Analyzing Vote Choice Data

Assignment 3 - Deadline: May 17, 2023

TA: Francesco Raffaelli (francesco.raffaelli@politics.ox.ac.uk)

1   R-PACKAGES

Before starting, let us review three `R-packages`, and their functions, that will be useful for this assignment - and in your research life!

- `plm` - This package is used to run linear (OLS) model with fixed effects. You must specify the estimation formula, the name of your dataset, the variable for which you are adding fixed effects (use term `index`), and add `model = "within"`. As usual, you may also use `modelsummary` to report your model.

  Figure 1.

  ```r
  modelname <- plm(formula = depvar ~ indepvar + covariates
                   data = datasetname,
                   index = "term",
                   model = "within")

  modelsummary(modelname, title = "Title", stars = TRUE, coef_rename = c("var1" = "Label1",
  "var2" = "Label2"))
  ```

- `bife` - This package is used to run logit models with fixed effects. Function `bife` is pretty straightforward: only the formula and the dataset name must be specified. You can then use function `modelsummary` to report the model in a nice, tidy table. The variable for which you are adding fixed effects should be added at the end of the formula, after the or-sign (the vertical bar).

  Figure 2.

  ```r
  modelname <- bife(formula = depvar ~ indepvar + covariates
                    data = datasetname)

  modelsummary(modelname, title = "Title", stars = TRUE, coef_rename = c("var1" = "Label1",
  "var2" = "Label2"))
  ```

- `lmer` - Finally, with this package you may run linear (OLS) models with random effects. You must specify the formula, adding as shown in the picture, the name of the dataset, and term `control`.

Figure 3.

```r
modelname <- lmer(formula = depvar ~ indepvar + covariates + (1 + firstterm | secondterm)
                  data = datasetname,
                  control = lmerControl(optimizer = "nloptwrap")
                  )

modelsummary(modelname, title = "Title", stars = TRUE, coef_rename = c("var1" = "Label1",
"var2" = "Label2"))
```

2   The Paper

This assignment focuses on fixed effects and clustered fixed effects, with a focus on settings with multiple countries. A good reference to understand these concepts and get through the assignment may be Schmidt-Catran and Fairbrother's *The Random Effects in Multilevel Models: Getting Them Wrong and Getting Them Right* (2015). The dataset, however, is unrelated to the paper *per se* and, this time, the assignment will not require you to replicate parts of it.

3   The Dataset

Dataset `ESS789.dta` contains some variables from waves 7, 8, and 9 of the European Social Survey (ESS) for 18 countries and covering years from 2014 to 2018. You need to upload it on R as you will use it for this assignment. In particular, the variables that we have selected for you are as follows:

- `cntry`: country

- `cregion`: region of each country

- `essround`: round of the ESS

- `ipequopt`: whether the respondent believes that it is important that people are treated equally and have equal opportunities (1-Important; 6-Not Important)

- `impfree`: whether the respondent believes that it is important to make own decisions and be free (1-Important; 6-Not important)

- `uemp5yr`: periods of unemployment experienced by the respondent in the five previous years

- `gndr`: respondent's gender

- `agea`: respondent's age

- `cons`: level of social conservatism in a given country in a given year

- `env`: level of green attitudes in a given country in a given year

- `eummbr`: whether a country is a member of the EU (0/1 dummy variable)

- `mnrchy`: whether a country is Constitutional Monarchy (0/1 dummy variable)

Upload the assignment on Canvas by the aforementioned deadline. Rename the pdf document obtained from the R-markdown as follows:

"AVCD-Assignment3-YOURLASTNAME"

4 Exercise 1

1. Let us do some data preparation:

   - Make sure that variable `gndr` is a dummy taking values 0/1

   - Rescale variables `ipequopt` and `impfree` so that higher values measure higher importance

   - Create variable `year` for each wave of the survey

   - Create a categorical variable `cohort` that measure in which decade the respondent was born. Make the variable have only 4 levels, one for each quartile of the year of birth distribution

2. Let us do some descriptives:

   - Look at the variables in the dataset: which ones vary at the individual level? Which at the country level? And which at the country-year level?

   - What's the mean value of variables capturing the importance of freedom and equality for respondents? Do they differ between countries with a Constitutional Monarchy and those without? And between EU members and non-members? Report your results in a nice, tidy table

   - Finally, for each observation, create a variable indicating how much more (or less) the respondent value freedom over equality

3. Which are the factors that better predict whether a respondent prefers freedom over equality? (Hint: build your dependent variable first)

   - Plot the coefficients and comment their significance

   - Plot how the predicted probabilities of preferring freedom over equality change for male and female respondents conditionally on their experience of unemployment

4. Estimate the model above using year-level fixed effects

   - What do the year-level fixed effects exactly do?

   - What are the variables that change? How? And why those in particular?

5. If you were asked at which other level you would add fixed effects, what would you answer?

5 Exercise 2

1. Re-estimate the model above using year-level fixed effects. This time, however, use a different dependent variable: the level of country's conservativism

- Plot the coefficients

2. Re-estimate the model above using country-level fixed effects. (Hint: what class if the variable for country? Is it the most appropriate?)

   - Plot the coefficients

   - What does it change with respect with the model with year fixed effects? Why?

3. Random Effects: estimate the model using random effects for years and country

4. Some Theory:

   - What do fixed effects account for? Specify: for years and countries/geographic regions

   - What do random effects account for?

   - Following Schimdt-Catran and Fairbrother, illustrate the structure of fixed effects