

Presentations

Resultssets to resultstables revisited

Roger B. Newson

Cancer Prevention Group, Wolfson Institute of Preventive Health, Queen Mary University London, UK

A resultsset is a dataset created as output by a Stata command. Multiple resultssets can be appended or merged or combined in other ways to make secondary resultssets. However, their usefulness is in that they can be converted to resultsplots and/or resultstables in documents in a variety of formats. We focus on resultstables in `.docx` documents. Converting resultssets to resultstables starts with decoding (using `sdecode` and its family of dependent packages) and ends with listing decoded variables to a document (using `docxtab` or `listtab`). However, intermediate steps may include `reshapeing` (`long` or `wide`), appending, merging, characterizing (to define column headers), inserting gap observations, and/or grouping rows into pages in multi--page tables. We illustrate this process using an example, outputting a multi-page table to a `.docx` document, and introducing the `ltop` package for grouping lines into pages..

Stata to Excel: From do-file to VBA

James Pike

Adelphi Real World

The introduction of Stata's `putexcel` command enhanced the integration between Stata and Excel, allowing users to export formatted results directly from one to the other. Via `putexcel`, complex outputs and spreadsheets are possible without copy-pasting or manual formatting. For many tasks, `putexcel` streamlines workflows and saves time. However, `putexcel` has limits. Some Excel features, such as conditional formatting, autofit of cells, text to columns, or removing excess formatting, cannot be performed. However, Excel's own language, Visual Basic for Applications (VBA), enables automation options that go beyond Stata's scope. Using `putexcel` and then VBA in Excel often means running a do file and then opening the resulting Excel file in Excel to run VBA macros.

We present a method of automation where we use Stata to write and execute VBA code via a Visual Basic Script (VBS) file. By generating a `.vbs` script from within Stata (using the `file` command) and running it (with the `shell` command), users can automate Excel tasks that require VBA, all in the comfort of the Stata environment. This approach creates new possibilities for a streamlined workflow.

Conditional average treatment-effects estimation using Stata

Di Liu

StataCorp

Treatment effects estimate the causal effects of a treatment on an outcome. The effect may be heterogeneous. Average treatment effects conditional on a set of variables (CATEs) help us understand heterogeneous treatment effects. By construction, they are useful to evaluate how different treatment-assignment policies affect different groups in the population.

In this talk, we will show how to use Stata 19's new command `cate` to answer questions such as the following:

1. Are the treatment effects heterogeneous?
2. How do the treatment effects vary with some variables?
3. Do the treatment effects vary across prespecified groups?
4. Are there unknown groups in the data for which treatment effects differ?
5. Which is best among possible treatment-assignment rules?

Data reduction for graphical and other purposes

Nicholas J. Cox

Department of Geography, Durham University

Reducing a dataset to another dataset containing summary or other statistics is an old problem, much addressed in Stata by official commands such as `collapse`, `contract` or `statsby` and by various community-contributed commands. Often an underlying principle is that a valuable command should do one thing well, so that a reduction command is just one step in a sequence that includes other analyses.

This presentation focuses on a bundle of new commands recently posted on SSC, `cisets`, `momentsets`, `ptilesets`, and `lmomentsets`. They have much in common, including support for obtaining results for multiple variables and for distinct groups of a single variable. Typically the next stage is some graphical representation, such as a customised variation on existing designs.

Examples of their use will be coupled with ruminations on the trade-offs in command design, for programmers and users alike, between versatility and simplicity. As in the rest of life, sometimes programmers need to step back before they can move forward in a better direction.

.

Using `locproj` to easily estimate nonlinear local projections

Alfonso Ugarte-Ruiz
BBVA Research

We review all the possible alternatives of specifying nonlinear impulse response functions (IRF) through local-projections that are available using the user-written command `locproj`. For instance, the command allows easily specifying shocks that include basic non-linearities such as state-dependent impacts, quadratic effects, interactions between continuous variables, etc.

Moreover, it allows non-linearities in the dependent variable, such as when we are interested in estimating the response of the probability of a binary outcome, or when we want to uncover nonlinear effects of a shock by letting the parameters of the local projection regressions vary across the conditional distribution of the dependent variable through the use of quantile regression.

We explain how to use all the available options in `locproj` to accommodate all these different methodological alternatives and discuss the advantages that the command offers, for instance, that the command facilitates introducing lags of the dependent or the shock variables when using the Stata command `qreg`, which in principle does not allow time-series operators.

Testing and Estimating Structural Breaks in Time Series and Panel Data in Stata

Jan Ditzen

Yiannis Karavias

Joakim Westerlund

Free University of Bozen

Identifying structural change is a crucial step in analysis of time series and panel data. The longer the time span, the higher the likelihood that the model parameters have changed as a result of major disruptive events, such as the 2007-2008 financial crisis and the 2020 COVID-19 outbreak. Detecting the existence of breaks, and dating them is therefore necessary, not only for estimation purposes but also for understanding drivers of change and their effect on relationships.

This talk will introduce an updated version of `xtbreak` and discuss use, options and capabilities of `xtbreak`. First, the relevant econometric theory will be revisited followed by empirical examples. Emphasis will be put on challenges using `xtbreak` in panel data, how to interpret results and speed improvements using Python.

Spatial Unit Roots in Regressions

P David Boll

Sascha O. Becker

University of Warwick

Hans-Joachim Voth

University of Zürich

Spatial unit roots can lead to spurious regression results. We present a brief overview of the methods developed in Müller and Watson (2024) to test for and correct for spatial unit roots. We also introduce a suite of Stata commands `spur` implementing these techniques. Our commands exactly replicate results in Müller and Watson (2024) using the same Chetty et al. (2014) data. We present a brief practitioner's guide for applied researchers.

Seamless Multi-Arm Multi-Stage (MAMS) designs with treatment selection and interim change of outcome: An update to `nstage`

Yumeng Liu

Babak Choodari-Oskooei

MRC Clinical Trials Unit at UCL

Multi-Arm Multi-Stage (MAMS) selection designs, as an extension of the standard MAMS designs, offer additional efficiencies that accelerate the evaluation of medical interventions in clinical trials. Standard MAMS designs use stagewise hypothesis testing to compare multiple experimental treatments against a common control at interim analyses, enabling early stopping for overwhelming efficacy or lack-of-benefit. MAMS selection designs further incorporate predefined rules to choose the best-performing treatments. Incorporating intermediate outcomes, introduced to significantly shorten the timing of interim analyses, naturally fits into the seamless trial design framework, which allows for outcome changes at early stages of trial.

Our existing `nstage` suite of commands calculates target sample sizes for MAMS designs with binary outcomes such as death or disease progression. The program also projects timelines for trial planning and computes overall operating characteristics (overall pairwise/familywise type I error rates, power, and expected sample sizes).

We have enhanced the program to support interim outcome changing and the interim rules for treatment selection, lack of benefit, and overwhelming efficacy. The updated `nstage` command is now more flexible, enabling changes to trial outcomes at interim stages, making it well-suited for seamless Phase II/III trial designs. It also supports treatment selection based on either Phase II (intermediate) or Phase III (primary clinical) outcomes.

We will describe the new MAMS design and the associated Stata command using a miscarriage MAMS platform trial in maternal health.

RAMPE: Randomisation Allocation Method Performance Evaluation

Cydney Bruce

Alan Montgomery

Reuben Ogollah

Christopher Partlett

Nottingham Clinical Trials Unit

When designing and conducting a randomised controlled trial, there are a variety of randomisation methods to choose from, but limited evidence on the performance of the methods under specific study designs. The `RAMPE` package contains 12 metrics designed to measure the balance and predictability of randomisation sequences in Stata. This will allow researchers to easily compare method performance using data that mirrors the specific trial that is being designed.

Balance metrics

Measured both as the greatest imbalance observed throughout recruitment, and the final imbalance once the target sample size is achieved.

`groupimbalance`

Measures the imbalance between the expected and observed ratio of participants in each treatment group.

`charimbalance`

Measures the greatest imbalance observed across a set of covariates and the average imbalance across covariates.

Predictability metrics

Measured as the proportion of correct guesses for a variety of prediction strategies. This is calculated for the whole sequence and assuming that recruiting sites only have information about previous allocations at their own site.

`alternation`

Recruiter assumes the next allocation is the one least recently allocated.

`backtheloser`

Recruiter assumes the next allocation is the one with the fewest previous allocations.

`predbalance`

Recruiter assumes the next allocation is the group with the smallest marginal total across randomisation covariates.

In this talk, I will describe each of the developed metrics in more detail, discuss the interpretation of each metric and demonstrate with an example how this package can be used in practice.

Optimal Policy Learning for Multi-Action Treatment and Risk Preference

Giovanni Cerulli
CNR-IRCREs

I present `opl_ma_fb` and `opl_ma_vf`, two community-distributed Stata command implementing first-best Optimal Policy Learning (OPL) algorithm to estimate the best treatment assignment given the observation of an outcome, a multi-action (or multi-arm) treatment, and a set of observed covariates (features). It allows for different risk preferences in decision-making (i.e., risk-neutral, risk-averse linear, risk-averse quadratic), and provide graphical representation of the optimal policy, along with an estimate of the maximal welfare (i.e., the value-function estimated at optimal policy). A practical example of the use of these commands is provided.

Optimal Policy Learning for Multi-Action Treatment and Risk Preference

Joao Santos Silva
Jeffrey H. Bergstrand
Matthew W. Clance
University of Surrey

In many applications, the outcome of interest is non-negative and has a mixed distribution with a long right-tail and a mass-point at zero. Applications using this sort of data are typical in health and international economics, but are also found in many other areas. The lower bound at zero implies that models for this kind of data are generally heteroskedastic, implying that the regressors will have different effects on different regions of the conditional distribution.

The traditional way to learn about heterogeneous effects in conditional distributions is to use quantile regression. However, the conditional quantiles of outcomes of this kind cannot be given by smooth functions of the regressors because the mass-point implies that some quantiles will be identically zero for certain values of the regressors. This complicates the estimation of quantile regressions for data of this kind and the interpretation of the estimated parameters. As an alternative, we can estimate Poisson-based expectile regressions using Efron's (1992) asymmetric maximum likelihood approach.

After highlighting the problems that afflict estimation of quantile regressions for this kind of data, we briefly introduce expectile regression as introduced by Newey and Powell (1987) and show how they can be estimated with non-negative data using Efron's (1992) approach. We then introduce the `appmlhdf` command and illustrate its use.

References

Efron, B. (1992): Poisson Overdispersion Estimates Based on the Method of Asymmetric Maximum Likelihood. *JASA*, **87**., 98–107.

Newey, W. K. and J. L. Powell (1987): Asymmetric Least Squares Estimation and Testing. *Econometrica*, **55**., 819–847.

Testing whether group-level fixed effects are sufficient in panel data models

David Vincent

David Vincent Econometrics Limited

This presentation introduces a new command, `xtfelevel`, which implements a Hausman-type test to assess whether controlling for fixed effects at a more aggregate (group) level is sufficient for consistently estimating the coefficients on unit-specific, time-varying variables in linear panel data models where units are nested within groups. The command builds on Papke and Wooldridge (2023), who develop a test of the null hypothesis that the probability limits of the fixed effects estimators for a coefficient of interest are the same, whether heterogeneity is controlled at the unit or group level. Rejection of the null suggests that unit-level fixed effects estimation is required.

`xtfelevel` extends this framework by comparing the unit-level fixed effects estimator with an IV estimator that allows the time-varying controls to be correlated with unit-level heterogeneity, while accounting for correlation between the variable of interest and group-level effects. This estimator yields results analogous to pooled OLS estimation of the Mundlak regression, where the time average of the variable of interest is first partialled out from the time averages of the controls. Under the null, the estimator can often be more efficient than the unit-level fixed effects estimator, especially when the variable of interest exhibits limited within-unit variation.

This extension addresses a limitation in applying the usual Mundlak device to obtain more efficient estimates, as discussed by Wooldridge (2019). When the variable of interest is uncorrelated with the unit-level heterogeneity but is correlated with the time-varying controls that are themselves correlated with those effects, excluding its time mean to improve efficiency can lead to omitted variable bias.

Power and sample size by simulation

Alex Asher
StataCorp

Stata's built-in `power` command accepts user-defined programs to calculate power, sample size, or effect size. Power can be estimated by simulation, even in complex scenarios where there is no closed-form expression. To estimate sample size given power, multiple simulations are needed. This talk describes how to use simulation to estimate power and sample size using the `power` command.

Learn how to do the following:

1. Write simulation programs that are compatible with all the features of `power`, `ciwidth`, and `gsdesign`.
2. Customize graphs and tables using an initializer.
3. Control Monte Carlo errors.
4. Estimate sample size using the bisection method.

Adventures with the profile log-likelihood

Ian White
Patrick Royston
UCL

`p11f`, written by Patrick Royston in 2007, computes and graphs the profile log-likelihood function for a wide variety of regression commands. This enables calculation of confidence intervals that do not rely on the standard Wald approximation that $(\text{estimate} - \text{true})/\text{SE}$ is Normally distributed: `p11f` confidence intervals are likely to perform better than Wald ones in smaller samples.

We believe `p11f` is an under-used command for analysis, and we also find it useful for understanding and explaining statistical methods. We aim to demonstrate its usefulness for teaching purposes and for understanding bias in two-stage meta-analysis.

We also describe some recent minor improvements in `p11f` (e.g. it is now a prefix command). The latest version is available on GitHub and SSC.

Bayesian meta-analysis is easier than you think

Robert Grant
Prof Gian Luca Di Tanna
Dr Joseph A R Santos
BayesCamp

Meta-analysis presents several methodological challenges when synthesizing evidence across studies, particularly in scenarios where conventional asymptotic approximations become unreliable. Bayesian methods offer a natural framework for evidence synthesis through their flexible treatment of uncertainty. The Bayesian paradigm accommodates sparse data structures, evidence beyond the study data, systematic biases, and missing study information. It leads to probabilistic outputs that directly address decision-makers' needs and allow easier interpretation.

We present findings from our comprehensive review of models and software in preparation for a new book, *Bayesian Meta-Analysis: a practical introduction*, and from a scoping review, and its ongoing update. This has shown the potential for many widespread problems in meta-analysis to be addressed in the near future.

We challenge the perception that Bayesian methods are inaccessible to non-statistical researchers, illustrating simple and flexible implementation in Stata. Bayesian meta-analysis extends naturally to network meta-analysis and living evidence synthesis from its foundations as a class of multilevel models. We also present practical guidance on prior specification and model validation to complete a reliable Bayesian workflow. Importantly, regulatory agencies and major journals increasingly recognize the value of Bayesian meta-analytic approaches, reflecting their growing adoption in high-impact research synthesis.

A simple approach to compute generalized residuals for nonlinear models

Arnab Bhattacharjee

Tibor Szendrei

Heriot-Watt University, UK

In models where the relationship between the outcome and the error term is linear, a residual can be computed by simply plugging-in the estimated coefficients and computing the difference between observed and predicted values of the outcome variable. These residuals can then be used for many different purposes, for example: (a) evaluating assumptions of orthogonality of errors (like, fixed and random effects); (b) examining the entire shape of the error distribution; and (c) computation and inference on externalities such as network effects. However, this simple approach does not work when the model is nonlinear in outcomes and errors. Here, different context-specific generalized residuals have been proposed, each having different properties for specific models. Note that, for the canonical linear or nonlinear Gaussian regression model, the above construction is simply a scaled version of the partial derivative of the log-likelihood contribution of an individual observation with respect to the outcome variable. This suggests a general construction of generalized regression by perturbing the outcome variable and computing contrasts. This approach is closely related to Huber's influence function and can be routinely computed using Stata for example and also parallelized for large datasets. We propose this general construction of generalized residuals and evaluate its use in several contexts: (a) quantile regression and evaluation of conditional quantiles at the tails (for example, growth at risk); (b) computing errors distributions (for example, binary regression and random effects models); and (c) computing network externalities in discrete choice and duration models. This delivers a unified approach with promising findings.

crosswalk: A new command for fast and flexible bulk recoding

Ben Jann
University of Bern

In this talk I will present the new `crosswalk` command, a data management utility for fast table-based recoding. The command comes with predefined crosswalk tables for common recoding tasks related to occupational classifications, e.g. to translate ISCO codes (International Standard Classification of Occupations) into ISEI scores (International Socio-economic Index of Occupational Status), OEP scores (Occupational Earning Potential), or ESeC classes (European Socio-economic Classification). However, it is also easy to define, manipulate, and apply custom recoding tables. In the talk I will briefly explain how `crosswalk` is implemented, present its syntax, and then illustrate its use with some applied examples.

blockops: A new Mata library for efficient operations on block matrices

Daniel Schneider

Max Planck Institute for Demographic Research

This presentation introduces a new Mata library called `blockops`. Its main feature is a class that divides a matrix into multiple submatrices. Operations on the original matrix are then carried out in terms of the submatrices. The library mainly serves two purposes: First, it provides a simple approach of dealing with special kinds of sparse matrices. Submatrices that consist entirely of zeroes are represented by a null pointer and do not partake in arithmetic operations. For suitable applications, this can lead to vast increases in speed with regards to matrix multiplication and matrix inversion. The second purpose is the application of a built-in or user-defined function to each submatrix, similar to, for example, R's `*apply()` functions. This can ease code generation and improve readability while maintaining Mata's favorable speed properties. Several examples are shown to demonstrate the usefulness of the new library for statistical calculations.

Posters

Imputation of four varieties of sweet potatoes when data cannot be pooled

Oyebola Elemide

Sulelmon Arisanyi Ayodeji

Federal College of Agriculture Akure Ondo state, Nigeria

Simulation studies are used in a variety of disciplines to evaluate the properties of statistical methods. They are an important and useful tool which enables researchers to compare different statistical methods and understand their properties.

We introduce the `simann` suite which has been created to assist the analysis of simulation results. A set of Stata programs that offer data manipulation, analysis and graphics to process, explore and visualize the results of simulation studies. (s): the question(s) the simulation study is designed to address. Data-generating Mechanism(s), DGM(s): Sets out how the simulated data will be created Estimated. Also called target. Sets out what quantity the analysis is trying to estimate.

Method(s): Sets out how each simulated data set will be analyzed to give an estimate of the estimand.

Performance Measure(s): Current imputation methods require the availability of individual data from the involved studies to impute missing data. This creates a need for methods that can impute data in one study using only information that can be easily and freely shared within a data network.

Estimates data set: results from analyzing multiple simulated data sets, with data relating to different statistics for each simulated data set., repetition (rep): simulation number For the input estimates data, there are 3 formats permitted by the `simann` suite: Long for both targets and methods:

User-friendly software for analyzing the results of simulation studies. Multiple estimates and performance measures graphs available, visual tools to help assess the performance of different methods.

Minimum wage and health status in Europe

Aleksandra Majchrowska

University of Lodz, Lodz, Poland

Sylwia Roszkowska

Jagiellonian University, Cracow, Poland

The importance of health for the economy is indisputable. In this study, we explore the impact of minimum wage increases on health status across European countries, highlighting the significance of income as a social determinant of health.

Empirical evidence suggests that raising the minimum wage can improve physical and mental health by reducing financial stress and enabling access to better nutrition, housing, and healthcare services. On the other hand, higher minimum wages can increase risky behaviours by allowing individuals to purchase unhealthy goods or can increase unemployment and poverty.

The study aims to analyse the relationship between the minimum wage and the health status and mortality in European countries and regions. We use the NUTS2 level data for EU countries in 2000-2020. We use the classic methods of panel data analyses and those allowing for endogeneity, including the GMM estimator. Various control variables, including cultural, environmental, and socio-economic factors, are used.

The results indicate a positive impact of minimum wage on the health status of citizens. In countries with higher minimum-to-average wage ratios, the health status is higher, and the mortality is lower due to lower income inequalities. Moreover, the health status is higher, and the mortality rate is lower in countries with lower regional disparities in the minimum-to-average wage ratio.

While some critics argue that minimum wage increases may lead to job losses or inflationary pressures, our research suggests that these effects are limited. Our results show that the potential health benefits are more consistent and long-lasting.

Macroeconomic Shocks and Financial Resilience in Emerging Markets: An Empirical Investigation of Morocco's Stock Market Dynamics

Hamza Kadiri

Khalid Belkhoutout

Mohammed V University, Morocco

This study examines the relationship between six macroeconomic variables (inflation, gross domestic product, interest rate, budget deficit, exchange rate, and foreign direct investment) and the performance of the Moroccan stock market, specifically the Moroccan All Shares Index (MASI), from 2005 to 2025. Employing an Autoregressive Distributed Lag (ARDL) approach, this study aims to determine the short- and long-run connections between these macroeconomic variables and stock performance in Morocco. The findings indicate that inflation exhibits an inverse relationship with stock performance in the short term, whereas GDP growth has a positive impact on stock performance with a lag effect. Interest rates are shown to exert a time-varying dual effect, being negatively correlated with stock performance in the short term, but positively correlated in the long term. Additionally, budget deficits are consistently found to have a significantly negative correlation with market performance. Conversely, FDI and exchange rates do not appear to have a substantial direct impact on the MASI index. To confirm these econometric findings and increase the predictive power, this study incorporates explainable artificial intelligence (XAI) techniques, specifically utilizing the XGBoost model. The results of the XAI analysis underscore GDP as the most significant predictor of MASI. Our research contributes to the literature by combining traditional econometric methods with XIA approaches to understand stock market dynamics in emerging economies such as Morocco. The implications of these findings are particularly relevant to policymakers and investors engaged in emerging markets.

Does aging matter in the impact of the minimum wage on inflation?

Sylvia Roszkowska

Jagiellonian University, Cracow, Poland

Aleksandra Majchrowska

University of Lodz, Lodz, Poland

We examine how demographic changes impact the transmission of minimum wage increases to inflation. The minimum wage growth can raise the prices of goods and services and accelerate inflationary processes. At the same time, a shrinking workforce and changes in its structure could lead to changes in the impact of minimum wage increases on the economy. We use the minimum wage augmented Phillips curve framework extended with the demographic variables. We employ the sample of 21 European Union countries in 2003-2023 and panel data techniques.

Our study proves that the strength of the minimum wage pass-through effects on inflation depends on demographic factors. Aging of the workforce and shrinking workforce size weakens the impact of minimum wage increase on inflation. Contrary, a lower proportion of the less educated working-age population strengthens the minimum wage pass-through effects on inflation. Our results have important implications for macroeconomic, minimum wage, and education policies.