PS 787 Assignment 1 (Sep 27, 2008; revised Sep 28, 2008)

The question for this assignment is how to estimate effects of a couple of variables on partisan vote choices in the 2000 U.S. general election—choosing between the Democrat and the Republican candidate running for the U.S. House of Representatives—even though candidates run unopposed in some districts. In the 2000 ANES data to be described, vote choices are described by the variable `vote`: 1, did not vote; 2, voted but not for a U.S. House candidate; 3, voted for Democratic candidate; 4, voted for Republican candidate. The question of primary interest is how a couple of variables relate to the choice between `vote`=3 and `vote`=4. The two explanatory variables of primary interest are party ID and the difference between `demdist` and `repdist` (defined below); note that incumbent status probably also makes a difference. A couple of nonsurvey variables are defined that might work as instruments in some approaches to trying to identify and estimate the effects of interest: `votebal`, defined using election returns, and `fecbal`, defined using FEC campaign contributions data. There are also a couple of variables that are frequently used to explain individual differences in voter turnout: `age` and `residence`.

This is a pretty complicated problem. I've never actually seen it posed this way or solved in the literature. We'll take increasingly (or differently) complicated cuts at it as the semester proceeds.

By Monday (Sept 29), the most I'll expect is for you to have thought about how this problem might be framed using a Heckman-like approach. What's the model specification? Are the data I've provided sufficient to estimate the model? Note that data appear in several files.

The data in `assign1b.csv` are from the 2000 American National Election Study (ANES) survey, supplemented with election returns data (compiled by the Clerk of the U.S. House of Representatives) and with itemized individual contributions data from the Federal Election Commission. For more information about the ANES and the survey, see the following:

`http://www.electionstudies.org/`
`http://www.electionstudies.org/studypages/download/datacenter_all.htm`

See the file `assign1doc.txt` for detailed information about the raw variables used to build the variables in `assign1b.csv`. For details about how the subset of data included in `assign1b.csv` was produced, see `assign1writeb.R`, `wrk2000b.R`, `2000election.R` and `fec00.R` and the intermediate files `dat2000b.csv`, `fec2000.csv` and `2000election.csv`.

In order (and by name), the variables in the columns of `assign1b.csv` are the following.

`[line number]` no variable name in the .csv header; not the same as the ANES observation number

`stcd` state-district abbreviation; STCD has the two-letter U.S. Postal abbreviation for the survey respondent's state followed by the two-digit (using leading zeros) U.S. House district number.

`weight` sampling weight

`stfips` state FIPS code

`age` respondent's age in years

`deminc` respondent's House district has a Democratic incumbent running

`repinc` respondent's House district has a Republican incumbent running

`demchal` respondent's House district has a Democratic challenger running

`repchal` respondent's House district has a Republican challenger running

`othinc` respondent's House district has a third party incumbent running

`othchal` respondent's House district has a third party challenger running

`openseat` respondent's House district has an open seat (no incumbent running)

`demopen` respondent's House district has an open seat with a Democratic candidate running

`repopen` respondent's House district has an open seat with a Republican candidate running

`othopen` respondent's House district has an open seat with a third party candidate running

`unopposed` respondent's House district has a candidate running unopposed

`pid` seven-point party ID; codes: 1 = Strong Democrat, 2 = Weak Democrat, 3 = Independent-Democrat, 4 = Independent-Independent, 5 = Independent-Republican, 6 = Weak Republican, 7 = Strong Republican

`libcon` seven-point libcon self-placement; codes: 1 = Extremely liberal, 2 = Liberal, 3 = Slightly liberal, 4 = Moderate, middle of the road, 5 = Slightly conservative, 6 = Conservative, 7 = Extremely conservative

`demlibcon` libcon placement of Democratic presidential candidate

`demdist` absolute difference between self and Dem libcon (NAs on `demlibcon` assigned to 4; for other complications see `wrk2000b.R`)

`replibcon` libcon placement of Republican presidential candidate

`repdist` absolute difference between self and Rep libcon (NAs on `replibcon` assigned to zero; for other complications see `wrk2000b.R`)

`demdist0` absolute difference between self and Dem libcon (NAs on `demlibcon` unchanged)

`repdist0` absolute difference between self and Rep libcon (NAs on `replibcon` unchanged)

`residence` length of residence in home in years (5 or more years is coded as 5)

`vote` respondents voting behavior: 1, did not vote; 2, voted but not for a U.S. House candidate; 3, voted for Democratic candidate; 4, voted for Republican candidate

`ALL` total number of votes cast for a House candidate in district election returns

`DEM` total number of votes cast for a DEM House candidate in district election returns

`REP` total number of votes cast for a REP House candidate in district election returns

`THIRD` total number of votes cast for a third-party House candidate in district election returns

`DEM.C` total dollars contributed to DEM challengers in district in itemized individual contribution data

`DEM.I` total dollars contributed to DEM incumbents in district in itemized individual contribution data

`DEM.O` total dollars contributed to DEM open seat candidates in district in itemized individual contribution data

`REP.C` total dollars contributed to REP challengers in district in itemized individual contribution data

`REP.I` total dollars contributed to REP incumbents in district in itemized individual contribution data

`REP.O` total dollars contributed to REP open seat candidates in district in itemized individual contribution data

`state` respondent's state

`votebal` proportion of votes cast for DEM candidates (election returns) in all other districts in the same state; set to .5 when a district is the only one in a state; formula is

$$\frac{\sum_{j \neq i}(\text{DEM}_j + .5)}{\sum_{j \neq i}(\text{DEM}_j + \text{REP}_j + 1)}$$

`voteObal` proportion of votes cast for DEM candidates (election returns) in all other districts in the same state that are openseats; set to .5 when there are no open seats in other districts in a state

`fecbal` proportion of money given to DEM candidates (itemized individual contributions) in all other districts in the same state; set to .5 when a district is the only one in a state; formula is

$$\frac{\sum_{j \neq i}(\text{DEM.I}_j + \text{DEM.C}_j + .5)}{\sum_{j \neq i}(\text{DEM.I}_j + \text{DEM.C}_j + \text{REP.I}_j + \text{REP.C}_j + 1)}$$

`fecObal` proportion of money given to DEM candidates (itemized individual contributions) in all other districts in the same state that are openseats; set to .5 when there are no open seats in other districts in a state