



## 7<sup>th</sup> German Stata Users Group Meeting

June 26, 2009

University Club, Bonn, Germany

**Organizer:** Nikos Askitas (IZA)  
Johannes Giesecke (WZB)  
Ulrich Kohler (WZB)

### Program

#### Friday, June 26:

09:00-09:30 **Registration**

09:30-10:15 **Between Tables and Graphs: A Tutorial**  
Nicholas J. Cox (Durham University)

The display of data or of results often entails the preparation of a variety of table-like graphs showing both text labels and numeric values. I will present basic techniques, tips, and tricks using both official Stata and various user-written commands. The main message is that whenever **graph bar**, **graph dot**, or **graph box** commands fail to give what you want, then you can knit your own customized displays by using **twoway** as a general framework.

10:15-11:00 **Causal Inference with Observational Data: Regression Discontinuity and Other Methods in Stata**  
Austin Nichols (Urban Institute, Washington DC)

This overview of implementing quasi-experimental methods of estimating causal impacts (panel methods, matching estimators, instrumental variables, and regression discontinuity) emphasizes practical considerations and Stata-specific approaches. I pay particular attention to the regression discontinuity method, which seems to be less well-known in the larger community of Stata users, but is the most well regarded of the quasi-experimental methods in those circumstances where it is appropriate. We will also examine some examples of hybrid applications of these estimation techniques.

11:00-11:15 **Coffee Break**

**11:15-11:45**      **Computing Poverty Measures with Survey Data**  
Philippe Van Kerm (CEPS/INSTEAD, Luxembourg)

I present some user-written commands for computing poverty measures from household survey data. Analytic standard errors that take into account survey design features are estimated. Standard errors are adjusted for the estimation of the poverty line as a fraction of the mean or median income. The approach based on influence functions is generally applicable to many estimators. This is illustrated on Laeken indicators with data from the EU SILC (Community Statistics on Income and Living Conditions).

**11:45-12:15**      **Using and Interpreting Restricted Cubic Splines**  
Maarten Buis (University of Tübingen)

Sometimes one wants to model the effect of a variable as a nonlinear smooth curve. A convenient choice for such a curve is a restricted cubic spline. This option has existed in Stata for a while through user-written programs, but as of Stata 10, the **mkspline** command in combination with the **cubic** option has been implemented in official Stata. In this talk, I will briefly introduce splines and restricted cubic splines, but I will mainly focus on what happens after one has estimated a model with a restricted cubic spline, and in particular how the **postrcspline** package can help in the interpretation of the results.

**12:15-12:45**      **Performing Within and Between Analysis (WABA) in Stata**  
Sven-Oliver Spieß (University of Mannheim)

The “fallacies of the wrong level” (e.g., ecological fallacy) are a concern in much of social-science research. When appropriate data are available, researchers frequently use intra-class correlations or ANOVAs to decide whether to use individual-level data or whether to aggregate and analyze data at some higher level, such as workgroups, schools, or locations to examine the associations among theoretical constructs. However, this approach disregards either lower-level information through aggregation or information on group membership by solely using individual-level data. Within and between analysis (WABA), on the other hand, provides a detailed picture of the correlation structure among constructs at different levels by partitioning correlations among constructs into between-groups and within-group components. WABA thus represents a more informative and efficient technique when interested in associations among constructs beyond mere within-group homogeneity. When dealing with more than two hierarchical levels, multiple within and between analysis indicates at which level associations are strongest. Therefore, WABA can also be fruitfully employed to inform model specification in HLM/MLM. In this presentation, I aim to outline the basic ideas and concepts of WABA and to introduce the user-written **wabanova** program to perform basic single and multiple WABA in Stata.



12:45-13:45 **Lunch**

13:45-14:30 **Using Mata to Work More Effectively with Stata: A Tutorial**  
Christopher F. Baum (Boston College)

Stata's matrix language, Mata, highlighted in Bill Gould's Mata Matters columns in the *Stata Journal*, is very useful and powerful in its interactive mode. Stata users who write do-files or ado-files should gain an understanding of the Stata-Mata interface: how Mata may be called upon to do one or more tasks and return its results to Stata. Mata's broad and extensible menu of functions offers assistance with many programming tasks, including many that are not matrix oriented. In this tutorial, I will present examples of how do-file and ado-file writers might effectively use Mata in their work.

14:30-15:00 **The Biplot Command and Software Development at StataCorp**  
Magdalena Luniak (TU Berlin)

In the summer of 2008, I had an opportunity to participate in the internship program at StataCorp. There I found that professional software development is much more than just programming. Programmers can learn a lot from the software development process at StataCorp.

One of the projects I worked on that summer is an enhancement of the **biplot** command. This is the topic of my presentation, and it will serve as a basis for sharing the experience I gained during my work at StataCorp.

15:00-15:15 **Coffee Break**

15:15-16:00 **Reproducible Research: Weaving with Stata**  
Bill Rising (StataCorp)

Reproducible research is one of many names for the same concept: writing a single report document that contains both the report and the commands needed to produce the results and graphics contained in the report. It is called reproducible research because any interested researcher can then reproduce the entire report from the one document. (Programmers call this same concept "iterate programming".) The utility of reproducible research documents extends far beyond research or programming. They allow rapid updates should there be additional data. They can also be used in teaching for generating differing examples or test questions, because different parameters will generate different examples. In this presentation, I will show you how to use a third-party application to embed Stata code, as well as its output, in either LaTeX or OpenOffice documents. I will also use example documents (including the talk itself) to show how you can update a report, its results, and its graphics by using new data or changing parameters.

**16:00-16:30**      **Creating LaTeX Documents from within Stata**  
Ben Jann (ETH Zurich)

In this presentation, I present a new user package called **texdoc**. **texdoc** can be used to create a LaTeX document from within Stata in a weaving fashion. This is especially useful if you want to produce a LaTeX document that contains Stata output, such as a *Stata Journal* article or solutions to statistics homework assignments. I will provide examples illustrating the usage of **texdoc**.

**16:30-17:00**      **Using Remote Access to Big Datasets Efficiently with Stata**  
Volker Lang (University of Tübingen)

In this talk, I discuss problems experienced and solutions developed with Stata, using remote access to a big dataset (around 10GB) of the Institute for Employment Research (IAB). I focus on two topics. The first problem is that of not directly controlling the data. The solution here is to implement good pre-documentation into the do-files to structure and improve the communication with the people hosting the remote access. Second, there are memory and running-time problems with using such a large dataset; I discuss this problem in relation to the first one. The solution here is the extensive use of sampling techniques. I present routines for entering such sampling procedures into remote-access do-files.

**17:00-17:15**      **Break**

**17:15-18:00**      **Report to Users**  
Bill Rising (StataCorp)

**18:00-18:30**      Wishes and Grumbles