Estimates 2

- The challenges
  - The income groups used by Gallup are different from those used in the Experian CAPE data and by the US Census Bureau
  - The household income of *individuals* age 18 or over who are US citizens is needed, but income is reported only at the *household or economic family level*
    - The distribution of the number of adults age 18 or over across households is needed, the closest available measure is the distribution of the number of household members across households
  - The joint distribution of predictor variables is needed rather that taking each variable individually for prediction

From the Preference Model to the Local Area Estimates 3

- The solutions

  - <u>Income groups</u>: "Port" the Gallup income group definitions to the Experian CAPE data through the use of non-parametric estimates of each local area's cumulative household income distribution (via an adjusted loess model), allowing for the estimation of the number of households in the area that fall into each Gallup income group

From the Preference Model to the Local Area Estimates 4

- The solutions, continued

  - <u>Going from household income to the number of adults who reside in households in a given income group</u>:

    - To obtain estimates of the distribution of the number adults in a household across households for a local area: (1) get the expected counts for a specific number of adults in households of a particular size and then sum for that number of adults across households of different sizes based on information obtain from the ACS microdata and (2) then reallocate households across different number of adults present groups to meet adding up restrictions on the number of adults and households in an area via quadratic programming

    - Use iterative proportional fitting to obtain estimates of the joint distribution of household income level and the number of adults age 18 and over residing in a household in an area, and then collapse the distribution on income to obtain an estimate of the marginal distribution of the number of individuals age 18 and over that reside in a household with a given income level. The prior distribution for this is obtained from the ACS Microdata.

From the Preference Model to the Local Area Estimates 5

- The solutions, continued

  - <u>The joint distribution of the predictor variables</u>: Obtain estimates of the joint distribution of the predictor variables combining individual level data from three different national Gallup Presidential preference polling surveys to construct a "prior" joint distribution of the variables, obtain the local area marginal distributions of the predictor variables data available from a combination of Experian CAPE, American Community Survey, and third-party political consultant voter political party affiliation data via iterative proportional fitting

The Data Used in the App

- Waves 10, 12, and 14 of the USA Today / Gallup 2012 Presidential Preference poll obtained from the Roper Center

- Experian CAPE 2012 Current Year Estimates

- American Community Survey 2010 5-Year Summary Census block group level data

- American Community Survey 2006 to 2010 Public Usage Microdata Sample for both households and individuals

- State level election returns for the 2004 and 2008 Presidential election

- Voter party identification data provided by a political consulting firm that wishes to remain anonymous