

Project 2 Rebuttal

Group 16, **TANG Tsz Hong, LAM Chung Wai, CHAN Koon Lam**

For the PhD reviewer:

They argued that we may use future data. However, we followed the original paper and use cross-sectional median to fill the NA values. The part regarding NA value in the report is some extra information that shows the proportion of missing values, or shows some common mistakes to be aware of when we fill NA values. Regarding PhD reviewer stated the result may be too good, table 1 is constructed in percentage R^2_{oos} , so the mean prediction performance is at most 4.851%, which is reasonable.

For reviewer 1, (LI Mingluo):

They argued that we might have some problem in the calculation of the Dividend price ratio and Earnings price ratio, stating that we should use the difference between the log of column D12 and log of index price as Dividend price ratio, and the difference between the log of column E12 and log of index price as Earnings price ratio. However, in our dataset the column D12 and E12 are already handled as dividend price ratio and earnings price ratio.

For reviewer 2, (LIU Enjie):

They argued that we use same year as the testing set all the time stated in our report. It is not true. The line in report "we decided to evaluate the final test period only" is indeed a typo, it was suggested in our draft report, and it should be deleted in the final report. In the final submission, all 30 rolling periods are used as test, valid and training set, you can check that in the .ipynb file. On the other hand, the title of Figure 1 is a typo, thank you for spotting out. We will put effort to reduce typo in the final project.

For reviewer 3, (XU Tongcan):

They argued that it might not be the most appropriate way to fill all missing values with the median for independent variables with high proportion of missing values. However, this method of replacing missing characteristics with the cross-sectional median is suggested by Gu et al. in the paper.