Results:

The Results can be differentiated into stock return predictions and portfolio return predictions.

First we look at the monthly out-of-sample stock-level prediction performance (Percentage R2oos) comparing linear machine learning methods (OLS, PLS, PCR, etc.) and non-linear methods (Neural Networks, elastic nets, etc.). The best performing method throughout the method landscape are the Neural Networks, peaking at a R2oos of 0,40%. They are followed by Trees and elastic nets with a R2oos of 0,34%. At the end of the field are linear methods with R2oos ranging from -3,46% to 0,27%. To substantiate these results the fit of each method was tested by applying the respective method to the top or bottom 1000 stocks. Results show the methods behaved in a similar way.

Looking at the annual horizon reinforces the results and ranking above. It shows that machine learning methods are also able to predict on a longer timeline rather than just capitalizing short-term inefficiencies.

Another topic to look at are the influence of covariates and macroeconomical variables.

To evaluate the importance and influence of individual characteristics, said values were set to 0 and their impact on R2 was measured per method. Generally, all models were relatively close regarding the most important predictor characteristics. It should be noted that most of the top characteristics are based on recent price trends. It is interesting to see that linear models are more influenced by momentum and reversals, whereas non-linear models take into account a broader set of characteristics. The analysis furthermore proved that the most influential characteristics were unaffected by noise, thus confirming the robustness of the results. This noise was simulated using “placebo”-characteristics.

In evaluation of macroeconomic variables, the analysis yields that all methods treat the aggregate book-to-market factor as crucial, while market volatility is relatively insignificant. When cascading down the different categories of methods it is to be noted that (generalized/penalized) linear methods favor bond market variables, whereas non-linear methods emphasize exactly those variables ignored by linear methods.

Next, we have a look at portfolio return predictions. Here, we differentiate between the forecast of pre-specified portfolios (e.g.: S&P 500) and machine learning portfolios.

The Reasoning to look at forecasting from a portfolio perspective is that this approach shows the robustness of the used models on a larger scale and offers a form of evaluation of mentioned models.

Regarding pre-specified portfolios the analysis confirmed our result from the individual stock perspective. Non-linear model perform the best and are the most consistent with neural networks being the best overall.

With regards to machine learning portfolios, let’s first have a look on how they were constructed. The machine learning portfolios are put together as follows: The stock return is predicted one-month-ahead out-of-sample. The portfolio is then constructed by buying the highest expected return stocks (decile 10) and selling the lowest expected return stocks (decile 1) on a monthly basis.

The data showed that, once again, neural networks dominate over trees and linear models. The main reason for this their very high accuracy of the predicted returns.

Concluding on above results...