Advanced statistical methods for Finance

(FINA0063)

Data analysis project & assessment

Academic year 2023 - 2024

1 Guidelines

The key learning outcome of this project (to be done by group of 2) consists in conducting a statistical analysis of financial data, with the help of the software R, and with the goal to solve or discuss a scientifically motivated research problem. Send your group composition by November 3rd.

At the end of the project, you will be expected to hand in a report (around 15 pages max) detailing your analysis. The date for sending the final report is January 5th. The report is graded and counts for 50% of the final grade.

You are expected to present your analysis during a seminar session, taking place on December 13th, from 1 pm to 5 pm. Your presentation should last between 15 and 20 min, and will be followed by 5 to 10 min of questions from the audience. The quality of your presentation, your answers to the questions and your overall participation will be graded. The oral presentation counts for 50% of the final grade.

To check if your topic is correctly defined, you have the opportunity to pitch your project (5 minutes) either on November 17th or November 24th, briefly explaining: your research question, the data you want to analyse and the model you intend to use. As a follow up, you can meet me once early December for a personalized meeting to discuss about the progresses of your project and receive an intermediate feedback. Take an appointment by email at jhambuckers@uliege.be.

2 Requirements

Your data analysis and your report are expected to meet the following requirements:

- Identify and detail a research question in the field of finance (e.g. related to corporate finance, asset pricing, risk management, macro-finance, financial forecasting etc.) that requires the analysis of an accessible dataset. The research question must be motivated/supported by 3 to 5 articles from peer-reviewed academic journals. **(2 pages)**

- The data analysis needs to make use of some advanced statistical/computational concepts seen in class, or exposed in the learning materials. Examples include binomial regression, Poisson regression, penalized estimation techniques (e.g. LASSO), volatility models, neural networks, simulation techniques or exteme value models.

- Access to the dataset needed for the completion of the project. Present its distinctive characteristics in light of your literature review. Illustrate these features with the help of graphs and descriptive statistics **(1 to 2 pages)**.

- Based on the characteristics of your data and your research question, motivate the advanced statistical method that you will use, respecting the mathematical conventions established during the course. In particular, link the financial/economic quantities of interest with the different aspects of the statistical model. Explain how the statistical analysis is useful to answer to the research question. **(2 to 3 pages)**.

- Conduct the data analysis. To do so, estimate the chosen model, and report the obtained results, also using tables and figures. If needed, conduct inferential procedures (e.g. hypothesis tests). Provide a financial/economic interpretation of your results in light of your research question. Stress the robustness of your results by using variants of your baseline model, subsets of your data or of time periods, etc. Discuss the limitations of your modelling approach and its correctness (*is the model well specified?*). Briefly discuss future work that could help improving your analysis. **(around 7 pages)**.

- Save all your code in a script file. Comment it appropriately. It must be possible for an external reader to run your code without your support. Then, compile all the needed files in a `.zip` or `.rar` archive file and send it at jhambuckers@uliege.be and phubner@uliege.be. You are expected to use the software R, but if dully motivated, you may also use Python. Basic data manipulation may be done with Excel.

The number of pages is indicative of the relative weights given to each part in the final grade. The determination of the grade is based on the correctness and the clarity of the explanations, the quality of the motivation underlying the project (both academically and from a logical standpoint), and finally from its technicality. The quality and clarity of the programming will also be a factor at play.

3 Topic by default: mutual funds out-of-sample performance and machine learning

As a default topic, you are expected to conduct a **mutual fund selection and out-of-sample analysis** similar to the one conducted in Kaniel et al. [2023]. Similarly to the steps outlined above, you are then expected to:

- Detail your research question (related to the interest of using ML methods for prediction of mutual fund performance). Motivate it by arguments developed in Kaniel et al. [2023] and at least 2 other papers of your choice. The general idea is to use ML methods to decide on an investment strategy between different mutual funds, in hope of obtaining a good financial performance (e.g. in terms of alpha of an asset pricing model).

- Using the Morning Star database at HEC Liege, access to the returns history of mutual funds. Contact `phubner@uliege.be` to obtain access to the database. Following the procedure detailed in Kaniel et al. [2023], compute the excess (or abnormal) returns of the funds.

- Access to a dataset of at least 15 candidate covariates that would help predicting excess funds' returns. Consider macro-financial variables, such as interest rates, volatility indicators but also funds' characteristics as advocated in Kaniel et al. [2023]. Holdings are usually not available to us. Thus, you are not expected to use thi information in your prediction exercise.

- Lagged values of these predictors are used to predict future excess returns (i.e. for predicting excess return $\mathbb{E}(r_t - r_t^f)$, only x_{t-j}, $j > 0$, is used).

- Illustrate the features of the data (variations over time, dependencies, correlations,...) with graphs and descriptive statistics.

- Explain and motivate the advanced statistical method that you will use. As methods, you have to combine some of the following approaches

  - Penalization methods to account for a large number of covariates.

- Nonlinear regression techniques such as neural network, regression trees, random forest, to predict excess returns.

You can use other methods in addition, if needed.

- Conduct your performance analysis, and out-of-sample prediction exercise, following a similar setting as the one outlined in Kaniel et al. [2023]. Rem: be pragmatic. You are not expected to fulfill every step. However, you are expected to stay as close as possible from their study.
- Conduct also a similar exercise with a linear regression model. Use this model to benchmark the results obtained with the more sophisticated model.
- Illustrate the obtained performance using graphics and summary statistics of performance measures. Stress the robustness of your results. Discuss the limitation of your modelling approach. Briefly discuss future work that could improve the performance of your investment strategy.
- Save your code in a script file. Comment it appropriately. It must be possible for an external reader to run your code without your support. Then, compile all the needed files in a `.zip` or `.rar` archive file and send it at `jhambuckers@uliege.be` and `phubner@uliege.be`.

References

Ron Kaniel, Zihan Lin, Markus Pelger, and Stijn Van Nieuwerburgh. Machine-learning the skill of mutual fund managers. *Journal of Financial Economics*, 150(1):94–138, 2023. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2023.07.004. URL `https://www.sciencedirect.com/science/article/pii/S0304405X23001253`.