

German Stata Conference 2021

Announcement and Program

June 25, 2021

Overview

Date/Venue/Cost

Date:	June 25, 2021
Venue:	Online
	Join Zoom Meeting
	Meeting ID: 610 2767 1921
	Passcode: 32692163
Costs:	Free

Registration is not necessary. Just log in. Note however that our licence allows hosting meetings with up to 200 participants. If the number of participants increases above 200, participants will be allowed to enter "first come, first serve".

Meeting

This year's German Stata Conference will be held Online on Friday, June 25 2021. Everybody from anywhere who is interested in using Stata is invited to join Online. The meeting will include presentations about causal models, general statistics, and data management, both by researchers and by StataCorp staff. The meeting will also include a "wishes and grumbles" session, during which you may air your thoughts to Stata developers.

In the attempt to emulate the atmosphere of a real Stata meeting, we offer topical informal break sessions. The *Speaker's Corner* is a break session to join the speaker(s) presented in the preceding slot. The *Coffee Bar* invites all participants to chat in English, and the *Smoker's Lounge* is for those who prefer small-talk in German. An advantage of the the virtual format is that smokers are allowed at the coffee bar, too.

Language

The conference language will be English because of the international nature of the meeting and the participation of non-German guest speakers. The Smoker's Lounge is in German, however.

Time table

9:00–9:15	Welcome Johannes Giesecke and Ulrich Kohler
9:15–10:00	Double-debiased machine learning in Stata Achim Ahrens, Christian B. Hansen and Mark Schaffer
10:00–10:30	kinkyreg: Instrument-free inference for linear regression models with endogenous regressors Jan F. Kiviet and Sebastian Kripfganz
10:30–11:00	Break-Session (with Speakers Corner, Coffee Bar, Smoker’s Lounge)
11:00–11:30	Twostep Multilevel Analysis Using Stata Johannes Giesecke und Ulrich Kohler
11:30–12:00	xtbreak: Estimating and testing break points in time series and panel data Jan Ditzen, Yiannis Karavias, and Joakim Westerlund
12:00–12:30	Break-Session (with Speakers Corner, Coffee Bar, Smoker’s Lounge)
12:30–13:00	Lunch Break
13:00–13:30	Playing nice with others: Initializing your work with external configurations Sven Oliver Spieß
13:30–14:15	Bayesian vector autoregressive models in Stata Nikolay Balov, StataCorp
14:15–14:45	Break-Session (with Speakers Corner, Coffee Bar, Smoker’s Lounge)
14:45–15:15	dstat: A unified framework for estimation of summary statistics and distribution functions Ben Jann
15:15–15:45	wikiviews—A Stata interface for the Wikipedia API Ulrich Kohler
15:45–16:15	Break-Session (with Speakers Corner, Coffee Bar, Smoker’s Lounge)
16:15–17:00	Treatment-effects estimation with lasso Di Liu, StataCorp
17:00–17:30	Open panel discussion with Stata developers
17:30	End of meeting

Abstracts

9:15–10:00 Double-debiased machine learning in Stata

[Achim Ahrens](#) (ETH Zürich), [Christian B. Hansen](#) (University of Chicago) and [Mark E. Schaffer](#) (Heriot-Watt University, Edinburgh)

Abstract: We introduce ddml, a package for double-debiased machine learning in Stata. ddml implements algorithms for causal inference aided by supervised machine learning. Five different models are supported; allowing for binary or continuous treatment variables as well as instrumental variables. ddml uses stacking regression as the default machine learner, but may be used in combination with other methods implemented in Stata.

10:00–10:30 `kinkyreg`: Instrument-free inference for linear regression models with endogenous regressors

Jan F. Kiviet (University of Amsterdam) and [Sebastian Kripfganz](#) (University of Exeter Business School)

Abstract: In models with endogenous regressors, a standard regression approach is to exploit just- or over-identifying orthogonality conditions by using instrumental variables. In just-identified models, the identifying orthogonality assumptions cannot be tested without the imposition of other non-testable assumptions. While formal testing of over-identifying restrictions is possible, its interpretation still hinges on the validity of an initial set of untestable just-identifying orthogonality conditions. We present the `kinkyreg` Stata program for kinky least squares (KLS) inference that adopts an alternative approach to identification. By exploiting non-orthogonality conditions in the form of bounds on the admissible degree of endogeneity, feasible test procedures can be constructed that do not require instrumental variables. The KLS confidence bands can be more informative than confidence intervals obtain from instrumental variable estimation, in particular when the instruments are weak. Moreover, the approach facilitates a sensitivity analysis for the standard instrumental variable inference. In particular, it allows to assess the validity of previously untestable just-identification exclusion restrictions. Further KLS-based tests include heteroskedasticity, function form, and serial correlation tests.

11:00–11:30 Twostep mulilevel analysis using Stata

[Johannes Giesecke](#) (Humboldt University Berlin) and Ulrich Kohler (University of Potsdam)

Abstract: This presentation describes `-twostep-`, a bundle of programs to perform multilevel analyses with the twostep approach in one step. The twostep approach to multilevel analysis means to estimate a parameter of interest in a unit level dataset (e.g. individuals within countries) which are feed as dependend variable into an analysis on the cluster level (e.g. countries). The twostep approach is sometimes seen as superior to the more standard one-step approach if the numbers of observation on the cluster level becomes small. Additionally, two-step multilevel analysis may be used as a companion of the one-step approach, for instance to check model or linearity assumptions. `-twostep-` is created specifically with this second use in mind.

11:30–12:00 `xtbreak`: Estimating and testing break points in time series and panel data.

[Jan Ditzen](#) (Free University Bozen-Bolzano), Yiannis Karavias (University of Birmingham) and Joakim Westerlund (Lund University)

Abstract: The recent events that have plagued the global economy, such as the 2008 financial crisis or the 2020 COVID-19 outbreak, hint to multiple structural breaks in economic relationships. I present `xtbreak` that implements the estimation of single and multiple break points and testing for structural breaks in time series and panel data. The

estimation and the tests follow the methodologies developed in Andrews (1993, *Econometrica*), Bai and Perron (1998 *Econometrica*) and Ditzen, Karavias and Westerlund (2021). For both time series and panel data regressions, five tools are provided: (i) a test of no structural change against the alternative of a specific number of changes, (ii) a test the null hypothesis of no structural change against the alternative of an unknown number of structural changes, (iii) a test of the null of s changes against the alternative of $s + 1$ changes, (iv) consistent break date estimators and (v) asymptotically valid confidence intervals for the break dates.

- Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61(4), 821–856.
- Bai, B. Y. J., & Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66(1), 47–78.
- Ditzen, J., Karavias, Y. & Westerlund J. (2021). Testing for Multiple Structural Breaks in Panel Data

13:00–13:30 Playing nice with others: Initializing your work with external configurations

[Sven Oliver Spieß](#) (DPC)

Abstract: Stata comes with ample internal features to set up and automate your workflows and analysis routines. However, interdisciplinary teams or interconnected workflows may give rise to the wish to separate easily adjustable settings from core procedures in a way that is accessible to those not fluent in Stata for configuration or review. This talk will consider three specific variants, namely external Stata macros, INI, and MS Excel and outline some general principles to facilitate discussion on good practices within the Stata community.

13:30–14:15 Bayesian vector autoregressive models in Stata

[Nikolay Balov](#) (StataCorp)

Abstract: Vector autoregressive (VAR) models are popular choices for studying the joint dynamics of multiple time series. They require no special structure because the outcome variables are regressed on their own lagged variables. One of the main problems with VAR models is the significant number of regression parameters, which is proportional to the number of lags. As a result, when fitted to small data, complex VAR models tend to show poor forecasting performance.

In Stata 17, we introduce a new command, `bayes:var`, for fitting Bayesian VAR models. Bayesian VAR models apply priors on the regression parameters and variance-covariance of the errors for a fine control over the posterior time-series process. By default, the prior on regression coefficients shrinks them toward a random-walk process that assumes no relationship between time-series variables. This assumption helps avoid overfitting

the data. The Bayesian approach also provides a systematic and unambiguous way of determining the number of lags.

We illustrate Bayesian VAR models on some real data and show model interpretations based on their impulse–response functions. We also compute Bayesian forecasts and compare them with classical forecasts

14:45–15:15 dstat: A unified framework for estimation of summary statistics and distribution functions

[Ben Jann](#) (University of Bern)

Abstract: I present a new Stata command that unites a variety of methods to describe (univariate) statistical distributions. Covered are density estimation, histograms, cumulative distribution functions, probability distributions, quantile functions, lorenz curves, percentile shares, and a large collection of summary statistics such as classical and robust measures of location, scale, skewness, and kurtosis, as well as inequality and poverty measures. Particular features of the command are that it provides consistent standard errors supporting complex sample designs for all covered statistics and that the simultaneous estimation of multiple statistics across multiple variables and multiple sub-populations is possible. Furthermore, the command supports covariate balancing based on reweighting techniques (inverse probability weighting and entropy balancing), including appropriate correction of standard errors. Standard error estimation is implemented in terms of influence functions, which can be stored for further analysis, for example, in RIF regressions.

15:15–15:45 wikiviews—A Stata interface for the Wikipedia API

[Ulrich Kohler](#) (University of Potsdam)

I present the user written Stata command `-wikiviews-` that allows flexible calls to the official Wikimedia API and to the database of its predecessor maintained by Peter Meissner. The program allows to create Stata datasets holding pageviews and related Statistics of long lists of Wikipedia pages from 2007 up to now.

16:15–17:00 Treatment-effects estimation with lasso

[Di Liu](#) (StataCorp)

Abstract: There is always an intrinsic conflict between the unconfoundedness assumption and the overlap assumption regarding the treatment-effects estimation. With high-dimensional controls, this conflict becomes even more vivid. This presentation shows how to overcome this conflict by using Stata 17's `-telasso-` command. `-telasso-` estimates the average treatment effects with high-dimensional controls while using `-lasso-` for model selection. This estimator is Neyman orthogonal because it is robust to the model-selection mistakes. It is also doubly robust, so only one of the models needs to be correctly specified.

17:00–17:30 Open panel discussion with Stata developers

Users air their wishes and grumbles, and StataCorp responds.

Scientific Organizers

The academic program of the conference is being organized by Johannes Giesecke (HU Berlin) and Ulrich Kohler (University of Potsdam)

TODO Logistics organizers

The logistics are being organized by Dittrich and Partner (<http://www.dpc.de>), the distributor of Stata in several countries including Germany, The Netherlands, Austria,