Virtual | 30–31 July 2020

Sign up for alerts

The Stata Conference was held on 30–31 July 2020.

Session 1: Methods and implementations

Better predicted probabilities from linear probability models with applications to multiple imputation

Abstract: Although logistic regression is the most popular method for regression analysis of binary outcomes, there are still many attractions to using least-squares regression to estimate a linear probability model. A major downside, however, is that predicted "probabilities" from a linear model are often greater than 1 or less than 0. That can be problematic for many real-world applications. As a solution, we propose to generate predicted probabilities based on a linear discriminant model, which Haggstrom (1983) showed could be obtained by rescaling coefficients from OLS regression. **...(Read more)**

Additional information:
us20_Allison.pdf

Paul Allison
Statistical Horizons LLC

Implementing quantile selection models in Stata

Abstract: This presentation describes **qregsel**, a community-contributed command to implement a copula-based sample-selection correction for quantile regression recently proposed by Arellano and Bonhomme (2017). We illustrate the use of **qregsel** with an empirical example using the data employed in the Stata base reference manual for the **heckman** command.

Additional information:
us20_Siravegna.pdf

Mariel Siravegna
Georgetown University

Expanding Stata's capabilities for sensitivity analysis

Abstract: Nonexperimental approaches to estimating treatment effects often balance observable characteristics to minimize potential for bias. Rosenbaum (2002) recommends a sensitivity analysis to test the assumption that a study is free from hidden bias once such balance is achieved. There are currently two Stata commands that can implement this sensitivity test: **mhbounds** and **rbounds**. **...(Read more)**

Additional information:
us20_Litwok.pptx

Daniel Litwok
Abt Associates

StataCorp presentation: Meta-analysis using Stata

Abstract: Meta-analysis combines results of multiple similar studies to provide an estimate of the overall effect. This overall estimate may not always be representative of a true effect. Often, studies report results that vary in magnitude and even direction of the effect, which leads to between-study heterogeneity. And sometimes the actual studies selected in a meta-analysis are not representative of the population of interest, which happens, for instance, in the presence of publication bias. Meta-analysis provides the tools to investigate and address these complications. Stata has a long history of meta-analysis methods contributed by Stata researchers. In my presentation, I will introduce Stata's new suite of commands, **meta**, and demonstrate it using real-world examples.

Additional information:
us20_Assaad (https:)

Houssein Assaad
StataCorp

Session 2: Financial data

Economic forecasting with multiequation simulation models

Abstract: Capturing interdependencies among many variables is a crucial part of economic forecasting. We show how multiple estimated equations can be solved simultaneously with the Stata **forecast** command and how to simulate the system through time to produce forecasts. This can be combined with user-defined exogenous variables, so that different assumptions can be used to create forecasts under different scenarios. Techniques for assessing the quality of both ex post and ex ante forecasts are shown, along with a simple example model of the U.S. economy.

Additional information:
us20_Price.pptx

Calvin Price
MUFG Bank

Applications of generalized structural equation modeling for enhanced credit risk management

Abstract: The integration of the generalized structural equation modeling (GSEM) framework to widely used statistical packages like Stata offers significant opportunities for credit risk management. GSEM techniques bring to bear a modular and all-inclusive approach to statistical model building. We illustrate the “game changing” potential of the GSEM framework with an application to credit risk stress testing and loss forecasting for a representative portfolio of mortgages originated over the past 20 years. **...(Read more)**

Additional information:
us20_Canals-Cerdá.pdf

José Canals-Cerdá
Federal Reserve Bank of Philadelphia

Event studies with daily stock returns in Stata: Which command to use?

Abstract: This presentation provides an overview on existing user-written commands for executing event studies. By conducting a review of articles that appeared in the past 10 years in 3 leading accounting, finance, and management journals and by assessing which commands could have been used to conduct these studies, I argue that currently only my command eventstudy2 provides sufficient flexibility to conduct a broad range of state-of-the-art event studies. **...(Read more)**

Additional information:
us20_Kaspereit.pdf

Thomas Kaspereit
Universite du Luxembourg

StataCorp presentation: Call Stata from Python

Abstract: Stata 16 introduced tight integration with Python, allowing users to embed and execute Python code from within Stata. In this talk, I will demonstrate new functionality we have been working on—calling Stata from within Python. We are working on providing two ways to let users interact with Stata from within Python: the IPython magic commands and a suite of API functions. With those utilities, you will be able to run Stata conveniently from Python environments, such as Jupyter Notebook/console, Jupyter Lab/console, Spyder IDE, or Python launched from a Windows Command Prompt, Unix terminal, etc.

Additional information:
us20_Zhao_Xu (https:)

Zhao Xu
StataCorp

Special Stata musical interlude by Dorry Segev and Allan Massie

Additional information:
(Teach Me More) Stata Code (https:)

Session 3: Programming

Implementing programming patterns in Mata to optimize your code

Abstract: Have you ever created a program that requires a nontrivial amount of data to be present or available (for example, look-up/value tables, data used for the program interface, etc…)? If you have, you'll likely have experienced the performance penalty that multiple I/O operations can cause. **...(Read more)**

Additional information:
us20_Buchanan1 (https:)

Billy Buchanan
Fayette County Public Schools

Text mining with n-gram variables

Abstract: Text data, such as answers to open-ended questions, are sometimes ignored because they are hard to analyze. Our Stata command **ngram** turns text into hundreds of variables using the "bag of words" approach. Broadly speaking, each variable records how often the corresponding word or word sequence occurs in a given text. This is more useful than it sounds. The program supports text in 12 European languages. (Schonlau, M, Guenther, and N Sucholutsky 2017)

Additional information:
us20_Schonlau.pdf

Matthias Schonlau
University of Waterloo

f_able estimation of marginal with transformed data

Abstract: The command **margins** is a very powerful command that can be used for the estimation of marginal effects for linear and non-linear models (using official or community-contributed commands), as long as the variables of interest are introduced linearly or as polynomials (using factor notation). When other types of transformations are used, Stata is usually unable to estimate marginal effecs correctly because it may not understand that, for example, log_x is actually log(x), considering it as an unrelated independent variable in the model. In this presentation, I provide a simple command, **f_able**, that enables **margins** to correctly estimate marginal effects when transformations other than polynomials are used in the model specification.

Additional information:
us20_Rios-Avila1.pdf

Fernando Rios-Avila
Levy Economics Institute

Two-dimensional Gauss–Legendre quadrature: Seemingly unrelated dispersion-flexible count regressions

Abstract: Many contexts in empirical econometrics require nonclosed form two-dimensional (2D) integration for appropriate modeling and estimation design. Applied researchers often avoid such correct but computationally demanding specifications and opt for simpler biased or less efficient modeling designs. The presentation will detail a new Mata implementation of the 2D version of a relatively simple numerical integration technique—Gauss–Legendre quadrature. **...(Read more)**

Additional information:
us20_Terza.pdf

Joseph Terza
IUPUI

Empirical application

Investigating factors that influence bicyclist injury severity in bicycle-motor vehicle crashes at unsignalized intersections in North Carolina

Abstract: In 2014, North Carolina implemented a strategic highway safety plan to reduce fatalities and serious injuries. The plan defined nine areas of focus to address safety issues; two main areas were investigated, unsignalized intersections and bicyclist safety. The purpose of this study was to evaluate (1) potential factors associated with bicyclist injury severity in bicycle-motor vehicle crashes at unsignalized intersections and (2) the impact of these factors on bicyclist safety. **...(Read more)**

Additional information:
us20_Covert.pdf

Shatoya Covert
Elizabeth City State University

Session 4: Panel data

Generalized method of moments estimation of linear dynamic panel-data models

Abstract: In dynamic models with unobserved group-specific effects, the lagged dependent variable is an endogenous regressor by construction. The conventional fixed-effects estimator is biased and inconsistent under fixed-T asymptotics. To deal with this problem, "difference GMM"

and "system GMM" estimators in the spirit of Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) are predominantly applied in practice. The Stata community widely associates these methods with the **xtabond2** command provided by Roodman (2009). **... (Read more)**

Additional information:
us20_Kripfganz.pdf

Sebastian Kripfganz
University of Exeter Business School

Pretesting for unobserved cluster effects and inference in panel-data sets

Abstract: This presentation addresses the question of how to estimate the standard errors in panel data when there are potentially unobserved cluster effects. We analyze the performance of statistical inference regarding the parameters of a panel-data model when it is first subjected to a pretest for the presence of individual and time unobserved cluster effects. **...(Read more)**

Ercio Munoz
CUNY Graduate Center

XTSEL: Selection of variables and specification in a panel-data framework

Abstract: We have developed two new commands that allow selecting the best predictor between a number of alternative explanatory variables (**xtselvar**) and the best specification between all possible combinations of a defined set of explanatory variables (**xtselmod**) in a panel-data framework. **xtselvar** helps us to select the best predictor between a number of alternative explanatory variables (candidates). **...(Read more)**

Additional information:
us20_Ugarte-Ruiz.pdf

Alfonso Ugarte-Ruiz
BBVA Research

Session 5: Flexible and SEM estimation

Smooth varying coefficient models in Stata

Abstract: Nonparametric regressions are a powerful statistical tool to model relationships between dependent and independent variables with minimal assumptions on the underlying functional forms. Despite its potential benefits, these types of models have two weaknesses: The added flexibility creates a curse of dimensionality, and procedures available for model selection, like cross-validation, have a high computationally cost in samples with even moderate sizes. **... (Read more)**

Additional information:
us20_Rios-Avila2.pdf

Fernando Rios-Avila
Levy Economics Institute

Invited talk: Using Stata to simulate the impact of COVID-19 on organ transplantation

Abstract: We present a case study demonstrating how the Epidemiologic Research Group in Organ Transplantation (ERGOT) at Johns Hopkins uses Stata to further the group's research goals. Recent applications include simulation to estimate the benefit or harm of delaying organ transplantation in the context of the COVID-19 pandemic and modular code design to facilitate rapid analysis of changes in the landscape of organ transplantation under COVID-19 across different organ types. We will discuss techniques for simulation and integration of **putdocx** and

frames to rapidly produce manuscript-ready output. Additionally we will provide an overview of the Stata class we teach at the Johns Hopkins School of Public Health and discuss the songs about Stata we have written to promote the class.

Additional information:
us20_Segev.pptx

Dorry Segev and Allan Massie
Johns Hopkins University

Session 6: Integration with other software

Reading an arbitrary number of files into Stata made easy

Abstract: The Statalist is filled with threads from users who all want to do the same thing. You probably have run into the issue yourself. You have dozens, hundreds, or thousands of files that you need to combine into a single dataset for analysis and want to figure out the most efficient way to do it. In this talk, I'll describe **readit**, a new command that solves this problem and can solve the same problem when used across multiple file types using the Python API introduced in Stata 16. The **readit** command can operate in a few different ways that provide significant flexibility built on the I/O capabilities of the pandas package in Python.

Additional information:
us20_Buchanan2 (https:)

Billy Buchanan
Fayette County Public Schools

Using Microsoft Excel to improve efficiency in working with large datasets in Stata

Abstract: Introduction: There is an ongoing growth in the availability of data and increased number of variables in large datasets such as medical claim files or national surveys. Stata supports various descriptive, exploratory, and analytical approaches to work with these data to identify and study various topics such as public and clinical health outcomes and issues. Given the high volume of various data generated daily, implementing cross-platform approaches to manage and manipulate data can improve efficiency of data-science professionals and academic researchers. **...(Read more)**

Additional information:
us20_Khanijahani.pdf
us20_Khanijahani.xlsx

Ahmad Khanijahani
Duquesne University

Applying symbolic mathematics in Stata using Python

Abstract: I present an applied example of blending theory and data using Stata 16's new Python integration. The SymPy library in Python makes a wide range of symbolic mathematical tools available to Stata programmers. For a recent project, I used theory and SymPy to derive a relationship between two labor supply elasticities in a structural model and separately used Stata to generate reduced-form estimates of these elasticities. I then used the Stata Function Interface to directly plug the empirical Stata estimates into my SymPy model, allowing easy and reproducible estimation of the theoretical relationship of interest. I discuss these methods and provide code for use by other researchers.

Additional information:
us20_Lippold.pdf

Kye Lippold

UC San Diego

Rosetta Stone: Stata To Python Pandas crosswalk

Abstract: Given Stata's recent updates that promote Python integration and the growing popularity of Python and Pandas as a data wrangling and analysis platform, this session will provide a Rosetta Stone-like crosswalk between Stata and Python. The content will demonstrate Python code that replicates common techniques often executed in Stata. This session will be best for Stata users who desire to leverage recently available Python integrations but who have yet to attain beginner-to-intermediate proficiency in Python.

Additional information:
us20_Nelson.pptx
us20_Stata_Pandas_crosswalk.do

Adam Ross Nelson
American University

Downloading and preparing survey data using the Qualtrics API in the Stata ecosystem

Abstract: Downloading and preparing survey data for analysis from online platforms such as Qualtrics is a time-consuming and error-prone task. The qualtrics.ado command interacts with the Qualtrics API to download, and clean, data quickly with less error. The program requires users to enter their Qualtrics credentials. **...(Read more)**

Additional information:
us20_Hoepfner.pdf
us20_qualtrics.ado
us20_qualtrics.sthlp
us20_qualtrics_example.do

Danial Hoepfner
Gibson Consulting Group Inc.

StataCorp presentation: Nonlinear dynamic stochastic general equilibrium models in Stata

Abstract: Dynamic stochastic general equilibrium (DSGE) models are used in macroeconomics for policy analysis and forecasting. A DSGE model consists of a system of equations—usually a nonlinear system of equations—that is derived from economic theory. I will show you how to easily solve, estimate, and analyze nonlinear DSGEs. We will explore how to obtain policy matrices, transition matrices, and impulse–response functions for nonlinear models.

Additional information:
us20_Schenck.pdf

David Schenck
StataCorp

Session 7: Empirical applications

The causal effects of parents' marital status on children's earnings

Abstract: this research, I examine how the marital relationship affects children's future economic status. I introduce the parental marital status hypothesis of children's earnings: **...(Read more)**

Additional information:
us20_Wen.pdf

Bob Wen

Clemson University

The social costs of crime over trust: An approach with machine learning

Abstract: In Peru, 55% of the population considers insecurity as the country's main problem. The present study seeks to contribute to the understanding of the social costs of crime in Peru by measuring the impact of patrimonial crime on trust in public institutions, **...(Read more)**

Additional information:
us20_Cozzubo.pdf

Angelo Cozzubo
University of Chicago

Open panel discussion with Stata developers and closing remarks

Registration

Registration is now closed.

In light of the change to a virtual platform because of COVID-19, we are pleased to announce all proceeds from registrations for the 2020 Stata Conference will be donated to the **CDC Foundation**.

Sign up for alerts

Scientific committee

#Stata2020

Stata

- New in Stata
- Why Stata?
- All features
- Features by disciplines
- Stata/MP
- Which Stata is right for me?
- Order Stata

Shop

- Order Stata
- Bookstore
- Stata Press books
- Stata Journal
- Gift Shop

Support

- Training
- Video tutorials
- FAQs
- Statalist: The Stata Forum
- Resources
- Technical support
- Customer service

Company

- Contact us
- Customer service
- Announcements
- Search