Cleaning large dataset in Stata with applications to asset pricing

Online course

Centre for Econometric Analysis
Delivered by: Soon Leong.

Fees:

£120 City students, alumni, Staff

£140 External students

£240 External rate

A 15% discount is available for groups of three or more participants

Course overview

In this online course, you will learn to use Stata for research that involves large and unbalanced panel dataset with missing observations. In particular, the course deals with a large number of firms (approx. *N=25,000*) with a large number of observations (approx. *T=2,000*).

Benefits

- You will be introduced to the Stata statistics software
- You will learn novel procedures to clean large and unbalanced datasets, including trimming, merging, appending and dealing with missing observations
- You will get a good understanding of empirical asset pricing using the Capital Asset Pricing Model (CAPM), Fama-French three factors model as well as (G)ARCH models
- You will be able to continue your learning with the course materials, the data you used, as well as files created throughout the course.

Target audience

This course is useful to anyone with general research interest dealing with large and unbalanced panel dataset.

Course prerequisites

The course requires elementary knowledge in Stata but assumes an interest in research and in learning to use Stata. Knowledge of the fundamentals of econometrics and asset pricing theory will help participants to obtain the maximum benefit from the course.

Contents

Online session – Four hours long

- Using Stata via the Graphical User Interface
- Working with do-files
- Loading data and checking for errors
- Dealing with string variables and dates in Stata
- Computing descriptive statistics to summarise your data
- Plotting distributions of start and end dates
- Decision making using descriptive statistics and graphical inspection
- Trimming large dataset to create a balanced panel
- Merging large datasets
- Appending large datasets
- Dealing with missing observations
- Checking and dealing with duplicates
- Extracting idiosyncratic volatility using CAPM
- Fama-French three factors model
- Unit root tests
- Estimating idiosyncratic volatility using (G)ARCH models.

For more info contact:
elisabetta. pellini@city.ac.uk
Bayes Business School
CEA@city.ac.uk
www.bayes.city.ac.uk/
cea-econometrics-courses

Recommended reading

The following textbooks and journal articles are recommended for this course:

Fu, F., 2009. Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economic*s 91, 24-37.

Soon Leong

Soon is a PhD candidate at Bayes Business School, City, University of London. He is associated with the Centre for Econometric Analysis under the sponsorship of "2017/2021 PhD Studentship in Memory of Ana Timberlake". Soon works in the field of applied and theoretical financial econometrics under the supervision of Professor Giovanni Urga on several research topics such as measuring and testing systemic risk, causality measures for VARMA and VAR models, multivariate Granger causality in variance, short- and long-run volatility spillover measures, and out-of-sample forecasting. Soon is also involved in some teaching assistance within Timberlake training courses at BayesBusiness School, Cambridge University and Oxford University, European Banking Authority, European Central Bank. He is proficient with intermediate and advanced econometric/statistical software such as EViews, Matlab, Oxmetrics and Stata.

Registration, payment and cancellation policy

Payment of course fees is required prior to the course start date.

In case a course is cancelled, registered participants will receive the full refund.

Registration closesseven calendar days prior to the start of the course.