

German Stata Conference 2024

Announcement and Program

June 7, 2024

Overview

Date/Venue/Cost

Date:	Conference: June 7, 2024, 9am-6pm Workshop: June 6, 2024, 10am-5pm
Venue:	GESIS—Leibniz-Institut für Sozialwissenschaften B6, 4-5 68159 Mannheim
Costs:	Conference only: 49.99 EUR (students: 35 EUR) Workshop only: 65 EUR (students: 50 EUR) Conference and workshop: 85 EUR (students: 70 EUR)
Web:	https://www.stata.com/meeting/germany24/ https://dpc-software.de/2024-german-stata-conference/

Deadline for Registration: June 4th 2024

Meeting

The 21st German Stata Conference will be held on Friday, June 7th 2024 in Mannheim at GESIS—Leibniz Institute for the Social Sciences. 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 Johannes Giesecke (Humboldt University Berlin), Ulrich Kohler (University of Potsdam), and Reinhard Pollak (GESIS). 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 one-day workshop on "DID estimation using Stata" by Felix Knau; see the detailed description below the program.

Time table

8:15–9:00	Registration
9:00–9:15	Welcome Reinhard Pollak
9:15–10:15	Recent developments in the fitting and assessment of of flexible parametric survival models Paul Lambert
10:15–10:45	Coffee
10:45–11:15	<code>cfbinout</code> and <code>xtdhazard</code> : Control-Function Estimation of Binary-Outcome Models and the Discrete-Time Hazard Model Harald Tauchmann and Elena Yurkevich
11:15–11:45	Multi-dimensional well-being, deprivation, and inequality Peter Krause
11:45–12:00	How to assess the fit of choice models with Stata? Wolfgang Langer
12:00–13:00	Lunch Break
13:00–14:15	Customizable tables Kristin MacDonald
14:15–14:45	Coffee
14:45–15:15	<code>geoplot</code> : A new command to draw maps Ben Jann
15:15–15:45	<code>repreport</code> : Facilitating reproducible research in Stata Daniel Krähmer
15:45–16:15	<code>mkproject</code> and boilerplate: automate the beginning Maarten Buis
16:15–16:45	Coffee
16:45–17:15	Data structures in Stata Daniel Schneider
17:15–18:00	Open panel discussion with Stata developers
18:00	End of meeting

Conference venue

GESIS—Leibniz-Institute for the Social Sciences
B6, 4–5
68159 Mannheim

How to get to the venue

The venue is within walking distance from Mannheim’s central station (1.3 km). Alternatively you can take any of the Trams **1** (Schönau), **4** (Bad Dürkheim), **5** (Weinheim via Heidelberg), and **6** (Rheingönheim). In all cases, exit at *Schloss* (2 stations) and walk the remaining 450m.

Note very carefully that Mannheim's city center does not have street names. Instead, city-blocks are named with an index similar to the fields on a chess board. The venue is in field B5. The houses in all the Blocks A to K are numbered in ascending order when you circumnavigate the blocks counter-clockwise, starting at the corner that is closest to the castle. This all have its logic, but you can of course also just ask Google to bring you to the entrance door.

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

Workshop: DID Estimation Using Stata

Date and Place

Thursday, June 6, 2024
10am–5pm

GESIS—Leibniz-Institute for the Social Sciences
B6, 4–5
68159 Mannheim

Topics

- Short introduction DID und TWFE: How to apply those in Stata (simple examples, `didregress`, `xtreg`, `reghdfe`)
- What does TWFE identify (and when this may be problematic): brief introduction to `twowayfweights`
- New DID methods robust to heterogeneous treatment effects: Static case, dynamic case (and maybe continuous treatment)

- More extensive (interactive) session on implementing corresponding commands: `did_multiplegt_dyn`, `csdid`, `eventstudyinteract`, `did_imputation` (with focus on `did_multiplegt_dyn`)

Presenter

Felix Knau
SciencesPo, Paris
Department of Economics

Felix Knau is a Research Assistant working on Clément de Chaisemartin's ERC CoG project Completing the revolution: Enhancing the reality, the principles, and the impact of economics' credibility revolution (REALLYCREDIBLE).

Fees

Workshop only: 65 EUR (students 50 EUR) Workshop and Meeting: 85 EUR (students 70 EUR)

Registration

Please register for the conference and/or the workshop by writing an E-mail to Natascha Hütter (natascha.huetter@dpc-software.de)

Abstracts

9:15–10:15 Recent developments in the fitting and assessment of flexible parametric survival models

Paul Lambert (University of Leicester, UK and Karolinska Institutet, Sweden)

Abstract: Flexible parametric survival models are an alternative to the Cox proportional hazards model and more standard parametric models for the modelling of survival (time-to-event) data. They are flexible in that spline functions are used to model the baseline and potentially complex time-dependent effects. I will give a brief overview of the models and the advantages over the Cox model. However, I will concentrate on some recent developments. This will include the motivation for developing a new command to fit the models (`stpm3`), which makes it much simpler to fit more complex models with non-linear functions, non-proportional hazards and interactions and simplifies and extends postestimation predictions, particularly marginal (standardized) predictions. I will also describe some new postestimation tools that help in the evaluation of model fit and validation in prognostic models.

10:45–11:15 `cfbinout` and `xtdhazard`: Control-Function Estimation of Binary-Outcome Models and the Discrete-Time Hazard Model

Harald Tauchmann (FAU Erlangen-Nuremberg) and Elena Yurkevich (FAU Erlangen-Nuremberg)

Abstract: We introduce the new community-contributed Stata commands `cfbinout` and `xtdhazard`. The former generalizes `ivprobit`, `twostep` by allowing discrete endogenous regressors and different link functions than the normal link, specifically logit and cloglog. In terms of the underlying econometric theory, `cfbinout` is guided by Wooldridge (2015). In terms of the implementation in Stata and Mata, respectively, `cfbinout` follows Terza (2017). `xtdhazard` is essentially a wrapper for either `cfbinout` or alternatively `ivregress 2sls`. When calling `ivregress 2sls`, `xtdhazard` implements the linear first-differences (or higher-order differences) instrumental variables estimator suggested by Farbmacher & Tauchmann (2023) for dealing with time-invariant unobserved heterogeneity in the discrete-time hazard model. When calling `cfbinout`, `xtdhazard` implements—depending on the specified link function—several nonlinear counterparts of this estimator that are briefly discussed in the online supplement to Farbmacher & Tauchmann (2023). Using `xtdhazard`—rather than directly using `ivregress 2sls`, `ivprobit`, `twostep`, or `cfbinout`—simplifies the implementation of these estimators, as generating the numerous instruments required can be cumbersome, especially when using factor-variables syntax. In addition, `xtdhazard` performs several checks that may prevent `ivregress 2sls` and `ivprobit`, `twostep`, respectively, from failing and reports issues like perfect first-stage predictions. An (extended) replication of Cantoni (2012) illustrates the use of `cfbinout` and `xtdhazard` in applied empirical work.

- Cantoni, D. (2012). Adopting a new religion: The case of Protestantism in 16th century Germany, *The Economic Journal* 122, 502–531.
- Farbmacher, H. and Tauchmann, H. (2023). Linear fixed-effects estimation with nonrepeated outcomes, *Econometric Reviews* 42(8): 635–654.
- Terza, J. (2017). Two-stage residual inclusion estimation: A practitioners guide to Stata implementation, *Stata Journal* 17(4): 916–938.
- Wooldridge, J. M. (2015). Control function methods in applied econometrics, *The Journal of Human Resources* 50(2): 420–445.

11:15–11:45 Multi-dimensional well-being, deprivation, and inequality

Peter Krause (DIW Berlin, SOEP)

Abstract: The presentation offers a brief summary for a set of Stata programs for extended multidimensional applications on well-being, deprivation, and inequality. The first section illustrates the underlaying motivation by some empirical examples on decomposed multi-dimensional results. The second section on multi-dimensional well-being and deprivation measurement illustrates the conceptual background—based on the Alkire/Foster MPI framework (and CPI, N. Rippin)—which is also applied to well-being measurement,

and extended by a parameter driven fixed-fuzzy approach—with several illustrations and further details on the options offered in the Stata deprivation and well-being programs. The third section on multi-dimensional inequalities refers to a multidimensional Gini-based row-first measurement framework with a special emphasize on multiple within- and between-group-inequalities—including conceptual extensions on horizontal between-group applications and further details on the options offered in the Stata inequality program. Section four summarizes and opens up for advice and discussion.

11:45–12:00 How to assess the fit of choice models with Stata?

Wolfgang Langer (Martin-Luther-University Halle-Wittenberg)

Abstract: McFadden developed the conditional multinomial logit model in 1974 using it for rational choice modeling. In 1993 Stata introduced it in version 3. In 2007 Stata extended this model to the `asclogit` or `ascprobit` being able to estimate the effects of alternative-specific and case-specific exogenous variables on the choice probability of the discrete alternatives. In 2021, Stata added the class of choice models extending it to random-effect (mixed) and panel models. As it stands, Stata only provides an post-estimation Wald chi-square test to assess the overall model. However, although McFadden developed an pseudo r-square to assess the fit of the conditional logit model already in 1974, Stata still does not provide it even in version 18. Thus, I developed `fit_cmlogit` to calculate the McFadden pseudo r-square using a zero model with alternative-specific constants to correct the uneven distribution of alternatives. Furthermore, it calculates the corresponding Likelihood-Ratio-chi-square test which is more reliable / conservative as the Wald test. The program uses the formulas of Hensher & Johnson (1981) and Ben-Akiva & Lerman (1985) for the McFadden pseudo-r square to correct the number of exogenous variables and faced alternatives. Train (2003) discussed these characteristics of the McFadden pseudo r-square in detail. Additionally it calculates the log-likelihood-based pseudo r-squares developed by Maddala (1983, 1988), Cragg & Uhler (1970) and Aldrich & Nelson (1984). The latter uses the correction formula proposed by Veall & Zimmermann (1994). An empirical example of predicting voting behavior in the German federal election study of 1990 demonstrates the usefulness of the program to assess the fit of logit choice models with alternative-specific and case-specific exogenous variables.

- Aldrich, J.H. & Nelson, F.D. (1984): Linear probability, logit and probit models. Beverly Hills, CA: Sage
- Ben-Akiva, M. & Lerman, S.R. (1985): Discrete choice analysis: Theory and application to travel demand. Cambridge, MA: MIT Press
- Cragg, G. & Uhler, R. (1970): The demand of automobiles. Canadian Journal of Economics, 3, pp.386-406
- Hensher, D.A. & Johnson, L.W. (1981): Applied discrete choice modelling. London: Croom Helm/Wiley

- Domencich, T.A. & McFadden, D. (1975): Urban travel demand. A behavioral analysis. Amsterdam u. Oxford: North Holland Publishing Company
- Maddala, G.S. (1983): Limited-dependent and qualitative variables in econometrics. Cambridge, U.K.: Cambridge University Press
- Maddala, G.S. (1992² (1988)): Introduction to Econometrics. New York, N.Y.: Maxwell
- MacmillanMcFadden, D. (1974): Conditional logit analysis of qualitative choice behavior. In: Frontiers of econometrics. Ed. P. Zarembka,, pp. 105-142. New York: Academic Press
- McFadden, D. (1979): Quantitative methods for analysing travel behaviour of individuals: some recent developments. In: Hensher, D.A.& Stopher, P.R.: (eds): Behavioural travel modelling. London: Croom Helm, pp. 279-318
- Train, K.E. (2003): Discrete choice methods with Simulations. Cambridge, U.K.: Cambridge University Press
- Veall, M.R. & Zimmermann, K.F. (1994): Evaluating Pseudo-R²'s for binary probit models. Quality&Quantity, 28, pp. 151- 164

13:00–14:15 Customizable tables

Kristin MacDonald (StataCorp)

Abstract: Presenting results effectively is a crucial step in statistical analyses, and creating tables is an important part of this step. Whether you need to create a cross-tabulation, a Table 1 reporting summary statistics, a table of regression results, or a highly customized table of results returned by multiple Stata commands, the tables features introduced in Stata 17 and Stata 18 provide ease and flexibility for you to create, customize, and export your tables. In this presentation, I will demonstrate how to use the `table`, `dtable`, and `etable` commands to easily create a variety of tables. I will also show how to use the `collect` suite to build and customize tables and to create table styles with your favorite customizations that you can apply to any tables you create in the future. Finally, I will demonstrate how to export individual tables to Word, Excel, L^AT_EX, PDF, Markdown, and HTML and how to incorporate your tables into complete reports containing formatted text, graphs, and other Stata results.

14:45–15:15 geoplot: A new command to draw maps

Ben Jann (University of Bern)

Abstract: `geoplot` is a new command for drawing maps from shape files and other datasets. Multiple layers of elements such as regions, borders, lakes, roads, labels, and symbols can be freely combined and the look of elements (e.g. color) can be varied depending on the values of variables. Compared to previous solutions in Stata, `geoplot`

provides more user convenience, more functionality, and more flexibility. In this talk I will introduce the basic components of the command and illustrate its use with examples.

15:15–15:45 `repreport`: Facilitating reproducible research in Stata

Daniel Krähmer (Ludwig-Maximilians-Universität München)

In theory, Stata provides a stable computational environment and includes commands (i.e., `version`) that are specifically designed to ensure reproducibility. In practice, however, users often lack the time or the knowledge to exploit this potential. Insights from an ongoing research project on reproducibility in the social sciences show that computational reproducibility is regularly impeded by researchers being unaware what files (i.e., datasets, do-files), software components (i.e., ados), infrastructure (i.e., directories), and information (i.e., ReadMe files) is needed to enable reproduction. This presentation introduces the new Stata command `repreport` as a potential remedy. The command works like a log, with one key difference: Instead of logging the entire analysis, `repreport` extracts specific pieces of information pertinent to reproduction (e.g., names and locations of datasets, ados, etc.) and compiles them into a concise reproduction report. Furthermore, the command includes an option for generating a full-fledged reproduction package containing all components needed for push-button reproducibility. While `repreport` adds little value for researchers whose workflow is already perfectly reproducible, it constitutes a powerful tool for those who strive to make their research in Stata more reproducible at (almost) no additional cost.

15:45–16:15 `mkproject` and `boilerplate`: automate the beginning

Maarten L. Buis (University of Konstanz)

There is usually a set of commands that are included in every `.do` file a person makes, like `clear all` or `log using`. What those commands are can differ from person to person, but most persons have such a standard set. Similarly, a project usually has a standard set of directories and files. Starting a new `.do` file or a new project thus involves a number of steps that could easily be automated. Automating has the advantage of reducing the amount of work you need to do. However, the more important advantage of automating the start of a `.do` file or project is that it makes it easier to maintain your own workflow: it is so easy to start "quick and dirty" and promise to yourself that you will fix that "later". If the start is automated, then you don't need to fix it.

The `mkproject` command automates the beginning of a project. It comes with a set of templates I find useful. A template contains all the actions (like create sub-directories, create files, run other Stata commands) that `mkproject` will take when it creates a new project. Since everybody's workflow is different, `mkproject` allows users to create their own template. Similarly, the `boilerplate` command creates a new `.do` file with boilerplate code in it. It comes with a set of templates, but the user can create their own.

This talk will illustrate the use of both `mkproject` and `boilerplate` and in particular how to create your own templates.

16:45–17:15 Data structures in Stata

Daniel C. Schneider (Max Planck Institute for Demographic Research, Rostock)

This presentation starts out by enumerating and describing the main data structures in Stata (e.g., data sets / frames, matrices) and Mata (e.g., string and numeric matrices, objects like associative arrays). It analyzes ways in which data can be represented and coerced from one data container into another. After assessing the strengths and limitations of existing data containers, it muses on potential additions of new data structures and on enriching the functionality of existing data structures and their interplay. Moreover, data structures from other languages, such as Python lists, are described and examined for their potential introduction into Stata / Mata. The goal of the presentation is to stimulate a discussion among Stata users and developers about ways in which the capabilities of Stata's data structures could be enhanced in order to ease and open up new possibilities for data management and analysis.

17:15–18:00 Open panel discussion with Stata developers

Abstract Contribute to the Stata community by sharing your feedback with StataCorp's developers. From feature improvements to bug fixes and new ways to analyze data, we want to hear how Stata can be made better for you.

Scientific Organizers

The academic program of the conference is being organized by Johannes Giesecke (HU Berlin), Ulrich Kohler (University of Potsdam), and Reinhard Pollak (GESIS—Leibniz-Institute for the Social Sciences)

Logistics organizers

The logistics are being organized by DPC Software GmbH, the distributor of Stata in several countries including Germany, The Netherlands, Austria, Czech Republic and Hungary (<http://www.dpc.de>).