PLSC 504: "Topics in Political Methodology"

Exercise Five
October 11, 2022

Part I

The first part of this exercise is a simulation-based exploration of one-way unit effects models. Broadly speaking, it's designed to answer a general question: Under what condition(s) will within- and random-effects models for panel/TSCS data give different (and therefore possibly misleading) results? In particular, we're doing to focus on the key assumption of "random effects" models, that $\text{Cov}(X_{it}, \alpha_i) = 0$.

While keeping the goal of answering this question in mind, do something along the lines of the following:

1. Begin with simulations of relatively small panel-style data (say, with $N = T = 10$).

   (a) Generate data with one-way unit-level effects – that is, as:

   $$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it} \tag{1}$$

   – where the "usual" OLS-type assumptions hold for the u_{it}s and where the assumptions of the "random-effects" model are met [that is, where $\text{Cov}(X_{it}, \alpha_i) = 0$].

   (b) Fit and compare within- and random-effects models on those data, examining and comparing the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ along with their estimated standard errors across the two types of models.

   (c) While you're at it, use your simulations to examine the results (here, P-values) of Hausman tests for the choice between within- and random-effects models.

2. Repeat the steps in (1) above, but vary the values of N and/or T, to illustrate if/how sample sizes (in both N and T) affect the conclusions you draw. In doing so, recall that statistical properties generally change in N / T in a geometric, rather than a linear, way.

3. Repeat the steps in (1) - (2), but relax the assumption that $\text{Cov}(X_{it}, \alpha_i) = 0$.[1] Assess how and to what extent a non-zero correlation between X_{it} and α_i impacts the results estimated by within- and random-effects models.

---

[1]Hint: While there are a lot of alternatives, a simple way to generate data where X_{it} and α_i are correlated is to start by generating your unit-level α_is, then generate your Xs as a function of the α_is plus some mean-zero random noise. So, for example:

```
N <- 10
T <- 10
A <- rep(rnorm(N),each=T)
X <- (A + rnorm(N*T))/2
```

will generate $NT = 100$ observations, each with a value $\alpha_i \sim N(0, 1)$ and a Standard-Normally-distributed X_{it} such that the expected correlation between X_{it} and α_i will be about $r = 0.67$. Increasing or decreasing the relative proportion of variation in X_{it} that is shared with α_i will yield different degrees of correlation between X_{it} and α_i; so, for example, in this particular case (with $N = T = 10$) setting

```
X <- (A + 1.6*rnorm(N*T))/2
```

will lead to a correlation between X_{it} and α_i of around $r = 0.5$.

Part II

For the second part of this exercise, we'll use "country-year" data comprising annual measurements (from 1945-2014) on several variables for approximately 180 countries in the international system ($N \approx 180$, $T = 70$, unbalanced). The data are available on the course github repository, in the "Exercises" folder. The variables in those data are:

- `Country`: The name of the country for that observation;
- `CountryCode`: A three-digit country identifier;
- `Year`: The year of the observation;
- `POLITY`: The country's POLITY IV score in that year, ranging from -10 (fully autocratic) to 10 (fully democratic);
- `PercentLiterate`: The percentage of the population of that country in that year who are literate;
- `UnivEnrollmentPerK`: University enrollment (per 1000 population) in that country during that year;
- `GDP` : The country's Gross Domestic Product that year (at factor cost);
- `TotalTrade`: The value of imports + exports for that country in that year, per capita, in constant U.S. dollars;

This part of the assignment is simple, in that it involves answering the following question: What, if anything, is the association between a country's degree of education (measured in terms of literacy and university enrollment) and the extent of its involvement in the international economy (defined as total trade)? Use the tools we learned in the October 10 class session, and include "control" (other) variables in the data as you see fit (though you need not go outside the provided dataset to do so).

This assignment is due *electronically*, as a *PDF file*, at 11:59 p.m. EST on Tuesday, October 18, 2022, and should include both answers to each part and the code necessary to replicate your analyses. You should submit your homework by emailing copies **both** to Dr. Zorn (`zorn@psu.edu`) and to Mr. Burnham (`mike.burnham@gmail.com`). This assignment is worth 50 possible points.