The Econometrics of Financial Returns. An Introduction

Carlo Favero

September 2018

Why Econometrics?

- Econometrics is the statistical analysis of economics data.
  - Government (decision on optimal policies: monetary, fiscal, health, education)
  - Business (strategic planning, pricing, advertising, sales, production, investment, inventories)
  - Financial Services (modelling returns and risk)

The Econometrics of Financial Returns

- Predicting the distribution of returns of financial assets is a task of primary importance for identifying desirable investments, performing optimal asset allocation within a portfolio, as well as measuring and managing portfolio risk.
- Horizon at which returns are defined matters for financial and statistical reasons
- Portfolio allocation, i.e. the choice of optimal weights to be attributed to the different financial assets in a portfolio, is typically based on a long horizon perspective,
- the measurement of risk of a given portfolio takes typically a very short-horizon perspective.
- A long-run investor decides her optimal portfolio allocation on the basis of the (joint) distribution of the returns of the relevant financial assets at low frequency. However, the monitoring of the daily risk of her portfolio depends on the statistical properties of the distribution of returns at high frequency.

The relevant dimensions of the data

- The statistical properties of the distribution of returns vary with the horizon at which they are defined. There are three relevant dimensions in financial returns
  - time-series
  - cross-section
  - the horizon at which returns are defined

The relevant dimensions of the data

- In general, we shall define $r^i_{t,t+k}$ as the returns realized by holding between time t and time $t+k$, the asset i.
  - the t index captures the time-series dimension
  - the i index the cross-sectional dimension
  - the k index the horizon dimension

The Econometric Modelling Process

- Econometrics uses the "past available data" to predict the future distribution of returns. This is a process the provide some information only if past data are capable of giving us some information on the future and involves several steps:
  - Data collection and transformation
  - Graphical and descriptive data analysis
  - Model Specification
  - Model Estimation
  - Model Validation
  - Model Simulation

Financial data

- In general financial data are not generated by experiments, we have only "observational data"
- Special issues arise routinely in economic and financial data (special days, seasonality, trends, cycles)
- We use the data by building, estimating and simulating models to predict returns and their distribution
- models need to be validated to minimize the risk of using a "wrong" model

Econometric Modelling of Financial Returns

- Econometric models of financial returns specify the distribution of a vector of variables $\mathbf{y}_t$ conditional upon other variables $\mathbf{z}_t$that are helpful in predicting them and do not depend on them.
- The mapping between $\mathbf{y}_t$ and $\mathbf{z}_t$ is determined by some functional relation and some unknown parameters.
- All the relevant variables are stochastic and they are therefore characterized by a density function.
- Linear Econometric Models specify conditional means of the $\mathbf{y}_t$ as linear functions of the $\mathbf{z}_t$.

- the data

$$D\left(\mathbf{y}_t, \mathbf{z}_t, \mathbf{w}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \mathbf{W}_{t-1,}\boldsymbol{\theta}\right)$$

- a general multivariate model

$$D\left(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}\right)$$

- decomposing a multivariate into conditional and marginal

$$D\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1\right) D\left(\mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_2\right)$$

- a general linear univariate conditional model

$$\begin{aligned} y_t &= \boldsymbol{\beta}_1' \mathbf{z}_t + u_{1t} \\ \mathbf{z}_t &= \boldsymbol{\beta}_2' \mathbf{z}_{t-1} + u_{2t} \end{aligned}$$

A simple example

- $$\begin{aligned} \left(r_t^i - r_t^{rf}\right) &= \beta_{0,i} + \beta_{1,i}\epsilon_t + u_t \\ \left(r_t^m - r_t^{rf}\right) &= \beta_{0,m} + \epsilon_t \\ \begin{bmatrix} u_t \\ \epsilon_t \end{bmatrix} & \quad u_{i,t} \sim n.i.d.\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_m^2 \end{bmatrix}\right) \end{aligned}$$

To do list

- estimate $\beta_{0,i}, \beta_{1,i}, \beta_{0,m}, \sigma_i^2, \sigma_m^2$
- simulate the model to predict future returns and their distribution
- validate the model

Why a model can be wrong ?

There are many ways in which the model can go wrong:

- other factors beyond the market are relevant in determining excess returns on asset i
- the excess returns on the market do depend on excess returns on asset i
- the model is non-linear
- the residuals are non-normal

This course

The objective of this course is to lead students to learn the econometrics of financial returns by developing skills along four different, but highly interrelated, dimensions:

- knowledge of the relevant theory
- knowledge of the relevant data;
- knowledge of the relevant econometric methods;
- capability of implementing empirical applications (coding).

The Exam

- Students will be assessed in a final computer based exam.
- The objective of the exam will be to evaluate the individual capability of students of using the inputs given to build the relevant output
- During the exam students will be required to modify the R codes that they have built during the course to generate answers to the questions posed in the exercises.
- Working on the exercise step by step and using all the inputs given is the best preparation strategy for the exam.
- The exams will be open books.

The Data

- We will consider time series data of monthly observations of different portfolios made available by Ken French from his website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_lil data are available in EXCEL format in the file FF_Data.xls. This files combines 3 data files from the website to make available monthly observation on 32 time series from July 1927 onward.
- We will also consider data made available by Bob Shiller from his website, http://www.econ.yale.edu/~shiller/data.htm, that consists of an updated version of the time-series used in the book Irrational Exuberance

Ken French Data

- The dataset contains the risk-free rate,Rf, and the five Fama-French factors : EXRET_MKT, SMB, and HML, RMW and CMA.
  - SMB (Small Minus Big) is the average return on the small stock portfolios minus the average return on the big stock portfolios
  - HML (High Minus Low) is the average return on the value portfolios minus the average return on the growth portfolios
  - RMW (Robust Minus Weak) is the average return on the robust operating profitability portfolios minus the average return on the weak operating profitability portfolios
  - CMA (Conservative Minus Aggressive) is the average return on the conservative investment portfolios minus the average return on the aggressive investment portfolios
- details on the construction of these portfolios are available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f_5_factors_2x3.html

Ken French Data

- The momentum factor, MOM, that comes from the file Momentum factor. It contains a momentum factor, constructed from six value-weight portfolios formed using independent sorts on size and prior return of NYSE, AMEX, and NASDAQ stocks. MOM is the average of the returns on two (big and small) high prior return portfolios minus the average of the returns on two low prior return portfolios. The portfolios are constructed monthly. Big means a firm is above the median market cap on the NYSE at the end of the previous month; small firms are below the median NYSE market cap. Prior return is measured from month -12 to - 2. Firms in the low prior return portfolio are below the 30th NYSE percentile. Those in the high portfolio are above the 70th NYSE percentile.

Ken French Data

- The last 25 time series come from the file 25_portfolios_5x5. This file contains value- and equal-weighted returns for the intersections of 5 ME portfolios and 5 BE/ME portfolios. The 25 time series we consider are to equal weighted returns.

The portfolios are constructed at the end of Jun. ME is market cap at the end of Jun. BE/ME is book equity at the last fiscal year end of the prior calendar year divided by ME as of 6 months before formation. Firms with negative BE are not included in any portfolio. PR11 are the returns on the portfolio made with the smallest firm and the lowest book equity, PR12 are the returns on the portfolio made with the smallest firm and the second lower book to equity and so on until PR55, which are the returns on the portfolio of the largest firms with the highest book to equity.