2022 German Stata Users Group Meeting

Announcement and Program

Conference Date	Friday, June 10, 2022 10:00 am – 4:30 pm
Workshop Date	Thursday, June 9, 2022 12:00 pm – 7:00 pm
Venue	Goethe-University Frankfurt am Main Campus Westend - Casino Norbert-Wollheim-Platz 60323 Frankfurt am Main
Cost	Meeting only: 45 EUR (students 35 EUR) Workshop only: 65 EUR (students 50 EUR) Workshop and Meeting: 85 EUR (students 70 EUR)

DEADLINE Workshop and User Meeting registration: June 1, 2022

Overview

Meeting The German Stata Users Group Meeting 2022 will be held on Friday, 10th June 2022 in Frankfurt at the Goethe-University Frankfurt am Main. We would like to invite everybody from everywhere who is interested in using Stata to attend this meeting. The academic program of the meeting is being organized by Alexander Schmidt-Catran, Christian Czymara (both Goethe-University Frankfurt), Johannes Giesecke (Humboldt University Berlin), and Ulrich Kohler (University of Potsdam). The conference language will be English due to the international nature of the meeting and the participation of non-German guest speakers. The logistics of the conference are being organized by DPC Software GmbH, distributor of Stata in several countries including Germany, The Netherlands, Austria, Czech Republic and Hungary (http://www.dpc-software.de).

Workshop On the day before the conference, there will be a workshop titled "Taking a page from Git: Reproducible research & dynamic documents with Stata" given by Sven Spieß. The workshop will be held at Campus Westend, PEG-Building, Room PEG 2G111. Details about the workshop are given below and at http://www.stata.com/meeting/germany22 or www.dpc-software.de

Conference Dinner There is (at additional cost) the option of an informal meal on Friday evening. Details about this event will be provided soon.

Registration and accommodations
Participants are asked to travel at their own expense. The conference fee covers costs for coffee, tea, and lunch. There will also be an optional informal meal at additional cost on Friday evening.

You can enroll by emailing Natascha Hütter (natascha.huetter@dpc-software.de) by writing, phoning, or faxing to

       Natascha Hütter

       DPC Software GmbH
       Prinzenstraße 2
       42697 Solingen
       Germany
       Tel: +49 (0)212 / 22 47 16 -21

       www.dpc-software.de

Program Schedule

09:45 - 10:00 **Registration**

10:00 - 10:15 **Welcome**

10:15 - 11:15 **Finite Mixture Models for Linked Survey and Administrative Data**
Stephen P. Jenkins, Fernando Rios-Avila (The London School of Economics and Political Science; Bard College)

11:15 - 11:45 **Coffee**

11:45 - 12:15 **A mixture of ordered probit models with endogenous switching between two latent classes**
Jochem Huismans, Jan Willem Nijenhuis, Andrei Sirchenko (Maastricht University)

12:15 - 12:45 **Network analysis using `nwxtregress`**
Jan Ditzen, William Grieser, Morad Zekhnini (Free University of Bozen-Bolzano; Texas Christian University; Michigan State University)

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

13:45 - 14:15 **Visualizing categorical data with hammock plots**
Matthias Schonlau (University of Waterloo)

14:15 - 14:45 **Measuring the accuracy of the probabilistic predictions of discrete-choice models: The `classify` command**
Jochem Huismans, Jan Willem Nijenhuis, Andrei Sirchenko (Maastricht University)

14:45 - 15:15 **Coffee**

15:15 - 15:45 **Difference-in-differences estimation using Stata**
Joerg Luedicke (StataCorp)

15:45 - 16:30 **Wishes and grumbles**

16:30 **End of the meeting**

Conference venue
Goethe-University Frankfurt am Main
Campus Westend – Casino Building
Room: Renate von Metzler-Saal (Cas 1.801)

Norbert-Wollheim-Platz
60323 Frankfurt am Main

How to get to the venue
From the main station, take trains S1-S9 via Hauptwache, switch to U1, U2, U3 or U8
and go to Holzhausenstraße. The campus is about 300 meters from there.

For more information on how to get to the campus, see https://www.goethe-university-frankfurt.de/73011530/Goethe_University_Locations?locale=en

For more information on how to get to the buildings, see https://www.uni-frankfurt.de/73177180/GU_Lageplan_Campus_Westend_0318_Mensa_Bib_ENGL.pdf
(the meeting will take place in Building 7 (Casino) and the workshop in Building 12
(PEG) of this plan)

Abstracts

09:45 - 10:00 **Registration**

10:00 - 10:15 **Welcome**

10:15 - 11:15 **Finite Mixture Models for Linked Survey and Administrative Data**
Stephen P. Jenkins, Fernando Rios-Avila (The London School of Economics and Political Science; Bard College)

Email: s.jenkins@lse.ac.uk

Abstract: Researchers use finite mixture models to analyze linked survey and administrative data on labour earnings (or similar variables), taking account of various types of measurement error in each data source. Different combinations of error-ridden and/or error-free observations characterize latent classes. Latent class probabilities depend on the probabilities of the different types of error. We introduce a set of Stata commands to fit a general class of finite mixture models to fit to linked survey-administrative data. We also provide post-estimation commands for assessment of reliability, marginal effects, data simulation, and prediction of hybrid earnings variables that combine information from both data sources.

11:15 - 11:45 **Coffee**

11:45 - 12:15 **A mixture of ordered probit models with endogenous switching between two latent classes**
Jochem Huismans, Jan Willem Nijenhuis, Andrei Sirchenko (Maastricht University)

Email: jochemhuismans@gmail.com, janwillemnijenhuis@gmail.com, andrei.sirchenko@gmail.com

Abstract: Ordinal responses can be generated, in a time-series context, by different latent regimes or, in a cross-sectional context, by different unobserved classes of population. We introduce a new command swopit that fits a mixture of ordered probit models with either exogenous or endogenous switching between two latent classes (or regimes). Switching is endogenous if the unobservables in the class-assignment model are correlated with the unobservables in the outcome models. We provide a battery of postestimation commands, assess by Monte Carlo experiments the finite-sample performance of the maximum likelihood estimator of the parameters, probabilities and their standard errors (both the asymptotic and bootstrap ones), and apply the new command to model the policy interest rates.

12:15 - 12:45 **Network analysis using `nwxtregress`**
Jan Ditzen, William Grieser, Morad Zekhnini (Free University of Bozen-Bolzano; Texas Christian University; Michigan State University)

Email: Jan.Ditzen@unibz.it

Abstract: Network analysis has become critical to the study of social sciences. While several Stata programs are available for analysing network structures, programs that execute regression analysis with a network structure are currently lacking. We fill this gap by introducing the nwxtregress command. Building on spatial econometric methods (LeSage and Pace 2009), nwxtregress uses MCMC estimation to produce estimates of endogenous peer effects, as well as own-node (direct) and cross-node (indirect) partial effects, where nodes correspond to cross-sectional units of observation, such as firms, and edges correspond to the relations between nodes. Unlike existing spatial regression commands (for example, spxtregress), nwxtregress is designed to handle unbalanced panels of economic and social networks as in Grieser et al. (2021). Networks can be directed or undirected with weighted or unweighted edges, and they can be imported in a list format that does not require a shapefile or a Stata spatial weight matrix set by spmatrix. Finally, the command allows for the inclusion or exclusion of contextual effects. To improve speed, the command transforms the spatial weighting matrix into a sparse matrix. Future work will be targeted toward improving sparse matrix routines, as well as introducing a framework that allows for multiple networks.

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

13:45 - 14:15 **Visualizing categorical data with hammock plots**
Matthias Schonlau (University of Waterloo)

Email: schonlau@uwaterloo.ca

Abstract: Visualizing data with more than two variables is not straight forward, especially when some variables are categorical rather than continuous. My hammock plots are one option to visualize categorical data and mixed categorical / continuous data. Hammock plots can be viewed as a generalization of parallel coordinate plots where the lines are replaced by rectangles that are proportional to the number of observations they represent. Hammock plots also incorporate optional univariate descriptors such as category labels into the graph. I will introduce my Stata program for hammock plots and give examples.

14:15 - 14:45 **Measuring the accuracy of the probabilistic predictions of discrete-choice models: The `classify` command**
Jochem Huismans, Jan Willem Nijenhuis, Andrei Sirchenko (Maastricht University)

Email: jochemhuismans@gmail.com, janwillemnijenhuis@gmail.com, andrei.sirchenko@gmail.com

Abstract: We introduce a new Stata command classify that produces a classification

table (confusion matrix) and computes various diagnostic metrics of forecasting performance and measures of the accuracy of probabilistic predictions, using the predicted probabilities of discrete outcomes or classes (binary, ordinal or categorical) estimated by any discrete-choice model. In the case of more than two outcomes the command computes the class-specific performance metrics for each class as well as their micro, macro and weighted averages.

14:45 - 15:15 **Coffee**

15:15 - 15:45 **Difference-in-differences estimation using Stata**
Joerg Luedicke (StataCorp)

Email: jluedicke@stata.com

Abstract: Difference-in-differences (DID) estimation has become a popular tool in the context of treatment-effects estimation and program evaluation. In this presentation, I will show how to use Stata's didregress and xtdidregress commands to estimate treatment effects with repeated cross-sectional as well as panel data. I will also discuss a variety of methods for calculating cluster-robust standard errors when the number of clusters is small. Finally, I will show how to use diagnostic tools for checking the parallel-trends assumption, which is an identifying assumption of DID.

15:45 - 16:30 **Wishes and grumbles**

16:30 **End of the meeting**

Workshop "Taking a page from Git: Reproducible research & dynamic documents with Stata"

Date and Place	Thursday, June 9, 2022 12:00 pm – 7:00 pm Campus Westend, PEG-Building, Room PEG 2G111
Presenter	Sven Spieß
Fees	Workshop only: 65 EUR (students 50 EUR) Workshop and Meeting: 85 EUR (students 70 EUR)
Register	natascha.huetter@dpc-software.de
Information	http://www.stata.com/meeting/germany22 www.dpc-software.de

Reproducibility has always been a hallmark of Stata. The popular version control system Git offers useful additions to the versioning features implemented in Stata with regards to keeping track of revisions of individual (do-)files over the course of evolving research projects. The advantages are even more substantive in "distributed" projects where collaborators don't necessarily work on a common infrastructure. Leveraging Git in combination with the power of dynamic documents furthers your ability to easily present and disseminate your most recent findings.

In this workshop we will first learn the basics of working with the free and open source version control system Git in conjunction with Stata. After having Git up and running, we will dive into Stata's facilities for creating dynamic documents to automatically reflect changes in our analyses and/or data.

Prerequisites
- Working knowledge of Stata
- Git installed on your system (https://git-scm.com/downloads); optionally: text editor with support for version control (e.g., VS Code, https://code.visualstudio.com)
- Free GitHub account (https://github.com/signup)
- Limited prior exposure to Markdown, HTML, & CSS is beneficial but not required

Lecturer
Sven Spieß is a Stata consultant with DPC Software.