Rebuttal 2 Group 14

1. The outcome of the report was different to the paper, the report did not clearly mention the reason why nor did they attempt to replicate the processing and calculations of the data and see if the output will be more similar to the paper.

A :

Actually we have attempt to replicate the processing and calculations of the data but we found that it took more than a dozen hours to run the code of just reading and processing data and even worse in run some ML model (The Python server died after spending a whole day running models).

In this case, we chose to make some adjustments to the data. In part 3 of our poster, we also stated that "Considering computer memory and operating speed, we made some adjustments by selecting features at first and then creating interactive features".

What' s more, we thought it was more important to learn some data processing techniques and apply machine learning models to solve problems than to copy the results of the report completely.

2. The report is too short to introduce more details of data processing and parameter selection of various models.

A :

(1) Data processing: (i). Eliminate some variables which do not appear in the paper and then we get 94 stock features and 8 macro features. (ii). Use one-hot encoding for nominal variables such as sci2. (iii). Dropped data without price and filled missing value with 0.

(2) Parameter selection: Referring to the parameter combinations given in the Appendix E (given in the table below), we tried all parameter combinations and used R^2oos to judge model performance in different parameter combinations. Since our assignment does not require the *section2.4 Portfolio forecast*, we did not write out the optimal combination of parameters in poster.

Table A.5: Hyperparameters For All Methods

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1 - NN5
Huber loss $\xi =$ 99.9% quantile	✓	-	-	✓	✓	-	✓	-
Others		K	K	$\rho = 0.5$ $\lambda \in (10^{-4}, 10^{-1})$	#Knots=3 $\lambda \in (10^{-4}, 10^{-1})$	Depth= $1 \sim 6$ #Trees= 300 #Features in each split $\in \{3, 5, 10, 20, 30, 50...\}$	Depth= $1 \sim 2$ #Trees= $1 \sim 1000$ Learning Rate LR$\in \{0.01, 0.1\}$	L1 penalty $\lambda_1 \in (10^{-5}, 10^{-3})$ Learning Rate LR$\in \{0.001, 0.01\}$ Batch Size=10000 Epochs=100 Patience=5 Adam Para.=Default Ensemble=10

Note: The table describes the hyperparameters that we tune in each machine learning method.

We have to admit that after choosing poster as our presentation method, it is indeed a problem that the poster is too short to show more details. To remedy this problem, we annotated our code in as much detail as possible.

3. Didn't calculate the out-of-sample R2 for top market-value stocks and the bottom ones which is required in the ppt that teaching assistant gave. Feature selection process has space to be improved. Maybe could do more than just calculating Pearson correlation coefficient.

   A：

   (1) The out-of-sample R2 for top market-value stocks and the bottom ones: We have to admit our mistake of thinking this part as not mandatory.

   (2) Feature selection process has space to be improved: Thank you for your suggestion. We will try more feature selection methods in the next assignment.

4. 因为计算机 **memory** 限制，采取实时的构建 **interaction feature** 的方式，很好的做法；对有显著性预测作用的 **feature** 进行归类，很好；没解释 **NN2** 与 **NN5** 差别大的原因。

   A：

   The reason for the large difference between NN2 and NN5： Let's observe the performance of NN1~NN5 in our replication study (see in the below chart), and we can see that the R^2_{oos} increased as neural networks become more and more complex. (NN5>NN4>NN3>NN2>NN1), which is also different from the result in paper (section 2.2: *"The results also show that in the monthly return setting, the benefits of "deep" learning are limited, as four- and five-layer models fail to improve over NN3.33"*). Why dose this happen in our study? We consider the following reason:

   Due to the selection of variables, the number of variables in our study is greatly reduced compared with the original paper, which is 30 stock features and 94 stock features, respectively. It is the fewer data and variables that makes the model perform better when it is complex.

And this is also the reason for the larger difference between NN2 and NN5: With hidden layers and neurons increased, the model tend to perform well significantly.

model	R^2	model	R^2
PLS	-0.274	NN1	-39.996
PCR	-0.058	NN2	-19.373
Elastic Net	0.00299	NN3	-1.381
RF	0.00378	NN4	-0.833
GBRT	0.00391	NN5	0.0035