

PS 787 Assignment 4 (Nov 16, 2007, due Nov 28)

The data in `nes1992.csv` are from the American National Election Study (ANES) survey of 1992. For more information about the ANES, see the following:

<http://www.electionstudies.org/>

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

The `.csv` file contains eleven variables:

1. `pvote`: respondent's presidential vote choice; codes: 0 = Republican, 1 = Democrat
2. `ideal`: respondent's policy ideal point (produced by combining and rescaling several survey items); the values satisfy $ideal \in [0, 1]$ with values near 0 being more "liberal" and values near 1 being more "conservative"
3. `dempos`: respondent's policy placement of the Democratic presidential candidate; $dempos \in [0, 1]$ with values near 0 being more "liberal" and values near 1 being more "conservative"
4. `reppos`: respondent's policy placement of the Republican presidential candidate; $reppos \in [0, 1]$ with values near 0 being more "liberal" and values near 1 being more "conservative"
5. `pidXX`: seven dummy variables to measure Party Identification: codes: `pidSD`, Strong Democrat; `pidD`, Weak Democrat; `pidID`, Independent - Democrat, ; `pidI`, Independent - Independent; `pidIR`, Independent - Republican; `pidR`, Weak Republican; `pidSR`, Strong Republican

Only those who voted for either the Democrat or the Republican are included.

This exercise is primarily about nonnested model selection tests. Specification issues also make a small appearance.

The issue is to model individuals' choices between the two candidates. The models of interest use the following linear predictor (coefficients are implicit):

$$z = 1 + \text{dummy}(\text{pid}) + \text{dadist} + \text{radist}$$

Here `dummy(pid)` represents the set of six dummy variable that distinguish the responses to `pid` other than "Strong Democrat." The variables `demdist` and `repdist` are intended to capture the implications of a spatial model for policy positions:

$$\text{dadist} = |\text{ideal} - \text{dempos}|$$

$$\text{radist} = |\text{ideal} - \text{reppos}|$$

The idea is that someone whose self-placement is farther from the placement attributed to a party's candidate is less likely to vote for that candidate.

We consider two binary regression models: probit and logit (see, e.g., Cameron and Trivedi 2005, 469–471).

The **R** program `pv192.R` estimates these models. For each model both the standard errors based on the observed information and sandwich standard errors are reported. Results from running that program via the command line (in Linux)

```
R CMD BATCH --no-save pv192.R
```

are in `pv192.Rout`.

Does either model appear to be correctly specified (or nearly so)? Does a nonnested test show one of the models to be significantly better than the other?

Present a detailed response to these questions supported by the data in `nes1992.csv` (you need not use **R** to do your analysis).

Background: The data used in this assignment is a subset of the data I used in the following article.

Mebane, Walter R., Jr. 2000. "Coordination, Moderation, and Institutional Balancing in American Presidential and House Elections." *American Political Science Review* 94 (March): 37–57.