THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

STATA TRAINING COURSE 2020/21

Dr. Mohan Bijapur

Syllabus (subject to revisions)

Pre-sessional online course

Introduction to STATA and data management tools

- Stata interface and files
- Using do-files
- Variables, locals and globals
- Changing directory
- Opening, browsing, saving, sorting, importing and exporting datasets
- Stata Help
- Saving results output in a log file
- Renaming, labelling, ordering, creating and replacing variables
- Merging and appending datasets
- Manipulating datasets *(reshape; collapse*)
- Using loops (*foreach; forvalues*)
- Summary statistics (*sum; _pctile; by(varlist)*)
- Useful functions for creating summary variables (*egen*)
- Creating graphs (line plots; scatter plots)

Core programme

Sessions 1 & 2: Regression analysis

Application: *Does Compulsory School Attendance Affect Schooling and Earnings? (Joshua D. Angrist and Alan B. Krueger, The Quarterly Journal of Economics, Vol. 106, No. 4 (Nov., 1991), pp. 979-1014*

- Data cleaning with large datasets
- OLS and output interpetation
- Hypothesis testing
- Residuals and fitted values (*predict*)
- Diagnostic tests

- Using robust and clustered standard errors
- Instrumental-variable estimators (*ivreg: (2sls, gmm*))
- Hausman test (in panel)
- Reporting output in tables (outreg2)

Sessions 3 - 5: Dynamic panel data models

Application: *Fertility and Economic development: a panel analysis*

- Stacking data in panel form
- Time series operators; creating date variables; encoding string variables; xtset
- Random effects estimator
- Fixed effects estimator
- Hausman test (in panel)
- Reghdfe and absorbing fixed effects
- Arellano–Bond estimator
- choice of instruments: endogenous vs. pre-determined vs. exogenous variables; (imposing maxlags)
- Sargan test of overidentifying restrictions
- testing for autocorrelation in first-differenced errors (if time permits)
- two-step Windmeijer bias correction (if time permits)
- Blundell – Bond System GMM estimator (if time permits)

Session 6: Time series methods

Application: *Monetary shocks and GDP*

- General discussion of time series
- tsset
- Autocorrelation and partial autocorrelation function
- Testing for unit roots (ADF)
- Vector autoregressions:
- reduced form VAR
- optimal lag structure (*varsoc*)
- Cholesky-orthogonalized impulse response functions
- forecast error variance decompositions
- Revision: Q&A