Paper Replication Report
- Empirical Asset Pricing via Machine Learning
by S. Gu, B. Kelly, D. Xiu

Hong Kong University of Science and Technology
MAFS6010Z- Artificial Intelligent Fintech
Project 2- Paper I

Group members	Student Number	Contribution
JIA Yaoyao	21027570	Write Neural Network Pytorch Demo; Run Random Forest, NN1 and NN4, complete relative work and drawings.
JIANG Xiaoyue	20987163	Run XGBoost, NN2 and NN3 and complete other related works like analyzing and drawing.
YANG Tianhao	21027489	Complete all data cleaning work, Run PLS and PCR and complete relative works, and complete model comparison
HUANG Yuxuan	21030955	Write ElasticNet model and Generalized Linear Model, complete corresponding paper writing and drawings.

I. Introduction

This report aims to replicate the results of a paper "Empirical Asset Pricing via Machine Learning", written by Shihao Gu, Bryan Kelly, and Dacheng Xiu. The paper performed a comparative analysis of 13 modes among 5 types of machine learning methods to predict assets' excess return, also called asset risk premiums, including linear models, dimension reduction (PLS and PCR), Generalized linear model, tree models and neural networks. The author highlighted their contribution twofold. Firstly, the paper establishes a new set of benchmarks for the predictive accuracy of machine learning methods in measuring the excess returns. In particular, the paper magnifies profits that can be achieved in prediction, instead of the part dominated by unforecastable news, and finds out the most informative predictor variables. Secondly, the paper synthesizes the empirical asset pricing literature with the machine learning field. A comparative overview is provided, which of machine learning methods applied to the two canonical problems of empirical asset pricing that predict returns by cross-section data and time series data. It emphasizes three outstanding advantages by applying machine learning methods in financial fields, especially when facing problems associated with prediction tasks, huge size of correlated conditioning variables and ambiguity about the functional forms. Conducting a large-scale empirical analysis, investigating around 30,000 individual stocks over 60 years from 1957 to 2016, authors include more than 900 baseline signals, consisting 94 characteristics for each stock, interactions of each characteristic with eight aggregate time-series valuables, and 74 dummy variables of industry sector. The paper concludes that machine learning methods can help improve empirical understanding of asset pricing. Neural networks and regression trees perform best among 13 methods. Besides, "shallow" learning outperforms "deep" learning, caused by the comparative dearth of data and low signal-to-noise ratio in asset pricing problems. The paper affirms the important role of machine learning in the field of fintech industry. Although the paper includes the portfolio forecast part, we only focus on replicating machine learning methods part and variable importance part. In order to dive into the comparative analysis among different machine learning models, we attempt to replicate 6 out of 13 models mentioned in this paper by processing the same data set and conducting a similar methodology. Considering the poor performance of the OLS model, we choose to reproduce Elastic Net (penalized linear), generalized linear model, PCR, PSL, random forest and neural networks methods.

II. Data Analysis and Processing

A. Data Download

First, we download the stock characteristics file 'GKX_20201231.csv' and the macro factors file 'PredictorDate2022.xlsx', and merge the two files according to date as our initial data. Then we get 97 stock characteristics and 17 macro factors in all. The meaning of the factors is shown in the table below.

Details of the Characteristics		
No.	Acronym	Firm characteristic
1	absacc	Absolute accruals
2	acc	Working capital accruals
3	aeavol	Abnormal earnings announcement volume
4	age	# years since first Compustat coverage
5	agr	Asset growth

Details of the Characteristics		
No.	Acronym	Firm characteristic
6	baspread	Bid-ask spread
7	beta	Beta
8	betasq	Beta squared
9	bm	Book-to-market
10	bm_ia	Industry-adjusted book to market
11	cash	Cash holdings
12	cashdebt	cash flow to debt
13	cashpr	Cash productivity
14	cfp	cash flow to price ratio
15	cfp_ia	Industry-adjusted cash flow to price ratio
16	chatoia	Industry-adjusted change in asset turnover
17	chcsho	Change in shares outstanding
18	chempia	Industry-adjusted change in employees
19	chinv	Change in inventory
20	chmom	Change in 6-month momentum
21	chpmia	Industry-adjusted change in profit margin
22	chtx	Change in tax expense
23	cinvest	Corporate investment
24	convind	Convertible debt indicator
25	currat	Current ratio
26	depr	Depreciation / PP&E
27	divi	Dividend initiation
28	divo	Dividend omission
29	dolvol	Dollar trading volume
30	dy	Dividend to price
31	ear	Earnings announcement return
32	egr	Growth in common shareholder equity
33	ep	Earnings to price
34	gma	Gross profability
35	grCAPX	Growth in capital expenditures
36	grltnoa	Growth in long term net operating assets
37	herf	Industry sales concentration
38	hire	Employee growth rate
39	idiovol	Idiosyncratic return volatility
40	ill	Illiquidity
41	indmom	Industry momentum
42	invest	Capital expenditures and inventory
43	lev	Leverage Bhandari
44	lgr	Growth in long-term debt
45	maxret	Maximum daily return

Details of the Characteristics		
No.	Acronym	Firm characteristic
46	mom12m	12-month momentum
47	mom1m	1-month momentum
48	mom36m	36-month momentum
49	mom6m	6-month momentum
50	ms	Financial statement score
51	mvel1	Size
52	mve_ia	Industry-adjusted size
53	nincr	Number of earnings increases
54	operprof	Operating profability
55	orgcap	Organizational capital
56	pchcapx_ia	Industry adjusted % change in capital expenditures
57	pchcurrat	% change in current ratio
58	pchdepr	% change in depreciation
59	pchgm_pchsale	% change in gross margin - % change in sales
60	pchquick	% change in quick ratio
61	pchsale_pchinvt	% change in sales - % change in inventory
62	pchsale_pchrect	% change in sales - % change in A/R
63	pchsale_pchxsga	% change in sales - % change in SG&A
64	pchsaleinv	% change sales-to-inventory
65	pctacc	Percent accruals
66	pricedelay	Price delay
67	ps	Financial statements score
68	quick	Quick ratio
69	rd	R&D increase
70	rd_mve	R&D to market capitalization
71	re_sale	R&D to sales
72	realestate	Real estate holdings
73	retvol	Return volatility
74	roaq	Return on assets
75	roavol	Earnings volatility
76	roeq	Return on equity
77	roic	Return on invested capital
78	rsup	Revenue surprise
79	salecash	Sales to cash
80	saleinv	Sales to inventory
81	salerec	Sales to receivables
82	secured	Secured debt
83	securedind	Secured debt indicator
84	sgr	Sales growth
85	sin	Sin stocks
86	sp	Sales to price
87	std_dolvol	Volatility of liquidity (dollar trading volume)
88	std_turn	Volatility of liquidity (share turnover)
89	stdacc	Accrual volatility

Details of the Characteristics		
No.	Acronym	Firm characteristic
90	stdcf	Cash flow volatility
91	tang	Debt capacity/firm tangibility
92	tb	Tax income to book income
93	turn	Share turnover
94	zerotrade	zerotrade Zero trading days

B. Data Processing

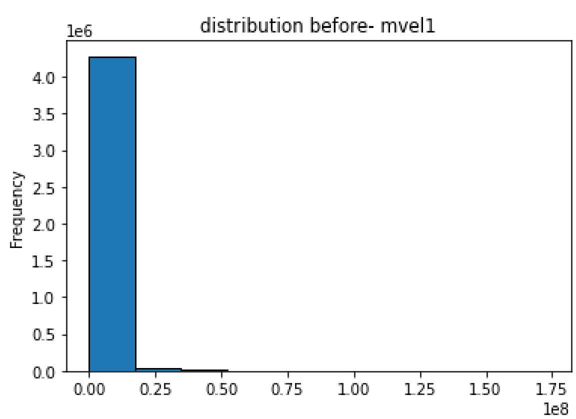
1. Delete Some Data

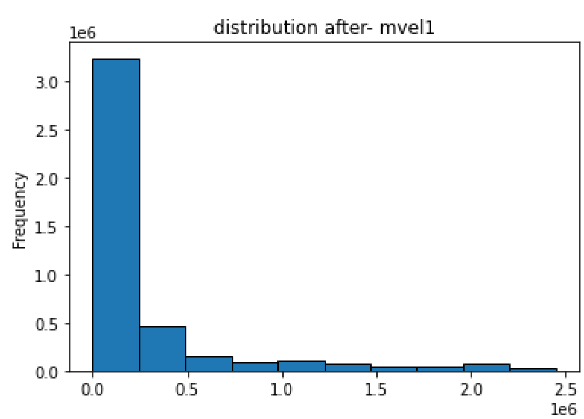
We noticed that the stock features used in the original paper did not include the four features of 'permno', 'SHROUT', 'mve0', and 'prc', and we had no way to find the meaning of these four characteristics. Since the subsequent data processing needs to be processed separately according to the meaning and characteristics of the data itself, taking into account the rationality of the data processing process, we finally decided to directly delete the data with these four characteristics.

2. Handling Extreme Value

Because we consider that there are too many missing data this time, processing the missing data first and then processing the extreme values of the data will have a greater impact on the original data distribution, which will bias our subsequent processing of extreme values, so our team decided that the extreme values in the data set should be processed first. Considering that 'sic2', 'convind' are dummy variables, and 'RET' is the label learned by our model, these 3 features are not considered in the extreme value processing process. When dealing with extreme values, we use the 3-MAD method. Since the sample is a time series and spans a large time range, the median of each feature is actually a time series data. Therefore, our team believes that the median of the entire sample should be directly used. The number of digits is unreasonable, which will bias the processing of extreme values. Therefore, in the actual processing process, we divide the initial data set by month, calculate the median of each feature for each month, and use the data of that month. 3-MAD method is used to remove extreme values.

The figure below takes the 'mvel1' data as an example





3. Handling Missing Value

Considering that the industry of the company, especially in the early stages, may change over time, we believe that the industry data of the company is more difficult to fill in if it is missing; at the same time, the stock return rate itself is the label we learn, so we finally remove all missing lines of 'sic2' and 'RET' directly. As for the beta value, we believe that it has a great correlation with the industry in which the company is located, and in most cases, the beta value of the company is generally predicted to be replaced by the average beta value of the industry in which the company is located. Therefore, We believe that the feature 'beta' and its associated 'betasq' need to be treated separately from other stock features. For other stock characteristics, we fill missing values using the mean of that feature. In the same way, the average value of each feature is actually a time series data. Therefore, our team believes that it is unreasonable to directly use the average value of the entire sample. Instead, we divided the initial data set by month. During the actual processing, we use the mean value of the current month to fill the missing values of the feature. If the feature data is missing values for the entire month, choose to fill it with the latest monthly mean value. Considering that the industry in which the company operates may change over time, if the beta value of the industry for an entire month is null, these rows will be deleted and the most recent monthly average will not be used for filling.

4. **Data Normalization**

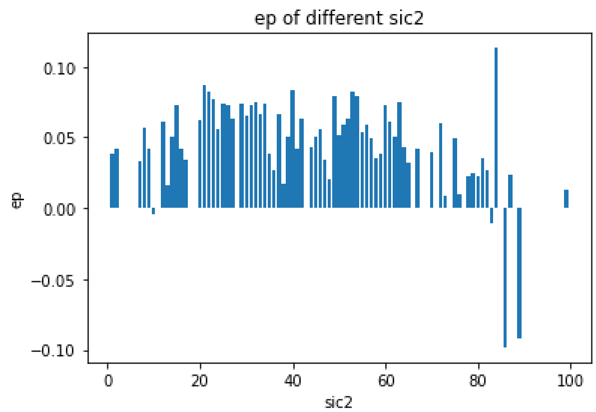
Normalization is generally required when we are dealing with attributes on a different scale, otherwise, it may lead to a dilution in effectiveness of equally important attributes(on a lower scale) because of other attributes having values on a larger scale. In simple words, when multiple attributes are there but attributes have values on different scales, this may lead to poor data models while performing data mining operations. So they are normalized to bring all the attributes on the same scale. We use the Z-score method to normalize all other features of the data except the two nominal variable columns 'sic2' and 'convind' and the final learned label 'RET' column.

5. **One-hot Encoding**

Among all nominal variables, 'convind' is a 0-1 variable, and the maximum value range of 'sic2' reaches 99. Since the number in 'sic2' only represents the industry in which the stock is located and does not represent the order, size or other meanings, we use one-hot encoding for it.

6. **Industry & Market Capitalization Neutralization**

When picking stocks based on equity factors, we attribute the returns of an individual stock to its exposure to a factor. However, there are some factors whose values have a certain correlation with the industry or market value of the stock. For example, when we choose stocks based on 'ep', we usually think that stocks with low valuations are cheaper and will have higher future returns, so we are more inclined to choose stocks with lower p/e values. However, through this method of stock selection, we found that most of the industries in which the stocks ultimately selected were banking or real estate, which means that there are differences in the average P/E ratios between different industries. Likewise, we are more likely to choose small-cap stocks over large-cap stocks because small-cap stocks have more room for growth and correspondingly higher expected returns. Therefore, if we select stocks based on non-neutral factors, the final selection will occur more among industries. If we want to select stocks within an industry, we need to exclude the industry and market value from the factors impact. At the same time, Industry & Market Capitalization neutralization of factors can also reduce the correlation between factors, thereby improving the accuracy of our return forecasts. In the actual cleaning process, we neutralized all stock characteristics except 'mvel1' (which is the market capitalization of the company's stock) to market capitalization and industry.



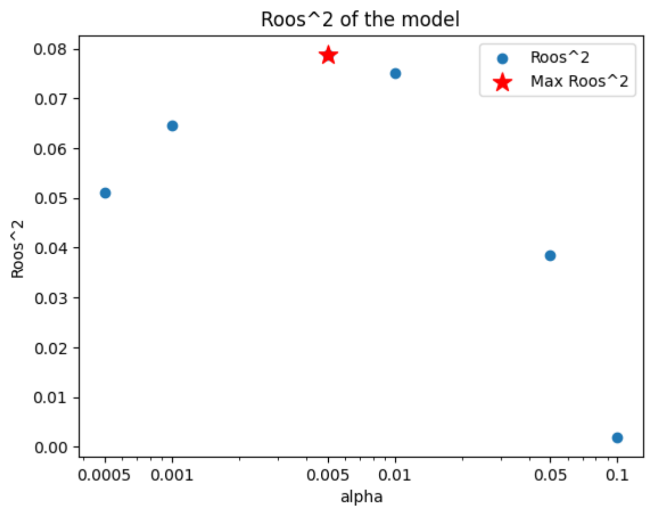
It can be seen that the difference in ep between different industries is very large, so it is necessary to neutralize it

The final cleaned data set ranges from 1985 to 2020, a total of 36 years. There are 186 features in total. The data set is then divided by time during training. The initial training set spans 13 years, from 1985 to 1997, the validation set spans 13 years, from 1998 to 2010, and the validation set spans 10 years, from 2011 to 2020. Each subsequent cycle increases the time span of the training set by 1 year, and the time span of the validation set remains unchanged until the time span of the test set reaches 1 year. 10 cycles in total.

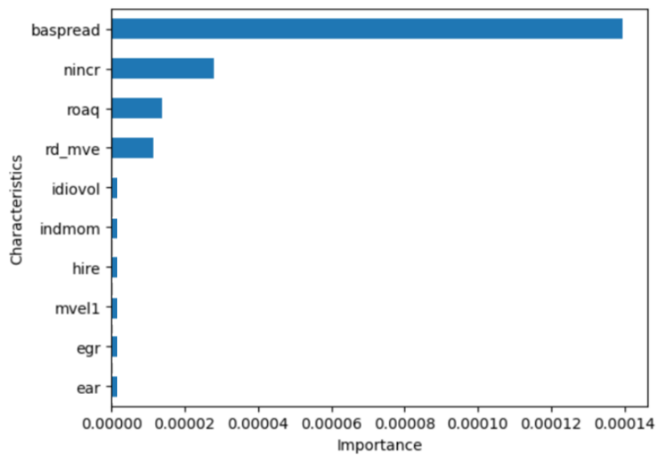
III. Model Replication

A. Penalized Linear - Elastic Net

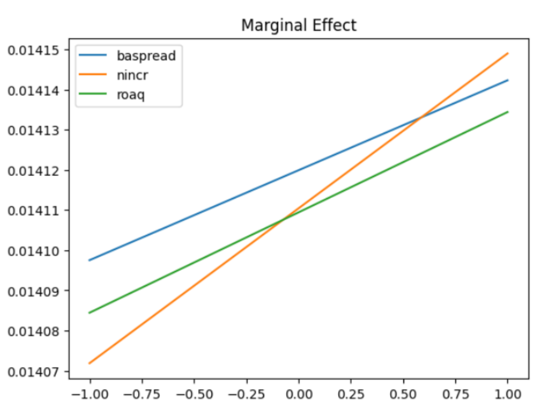
Considering that the simple linear model tends to outfit noise rather than extracting signals, we employed a penalized linear model. The difference between penalized linear model and simple linear model is to append a penalty to the objective function in order to impose parameter parsimony, in other words to reduce the number of estimated parameters. Following the paper, we also focus on the “elastic net” penalty, involving two nonnegative hyperparameters, alpha and l1_ratio. According to the paper supplement, we set l1_ratio as 0.5, and tuning alpha within [0.0005, 0.001, 0.005, 0.01, 0.05].
The output of tuning:
Best alpha: 0.005
Best l1_ratio: 0.5
Average R-squared: 0.07877294439374462 (validation set)

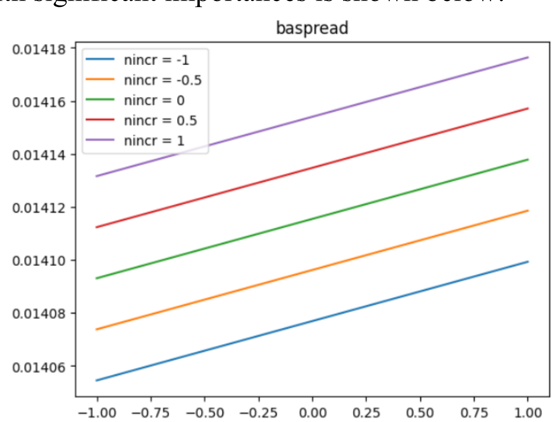


The test set Roos = 0.09531373.
Top 1000_Roos: 0.210843.   Bottom 1000_Roos: 0.032589
The top ten important variables are as followed:



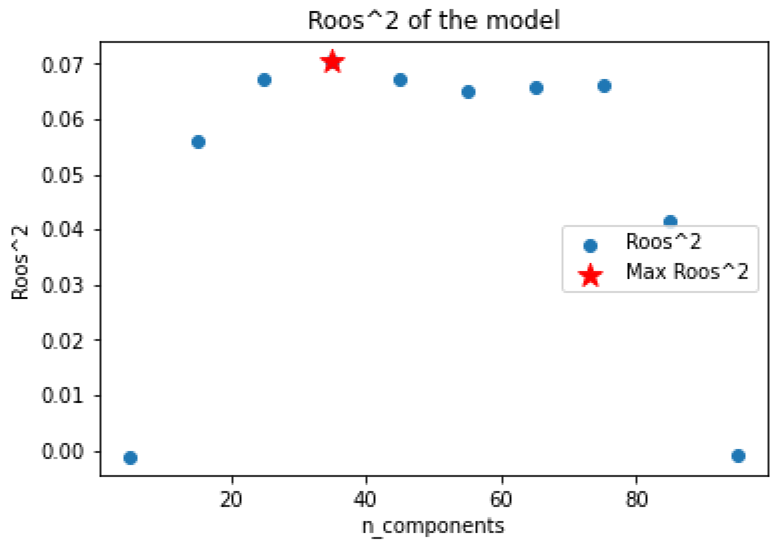
The marginal effect of three particular variables with significant importances is shown below:





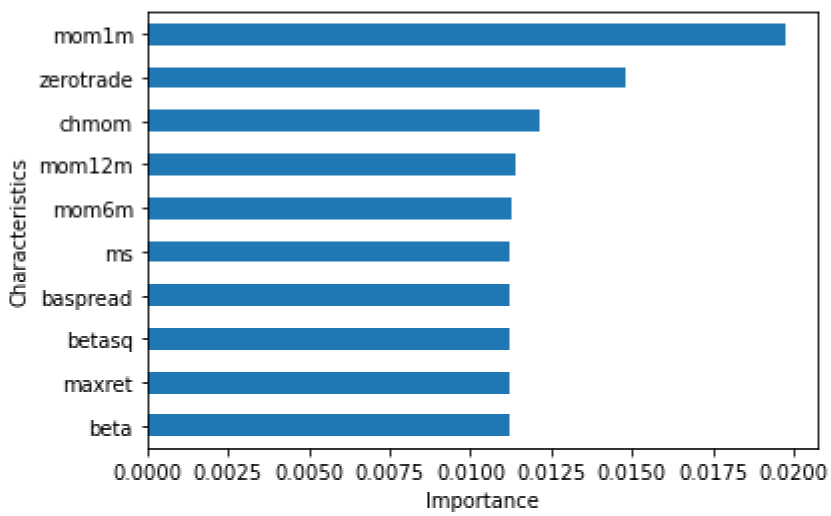
B. Dimension reduction: PCR

Set the value range of parameter n_components between 5 and 95, and calculate the R-squared on the validation set for parameter selection. The final optimal parameter is n_components = 35.



The Roos^2 of the model on the test set is 0.0773. For the data of the top 1,000 stocks by market capitalization each month, the Roos^2 of the model is 0.239. And for the data of the bottom 1,000 stocks by market capitalization each month, the Roos^2 of the model is 0.0113.

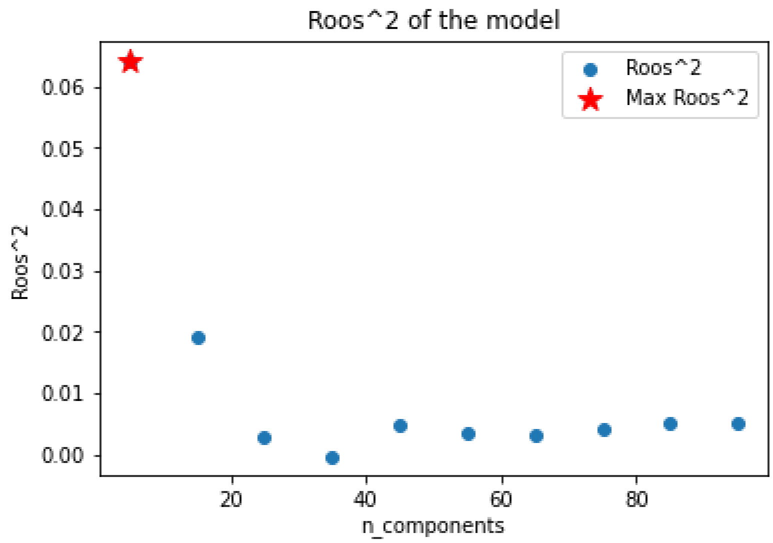
The top ten important variables in the model and their importance are shown in the following figure:



Due to the characteristics of the PCR model, setting many data to constant values will have a greater impact on model predictions, so there is no way to draw the marginal impact of variables on expected returns.

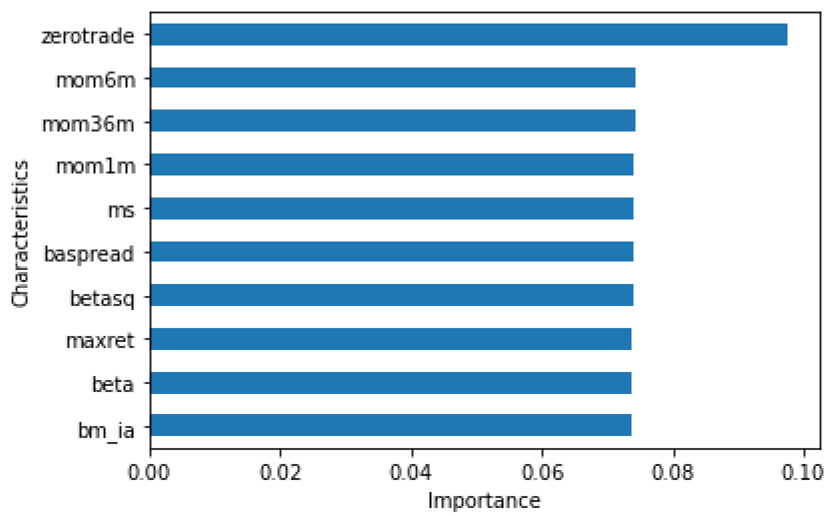
C. Dimension reduction: PLS

Set the value range of parameter n_components between 5 and 95, and calculate the R-squared on the validation set for parameter selection. The final optimal parameter is n_components = 5.

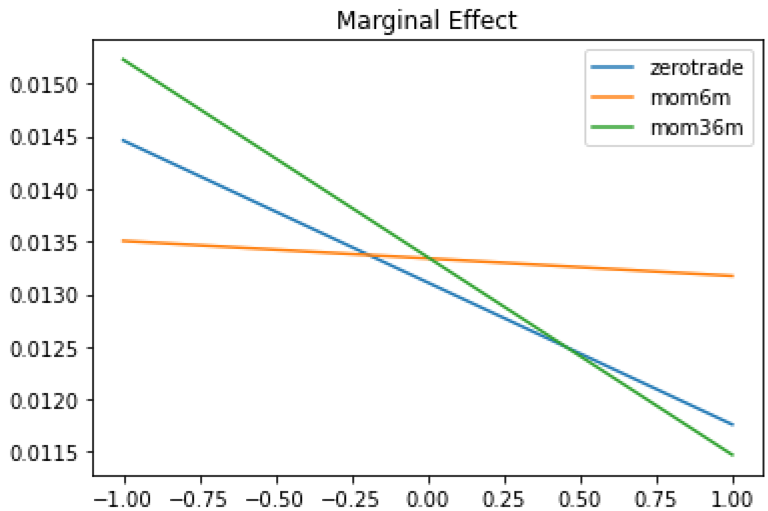


The Roos^2 of the model on the test set is 0.068. For the data of the top 1,000 stocks by market capitalization each month, the Roos^2 of the model is 0.102. And for the data of the bottom 1,000 stocks by market capitalization each month, the Roos^2 of the model is -0.416.

The top ten important variables in the model and their importance are shown in the following figure:

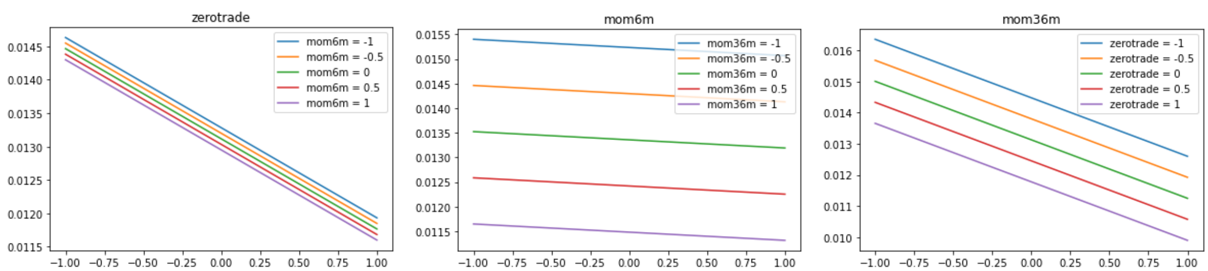


For the top 3 important variables, we examine their marginal impact on yield prediction.



It can be seen that all three factors have a negative impact on stock returns. Since the model is linear, the impact on expected returns is also linear.

Now examine the impact of the interaction between variables on expected returns。



As can be seen from the above figure, as the two variables decrease at the same time, the expected return will also decrease. This is mainly because the prediction model is a linear model and there are no interaction terms between variables.

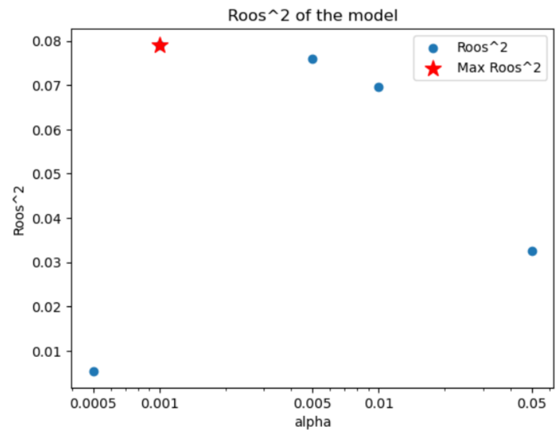
D. Generalized linear

Given the "true" model is complex and may not be a first-order approximation, the paper introduces the generalized linear model. The GL model applies nonlinear transformations of the original predictions as new additive terms based on the linear model. According to the appendix, we adapt a 3-term spline series expansion of the predictors, which are (1, z, (z-c_1)^2), where c_1 are knots. We set c_1 as *min(z)+[max(z)-min(z)]/K*. We conduct a least squares objective function with the Huber robustness modification. With the Huber robustness regression, the only hyperparameter needed tuning is the alpha, the penalty coefficient. The range we set is [0.0005,0.001,0.005,0.01,0.05].

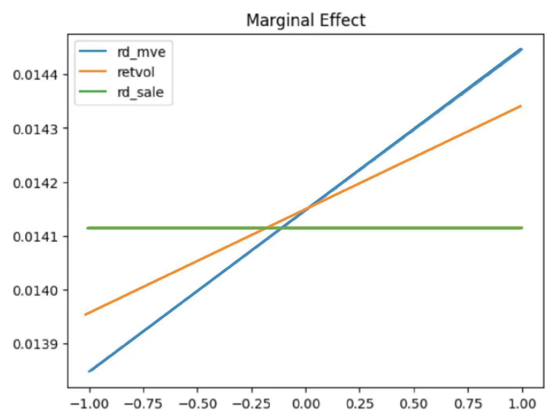
The output of tuning:

Best alpha: 0.001

Average R-squared: 0.079054637 (validation set)



The test set Roos = 0.078957324
Top 1000_Roos: 0.1784032.   Bottom 1000_Roos: 0.0253489
The marginal effect of three particular variables with significant importances is shown below:

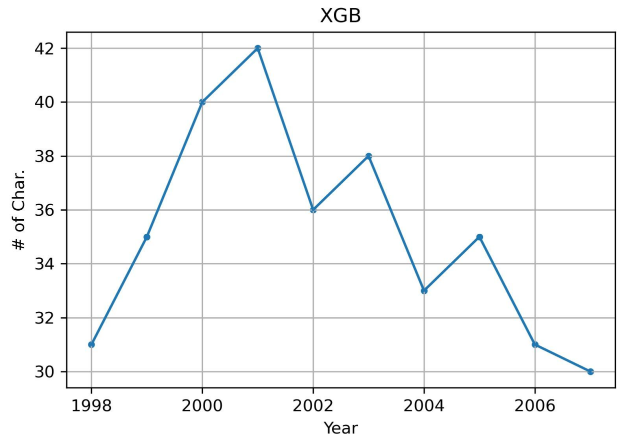


E. XGBoost

XGB_Roos: 0.077743
(parameters: learning_rate=0.01, max_depth=3, n_estimators=200)
Top 1000_Roos: 0.204844, Bottom 1000_Roos: 0.037181

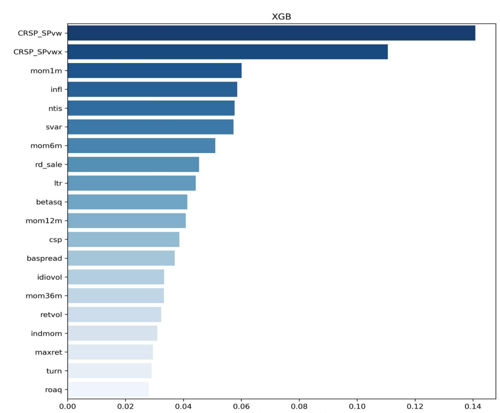
The process of choosing parameters
According to the paper supplement, I establish the range of parameters, learning_rate=[0.01, 0.1, 0.2], max_depth=[3, 4, 5] and n_estimators=[100, 200, 500]. After training models with different time divisions, I finally choose the parameter combination (learning_rate=0.01, max_depth=3, n_estimators=200) with the best validation Roos.

The figure of time-varying model complexity shows the number of distinct characteristics entering into the trees. In 2001, the number of selected features reached the maximum value and after 2003, the number of selected features tended to decrease.



Time-varying model complexity

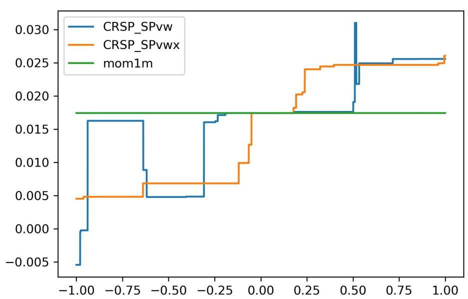
The following figure demonstrates the variable importance for the top-20 most influential variables in this model. And the features importance is normalized to sum to one.

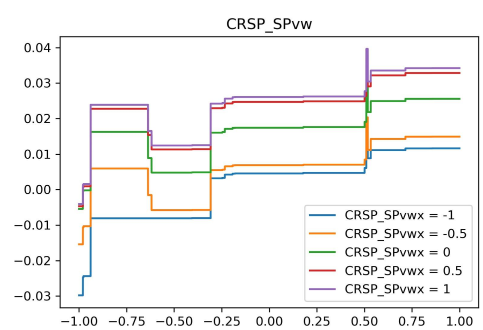


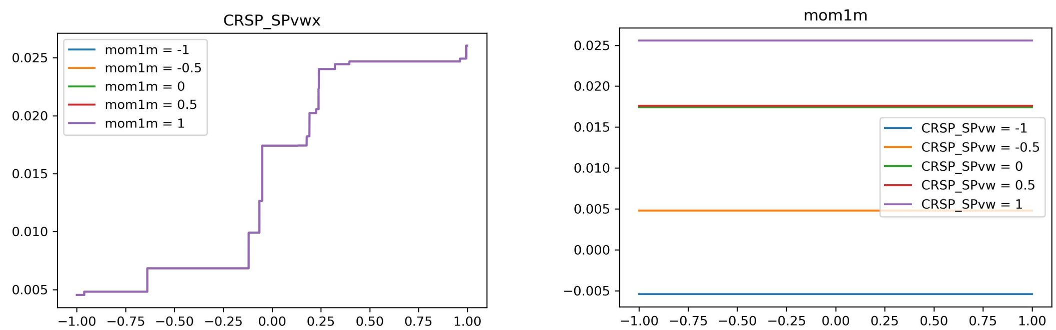
c

Features importance by model

I take three features (CRSP_SPvw, CRSP_SPvwx, mom1m) which are the most influential in this model and investigate the marginal association between the characteristics and their interaction effects.







Marginal association between expected returns and characteristics (XGBoost) & Expected returns and characteristic interactions (XGBoost)
The above figures report the variation of expected returns with the variation of a pair of characteristics over their support [-1,1].

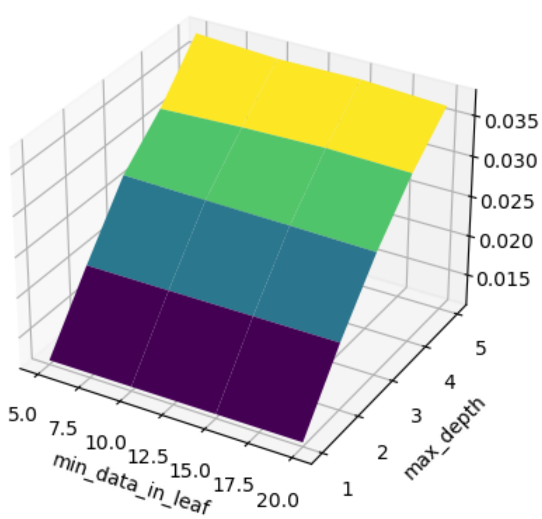
F. Random Forest

Best parameter and its out of sample R_square:
min_data_in_leaves=5, max_depth=5

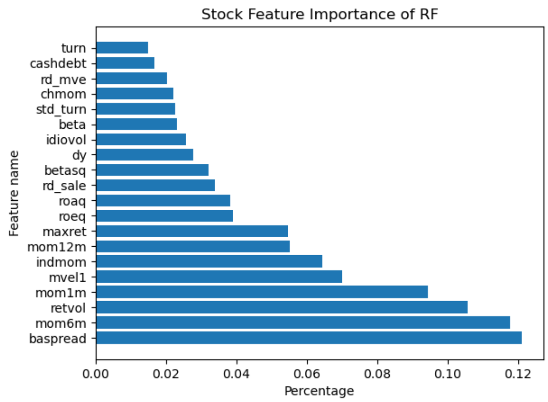
Validation_Roos=0.03759602, Test_Roos=0.04701205.
Top1000_Roos=0.1291411068159044, Bottom1000_Roos=0.03111660250166659

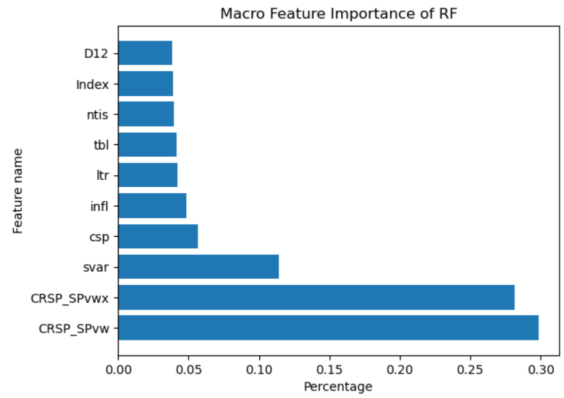
The process of choosing parameters:

According to the paper supplement, I set estimators=300, and min_data_in_leaves=[5,10,15,20], max_depths=[1,2,3,4,5]. After training different models, choose the one with the best validation Roos.



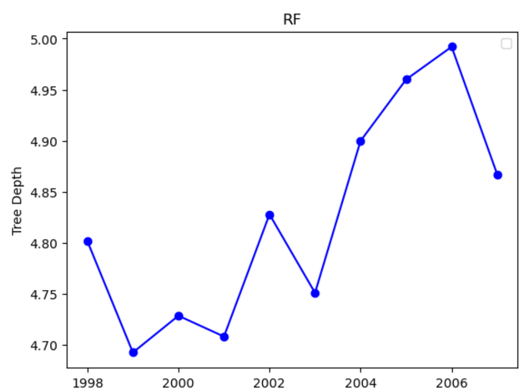
Validation Roos of Different Parameters of Random Forest
I also calculate top 20 stock features' importance and top 10 macro factors' importance as following:





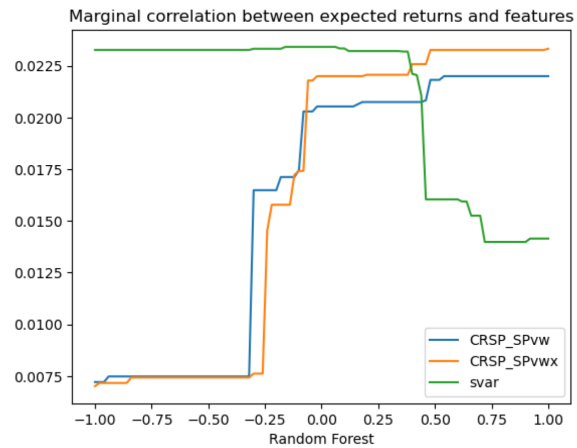
Feature Importance(top 20 stock features) of RF Feature Importance(top 10 macro features) of RF

To evaluate the time-varying model complexity, we choose the average tree depth of the random forest. In general, tree depth gets larger when the train set enlarges.



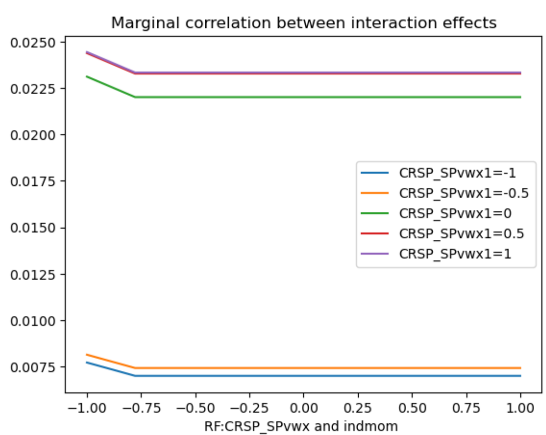
Time-varying Model Complexity of Random Forest using average tree depth

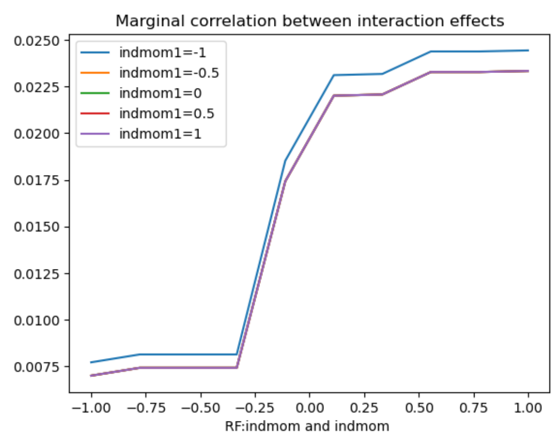
By setting other features the median number, putting a specific feature valued from -1 to 1 into the model will give the following expected returns (using svar, CRSP_SPvw and CRSP_SPvwx for example), and it proves that these three features have influence on the return result.



Marginal correlation between expected returns and features(Random Forest)

According to the paper, some features have a pairwise interaction effect, therefore I fix feature CRSP_SPvwx to [-1,-0.5,0,0.5,1], and fix other features equal to median feature number, and get a return corresponding to indmom from -1 to 1. On the other side, fix indmon to [-1,-0.5,0,0.5,1] and change CRSP_SPvwx.





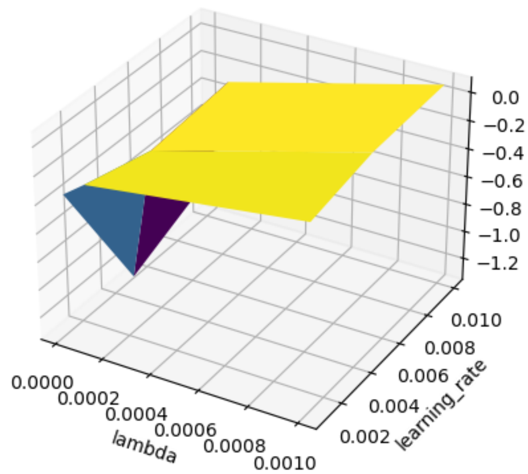
And according to the pictures, the feature indmom and CRSP_SPvwx are relative. And because this is a decision tree model, the interaction may not be continuous.

G. Neural networks

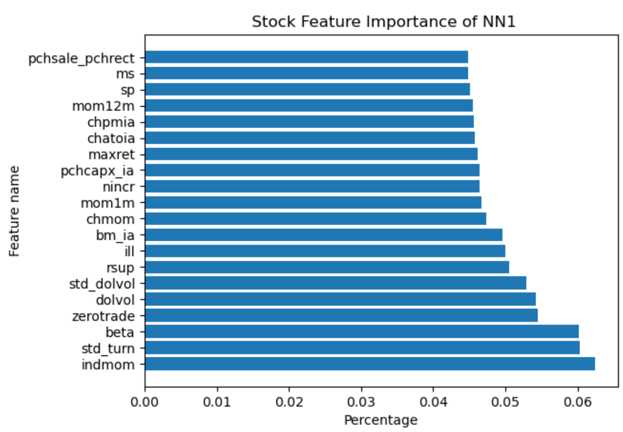
1. NN1

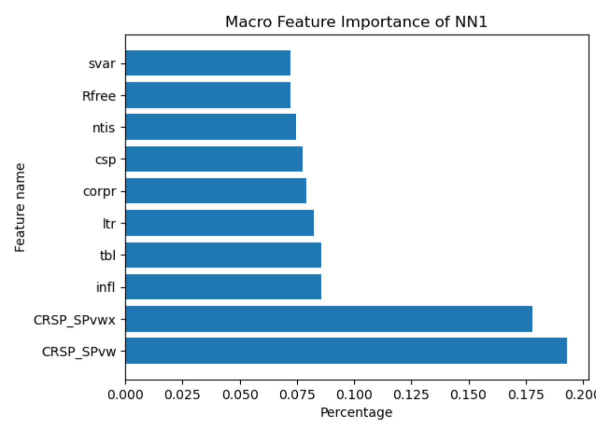
Best parameter:
lambda=0.001,learning_rate=0.01
Validation_Roos=0.07561129, Test_Roos=0.06446860893587911
Top1000_Roos=0.1846867280684258, Bottom1000_Roos=0.026047546382414355

The process of selecting parameters:
According to the paper supplement, I set L1 lambda=[0.00001,0.0001,0.001], learning_rate=[0.001,0.005,0.01]. After training different models, choose the ones with the best validation Roos.



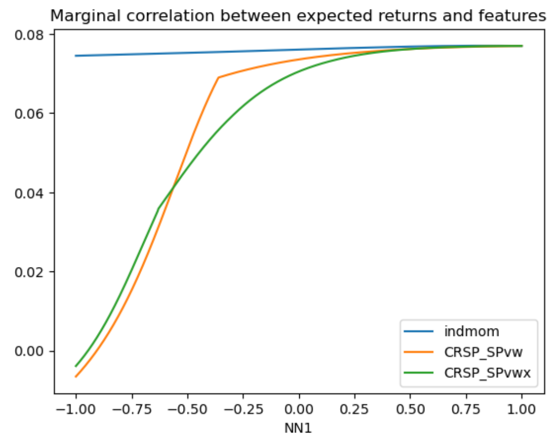
Validation Roos of Different Parameters of NN1





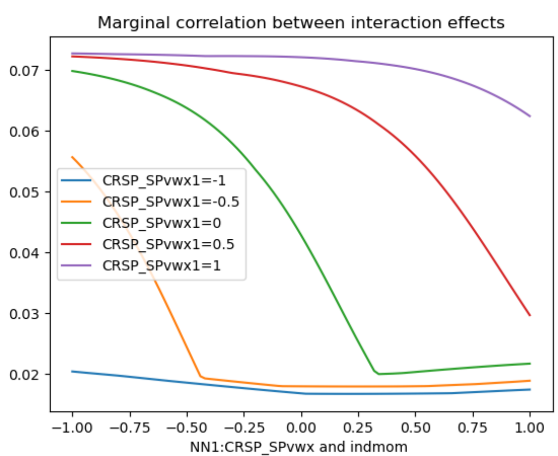
Feature Importance(top 20 stock features) of NN1 Feature Importance(top 10 macro features) of NN1

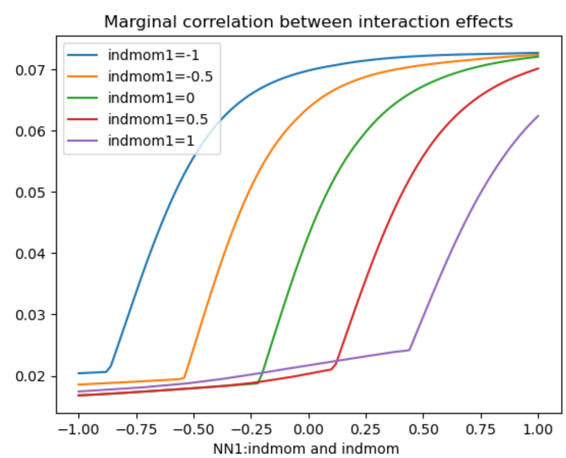
By setting other features the median number, putting a specific feature valued from -1 to 1 into the model will give the following expected returns (using indmom, CRSP_SPvw and CRSP_SPvwx in model 9 for example):



Marginal correlation between expected returns and features(NN1)

According to the paper, some features have a pairwise interaction effect, therefore I fix feature CRSP_SPvwx to [-1,-0.5,0,0.5,1], and fix other features equal to median feature number, and get a return corresponding to indmom from -1 to 1. On the other side, fix indmon to [-1,-0.5,0,0.5,1] and change CRSP_SPvwx.





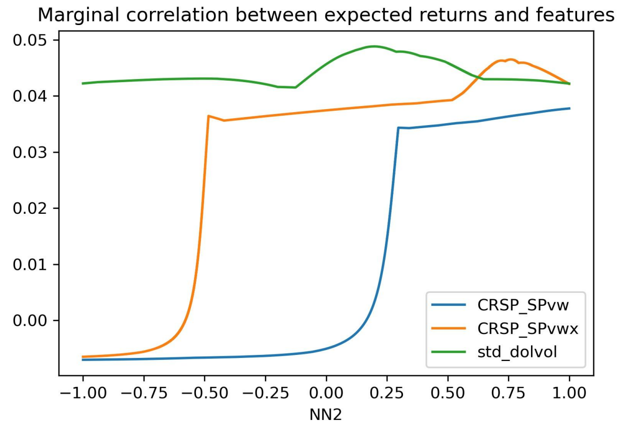
And according to the pictures, the feature indmom and CRSP_SPvwx are relative.

2. NN2

NN2_Roos: 0.069434
(best parameters: lambda_=0.001, learning_rate=0.02)
Top1000_Roos: 0.163829, Bottom1000_Roos:0.034273

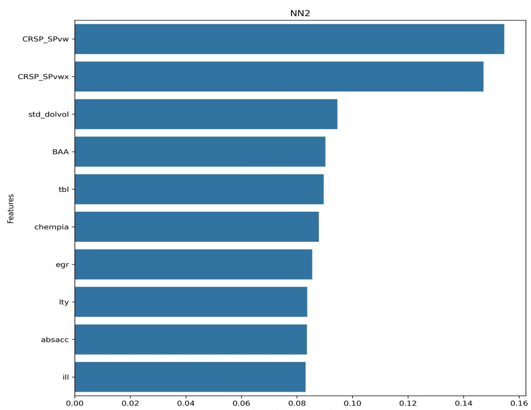
According to the paper supplement, I set L1 lambda=[0.0001,0.001,0.01], learning_rate=[0.005,0.01,0.02]. After training different models, choose the parameter combination (lambda=0.001, learning_rate=0.01) with the best validation Roos.

We set other features the median number and put a specific feature valued from -1 to 1 into the model, which gives the following expected returns. (choose the features CRSP_SPvw, CRSP_SPvwx, std_dolvol)

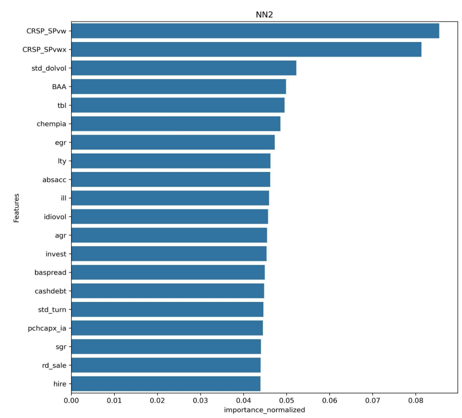


Marginal correlation between expected returns and features(NN2)

The following figures demonstrate the variable importance for the top-10 and top-20 most influential variables in this model. And the features importance is normalized to sum to one.



Feature Importance(top 10 stock features) of NN2



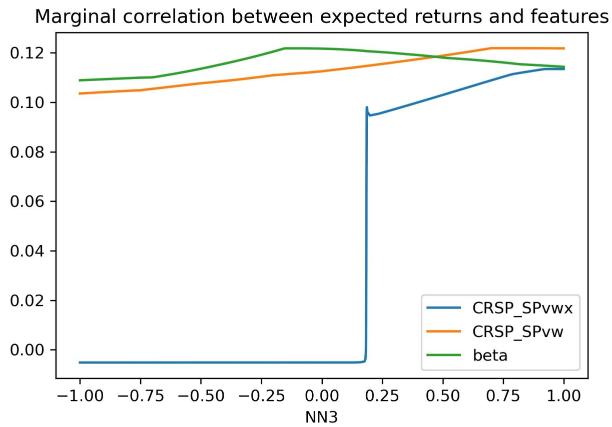
Feature Importance(top 20 macro features) of NN2

3. NN3

NN3_Roos: 0.063835
(best parameters: lambda_=0.001, learning_rate=0.01)
Top1000_Roos: 0.151992, Bottom1000_Roos:0.022985

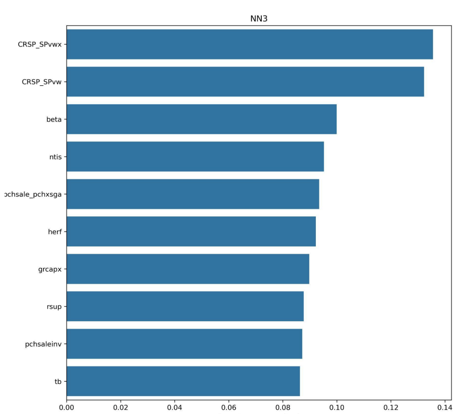
According to the paper supplement, I set L1 lambda=[0.0001,0.001,0.01], learning_rate=[0.005,0.01,0.02]. After training different models, choose the parameter combination (lambda=0.001, learning_rate=0.02) with the best validation Roos.

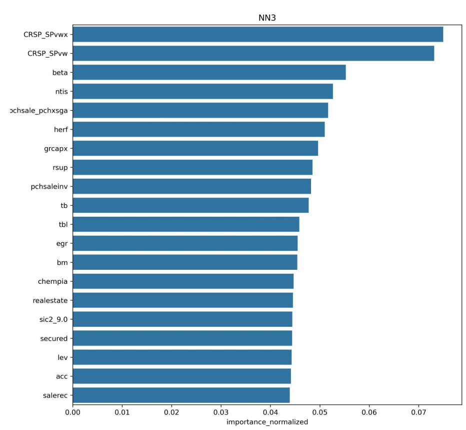
We set other features the median number and put a specific feature valued from -1 to 1 into the model, which gives the following expected returns. (choose the features CRSP_SPvwx, CRSP_SPvw, beta)



Marginal correlation between expected returns and features(NN3)

The following figures demonstrate the variable importance for the top-10 and top-20 most influential variables in this model. And the features importance is normalized to sum to one.



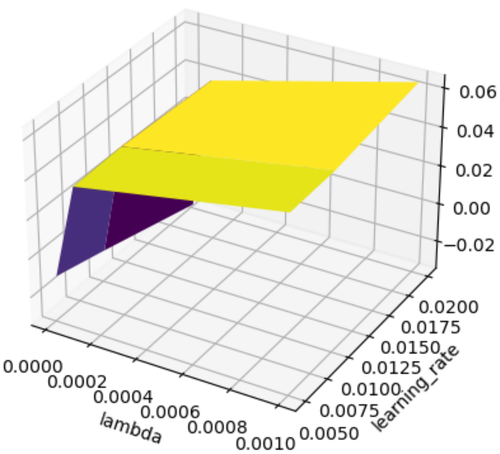


Feature Importance(top 10 stock features) of NN3 Feature Importance(top 20 macro features) of NN3

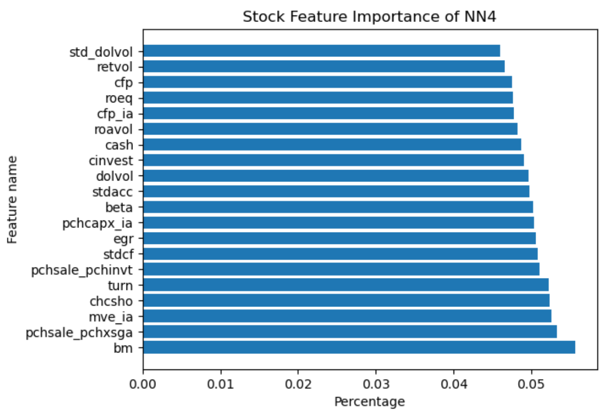
4. NN4

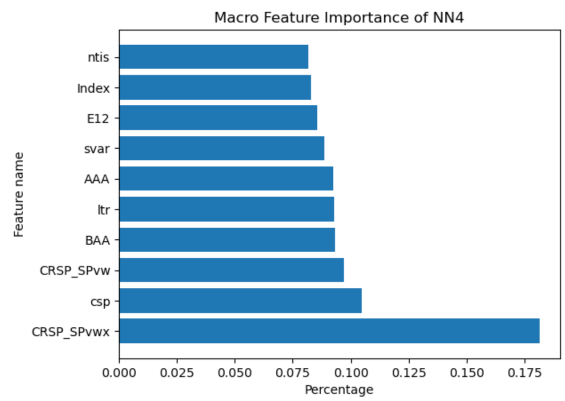
Best parameter:
lambda=0.001,learning_rate=0.02
Validation_Roos=0.06450469, test_Roos=0.05900669924857046.
Top1000_roos=0.14593052149665764, Bottom1000_roos=0.01900617589363074

The process of selecting parameters:
According to the paper supplement, I set L1 lambda=[0.00001,0.0001,0.001], learning_rate=[0.005,0.01,0.02]. After training different models, choose the ones with the best validation Roos.



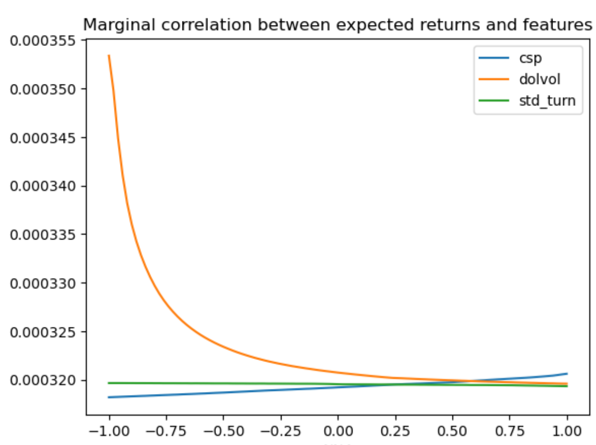
Validation Roos of Different Parameters of NN4





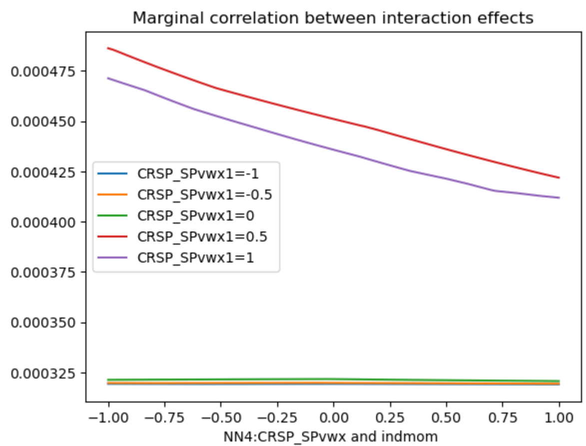
Feature Importance(top 20 stock features) of NN4 Feature Importance(top 10 macro features) of NN4

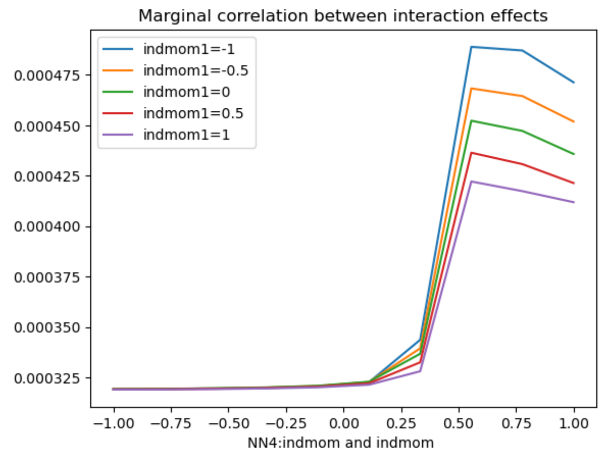
By setting other features the median number, putting a specific feature valued from -1 to 1 into the model will give the following expected returns(using dolvol, csp and std_turn in model 8 for example):



Marginal correlation between expected returns and features(NN4)

According to the paper, some features have a pairwise interaction effect, therefore I fix feature CRSP_SPvwx to [-1,-0.5,0,0.5,1], and fix other features equal to median feature number, and get a return corresponding to indmom from -1 to 1. On the other side, fix indmom to [-1,-0.5,0,0.5,1] and change CRSP_SPvwx.





And according to the pictures, the feature indmom and CRSP_SPvwx are relative.

IV. Conclusion

A. Model Comparison

Compare model prediction performance based on the Diebold-Mariano test statistic. A value greater than 0 means that the model corresponding to the row of the value is better than the model corresponding to the column of the value.

	EL	PCR	PLS	XGB	RF	NN1	NN2	NN3	NN4
EL		**1.97**	**2.08**	0.38	**2.44**	**1.73**	**1.99**	**2.56**	**2.33**
PCR			1.24	1.56	**2.15**	-0.43	0.58	0.77	1.03
PLS				-0.47	**1.78**	**-0.03**	-1.62	-1.52	**-1.99**
XGB					**1.79**	-0.33	-0.17	0.03	-0.78
RF						-0.43	-0.28	-0.09	-0.16
NN1							**1.69**	0.98	0.43
NN2								0.03	-0.78
NN3									-0.36
NN4									

It can be seen that among many prediction models, the elastic network model has an unexpectedly good prediction level, and the random forest performs the worst.

B. Summary of Importance

In the table below, we show the importance of macro factors in different models. The higher the factor, the greater the sum of its importance in all models.

	EN	PLS	PCR	XGB	RF	NN1	NN2	NN3	NN4
CRSP_SPvwx	0.7164	0.0825	0.1567	0.2374	0.2321	0.1290	0.1463	0.0849	0.1210
CRSP_SPvw	0.2547	0.0245	0.1176	0.1069	0.2458	0.1397	0.0885	0.0493	0.0647
ltr	0.0257	0.0741	0.0392	0.1258	0.0349	0.0598	0.0680	0.4219	0.0620
infl	0.0000	0.0632	0.0470	0.0653	0.0399	0.0620	0.1343	0.0822	0.0493
svar	0.0000	0.0655	0.0768	0.0516	0.0940	0.0523	0.0716	0.0438	0.0591
csp	0.0006	0.0662	0.0549	0.0394	0.0465	0.0561	0.0482	0.0274	0.0698
Index	0.0000	0.0642	0.0705	0.0446	0.0324	0.0327	0.0500	0.0384	0.0552
D12	0.0000	0.0723	0.0690	0.0444	0.0314	0.0371	0.0506	0.0301	0.0446
b/m	(0.0001)	0.0718	0.0525	0.0439	0.0226	0.0404	0.0590	0.0362	0.0467
Rfree	0.0027	0.0717	0.0353	0.0377	0.0263	0.0523	0.0470	0.0318	0.0480
ntis	0.0000	0.0660	0.0102	0.0310	0.0328	0.0539	0.0458	0.0329	0.0546
AAA	0.0000	(0.0016)	0.0768	0.0321	0.0221	0.0455	0.0524	0.0329	0.0616
corpr	0.0000	0.0737	0.0188	0.0312	0.0304	0.0573	0.0391	0.0241	0.0448
tbl	0.0000	0.0655	0.0627	0.0269	0.0341	0.0620	0.0054	0.0011	0.0492
lty	0.0000	0.0725	0.0345	0.0304	0.0287	0.0402	0.0271	0.0181	0.0502
BAA	0.0000	(0.0032)	0.0596	0.0262	0.0219	0.0390	0.0464	0.0285	0.0621
E12	0.0000	0.0711	0.0180	0.0252	0.0241	0.0405	0.0205	0.0164	0.0570

It can be seen that 'CRSP_SPvwx' and 'CRSP_SPvw' are more important.
The figure below is a heat map representing different features in different models