PS 787 Assignment 2 (Nov 2, 2008; revised [2] Nov 15, 2008)

The question for this assignment is how to estimate effects of a couple of variables on partisan vote choices and on turnout and rolloff decisions in the 2000 U.S. general election—choosing between the Democrat and the Republican candidate running for the U.S. House of Representatives. We'll ignore (i.e., omit) cases where a candidate is running unopposed. The data are the same 2000 ANES data used for assignment 1. Specifications for the vote choice and turnout regressors can be the same as in the previous assignment (what to use for the rolloff option will require some thought).

The minimum is to use a GEV (nested logit) model for the four choices, but you may also wish to try out an MNP model. GEV models will be easier to compute and have other virtues. For the GEV model there are several possible nesting patterns. Choose one or try a couple to set up doing nonnested tests with them.

Files `assign2.gev.R`, `assign2.gev1.R` and `assign2.gev2.R` contain specifications for estimating three different GEV models. In `assign2.gev.R` the four choice categories are grouped

Don't Vote, (Rolloff, Vote Republican, Vote Democratic)

in `assign2.gev1.R` the grouping is

Don't Vote, Rolloff, (Vote Republican, Vote Democratic)

and in `assign2.gev1.R` the grouping is

(Don't Vote, Rolloff), (Vote Republican, Vote Democratic)

The first two specifications try to estimate one similarity parameter while the third one tries to estimate two.

To be a bit more precise, there are three matrices of attributes: X_1, X_2 and X_3, and correspondingly three coefficient vectors b_1, b_2 and b_3. These define four linear predictors, which represent the observed attributes for each of the four choice categories:

$$\begin{aligned}
Z_T &= X_1 b_1 && \text{(don't vote)} \\
Z_O &= X_2 b_2 && \text{(rolloff)} \\
Z_R &= X_3 b_3 && \text{(Republican vote)} \\
Z_D &= -X_3 b_3 && \text{(Democratic vote)}
\end{aligned}$$

Using $v_T = \exp(Z_T)$, $v_O = \exp(Z_O)$, $v_R = \exp(Z_R)$ and $v_D = \exp(Z_D)$, the three models have likelihoods based respectively on the following functions.

$$G_1 = v_T + \left(v_O^{1/(1-\tau)} + v_R^{1/(1-\tau)} + v_D^{1/(1-\tau)}\right)^{1-\tau} \quad (1)$$

$$G_2 = v_T + v_O + \left(v_R^{1/(1-\tau)} + v_D^{1/(1-\tau)}\right)^{1-\tau} \quad (2)$$

$$G_3 = \left(v_T^{1/(1-\sigma)} + v_O^{1/(1-\sigma)}\right)^{1-\sigma} + \left(v_R^{1/(1-\tau)} + v_D^{1/(1-\tau)}\right)^{1-\tau} \quad (3)$$

For each observation i the probability p_{ik} of choosing alternative k in these models is derived using

$$p_{ik} = \frac{v_{ik}}{G_i} \frac{\partial G_i}{\partial v_{ik}}$$

Files `assign2.geva.R`, `assign2.gevb.R`, `assign2.gev1a.R`, `assign2.gev1b.R`, `assign2.gev2.R` and `assign2.gev2b.R` contain variations of the preceding specifications that use different choices for X_2. All the models estimate with little difficulty except those in `assign2.gev2.R` and `assign2.gev2b.R`.

The **R** code should not be that difficult to modify to use alternative specifications for the observed attributes.

To estimate a model based on

$$G_4 = v_T + \left[v_O^{1/(1-\sigma)} + \left(v_R^{1/(1-\tau)} + v_D^{1/(1-\tau)} \right)^{(1-\tau)/(1-\sigma)} \right]^{1-\sigma}$$

it is necessary first to do the algebra to derive the choice probabilities.