

MAFS6010Z: Project2 Review for Group 16

Liu Enjie eliuac@connect.ust.hk (20745838)

● Summary of the report:

This report provides a detailed introduction to their replication of the paper *Empirical Asset Pricing via Machine Learning*. (1) Data Processing: The report mainly focused on the handling of missing values: they filled the NA with the cross-sectional median at each month for each stock, respectively. (2) Variable: The 94 stock features and 8 macroeconomic factors were used in their model analysis separately, which means that the interaction between variables was discarded due to some computational constraint. (3) Model fitting and analysis: Six models were used to fit the data which are OLS, OLS-3, PLS, PCR, ENet, RF and GBRT. By calculating R^2_{oos} through recursive evaluation method, the different performance of all 6 models were visually presented. Feature importance, calculated by sum of squared partial derivatives (SSD), were also visually showed in charts and tables.

● The strengths of the report:

1. Detailed report writing: Each step is detailly explained in concise writing, and results can be easily understood by the reader through tables and diagrams.

- **The weaknesses of the report:**

1. Technical defects: It is not desirable to ignore the interaction between independent variables. I think the problem of computing resources and running speed can be improved by variable selection of 94 stock characteristics before calculating interaction.
2. Some errors and lack of clarity part in report writing: There are some confusing points in the report. (1) For example, “*Percentage of NaN values in application train*” is the title of Figure1, which actually shows the model performance. (2) And in part 4.2, “*Because of huge computation time, here we decided to evaluate the final test period only, instead of testing all 30 rolling periods*”, I wonder if this means using the same year as the testing set all the time. If so, I think it is also a technical flaw.

- **Score**

Evaluation on Clarity and quality of writing: Score: 4

Evaluation on Technical Quality Score: 3

Overall rating: Score: 3.5

Confidence on your assessment: Score: 3