

PS 787 Assignment 3 (Nov 10, 2007, due Nov 21)

The data in `nes8804e.csv` are from the American National Election Study (ANES) surveys of 1988, 1992, 1996, 2000 and 2004. The data come from the ANES Cumulative Data File. For more information about the ANES and the surveys, see the following:

<http://www.electionstudies.org/>
http://www.electionstudies.org/studypages/download/datacenter_all.htm
<http://www.electionstudies.org/nesguide/nesguide.htm>
http://www.umich.edu/~wmebane/gov317/anes_cdf_int/anes_cdf_int.html
http://www.umich.edu/~wmebane/gov317/anes_cdf_var/anes_cdf_var.html

See the file `nes2000.vars` for detailed information about the variables named VCFnnnn in the following. The `.csv` file contains eleven variables:

1. **year**: year of study the observation comes from
2. **wgt**: sampling weight
3. **anysay**: recoded version of VCF0613 (“Do People like R Have Any Say in What Government Does”); codes: 0 = agree, 1 = neither, DK, NA, 2 = disagree
4. **age**: recoded version of VCF0101 (“Respondent Age”)
5. **residence**: recoded version of VCF9002 (“R Length of Residence in Home”); codes: 1 = 4 years or less, 2 = 5-9 years, 3 = 10-19 years, 4 = 20-29 years, 5 = 30 or more years (2004: includes 76 years or more), 6 = ‘All of life’ (regardless of number of years)
6. **pid**: recoded version of VCF0301 (“7-pt Scale Party Identification”); codes: 1 = Strong Democrat, 2 = Weak Democrat, 3 = Independent - Democrat, 4 = Independent - Independent, 5 = Independent - Republican, 6 = Weak Republican, 7 = Strong Republican
7. **vote**: recoded version of VCF0704 (“Party of R Vote: President- Major Candidates”); codes: 1 = Democrat, 2 = Republican, 3 = Major third party candidate (Perot 1992,1996), 0 = Did not vote
8. **libcon**: recoded version of VCF0803 (“R Placement: Liberal-Conservative Scale”); codes: 1 = Extremely liberal, 2 = Liberal, 3 = Slightly liberal, 4 = Moderate, middle of the road, 5 = Slightly conservative, 6 = Conservative, 7 = Extremely conservative
9. **demlibcon**: recoded version of VCF9088 (“Democratic Presidential Cand: Liberal-Conservative Scale”); codes: 1 = Extremely liberal, 2 = Liberal, 3 = Slightly liberal, 4 = Moderate, middle of the road, 5 = Slightly conservative, 6 = Conservative, 7 = Extremely conservative
10. **replibcon**: recoded version of VCF9096 (“Republican Presidential Cand: Liberal-Conservative Scale”); codes: 1 = Extremely liberal, 2 = Liberal, 3 = Slightly liberal, 4 = Moderate, middle of the road, 5 = Slightly conservative, 6 = Conservative, 7 = Extremely conservative

11. **econ**: recoded version of VCF0872 (“R Opinion: Better or Worse Economy in Next Year”); codes: $-1 =$ Get worse, $0 =$ Stay about the same, DK, NA, $1 =$ Get better

If any of VCF0803, VCF9088 or VCF9096 equals 0, 8 or 9, then **libcon**, **demlibcon** and **replibcon** all equal 0. Observations with code = 0 on variable VCF0702 (“Did R Vote in the November Elections”) are omitted. Also omitted are observations with missing data for age, residence or pid.

This exercise is primarily about discrete choice models. Specification issues also make a small appearance, and there is a little bit about hypothesis tests.

The issue is to model individuals’ choices among the three responses to **vote**. The models of interest use the following linear predictors (coefficients are implicit):

$$\begin{aligned} z_0 &= 1 + \text{anysay} + \text{age} + \text{residence} + \text{dummy(pid)} + \text{demdist} + \text{repdist} \\ z_1 &= 1 + \text{dummy(pid)} + \text{demdist} + \text{repdist} \\ z_2 &= 0 \end{aligned}$$

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:

$$\begin{aligned} \text{demdist} &= |\text{libcon} - \text{demlibcon}| \\ \text{repdist} &= |\text{libcon} - \text{replibcon}| \end{aligned}$$

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.

One model for these choices is the multinomial logit (MNL) model. In this case, using $p_j \equiv \text{prob}(\text{vote} = j)$,

$$\begin{aligned} p_0 &= \frac{\exp(z_0)}{\exp(z_0) + \exp(z_1) + \exp(z_2)} \\ p_1 &= \frac{\exp(z_1)}{\exp(z_0) + \exp(z_1) + \exp(z_2)} \\ p_2 &= \frac{\exp(z_2)}{\exp(z_0) + \exp(z_1) + \exp(z_2)}. \end{aligned}$$

Define a set of dummy variables to indicate each respondent’s choice: for each observation i ,

$$y_{ji} = \begin{cases} 0, & \text{vote}_i \neq j \\ 1, & \text{vote}_i = j \end{cases}$$

the loglikelihood is

$$l_{\text{MNL}i} = \sum_{j=0}^2 y_{ji} \log p_{ji} .$$

An alternative model is the generalized extreme value (GEV) model, with correlated random elements specified between the two choices where the respondent reports a vote for

a candidate. Using σ to denote the GEV model similarity parameter, $0 \leq \sigma < 1$, and defining

$$J = \exp(z_1)^{1/(1-\sigma)} + \exp(z_2)^{1/(1-\sigma)}$$

the choice probabilities $q_j \equiv \text{prob}(\text{vote} = j)$ are

$$\begin{aligned} q_0 &= \frac{\exp(z_0)}{\exp(z_0) + J^{1-\sigma}} \\ q_1 &= \frac{J^{-\sigma} \exp(z_1)^{1/(1-\sigma)}}{\exp(z_0) + J^{1-\sigma}} \\ q_2 &= \frac{J^{-\sigma} \exp(z_2)^{1/(1-\sigma)}}{\exp(z_0) + J^{1-\sigma}} . \end{aligned}$$

The loglikelihood is

$$l_{\text{GEV}i} = \sum_{j=0}^2 y_{ji} \log q_{ji} .$$

The two models are nested, so a possible test statistic to assess whether the added parameter in the GEV model makes a significant difference is

$$G = -2(l_{\text{MNL}i} - l_{\text{GEV}i}) .$$

The **R** program `nes8804mnlgev.R` estimates these MNL and GEV models. For each model both the standard errors based on the observed information and sandwich standard errors are reported. The statistic G is computed as `2*(m1MNL$min-m1GEV$min)`; . Results from running that program via the command line (in Linux)

```
R CMD BATCH --no-save nes8804mnlgev.R
```

are in `nes8804mnlgev.Rout`.

Do the estimates suggest IIA is a reasonable assumption for these choices with the specified covariates? Beyond the statistic G , what other indications point to a choice between the MNL and GEV specifications? Does either model appear to be correctly specified (or nearly so)?

Do the results tend to confirm the basic implication of the spatial model?

What difference does it make if economic expectations are added to the linear predictor? A huge body of literature finds that individuals who think the national economy will get better tend to be more likely to vote for the presidential candidate of the party that currently occupies the White House. For the years included in the current data, the Republican party was the incumbent party in 1988, 1992 and 2004. In the other years the Democratic party was the incumbent party. Do estimates derived using the current data tend to confirm the usual finding? Does including economic expectations change the impression the first models suggested about the basic implication of the spatial model?

For all these questions, present a detailed response supported by the data in `nes8804e.csv` (you need not use **R** to do your analysis).

In case the variation in the sampling weights across observations in some years is a concern, the **R** program `nes8804mnlgevW.R` estimates the models while taking the weights

into account. The program does this by weighting observations in the loglikelihood. That is, using $w_i > 0$ to denote the sampling weight for observation i , with $n = \sum_{i=1}^n w_i$ (n denotes the sample size for a particular year), the respective loglikelihoods are modified to be

$$l_{\text{MNL}i} = \sum_{j=0}^2 w_i y_{ji} \log p_{ji}$$
$$l_{\text{GEV}i} = \sum_{j=0}^2 w_i y_{ji} \log q_{ji} .$$

Background: File `wrk8804e.R` contains the **R** program that built `nes8804e.csv` from the raw data extract file `nes8804e.dat`. See that file for the exact set of observations excluded because of missing data.

The models used in this assignment are simplified versions of the kind of models I used in the following two articles.

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

Mebane, Walter R., Jr., and Jasjeet S. Sekhon. 2002. "Coordination and Policy Moderation at Midterm." *American Political Science Review* 96 (March): 141–157.