SC706: Longitudinal Data Analysis Instructor: Natasha Sarkisian

**Assignment 2**
**Due December 10 at 5PM (by email)**

For this assignment, you will utilize an imputed subset of the Cross-National Time Series, 1815-1973 dataset, collected by Arthur S. Banks. The data in this subset are from years 1968-1973 for 113 countries. The dataset is located at http://www.sarkisian.net/sc706/crossnat.dta and the codebook can be found at http://www.sarkisian.net/sc706/codebook_crossnat73.doc.

1) Devise a theoretical argument you can test through linear regression on these data, with approximately 3-6 independent variables in addition to the time variable (v2).
2) Examine your variables using xtsum (or xttab and xttrans if appropriate). Use xtreg to examine the amount of variance in your dependent variable that is due to level 2 variation. Examine univariate normality, bivariate linearity, and univariate outliers and apply remedies if needed.
3) Fit an OLS regression model and examine residuals for heteroscedasticity, both graphically and using hettest command. Briefly describe what you see. Estimate OLS with robust standard errors and with cluster correction; compare the results.
4) 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 country-specific residuals are zero, and hausman test to test that RE model is correctly specified. Based your visual inspection of FE and BE and 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 robust standard errors and adjusting for clustering. Generate residuals for that model (level 1 and level 2) and examine them for normality, linearity, and outliers; if necessary, modify your model.
5) Test for autocorrelation of residuals using xtserial, and, based on the decision made in #4, 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 $\rho$ as well as the modified Durbin-Watson and LBI statistics.
6) Estimate population-averaged model using xtreg, pa as well as xtgee. Examine correlations among residuals you obtained from the OLS model (to calculate correlations across time points, you have to reshape wide the file containing predicted residuals – don't forget to save your working file before reshaping!). Decide what kind of covariance structure would be most appropriate and estimate the corresponding model using xtgee.
7) Use GLS models to assess whether there is heterogeneity of residuals; if necessary, introduce adjustments for autocorrelation as well.
8) Present the results of various models generated by OLS (with and without adjustments for clustering and robust SE), FE and RE (with and without adjustments for clustering and robust SE), xtreg with AR, population averaged xtreg, xtgee with the covariance structure you selected, and xtgls in one table as multiple columns (but omit those models that are clearly inappropriate) and discuss differences in findings. Identify which model or models you would choose for presentation in a journal article and discuss why. Briefly interpret the results of those models.
9) Graphically examine trajectories for your outcome variable for different countries and conclude if you think slopes vary across countries. Estimate mixed effects model allowing the slope of time to vary. Assess whether the linear trend for time is appropriate and modify your model if necessary. If no significant variance in the effect of time is discovered, remove that random

component from the model. Introduce time-varying predictors, allow their slopes to vary as well and assess if there is significant variance for each. If there is significant variance in any of the slopes, create level 2 variables by generating averages of your time-variant predictors and use those to explain variance in the intercept as well as slopes (by using these level 2 variables as predictors as well as by generating cross-level interactions). Finally, reduce the number of included predictors and cross-level interactions by keeping only those that are statistically significant, and test whether such a reduction is appropriate using LR test and BIC. For your final model, calculate % variance explained at each level as well as total. Generate residuals (level 1 and level 2) and assess normality, linearity, and outliers; if necessary, modify your model.
10) Present the results of the model with random slopes for time (with no other variables) as well as your final mixed model in a separate table. Make sure to present both the coefficients and the variance components (random components). Briefly interpret the results.

Journal write-up component (optional):
Select a model or models to write up the results like you would for a journal publication. First, include an Introduction that will provide a short substantive description of your theoretical argument, your research questions and hypotheses (1 page max.). Second, include a brief Data and Methods section (1-2 pages) describing the variables and the analysis methodology. Include any discussion of diagnostics and modifications in this section, either in the text or in the footnotes. Also, include a table with summary statistics for the variables you use (means, standard deviations, number of observations). Third, provide a 1-3 page description of the results including tables (in journal format) and any graphs assisting in the interpretation of results. Finally, include a brief conclusion summarizing your findings and discussing contributions and limitations of your research (1 page max.). The page limit for this write-up is 7 pages double-spaced.

**Assignment 2 Grading Sheet**
**Total Preliminary Grade: out of 100**

1. Theoretical argument (5 points):
2. Univariate and bivariate data examination (5 points):
3. OLS model (5 points):
4. FE and RE models (15 points):
5. Autocorrelation test and xtregar model (10 points):
6. Population-averaged models (10 points):
7. GLS models (10 points):
8. Table of results and interpretation (10 points):
9. Mixed effects models (15 points):
10. Table of mixed effects results and interpretation (10 points):
11. Log organization (5 points):

**Journal-style Write-up Grading Sheet**
**Total Preliminary Grade: out of 100**

1. Introduction (15 points)
2. Data and methods (30 points)
3. Tables and graphs (15 points)
4. Description of results (30 points)
5. Conclusion (10 points)