Essay 2: Data Transformation and Interpretation

Due: 20. March at 12:00 to Thomas via Blackboard

The purpose of this assignment is to evaluate your ability to use complex data structures, identify research questions, and interpret regression. You will need to use both the country-level Quality of Government dataset and the individual-level, multi-country European Social Survey (ESS) Round 6 data to test a hypothesis involving conditional effects (i.e., interaction).

Your task:

1. We would like you to analyze individual-level data from the ESS as a whole (i.e., ignoring the fact that individuals come from different countries). Looking at the available variables in the ESS and QoG datasets, select an outcome variable, ask a causal research question about that outcome, and state an empirically testable hypothesis about a potential cause of that outcome. This hypothesis should include an expectation about conditional effects (i.e., an interaction term between two independent variables).

2. Identify variables that may confound the relationship between these variables, considering both individual-level and country-level variables, and arrive at a causal model you would like to understand using linear regression analysis. Note: Most of you will have a version of Stata that can only contain datasets with up to 2047 variables. Because the QoG dataset contains 2037 variables, you will need to use `keep` and `drop` to reduce the size of the dataset.

3. Select the relevant country-level variables from the QoG dataset and merge these country-level variables into the individual-level ESS dataset. You may also want to consider aggregating individual-level variables from the ESS into country-level variables (thus using `collapse` and multiple `merge` operations).

4. Using the merged country-level and individual-level data, estimate the coefficients for your regression model.

5. Using the coefficients, the `margins` command, and any other relevant tools, properly interpret the statistical significance and substantive size of the coefficients.

6. Using `marginsplot` or any other plotting commands, display one or more relevant relationships and interpret them appropriately.

7. Reestimate your model(s) now including country-level fixed effects (i.e., including a factor variable for country in your regressions). Reinterpret the results.

Data:

- QoG data: `http://www.qog.pol.gu.se/data/datadownloads/qogstandarddata/`

- ESS6 data: `http://www.europeansocialsurvey.org/data/download.html?r=6`