

PLSC 503: “Multivariate Analysis for Political Research”

Exercise Seven

Part I

For the simulation part of this exercise, consider a data-generating process for a latent variable Y^* of the form:

$$Y_i^* = 0 + 2X_{1i} - 4X_{2i} + 3X_{3i} + u_i \quad (1)$$

with

$$Y_i = \begin{cases} 0 & \text{if } Y_i^* \leq 0 \\ 1 & \text{if } Y_i^* > 0, \end{cases}$$

$X_1, X_2, X_3 \sim \text{i.i.d. } U(0, 1)$, and $u_i \sim \text{Logistic}(0, 1)$.¹ Using simulations:

1. Estimate the fraction of cases where the OLS regression

$$Y_i = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \gamma_3 X_{3i} + e_i$$

yields predictions \hat{Y}_i which are either less than zero or greater than one. What can you say about this fraction? Why is it what it is?

2. Illustrate, describe, and discuss the relationship between the estimated logit coefficients

$$\Pr(Y_i = 1) = \Lambda(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})$$

and the regression coefficients in the data-generating process in (1).

3. Reexamine the relationship in question 2, changing the data-generating process to one where $u_i \sim \text{Logistic}(0, 4)$. How does this change the nature of the relationship? Why? Why does it matter (or not)?

¹Note that here the second term is the “scale parameter” of the [logistic distribution](#); the standard logistic distribution has a scale parameter of 1.0.

Part II

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 been discussing for the past couple classes.

The data (`PLSC503-2020-ExerciseSeven.csv`) are panel / time-series cross-sectional² observations on 213 countries in the international system between 1964 and 1995, inclusive ($N = 213$, $T = 32$). The main covariate of interest, `RefsDummy`, is coded 1 in country-years in which that state experienced an influx of refugees from another country, and 0 in years when it did not. Included covariates³ are:

- `AidPC` – Bi/multilateral aid, in constant per capita dollars,
- `logPop` – $\ln(\text{Population})$,
- `logGDP` – $\ln(\text{GDP per capita})$, in constant 1995 dollars,
- `POLITY` – the -10 (maximally autocratic) to 10 (maximally democratic) POLITY IV regime score,
- `transition` – an indicator variable coded 1 in a year in which that country experienced a transitional government, and 0 otherwise, and
- `civilwar` – a similar indicator variable for the occurrence of a civil war in that country.

Your assignment is simple: Estimate a model (or more than one) of your choosing, where the response variable is `RefsDummy`. Feel free to specify the model as you see fit. Interpret your findings, using the techniques discussed in class. This assignment is due (electronically, obviously) at 5:00 p.m. EST on Monday, April 20, 2020, and is worth the usual 50 points.

²For purposes of this exercise, you should ignore any and all potential panel/TSCS issues that might arise. One thing at a time...

³Note that all covariates are measured for the “host” (or “target”) country, *not* for the “sending” country; these data contain no information on the latter.