Introduction to (Finance) Data Analytics and Big Data

Professor: Javier Gómez Biscarri
E-mail: javier.gomez@upf.edu
Office hours: by appointment

Course Description

The world of finance is especially data intensive: most of the main financial decisions taken by investors or by corporations (investment and financing decisions) rely on numbers which need to be derived after the analysis of data. For example, the simple cash-flow forecasts needed for the calculation of any NPV analysis are, by definition, *predictions* of unknown quantities which are obtained from some statistical analysis which extracts patterns and structure from available data.

In recent years, two trends have significantly informed the development and application of data analysis in Finance (but also across all areas in the business world). One is the realization that *data analytics* must become a core activity of businesses and corporations. In other words, the analysis of data must be an ongoing activity, carried out not just at specific moments of time but constantly, aimed at the development of *business intelligence*, i.e., of insights that can inform all future decisions. The second one is the increased availability of larger and larger datasets (*big data*) that allow for richer analyses but also require an adaptation of traditional analyst-driven statistical techniques.

In this course we will, first, present the main classical techniques of analysis of financial data with an emphasis on transforming the results of the analysis into conclusions that can inform decision making. Second, we will give a brief introduction to big data, showing some examples and reviewing the main challenges that big data introduces into data analysis. This latter part should serve as an introduction to the core course "Data Science for Finance" which will specifically focus on big data analysis.

Competencies

The main competencies that the course will develop are:

- The application of financial tools –qualitative, econometric and statistical- to the calculation of the value of financial and real assets, as well as company valuation.
- The application of statistical tools to inform the loan granting activity of a financial institution.
- The understanding of how new A.I. and M.L. techniques can contribute to the digital economy.

Learning Outcomes

The main learning outcomes of the course are:

- A mastery of the tools neded for a proper analysis of performance of firms and banks.
- Knowledge of the appropriate qualitative, econometric and statistical techniques needed for each financial decision.
- The ability to value some assets and their risk-return characteristics.
- Knowledge of techniques for the evaluation of credit risk.
- The ability to define and frame a problem in a coherent way, as well as to design a rigorous process of analysis.
- The ability to communicate the results of the analysis in a coherent report.

Objectives

The main objectives of the course are:

1) To learn to apply the basic techniques of classical statistical analysis of cross-sectional and time-series financial data and to identify which technique is more appropriate for each specific type of data and objective of the analysis.
2) To learn how to transform financial data into knowledge useful for financial decision making and how to communicate such knowledge more effectively by linking it to sound analysis.
3) To introduce the analysis of "big data" contexts, which are becoming more and more frequent in business and financial settings.

Methodology

There will be 20 one and half-hour theory sessions where we will discuss the concepts and methodologies of financial data analysis, including a (short) introduction to big data issues (which are the object of a subsequent core course). These theory sessions will consist of lectures where the professor will cover the formal aspects of the methodologies and the interpretation of results, and of discussions of practical applications of the methodologies to specific finance –and, sometimes, non-finance but finance related- examples. Students are expected to take their own notes from the lecture parts and to engage actively in the theory sessions, especially in the discussion of examples. Class slides will be available on eCampus, but these sets of slides will be synthetic: lectures may not follow the slides literally and will expand on the main issues. Every week we will also have a short discussion of a simple case of "bad use of data analysis" from real life. These short discussions are intended to stress the importance of good data analytics.

There will be 6 seminar classes where students will practice data analysis and interpretation (including discussion of the data analysis projects after submission). These seminars should help review the main concepts covered in class, get hands-on practice with the software of analysis (Python) and discuss issues related to model building and interpretation of results. The latter should be key components of the preparation of the final exam.

Evaluation criteria

Regular Assessment

To pass the course, you need at least 50 points out of 100, according to the following distribution and conditions:

Three Data Analysis projects: 40 points. Two of these data analysis projects (1 and 3, 15 points each) will be individual and one (2, 10 points) will be done in groups of 2-3 people. Details for each project will be provided in separate documents. Deadline for project submission is given in the project documents and in the class syllabus. Late submissions will be penalized 15% of the grade per day (or fraction of a day) late.

Final sit-in exam: 60 points. In order to pass the course, you must earn at least 24 points in this exam.

Re-take

If the student has not passed the course, the final exam can be retaken during the retake period. The final grade will be computed using the grade in the retake exam.

Students in remote

Students who are following the course remotely due to justified reasons will be evaluated with the same grade items. If the exams are finally done in remote, these exams will be taken at the same time as the other students and submitted following the directions given by the professor on the day of the exam.

Students are required to attend 80% of classes. Failing to do so without justified reason will imply a Zero grade in the participation/attendance evaluation item and may lead to suspension from the program

Students who fail the course during the regular evaluation are allowed ONE re-take of the evaluation, in the conditions specified above. If the course is again failed after the retake, the student will have to register again for the course the following year.

In case of a justified no-show to an exam, the student must inform the corresponding faculty member and the director(s) of the program so that they study the possibility of rescheduling the exam (one possibility being during the "Retake" period). In the meantime, the student will get an "incomplete", which will be replaced by the actual grade after the final exam is taken. The "incomplete" will not be reflected on the student's Academic Transcript.

Plagiarism is to use another's work and to present it as one's own without acknowledging the sources in the correct way. All essays, reports or projects handed in by a student must be original work completed by the student. By enrolling at any UPF BSM Master of Science and signing the "Honor Code," students acknowledge that they understand the schools' policy on plagiarism and certify that all course assignments will be their own work, except where indicated by correct referencing. Failing to do so may result in automatic expulsion from the program."

Outline of topics and (tentative) calendar

Session	Content	Week
Introduction to data Analytics in Finance		
1	Introduction to the course: admin stuff, evaluation, plan & syllabus	1
2	The tools of data analytics: Python for financial data analysis	1
3	Data analytics in finance: prediction and policy; sampling and inference	2
4	Financial data: main variables, specific features	2
Cross-sectional data analytics		
5	Regression analysis for prediction (1)	3
6	Regression analysis for prediction (2)	3
7	Regression analysis for prediction (3) – **Data project 1**	4
8	Classification analysis (1)	4
9	Classification analysis (2)	5
10	Classification analysis (3) – **Data project 2**	5
Time series data analytics		
11	Basic time series models for prediction (1)	6
12	Basic time series models for prediction (2)	6
13	Basic time series models for prediction (3) – **Data project 3**	7

14	Nonstationary time series and the random walk	7
15	Multiple time series	8
15	The volatility of financial time series	8
When N/T or k become very large: big data		
17	An introduction to big data: the relationship with classic data analysis – model selection	9
18	Big N / big k datasets: recommendation analysis	9
19	Text analysis: Twitter in action	10
20	Summary, questions, etc.	10
	FINAL EXAM	

The day/time of the practice sessions will be conveniently announced.

Reading Materials/ Bibliography/Resources

Textbooks

Heiss, F., Brunner, D., *Using Python for Introductory Econometrics*. http://www.UPfIE.net.

Brooks, C., *Introductory Econometrics for Finance*. Cambridge University Press (4th edition).

Brooks, C., Tao, R., *Python Guide to accompany Introductory Econometrics for Finance*. Available online.

Ma Weiming, J., *Mastering Python for Finance* (2nd edition). Packt.

Class slides provided by professor

Bio of Professor

Prof. Gómez Biscarri holds an M.A. in Economics and a Ph.D. in Business Economics from UCLA. He is Serra Húnter Associate Professor of accounting and finance at Universitat Pompeu Fabra and affiliated professor of the Barcelona Graduate School of Economics and of the Barcelona School of Management. He also teaches for IESE Business School and Deusto Business School. His current research focuses on the impact of accounting on the banking sector, specifically on earnings management behavior and on the effect of depositor discipline and new accounting regulation on banks' strategies.