SOCY7706: Longitudinal Data Analysis
Instructor: Natasha Sarkisian
Assignment 3
Due: November 18 at 11:59pm

For this assignment, you will conduct data management, run all the necessary analyses, conduct diagnostics and apply remedies, and write a brief interpretation of your findings. You will submit your do file and your annotated log that will contain the output (with your brief comments) for all of the tasks that you will perform for this assignment. Make sure to provide brief comments throughout your do-file and log file that will clarify your steps and decisions; also, make sure to paste graphs to the corresponding locations in the log. There is no page limit for your annotated log but please edit it to contain only the relevant syntax, output, and graphs (i.e., omit any unproductive steps).

1) Read the following article, and then, for your assignment, conduct a study that will explore the relationship between education finance policies and higher education access using the same data but for an expanded period of time and potentially a different set of countries.

> Lijing Yang and Brian McCall. 2014. "World education finance policies and higher education access: A statistical analysis of World Development Indicators for 86 countries." *International Journal of Educational Development* 35: 25–36.

2) The dataset for this assignment is available on the course website: wdi_education.zip. I limited the dataset to the variables that you will need; you will, however, need to construct a lagged version of one of the variables (see the methods section of the article) as well as create a dichotomy for developed vs less developed countries (use their list of countries as a guide, but make your own decision which countries to include in the analysis based on data availability; note that for this assignment, we will not do any imputation.) Convert the dataset from wide into long; keep only the countries you want to use, construct the variables that you need and xtset the dataset.

2) Examine your variables using xtsum, xttab, and xttrans, depending on the type of variable. Examine univariate normality, bivariate linearity, and univariate outliers and apply remedies if needed.

3) Estimate a fixed-effects (FE) model. Estimate a between-effects (BE) model and compare the results to the FE model. Estimate a random-effects (RE) model, use xttest0 to test the hypothesis that all unit-specific residuals are zero, and hausman test to test that RE model is correctly specified. Based on your inspection of FE and BE coefficients as well as hausman test, decide whether FE or RE model is more appropriate. If RE model is not appropriate, estimate a model examining separately within and between effects in a random effects model. Estimate your final model with standard errors adjusted for clustering.

4) Generate residuals for your model (level 1 and level 2, as appropriate for the model type) and examine them for normality, linearity, and outliers; if necessary, modify your model. Evaluate additivity and multicollinearity as well.

5) Test for autocorrelation of residuals using xtserial, and, based on the decision made in #3, estimate either a FE or RE model allowing for the autoregressive error term (using xtregar). Evaluate the strength and direction of autocorrelation by examining the estimate of ρ as well as the modified Durbin-Watson and LBI statistics.

6) Estimate your final model of choice (FE or RE) with fixed effects of time included. Evaluate whether a linear or quadratic trend for time might be appropriate (i.e., compare model fit with dummies for time vs a trend variable).

7) Present the results of various models generated (FE and RE, with and without adjustments for clustering), a model with AR adjustment (either RE or FE), and model including time, either as dummies or as a trend, in one table as multiple columns (omit those models that are clearly inappropriate or unnecessary) and discuss differences in findings. Identify which model or models you choose for presentation in a journal article and discuss why. Write up a brief interpretation of findings for that model/ those models.