Data Science and Machine Learning in Finance (ACCFIN5246) 2022-23

Lecturer: Hormoz Ramian (HR)
Office: Adam Smith Business School
Email: hormoz.ramian@glasgow.ac.uk
Office Hours: Friday 9-10am
Lecture time: Friday 3-5pm
Venue: MyGlasgow
Course GTAs: Tongtong Wang (TW); Hadi Movaghari (HM)
Course material are available on Moodle
Course Unit: 20 Credits

1 Overview

Addressing real-world economic and financial problems via information embedded in data is an active area of academic and professional interests. This course contributes towards this goal with the following two approaches. First, the course provides a foundation to methodically structure large-dimensional datasets and summarise information into interpretable outcomes. The second part of the course examines how the combination of large datasets accompanied by statistical learning and artificial intelligence techniques are helping practitioners to make more efficient economic and financial decisions. The course is delivered based on a balanced combination of descriptive contents required to formulate financial and economic problems into quantifiable objects of interest, analytical derivations and statistical techniques and software programming.

The course is structured based on four pillars. All pillars are equally weighted in terms of course contents and examinations but also in terms of their importance towards developing a foundation for future careers in data science and machine learning in finance within or outside academia:

- Data
- Methodological Frameworks
- Software Implementations
- Finance Theory and Applications

2 Pre-requisites

Intermediate statistics or econometrics including *ACCFIN5039 Principles of Financial Econometrics* or *ACCFIN5217 Financial Econometrics*. Prior familiarity with calculus and finance *(ACCFIN5205, ACCFIN5236)* would be helpful but not required.

3 Course Contents

The course is delivered via weekly sessions:

1. Data Science: Definition, Aim and Operations, Data Characterization

2. Data Science: Descriptive Analytics and Predictions
3. Statistical Models and Specifications
4. Constrained Linear Methods
5. Nonlinear Methods (Semi-Linear Least Squares, NLS, MLE)
6. Reduction and Regularization (Ridge and Lasso-family)
7. Learning-based Specification
8. Classification
9. Principal Component Analysis

4 Semester Plan

The dates and venues below are subject to change, please review this timetable on weekly basis: MyGlasgow

Week	Week Commencing	Main Sessions (HR) Fridays 3-5	Tutorial Classes (HM) (i) Thursdays 4-5 (ii) Fridays 5-6 (iii) Mondays 9-10	Software Lab (TW) (i) Thursdays 5-6 (ii) Mondays 4-5 (iii) Mondays 5-6	Office Hours	Assessed Coursework
1	09 Jan	Data Science: Definition, Aim and Operations, Data Characterization	—	—	HR: Fridays 9-10am, TW: Friday 13-14pm, HM:	—
2	16 Jan	Data Science: Descriptive Analytics and Predictions	—	—		—
3	23 Jan	Statistical Models and Specifications	Probability Fundamentals and Statistical Distributions	Numerical Sructures		—
4	30 Jan	Constrained Linear Methods	Matrix Algebra 1	Basic Operations, Import & Export Data, Visualisation		Quiz Monday 30 Jan (at 9am within 24hrs)
5	06 Feb	Nonlinear Methods (Semi-Linear Least Squares, NLS, MLE)	Matrix Algebra 2	Loop		Group Assignment: Deadline 9 February
6	13 Feb	Reduction and Regularization (Ridge and Lasso-family)	Problem Set 1a	Regression		—
7	20 Feb	Learning-based Specification	Problem Set 1b	Distribution, Max Likelihood Estimation		—
8	27 Feb	Classification	Problem Set 2a	Optimization		—
9	06 Mar	Principal Component Analysis	Problem Set 2b	Numerical Calculation & Symbolic Operation		—
10	13 Mar	Applications	Problem Set 2c	Review and Questions		—
11	20 Mar	—	—	—		Individual Assignment Deadline 24 March

5 Tutorials and GTA Support

You are expected to have covered the material ahead of the tutorials. There are two weekly tutorial classes delivered by the course GTAs, starting in week 3.

5.1 Analytical Tutorials

The classes are arranged to practice analytical problem sets. The first three weeks provide a brief summary of matrix calculus and statistical inference:

- Hadi Movaghari
- TA Office Hours

5.2 Software Labs

The classes are arranged to build up computational foundations to work with data and methodological frameworks:

- Tongtong Wang
- TA Office Hours Fridays 1-2pm

6 Computations

6.1 Financial Datasets and Empirical Exercises

The course contents, practice problem sets and assessment components are based on real-world financial data. It is a requirement that all class participants set up their accounts with the data platforms described below:

- Register your accounts on Financial Analysis Made Easy (FAME) via the university library and additionally Wharton Research Data Services directly on their platform using the university email address.
- This registration is then activated by the business database administration within one week. Please initiate the registration in the first week of the course before we progress towards further course contents and assignments.
- Key statistics and learning outcomes arising from the activities related to the data will be part of the exam. Treat the empirical exercises as an essential part of the learning experience
- As a financial analyst or a research financial economist, you will work with the very same data providers repeatedly. Developing an understanding of the empirical counterparts of theories will be an important takeaway for future careers in finance.

Platform	Reference
FAME	gla.ac.uk/myglasgow/library/specificsearch/databasesbyname/
WRDS	wrds-www.wharton.upenn.edu/register/
Bank of England:	bankofengland.co.uk/boeapps/database/default.asp
Office of National Statistics	ons.gov.uk/
Companies House	gov.uk/get-information-about-a-company

6.2 Software Packages and Implementations

Computational and methodological frameworks are implemented in Matlab. An additional spreadsheet is needed for supplementary data transformation and visual inspection, e.g. Libre Office, AWK or Excel (with Analysis ToolPak and Solver Add-in packages enabled). Please make sure you have set up both packages during the first week of the course to be able to practice exercises, replicate examples and complete assignments.

6.3 Computational Requirements

All course material and exercises are designed such that the learning outcomes are achieved based on any computer. However, you may also prefer to consider exploring the following available options to enhance computational capacity and further familiarising yourselves with professional computing systems.[1]

7 Problem Sets

There will be two practice problem sets with answers during the course.

[1] Google Cloud: Machine template is KX8765D, you will need to set up a new machine following the template ID which provides limited free service for the purpose of the class exercises.

8 Assessments

Assessment	Weight
Quiz	15%
Term Project (Group Assignment)	25%
Final Exam (Individual Assignment	60%

8.1 Quiz (15%)

This is an individual assessment including multiple-choice questions and covering course contents during the first four weeks including methodological learning outcomes, key facts and statistics, and empirical exercises. The quiz takes place on 30 January (duration of the quiz is 60 mins) - the assessment is accessible via Moodle at 9am and is required to be started within a 24 hours window.

8.2 Term Project (25%)

This assessment is completed by groups including a problem sheet requiring methodological derivations and empirical exercises followed by an interpretation of results. The problem sheet will be posted within two weeks prior to deadline and will be due according to the timetable above.

8.3 Final Exam (60%)

his is an individual assessment including a problem sheet requiring methodological derivations, and numerical computations followed by interpretation of results based on the course contents throughout the semester. he problem sheet will be posted within two weeks prior to deadline and will be due according to the timetable above.

Feedback Answers to the assignments will be provided in the subsequent week after the deadline and after everyone's submissions are received. Aside from the assessed assignments indicated above, the course includes two practice problem sets with solutions. These are distributed to practice theories and implementations during the semester. Students are expected to attend the office hours and tutorial workshops for reviewing specific queries.

Grading Policies: Grading is based on meeting the course intended learning outcomes examined in each assignment and following the University's Schedule A. Grades are rewarded based on both the input and output presented in each part thus demonstrating intermediate steps building up towards an overall answer are required and graded. Problem set and assignments require accessing real-world financial data from the professional platforms, thus class participants are required to register and activate their accounts with data providers by following the information provided.

Past Papers Past exam papers are available via the university portal. These can serve as a basis for preparation, however, note that the exam and course contents are subject to changes on an annual basis.

9 Reading List

There is no single textbook that covers all course material, thus selected chapters from the following references are assigned to each lecture topic. All reading items are required readings and considered

examinable material. The library has multiple copies of the main textbooks and provides online reading access:

9.1 Textbooks

1. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, *by Trevor Hastie, Robert Tibshirani, Jerome Friedman* (Second Edition)
2. Applied Data Science, *by Martin Braschler, Thilo Stadelmann, Kurt Stockinger*
3. *Profit-Driven Business Analytics: A Practitioner's Guide to Transforming Big Data into Added Value*, *by Wouter Verbeke, Bart Baesens, Cristián Bravo*: https://go.exlibris.link/7VxtBrRr
4. Applied Statistics I: basic bivariate techniques, *by Rebecca Warner*: https://go.exlibris.link/TzYSCMHg

9.2 Software References

- A handout is distributed at the first week
- MATLAB: a practical introduction to programming and problem solving, *by Stormy Attaway*: Online version available via the university library

9.3 Journal Articles & Supplementary Readings

Further to the textbooks, there will be journal article readings cited throughout the course. Journal articles indicated as 'required reading' should also be studied in conjunction with textbook reading and form part of the assessments:

- Longitudinal firm-level data: problems and solutions, by Karen Geurts (2016)
- Data Science and its Relationship to Big Data and Data-Driven Decision Making, by Foster Provost and Tom Fawcett, https://doi.org/10.1089/big.2013.1508
- Financial constraints, investment, and the value of cash holding, Denis, Sibilkov, The Review of financial studies
- Measuring economic growth from outer space, Henderson, Storeygard, and Weil, American economic review
- Endogeneity in Corporate Finance, Parts 1 & 2 (pages 494-511) by Roberts and Whited, Handbook of the Economics of Finance,
- Experimental Set-up and Ideals, Mastering 'Metrics: The Path from Cause to Effect, Angrist and Pischke (2014)
- Cash Holding and Credit Risk, Acharya et. al, Review of Financial Studies, 2012 (Introduction, Empirical Analysis, Concluding Remarks)