SOCY7706: Longitudinal Data Analysis
Instructor: Natasha Sarkisian
Assignment 4
Due: December 20 by 5pm

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 variation in trajectories of CESD over time using the same data but using somewhat different analytic strategy (no typology created), using an expanded period of time, and including some time-varying predictors.

> Lincoln, Karen D. and David T. Takeuchi. 2010. "Variation in the Trajectories of Depressive Symptoms: Results from the Americans' Changing Lives Study." *Biodemography Soc Biol. 56(1)*: 24–41.

2) The dataset for this assignment is available on the course website: acls_cesd.zip. I limited the dataset to the variables that you will need (please use all -- CESD as your dependent variable, other variables as predictors); you will, however, need to convert the dataset from wide into long and xtset the dataset as well as decide on how to handle categorical predictors. Also decide whether you want to use wave number or actual year as your time variable (the years are 1986, 1989, 1994, 2002, 2011) and xtset the data correctly.

3) Examine the time-varying 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. Examine time trajectories of your dependent variable using xtline. Decide how to center each of the independent variables, create corresponding variables.

4) Follow the step-by-step model building algorithm for mixed effects models. That is, start by estimating the model with only the time trajectory; explore the trajectory shape (e.g., linear, quadratic) and variance in trajectories. Then add time-varying covariates and examine variance in their slopes. Note, however, that you may not be able to enter simultaneously and evaluate all the random slopes at once, so test them one by one at first. Use significance tests/BIC comparisons to decide which covariates should stay in the model and which slopes should be random. Next, add time-invariant variables – first, to explain variance in intercepts, and second, to explain variance in slopes. Use likelihood ratio tests and BIC comparisons to trim the model so that it is parsimonious. Use meaningful centering choices at all stages.

5) Generate residuals for your model and examine them for normality, linearity, and outliers; if necessary, modify your model. Evaluate additivity and multicollinearity as well. Estimate your final model with standard errors adjusted for clustering and using bootstrapping; if these standard errors differ substantially with those obtained without such adjustments, then use the model with adjustments as your final model.

6) Calculate percentage of variance that your final model explains on level 1 as well as for each of level 2 components.

7) Present the results of your final model in a table and write up a brief interpretation of your findings – interpret both the fixed effects (coefficients) and the random components, including correlations among random slopes, if relevant. Generate at least one graph of predicted values (using margins) to assist your interpretation.