

Advanced Microeconometrics with STATA
by
Centre for Monetary and Financial Economics

Winter 2015

Instructor: David T. Jacho-Chávez, Robert J. Petrunia and Marcel C. Voia

Time: Classes will meet on February 23, 24, and 26, 2015, from 09:30-16:30.

Location: Industry Canada – CD-Howe Building – 131-A

Course Objectives: The purpose of this course is to provide professionals with the necessary tools to extract data from large databases, apply and develop econometric methods using STATA programming tools. The course is designed for new and intermediate STATA users who want to acquire advanced econometric skills and programming in STATA. Besides tools for data analysis, this course exposes participants to current empirical work along with microeconomic topics and techniques common to the analysis using large datasets. In addition to the emphasis on the statistical inference of these models, we will stress their empirical relevance. After taking this course, the participants should be able to:

1. Perform database management and estimation tasks using STATA.
2. Leverage STATA programming routines and user-contributed .ado files.
3. Understand empirical research using microeconometrics, and choose appropriate models and estimators for given economic applications.
4. Interpret model estimates and diagnose potential problems with models and know how to remedy them.

Pre-requisite: Undergraduate Econometrics and Matrix Algebra

Textbook: Participants will be provided with a complete set of lecture notes.

However, some participants may find the following books a useful resource: Cameron, A. C. and P. K. Trivedi (2005) "Microeconometrics using Stata," Revised edition, Stata Press.

An excellent self-learning resource is: <http://www.ats.ucla.edu/stat/stata>

Software: STATA – Data Analysis and Statistical Software. The computer room and the software access will be provided by Industry Canada with the acknowledgement of StataCorp.

Registration Fee for STATA Course with Instructors is \$3,000 per applicant. Industry Canada will cover seven seats for a total of \$21,000 plus HST (\$2,730), which equals \$23,730.

Course Outline

Section One: Data handling and efficient use of STATA using efficiently multiple files in an analysis, linking, shortcuts, macro. Topics for this section may include:

1. STATA basics:
 - (a) Advantages and disadvantages.
 - (b) File types: dta, do, log.
 - (c) Command structure - commenting.
 - (d) Pre-Data loading: memory, matrix size (matsize), maximum variables (maxvar).
2. Data handling
 - (a) Loading data: STATA datasets, insheet, infile, infix, using a data dictionary.
 - (b) Reducing memory requirements (compress).
 - (c) Adding variable description (label).
 - (d) Missing values.
 - (e) Merging additional datasets.
 - (f) Stata SQL.
3. Macros and loops:
 - (a) Local macros.
 - (b) Global macros.
 - (c) foreach, while loops.
 - (d) Matrices and scalars.
4. Basic Data manipulation:
 - (a) New or redefine variables: generate and replace, and conditioning statements.
 - (b) Indicator/Dummy variables and the xi command.
 - (c) Renaming variables.
 - (d) destring, tostring.
 - (e) longitudinal data: wide versus long forms.
5. Summary statistics:
 - (a) Summarize.
 - (b) Creating summary statistics data: collapse.
 - (c) Creating summary statistics data using matrix manipulation.
 - (d) Adding summary statistics to current data: egen.
6. Introduction to analysis
 - (a) Tabulations.
 - (b) Regression: OLS/GLS/IV.
 - (c) Graphics.
 - (d) Post estimation.

Section Two: Popular linear and nonlinear models in cross-sectional Econometrics and causal inference. Topics for this section may include:

7. Bootstrap
 - (a) Bias correction, confidence interval construction & hypothesis testing.
8. Quantile Regression
 - (a) Quantiles & quantile functions.
 - (b) Quantile regression model, estimation & estimates interpretation.
9. Binary Discrete Choice Models:
 - (a) Binomial discrete response models: Logit & Probit.
 - (b) Marginal effects.
10. Policy Evaluation and Matching Estimators:
 - (a) The Selection-on-Observables assumption.
 - (b) Average Treatment Effect, Average Treated on the Treated effect, Quantile
 - (c) Treatment Effect.
 - (d) Difference-in-Difference estimation.
 - (e) Case studies illustrating how to choose a method of matching
 - (f) Approches to executing the matching, e.g. programming from scratch versus using an embedded program in STATA.

Section Three: Panel data methods. Topics of this section may include:

11. How to handle Panel Data in STATA (unit of analysis, data organization, data operations).
12. Describing panel data (between and within group components of variation, transition tables).
13. Linear models for static panels:
 - (a) Pooled, random-effects and fixed-effects.
 - (b) Endogeneity in static panels.
14. GMM/Dynamic Panel Data Estimators with Emphasis on Arellano-Bond / Blundell-Bond.
 - (a) When to use GMM.
 - (b) How to execute GMM in Stata, including commands to ensure consistent standard errors in the presence of issues on the error term
 - (c) How to estimate Arellano-Bond / Blundell-Bond types of models
15. Guest Instructor: Kim P. Huynh (Bank of Canada).
 - (a) Applications to high-dimensional fixed effects i.e. firm-worker matched datasets.
 - (b) Frisch-Waugh-Lovell Method.
 - (c) a2reg, gpreg and reg2hdfe.

Disclaimer: The course outline provides a set of suggested topics. Topics may be expanded or omitted depending on time constraints.