

MODULE SPECIFICATION • ELECTIVE

Software for Econometrics and Big Data

Stata programming, do-files, ado-files and the Mata language — building the tooling to manage and analyse high-dimensional datasets.

10	15h	85h	100h
CREDITS	CONTACT HOURS	PRIVATE STUDY	TOTAL STUDY

COURSE CONVENOR

NAME Dr Vincent O’Sullivan

OFFICE By appointment — please email to arrange a convenient time

COURSE STRUCTURE

Econ 7917 is a 10-credit course and students are expected to input approximately 100 hours of study. The total number of contact hours is 15 hours, leaving 85 hours for private study. Delivery is by lectures — 15 hours across the first three weeks of term, comprising 10 hours of lectures and 5 hours of tutorials. Optional clinics run on the last day of the course.

During private study, balance the course material (the primary source) against the recommended textbooks: think critically about how these fit into the body of knowledge, perform the exercises, complete coursework and revise. You can expect to perform well only by working consistently through the year.

MODULE PROGRAMME

LECT.	TOPIC	SESSION CONTENT
01	Programming: Some Fundamental Concepts	- Do and ado files in Stata - Data types: working with dates and time, time series operators, factor variables - Debugging — the capture command - Protecting your data with preserve and restore
02	Guidelines for writing programs in Stata	- Basic guidelines for writing do-files - Enhancing speed and efficiency - How to seek help for Stata programming
03	Programming Stata Do-Files I	- Transformation of string and numeric variables - Recoding discrete and continuous variables - Recoding missing variables - Generate and egen functions - Local and global macros - Extended macro functions and list functions

LECT.	TOPIC	SESSION CONTENT
04	Programming Stata Do-Files II	- Scalars and matrices - Temporary variables and tempnames - Looping commands: forvalues and foreach - By, statsby, rolling and other prefix commands - Transforming data with reshape - Combining datasets with append and merge
05	Programming Concepts and Tools in Action I	- Tabulating and computing statistics across a set of variables and groups - Computing durations and aggregating transactions data - Summarising group characteristics over observations
06	Programming Concepts and Tools in Action II	- Adding aggregate characteristics to micro observations - Reusing results: return and ereturn commands - Accessing estimation results and margins
07	Automation — Do-File Programming I	- Storing results in matrices - Stata's post and postfile commands
08	Automation — Do-File Programming II	- Automating estimation output — the user-written commands estout and tabout, saveresults - Automating graphical output
09	ADO-File Programming I	- Structure of an ado-file - Syntax and return statements - Using program options - Including a subset of observations - Generalising commands to handle multiple variables - Rendering commands byable
10	ADO-File Programming II	- Writing an e-class program - Programs for egen, ml, nl - Programs for gmm - Guidelines for ado-file programming style
11	Fundamentals of Mata programming	- Mata fundamentals - A simple Mata function - Components of a Mata function - Calling Mata functions - Mata's st_interface functions

KEY SKILLS

By the end of this course, students should have some knowledge and understanding of:

- Understand the principles of Stata programming
- Write Stata do-files to automate the workflow
- Write ado-files to extend Stata with new commands
- Understand the basics of the Mata programming language
- Manage large datasets with the Stata software, and program using the Mata language
- Perform statistical and econometric analysis of high-dimensional datasets

**DESIRED OUTCOMES**

By the end of this course, students should be able to:

- Demonstrate a sound knowledge of applied econometric principles and basic quantitative techniques
- Present, interpret and analyse information in numerical form, and use econometric and other packages effectively
- Understand the relevance of different econometric approaches to specific applications in economics
- Communicate and present complex arguments with clarity and succinctness
- Plan and manage time effectively

ASSESSMENT

WEIGHTING	100% coursework
ISSUED	End of week 6 of term
DUE	End of week 10 — four weeks for completion
SUBMISSION	Electronically via the Moodle site for this course
FEEDBACK	Returned within 4 weeks of the deadline, as marker's notes appended to the PDF of your coursework

Marking criteria and penalties are set out in the Economics Undergraduate Handbook and the general course information paper, available via the Current Student page of the university website under Academic Regulations.

COURSE TEXT & RECOMMENDED READING

Lecture slides are made available on the course web pages, together with additional material and course announcements.

- Mitchell, M. N. (2010). Data Management using Stata: A Practical Handbook. Stata Press
- Baum, C. F. (2016). An Introduction to Stata Programming, 2nd edition. Stata Press