

# Advanced Microeconometrics with STATA

WINTER 2014

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

**Time:** Classes will meet on 12/03/2014, 13/03/2014 and 14/03/2014 from 09:00-16:00.

**Location:** Room C869 Loeb Building, Carleton University.

**Course Objectives:** The purpose of this course is to provide professionals with the necessary tools to manage and work with large administrative databases using STATA programming tools. The course is designed for new and intermediate Stata users who want to acquire advanced skills in data management and programming in STATA. Besides tools for data management, this course exposes participants to current empirical work along with microeconomic topics and techniques common to the analysis using large administrative 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.
5. Possess a sufficient grounding in econometric theory to conduct research with administrative data.

**Pre-requisite:** Undergraduate Econometrics and Matrix Algebra

**Textbook:** Participants will be provided with a complete set of lecture notes each week. 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. Other books you may find useful for additional reading in certain topics are:

Cameron, A. C. and P. K. Trivedi (2005) *“Microeconometrics: Methods and Applications,”* 1st edition, Cambridge University Press.

Wooldridge, J. M. (2010) *“Econometric Analysis of Cross Section and Panel Data,”* 2nd edition, MIT Press.

**Software:** *Stata* – Data Analysis and Statistical Software.

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

# Course Outline

## Day 1 Wednesday 12 March 2014

**Whole section to be completed...** Instructors: Robert J. Petrunia (Lakehead University)

Data handling and efficient use of Stata using efficiently multiple files in an analysis, linking, shortcuts, macro. Topics for this day 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. **From Chavez**  
**OLS/GLS/IV.**
  - (c) Graphics.
  - (d) Post estimation.

## Day 2: Thursday, 13 March 2014

**Instructor:** David T. Jacho-Chávez (Emory University)

Popular linear and nonlinear models in cross-sectional Econometrics and causal inference. Topics for this day include:

7. Discrete Choice Models:
- Binomial Discrete Response Models: Logit & Probit.
  - Multinomial Discrete Response Models: Ordered Probit & Multinomial Logit.
  - Marginal Effects.
8. Censored & Truncated Regression Models:
- Censored & Truncated Normal Random Variables.
  - Maximum Likelihood Estimation.
  - Marginal Effects.
9. Sample Selection Models:
- Maximum Likelihood Estimator.
  - 2-Step Estimator.
  - Marginal Effects.
10. Treatment Effects.

## Day 3: Friday, 14 March 2014

**Instructor:** Marcel C. Voia (Carleton University)

Linear and nonlinear panel data methods. Topics of this day include:

14. How to handle Panel Data in STATA (unit of analysis, data organization, data operations).
15. Describing panel data (between and within group components of variation, transition tables).
16. Modelling Approaches:
  - (a) Linear models for static panels (pooled, random effects, fixed effects, endogeneity in static panels) and for dynamic panels.
  - (b) Nonlinear models for binary outcomes (conditional fixed-effects logit and Random effects logit and probit) and dynamic binary logit/probit models.
  - (c) Nonlinear models for ordered response models.
  - (d) Count data models.

I think we should stop here!

17. Incomplete panels and sample selection.
18. Policy evaluation using panel data.

**Disclaimer:** *The course outline provided is a guide. Some topics may be expanded or omitted depending on time constraints.*