GSERM 2020: Regression for Publishing

Exercise Two

The empirical focus of this exercise is refugees, specifically, what causes countries to experience episodes of refugee inflows. Statistically, the point is to demonstrate mastery of models for binary response variables, of the sort we've discussed recently.

The data (`GSERM-2020-ExerciseTwo.csv`, available on CANVAS and at the course github repository) are panel / time-series cross-sectional[1] observations on 213 countries in the international system between 1964 and 1995, inclusive ($N = 213$, $T = 32$).

The main outcome of interest, `RefsDummy`, is coded 1 in country-years in which that state experienced an influx of refugees from another country (or countries), and 0 in years when it did not. In addition to country and year indicators, included covariates[2] are:

- `AidPC` – Bi/multilateral aid, in constant per capita dollars.

- `logPop` – ln(Population).

- `logGDP` – ln(GDP per capita), in constant 1995 dollars.

- `POLITY` – the -10 (maximally autocratic) to 10 (maximally democratic) POLITY IV regime score; scores of -10 indicate a fully autocratic regime in the host country, while a score of 10 is a fully democratic regime.

- `transition` – an indicator variable coded 1 in a year in which that country experienced a transitional government, and 0 otherwise.

- `civilwar` – a similar indicator variable for the occurrence of a civil war in that country in that year.

Your assignment is once again simple: Estimate a model (or more than one) of your choosing, where the response / dependent variable is `RefsDummy`. Feel free to specify the model as you see fit. Present your results, and interpret your findings, using the techniques discussed in class.

This homework is due at or before 11:59 p.m. CET on **Tuesday, June 23, 2020**. Your assignment should be submitted as a .PDF file, e-mailed to `zorn@psu.edu`.

---

[1]For purposes of this exercise, you should ignore any and all potential panel/TSCS issues that might arise. One thing at a time...

[2]Note that all covariates are measured for the "host" (or "target") country, *not* for the "sending" country; these data contain no information on the latter.