



National Research University Higher School of Economics
University of London Parallel Degree Program

Undergraduate Program in International Relations

Data Analysis in Economics and Finance Syllabus (Fall 2019)

Abstract

In this finance-oriented intermediate R course, you will learn how to apply logistic regression to a real-world financial data and how to construct and backtest an optimal investment portfolio. By the end of the course, you will be familiar with the basics of manipulating financial datasets to perform predictive analytics in R.

Pre-requisites

- basic courses in math, statistics and probability theory;
- basic knowledge of RStudio and *tidyverse* and *ggplot2* R packages;
- basic course in economics.

Course objectives

To provide an introduction to applications of R in finance and enable students to carry out a financial research in a reproducible fashion.

Learning Outcomes

- application of logistic regression to classification problems;
- mastering R packages for empirical finance (*quantmod*, *PerformanceAnalytics*, etc.);
- application of efficient financial data wrangling algorithms;
- computing optimal asset allocations based on modern approaches.

Duration

Fall 2019 (Module 1).

Course outline

Week 1. Review of the basic data manipulation and visualization R packages: *tidyverse*, *ggplot2*. Summary statistics of a dataset, basics of linear regression models.

Week 2. Introduction to logistic regression. Maximum likelihood estimation. Evaluation of model significance. P-value, confidence intervals, pseudo-R-squared.

Week 3. Evaluation of model predictive accuracy. Contingency table. ROC – curve. Selecting an optimal separation threshold.

Week 4. Modeling banks' probability of default. Selecting an optimal set of explanatory variables. Out-of-sample verification of the model.

Week 5. Portfolio performance evaluation. Financial data sources. Performance metrics. Selection of portfolio benchmarks. Backtesting and its biases. Overfitting and p-hacking.

Week 6. Naïve 1/N portfolio. Selecting a universe of stocks. Downloading and transforming data. Simulating trading process. Benchmarking against S&P 500.

Week 7. Markowitz Portfolio Theory. Portfolio returns, covariance matrix, mean-variance analysis. The efficient frontier. Rolling covariations. Instability of the covariance matrix.

Required readings

1. Cady, F. (2017). The data science handbook. John Wiley & Sons.
2. Lantz, B. (2015). Machine learning with R. Packt Publishing Ltd.
3. Fabozzi, F. J., & Pachamanova, D. A. (2016). Portfolio construction and analytics. John Wiley & Sons.
4. Hitchner, J. R. (2010). Financial Valuation,+ Website: Applications and Models (Vol. 545). John Wiley & Sons.
5. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.

Optional readings

1. Markowitz, H. (1952). Portfolio selection. The journal of finance, 7(1), 77-91.
2. López de Prado, Marcos, Building Diversified Portfolios that Outperform Out-of-Sample (May 23, 2016). Journal of Portfolio Management, 2016, Forthcoming. Available at SSRN: <https://ssrn.com/abstract=2708678> or <http://dx.doi.org/10.2139/ssrn.2708678>
3. DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. The review of Financial studies, 22(5), 1915-1953.
4. Chang, A., & Li, P. (2015). Is economics research replicable? Sixty published papers from thirteen journals say 'usually not'. Available at SSRN 2669564.
5. Focardi, S. M., & Fabozzi, F. J. (2012). What's wrong with today's economics? The current crisis calls for an approach to economics rooted more on data than on rationality. The Journal of Portfolio Management, 38(3), 104-119.

Grading system

$0.25*(\text{in-class assignment 1}) + 0.25*(\text{in-class assignment 2}) + 0.5*(\text{group project})$.

In-class assignments (10 pts/each) – week 3, 6

An in-class assignment will be given twice during the course. In-class assignments are short problem sets that are to be solved in R. Each problem set concerns a particular topic. Problem set 1 deals with the basics of probability theory, problem set 2 involves working with ggplot2 package.

Sample problems:

- Variable M contains a sample of 100 integers in the range of 1 to 1000. What is the probability that a randomly picked number from this population is greater than 100, but less than 400?
- Suppose you are a junior researcher in the World Bank, and you have been asked to model a relationship between unemployment rate (expressed as % of total labor force) and inflation (expressed as an annual % change of consumer prices). Estimate a linear regression between inflation and unemployment rate in 2000 (assume that inflation is the dependent variable) and use ggplot2 to plot data points along with the regression line. Is the model significant?

Group project (10 pts)

Maximum group size: 3 students.

Group project evaluation criteria:

- the purpose of the study is clearly stated (1 point);
- all steps of the research process are described in a clear and concise way (2 points);
- research outcomes are clearly defined (2 points);
- includes intuitive visualizations of research outcomes (2 points);
- all members of the project team are able to explain the code used for computations (1 point);
- code is properly structured (1 point);
- meets submission timeline (1 point).

Methods of instruction

The course is delivered via seminars. Material for this course will be presented using simulations in R and demonstrations of the source code.

Special Equipment and Software Support

R, RStudio

Instructors: Mikhail Vladimirovich Kamrotov (kamrotov@gmail.com)

Office hours: by appointment

Classroom policies:

- Hand-in assignments policy: All home assignments should be submitted electronically via instructor's email on the due date. No deadline extensions are possible.
- Cheating policy: In case of any kind of plagiarism (with the detected source), the assignment is evaluated as zero without the chance to make up for it. In case of two written assignments with the similarity index of 50% and higher from two students, both get a zero for the assignment.