

XIV ITALIAN STATA USERS GROUP MEETING

PROGRAM

9.00 - 9.30 Registration

9.30 - 10.30 **SESSION I - INVITED SPEAKER**

MANIPULATION OF SVG GRAPHS WITH STATA • Tim Morris, MRC Clinical Trials, University College London

Stata 15 includes a newly-documented option to export graphs in SVG (scalable vector graphics) format. SVG is a great storage format because instead of describing individual pixels it describes the shapes to be rendered by an application. While other export formats also do this, SVG opens up many more possibilities. Unlike any other vector format used by Stata, SVG is supported by modern web-browsers. It is also relatively human-readable as .svg files are plain-text XML code. This opens up a world of opportunity to manipulate SVG files themselves and go beyond the graphics Stata can currently create.

This talk gives an introduction to how SVG can be manipulated via several examples: invoking transparent elements 'by-hand', hexagonal binning, embedding images to appear behind semi-transparent graphs, and including interactive elements in graphs. I hope that this will encourage creative ideas from other users for further extending Stata graphics via SVG.

10.30 - 10.50 Coffee Break

10.50 - 12.00 **SESSION II - COMMUNITY CONTRIBUTED, I**

Theory and Practice of TFP Estimation: the Control Function Approach Using Stata • Gabriele Rovigatti, University of Chicago Booth School of Business and Vincenzo Mollisi, Free University of Bozen-Bolzano

Alongside Instrumental Variable (IV) and Fixed Effects (FE), the Control Function (CF) approach is the most widely used in production function estimation. Olley-Pakes (OP henceforth), Levinsohn-Petrin (LP), Akerberg-Caves-Frazer (ACF) have all contributed to the field proposing two-steps estimation procedures, while Wooldridge showed how to perform a consistent estimation within a single step GMM framework. In this paper we propose a new estimator, based on Wooldridge's, using dynamic panel instruments à la Blundell-Bond and we evaluate its performance by Monte Carlo simulations. We also present a new Stata module - `prodest` - for production function estimation, show its main features and key strengths in a comparative analysis with other user-written Stata commands. Lastly, we provide evidence of the numerical challenges faced when using OP/LP estimators with ACF correction in empirical applications and document how the GMM estimates vary depending on the optimizer/starting points employed.

Simulating dynamic panel data in Stata: New features of the `xtarsim` command • Giovanni S. F. Bruno, Bocconi University

The Stata command `xtarsim`, which I developed in 2005, simulates dynamic panel data models with exogenous regressors and i.i.d. errors. I have now extended the command in order to also simulate models with various types of endogenous or predetermined regressors and with MA(1) errors. This paper illustrates the new version of `xtarsim` and presents Monte Carlo applications.

CBPS: a Stata command to implement Covariate Balancing Propensity Score • Filip Premik, University of Minnesota

A dual nature of propensity score manifests itself in being both the conditional probability of treatment assignment and covariate balancing score. The standard approach in propensity score estimation exploits the first feature leaving balancing properties to be checked after estimation. Imai and Ratkovic2014 focus also on the second feature and propose Covariate Balancing Propensity Score (CBPS) estimator that automatically balances the conditional distribution of covariates. Being stated within GMM framework, CBPS is a simple way to obtain propensity score estimates or weights to be used in subsequent estimations. Monte Carlo studies indicate its good performance among others in reducing bias of treatment effects estimates. The paper reviews the method and introduces Stata user written package CBPS which implements the estimator.

**12.00 - 13.00 SESSION III - EXPLOITING THE
POTENTIAL OF STATA 15, I****A JOURNEY TO LATENT CLASS ANALYSIS (LCA) • Jeff
Pitblado, Director of Statistical Software, StataCorp**

Stata's estimation commands have evolved in how they account for groups in the sample. Since the early days of Stata, fitting models with group-specific parameters is simply a matter of using the `if` clause to condition on group membership. Inference between group-specific parameters was made possible with the introduction of `suest` in Stata 8. In Stata 12, we introduced `sem` and group analysis for structural equation models (SEMs). Stata 15 introduces two kinds of group analysis for generalized SEMs. For observed groups, `gsem` has the new `group()` option. For latent groups, `gsem` has the `lclass()` option and the ability to perform LCA.

13.00 - 14.15 Lunch

**14.15 - 16.00 SESSION IV - EXPLOITING THE
POTENTIAL OF STATA 15, II****Calcolo dell'aderenza alle terapie farmacologiche a
partire dai flussi amministrativi correnti • Jacopo Lenzi,
Alma Mater Studiorum University of Bologna**

Il Medication Possession Ratio (MPR) e la Proportion of Days Covered (PDC) sono le più note misure di aderenza alle terapie farmacologiche derivanti dai flussi amministrativi sanitari. Entrambe esprimono in termini percentuali quanta parte del follow-up individuale del paziente è coperto dal farmaco in studio. L'obiettivo di questo contributo è fornire alcuni consigli per calcolare tali indicatori di farmaco-aderenza usando il software

Stata, e presentare alcune applicazioni pratiche su popolazioni di pazienti affetti da patologia cardiovascolare.

**Matching tra due coorti consecutive estratte da un registro di
patologia • Dino Gibertoni, Alma Mater Studiorum University
of Bologna**

Utilizzando i dati di un registro di patologia si possono realizzare studi longitudinali che indagano i determinanti di uno o più esiti di interesse per la patologia. In particolare, si può usare una coorte estratta dal registro per eseguire la validazione temporale di un modello predittivo che era stato sviluppato a partire da una coorte di pazienti dello stesso registro arruolati in un periodo precedente. Nella presentazione verrà illustrato il programma Stata che è stato scritto per realizzare un matching a due step, deterministico e probabilistico, con il quale ottenere due coorti omogenee di pazienti in cui confrontare gli esiti di patologia.

**Variabili Socio-Demografiche e Economiche del Voto per
Brexit: tendenze territoriali • Salvatore Leonardo Alaimo,
Sapienza University of Rome**

Il 23 giugno 2016 si è tenuto il referendum relativo alla permanenza della Gran Bretagna nell'Unione Europea ("British exit" o "Brexit").

Già nel 1975 i cittadini britannici erano stati chiamati a esprimersi in merito alla permanenza della Gran Bretagna nell'allora Comunità economica europea. L'affluenza al c.d. "Common market referendum" era stata circa del 65% e il Remain si era affermato in maniera netta (67%) in tutte le circoscrizioni, restituendo l'immagine di una nazione fortemente europeista. I risultati del referendum del giugno 2016 sono stati diametralmente opposti. Anche in questo caso l'affluenza è stata molto elevata (72%), ma a prevalere è stato il Leave con il 52% delle preferenze. Viene fuori l'immagine di un Paese spaccato in due, diviso fra europeisti e antieuropeisti. Cos'è cambiato? .../...

**Socio-economic factors and ideal age of union formation.
A quantile regression approach • Brian W. Mandikiana and
Mahjabeen Ramzan, Qatar University**

There has been wide discussion about the relationship between marriage timing and socio-economic factors. However, empirical evidence on ideal age at first marriage for women, particularly in traditional societies where arranged marriages are common is still limited. We use a nationally representative Qatari Women Survey dataset to test the relationship between ideal age at first marriage and socio-economic factors. Given the presence of heterogeneity, in preferred ideal age at first marriage for young women, we apply the quantile regression method to make inference. Using the quantile regression methodology, we find that the influence of socio-economic factors differs across the ideal age distribution. In particular, there is a positive relationship between ideal age at first

marriage and post-secondary education. Based on the results, we offer suggestions to improve an understanding of the complex process of family formation.

16-00-16.15 Coffee Break

16.15-17.00 **SESSION V - COMMUNITY CONTRIBUTED, II**

DDID - Pre and post-treatment estimation of the Average Treatment Effect (ATE) with binary time-varying treatment

• Giovanni Cerulli, IRCrES-CNR

ddid estimates Average Treatment Effects (ATEs) when the treatment is binary and varying over time. Using ddid, the user can estimate the pre- and post-intervention effects by selecting the pre and post intervention periods, also by plotting the results in a easy-to-read graphical representation. Also, in order to assess the reliability of the causal results achieved by the user's specified model, ddid allows to test both the "common trend" assumption, and the degree of "balancing" achieved by the user's specified model. Thus, the model estimated by ddid can be seen as a generalization of the Difference-In-Differences (DID) approach to the case of many post- and pre-intervention times.

Combining Large Datasets of Patents and Trademarks • Grid Thoma, University of Camerino

Using the STATA entity names in patents and trademarks have been linked by the means of a matching algorithm which accounts for differences due to the position of the same word between otherwise identical strings. In particular, I have taken into the consideration the string similarity J index proposed by Thoma, Torrisi, Gambardella,

Guellec, Hall and Harhoff (2010), which computes the fraction of common words after breaking up the strings into words at the blank spaces. .../...

17.00 - 17.15 **SESSION VI - REPORT TO USERS WISHES AND GRUMBLES • JEFF PITBLADO, STATA CORP**

The "Wishes and Grumbles" session offers participants the opportunity to interact directly with StataCorp: providing participants with a forum in which to highlight any limitations encountered or suggest eventual improvements to the software.

20.15 Meeting Social Dinner (Optional)

An optional conference dinner will be held at OSTERIA CIPOLLA ROSSA, Via De' Conti 53/r, www.osteriacipollarossa.com.

We will be meeting directly at the restaurant at 20.15. Kindly note that for logistic purposes, we ask that you indicate on the registration form, whether you will be attending the conference dinner.

USING SIMULATION STUDIES TO EVALUATE STATISTICAL METHODS

INSTRUCTOR

Tim Morris, MRC Clinical Trials Unit at University College London

COURSE DESCRIPTION

Simulation studies are an invaluable tool for statistical research. They help us to understand the properties of statistical methods and to compare different methods. To perform a meaningful simulation study, careful thought needs to be put into planning, coding, analysis and interpretation.

This course will help participants to:

- Understand the rationale for simulation
- Appreciate the importance of careful planning and have a practical framework for planning their own simulation studies
- Have the tools to code and debug simple simulation studies in Stata
- Know how to analyse simulation studies producing estimates of uncertainty
- Present methods and results for publication

Examples will be taken from my experiences in medical statistics. The principles are the same for simulation studies in other applied areas but the examples may be less relevant.

Practical sessions will be taught using Stata. Participants should be familiar with Stata. For example, they should know how to generate data, run regression commands and produce simple graphs.

COURSE OUTLINE

9.00 Introductions

Lecture 1: Planning a simulation study

Practical 1: Plan a simulation study

11.00-11.15 Coffee Break

Lecture 2: Coding a simulation study in Stata

Practical 2: Coding a simulation study in Stata

13.00-14.00 Lunch

Lecture 3: Analysing simulation studies

Practical 3: Analyse your simulation study and feed back results

Lecture 4: Reporting simulation studies and course summary

16.30 End of Course

GENERAL INFORMATION

The meeting will take place in Florence at the Hotel Brunelleschi, Piazza Santa Elisabetta 3, on the 16th and 17th of November 2017.

Workshop participants will be provided with: a printed copy of the course handouts, the Stata do files and databases used through the workshop; together with a temporary licence of Stata 15 valid for 30 days. Participants will need to bring their own laptop with them.

The Maximum number of participants permitted will be restricted to 15.

Individuals interested in attending the Italian Stata User's Meeting must return their completed registration forms either by email (formazione@tstat.it) or by fax (+39 0864 206014) to TStat by the 5th of November 2017.

Further details regarding our registration procedures, including our commercial terms and conditions, can be found at <https://www.tstat.it/utenti/sug17>.

CONTACT INFORMATION:

Monica Gianni

Via Rettangolo, 12-14 | 67039 Sulmona (AQ)

Tel. +39 0864 210101 | Fax +39 0864 206014

training@tstat.eu | www.tstattraining.eu



The logo features the word 'TStat' in a stylized font. The 'T' is red and the 'Stat' is black. A white, dotted line forms a circular path around the 'T' and 'S', suggesting a simulation or data cycle.