II SESSION – USER WRITTEN COMMANDS AND ROUTINES, I

***datanet:* a Stata procedure to facilitate dataset organization for network analysis** Giovanni Cerulli and Antonio Zinilli, CNR-CERIS National Research Council of Italy, Unit of Rome.

In recent years much interest has been focused on network analysis. This study presents and applies to real data a new user-written Stata program called datanet, which facilitates the dataset organization for network analysis' purposes. Given a fixed number of units (or nodes) belonging to the same group (there will be a variable denoting group membership), possibly connected one each other or possibly not, this routine creates a new dataset containing all their possible couplings to then be easily exploited using Stata network analysis' commands. To our knowledge, no routine has been developed in Stata so far that executes this type of procedure. Moreover, this presentation will also review how to perform some basic network analysis in Stata, and discuss further network analysis' applications we deem of worth developing in Stata in the near future.

Approximate Bayesian logistic regression via penalized likelihood estimation with data augmentation – Andrea Discacciati, Karolinska Institutet Stoccolma.

Data augmentation is a technique for conducting approximate Bayesian regression analysis. This technique is a form of penalized likelihood estimation where prior information, represented by one or more specific prior data records, generates a penalty function that imposes the desired priors on the regression coefficients. We present a new command, penlogit, that fits penalized logistic regression via data augmentation. We illustrate the command through an example using data from an epidemiological study.

Efficient and effective management of big databases in Stata – Giovanni Marin, CNR-CERIS Milano.

Il modulo si occuperà della gestione della descrizione delle modalità di gestione di grandi dataset in Stata in maniera efficace ed efficiente. Il corso sarà basato sia su considerazioni generali delle problematiche connesse alla gestione di grandi dataset che su esempi pratici su specifici datasets. Il modulo tratterà i seguenti temi:

- informazioni di base sui formati delle variabili in Stata;
- modalità di importazione di dataset in formati diversi dal .dta;
- reshaping di dataset in Stata;
- nozioni di base su database relazionali;
- metodi per combinare più dataset;
- suggerimenti per ottimizzare l'archiviazione di grandi dataset.

***sftfe*: a Stata command for fixed-effects stochastic frontier models estimation** – Federico Belotti, CEIS Università degli Studi di Roma Tor Vergata e Giuseppe Ilardi, Banca d'Italia.

The classical stochastic frontier panel data models provide no mechanism to disentangle individual time invariant unobserved heterogeneity from ine_ciency. Greene (2005a,b) proposed the so-called "true" fixed-e_ects specification that distinguishes these two latent components and allows for time varying ine_ciency. However, due to the incidental parameters problem, the maximum likelihood estimator proposed by Greene leads to biased variances estimates in short panels. sftfe allows the estimation of this model via three alternative estimators (Belotti & Ilardi, 2012; Chen et al., 2014) that, by relying on a data transformation, achieve consistency for n ! 1 with fixed T. Of special notes is that sftfe allows the underlying mean and variance of the ine_ciency to be expressed as functions of exogenous covariates. Furthermore, the new command allows the estimation of a "true" fixed-e_ects model in which the inefficiency is assumed to follow a first-order autoregressive process. These features may be considered relevant from the methodological point of view since both model parameters and ine_ciency estimates may be adversely a_ected when ine_ciency heterogeneity, heteroskedasticity and serial correlation are neglected. They are also important from the empirical perspective because they allow to test specific hypothesis of interest and policy implications, avoiding biased two-step procedures.

***ntreatreg:* a Stata module for estimation of treatment effects in the presence of neighborhood interactions** – Giovanni Cerulli, CNR-CERIS National Research Council of Italy, Institute for Economic Research on Firm and Growth.

This paper presents a parametric counter-factual model identifying average treatment effects (ATEs) by conditional mean independence when externality (or neighborhood) effects are incorporated within the traditional Rubin-potential outcome model.As such, it tries to generalize the usual control-function regression, widely used in program evaluation and epidemiology, when the stable unit treatment value assumption (SUTVA) is relaxed. As a by-product, the paper also presents ntreatreg, a user-written Stata routine for estimating ATEs when social interaction may be present. Finally, an instructional application of the model and of its Stata implementation (using ntreatreg) through two examples (the first on the effect of housing location on crime; the second on the effect of education on fertility) is shown and results compared with a no-interaction setting.

III SESSION – USER WRITTEN COMMANDS AND ROUTINES, II

Dynamic Documents in Stata: *MarkDoc, Ketchup,* and *Weaver* – E. F. Haghish, University of Freiburg Germany.

For Stata users who do know LaTeX, writing a document that includes text, graphs, and Stata syntax and output has been a tedious and unreproducible manual process. To ease the process of creating dynamic documents in Stata, many Stata users have wished to see two additional features in Stata, which are literate programming and combining graphs with logfiles in a single document. Weaver, MarkDoc, and Ketchup are three user-written Stata packages that allow you to create a dynamic document that includes graphs, text, and Stata codes and outputs and export it in a variety of file formats including PDF, Docx, HTML, LaTex, OpenOffice/LibreOffice, EPUB, etc. I will also discuss further details about the specialties of these packages and their potential applications.

The A to Z of How to Create Thematic Maps of Italy using *spmap* – Maurizio Pisati, Università di Milano Bicocca.

The purpose of this talk is to present a step-by-step tutorial on how to draw thematic maps of Italy using the Stata user-written command -spmap- and spatially-referenced data freely available on the Internet.

V SESSION – APPLIED RESEARCH USING STATA

Average partial effects in multivariate probit models with latent heterogeneity: Monte Carlo experiments and an application to immigrants' ethnic identity and economic performance – Giovanni Bruno, Università Commerciale Luigi Bocconi Milano and Orietta Dessy, Università Ca' Foscari Venezia.

We extend the univariate results in [Wooldridge, J. M. (2005): "Unobserved heterogeneity and estimation of average partial effects," in Identification And Inference For Econometric Models: Essays In Honor Of Thomas Rothenberg, ed. by D. W. K. Andrews, and J. H. Stock. Cambridge University Press, New York] to multivariate probit models, proving the following. 1) Average partial effects (APEs) based on joint probabilities are consistently estimated by conventional multivariate probit models under general forms of conditionally independent latent heterogeneity (LH), as long as the only constraints beyond normalization, if any, are within-equation homogenous restrictions. The normalization of choice is not neutral to consistency in models with cross-equation parameter restrictions beyond normalization, such as those implemented by Stata's asmprobit or in the panel probit model: if the normalization is through an error covariance matrix in correlation form, consistency breaks down, unless the LH components are truly homoskedastic. This is substantial since an error covariance matrix in correlation form is the only normalization permitted by Stata's biprobit and mvprobit or Limdep's BIVARIATE PROBIT and MPROBIT. Covariance restrictions beyond normalizations generally conflict with an arbitrary covariance matrix for the LH components. The multinomial probit model with i.i.d. errors, implemented by Stata's mprobit, is a case in point. 2) Conditional independence of the LH components is not generally sufficient for consistent estimation of APEs on conditional probabilities. Consistency is restored by maintaining an additional independence assumption. This holds true whether or not the response variables are used as regressors. 3) The dimensionality benefit observed by [Mullahy, J. (2011): "Marginal effects in multivariate probit and kindred discrete and count outcome models, with applications in health economics," NBER WP SERIES 17588, NBER] in the estimation of partial effects extends to APEs. We exploit this feature in the design of a simple procedure estimating APEs, which is both faster and more accurate than simulation-based codes, such as Stata's mvprobit and cmp. To demonstrate the finite-sample implications of our results, we carry out extensive Monte Carlo experiments with bivariate and trivariate probit models. Finally, we apply our procedure in (3) to Italian survey data of immigrants in order to estimate the APEs of a trivariate probit model of ethnic identity formation and economic performance.

A review of propensity score: principles, methods and application in Stata – Rino Bellocco, Università di Milano Bicocca, Karolinksa Institutet Stoccolma e Alessandra Grotta, Karolinska Institutet Stoccolma.

This talk introduces the principles of propensity score theory and reviews available programs to implement propensity score methods in Stata, with particular focus on **psmatch2** and **teffects psmatch**. An application on real data will be shown.

Social Mobility and Mortality in southern Sweden (1815-1910) – Paolo Emilio Cardone, Università degli Studi di Roma La Sapienza.

Aim of this research project is to seek the influence of how intra social group mobility affected mortality patterns in Sweden, covering the transition from preindustrial to a breakthrough industrial society. According to previous studies (see, e.g., Bengtsson: 2010; Bengtsson and Van Poppel: 2011; Bengtsson and Dribe:2011; Dribe, Helgertz, Van de Putte: 2013) Social Economical Status (SES) does not affect substantially life expectancy of Swedish population in the XIXth century, instead of this, other variables, such as public health measures or education, were key factors. However, a new question emerge for us: Could it be possible that other socio-economic factors, such as the intergenerational mobility, may affect positively life expectancy?
In order to achieve this goal, a dataset between 1815 and 1910 from the Scanian Economic-Demographic Database (SEDD) is going to be used. The database is based on local population registers for five rural Scanian coast parishes (Hög, Kävlinge, Halmstad, Sireköpinge, and Kågeröd). Analysis is based on three periods according to historical criterion (preindustrial period: 1815-1869; early industrial period: 1870-1894 and the first part of the breakthrough of industrialization: 1895-1910).
In our study, intra social mobility is going to be defined as the chances of an individual, between ages 30 and 49, experiences a change of his SES according to SOCPO codification. SOCPO is comprised by 5-category classification scheme. Our main reason for using it is that while it focuses on social power, it is also highly correlated with education and income, as well as is that this classification can be used both for rural and industrial societies. Therefore, a Cox Proportional Hazard model is going to be applied in order to estimate the influence of social mobility, controlling for age and other possible determinant variables. We are going to estimate a model for each SOCPO category. This model includes social mobility status (a categorical variable in which *1* is when the individual experiences the upwards mobility event and *0* otherwise), age, sex, year of birth, parish of residence and position in the household. Thus, after these analyses, we expect to find a significant and positive relationship between social economic mobility and mortality.