Empirical Asset Pricing via Machine Learning

Shihau Gu, Bryan Kelly, and Dacheng Xiu (2020)

Presented by Tommy Brown

Motivation

Can we use Machine Learning to measure equity premiums?

"A portfolio strategy that times the S&P 500 with neural network forecasts enjoys an annualized out-of-sample Sharpe ratio of 0.77 versus the 0.51 Sharpe ratio of a buy-and-hold investor. And a value-weighted long-short decile spread strategy that takes positions based on stock-level neural network forecasts earns an annualized out-of-sample Sharpe ratio of 1.35, more than doubling the performance of a leading regression-based strategy from the literature."

What is Meant by Machine Learning?

a) A diverse collection of high-dimensional models for statistical prediction
   - Enhanced robustness and flexibility relative to more traditional models

b) Methods for model selection and mitigation of overfit
   - Emphasis on OOS performance

c) Efficient algorithms for searching among a vast number of potential model specifications
   - Approximate optimal specification with manageable cost

Why Machine Learning?

- Great at prediction! Similar to the conditional risk premium
- Enormous number of predictor variables or “Zoo of Factors”
- Well suited to the ambiguous functional form of returns

(Aside) Disadvantages to Machine Learning?

- Not necessarily causal
- Can lose measurable economic significance (may not have coefficients!)

Data

Predictor Variables: 94 characteristics, interacted with 8 time series variables, with industry sector dummies, for a total of 900 signals

Time Frame: 1957-2016

18 years of training, 12 years of validation, 30 years for OOS testing

Methods

OLS:

Linear g(·), minimize: $$\mathcal{L}(\theta)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(r_{i,t+1}-g(z_{i,t};\theta)\right)^2$$

Why will this perform so poorly?

In tables, will only include 3 factor with Huber loss (heavy tails):

$$\mathcal{L}_H(\theta)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}H\left(r_{i,t+1}-g(z_{i,t};\theta),\xi\right), \qquad (6)$$

where

$$H(x;\xi)=\begin{cases} x^2, & \text{if} \quad |x|\leq\xi; \\ 2\xi|x|-\xi^2, & \text{if} \quad |x|>\xi. \end{cases}$$

Methods

Penalized Linear: Elastic Net

$$\mathcal{L}(\theta;\cdot)=\mathcal{L}(\theta)+\phi(\theta;\cdot). \tag{7}$$

There are several choices for the penalty function $\phi(\theta;\cdot)$. We focus on the popular "elastic net" penalty, which takes the form

$$\phi(\theta;\lambda,\rho)=\lambda(1-\rho)\sum_{j=1}^{P}|\theta_j|+\frac{1}{2}\lambda\rho\sum_{j=1}^{P}\theta_j^2. \tag{8}$$

Add a penalty to loss function

Acts as "variable selection\shrinkage" depending on ρ , combination of ridge and lasso

Methods

PLS: Partial Least Squares

PCR: Principal Components Regression

Vectorize the linear model:

$$R = Z\theta + E, \tag{9}$$

Reduce dimensionality using a weighting matrix:

$$R = (Z\Omega_K)\theta_K + \tilde{E}. \tag{10}$$

Coefficients are estimated like OLS

Methods

GLM: General Linear Model

- Similar to the simple linear form, but with spline series expansion

- Uses the least squares objective function with penalization "group lasso"

$$\phi(\theta;\lambda,K)=\lambda\sum_{j=1}^{P}\left(\sum_{k=1}^{K}\theta_{j,k}^{2}\right)^{1/2}. \qquad (14)$$

- Selects either all or none of a group of spline terms

Methods

<u>Boosted Regression Trees and Random Forests</u>:

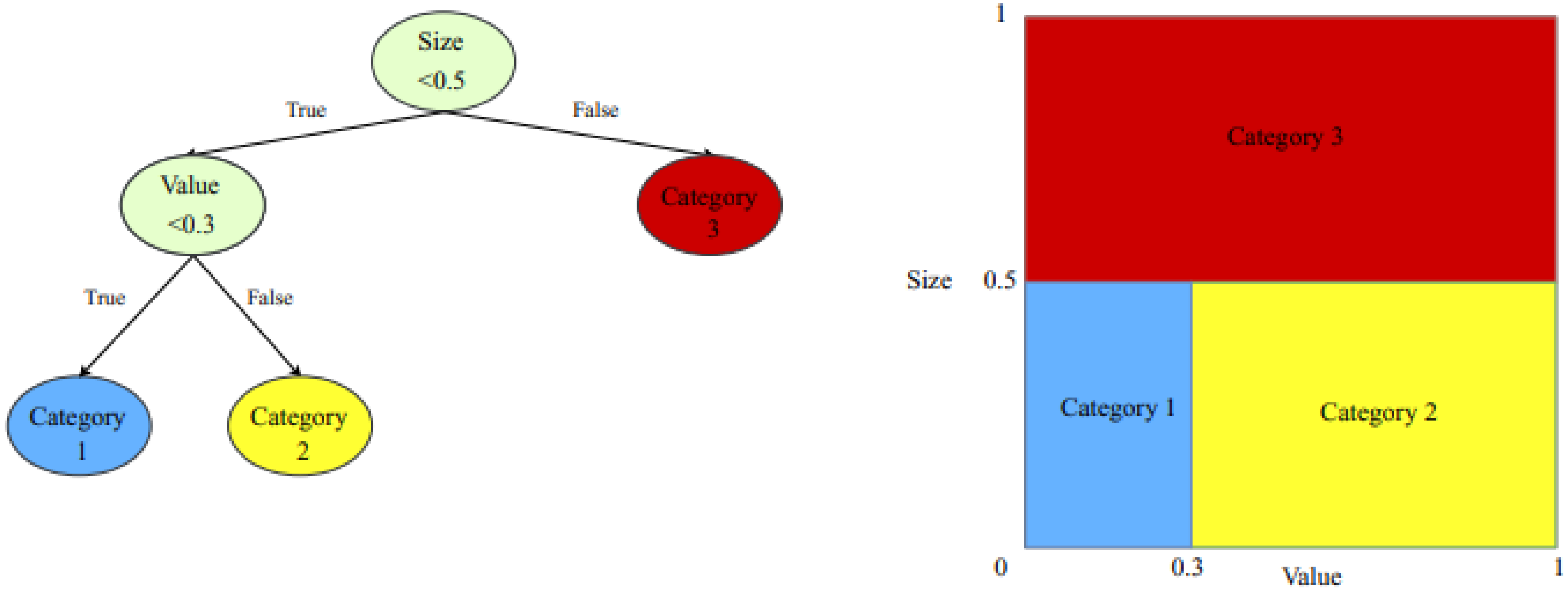


Figure 1
Regression tree example
This figure presents the diagrams of a regression tree (left) and its equivalent representation (right) in the space of two characteristics (size and value). The terminal nodes of the tree are colored in blue, yellow, and red. Based on their values of these two characteristics, the sample of individual stocks is divided into three categories.

Methods

Boosted: Many shallow trees, as an ensemble, create a strong learner

RF: Uses different bootstrap samples of the data, fits to different trees, then averages their forecasts. Variation will decorrelate trees.

Neural Networks

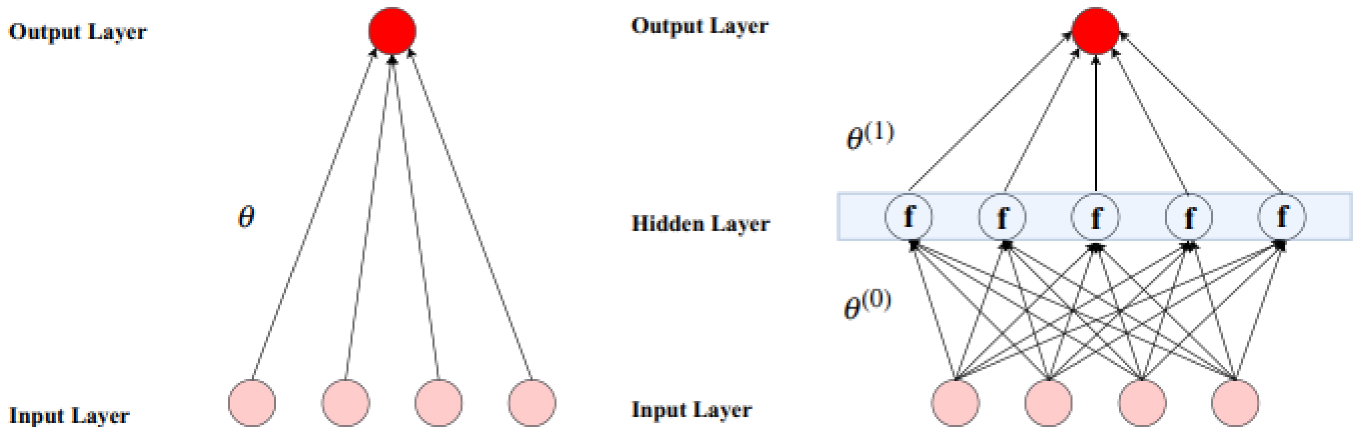


Figure 2
Neural networks
This figure provides diagrams of two simple neural networks with (right) or without (left) a hidden layer. Pink circles denote the input layer, and dark red circles denote the output layer. Each arrow is associated with a weight parameter. In the network with a hidden layer, a nonlinear activation function f transforms the inputs before passing them on to the output.

Evaluation

Performance Evaluation: OOS R squared

$$R^2_{\text{oos}} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \widehat{r}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{i,t+1}}, \tag{19}$$

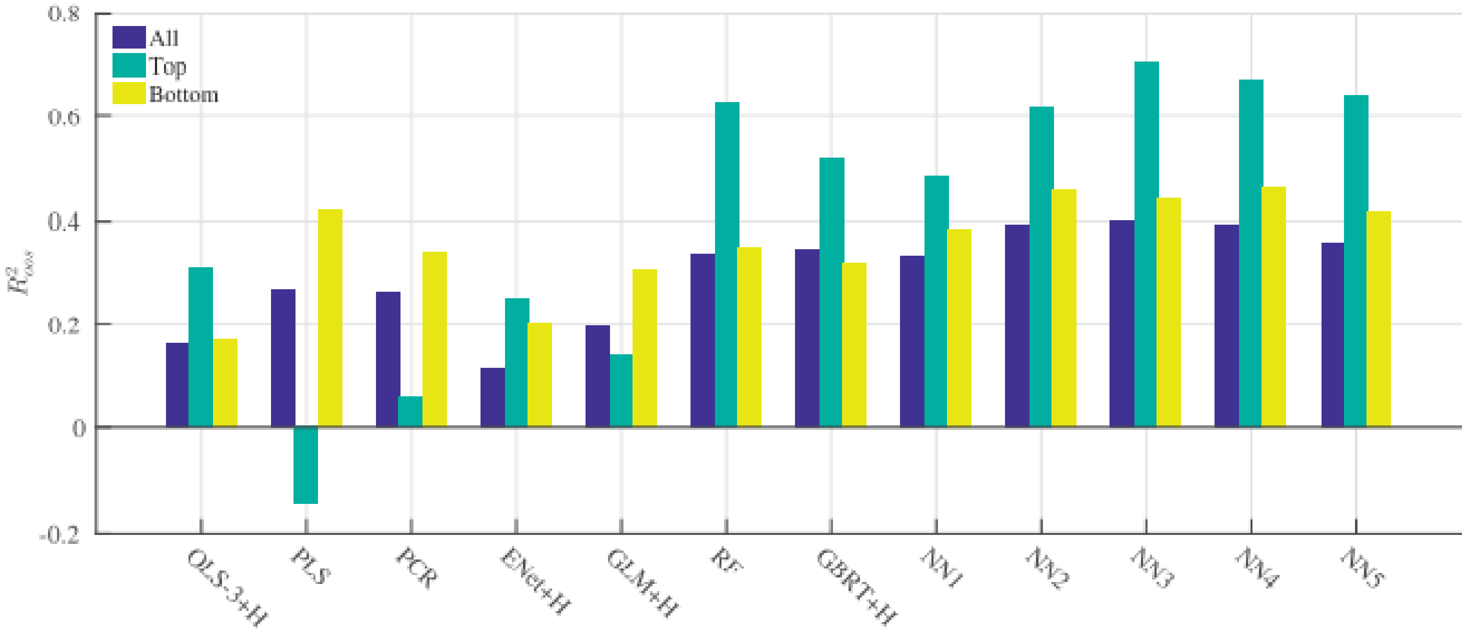
Variable Importance:

1. Calculate the reduction in panel predictive R squared from setting the variable of interest to zero and hold all other variables constant
2. Sum of Squared Partial Derivatives

$$SSD_j = \sum_{i,t\in\mathcal{T}_1}\left(\left.\frac{\partial g(z;\theta)}{\partial z_j}\right|_{z=z_{i,t}}\right)^2$$

Table 1
Monthly out-of-sample stock-level prediction performance (percentage R^2_{oos})

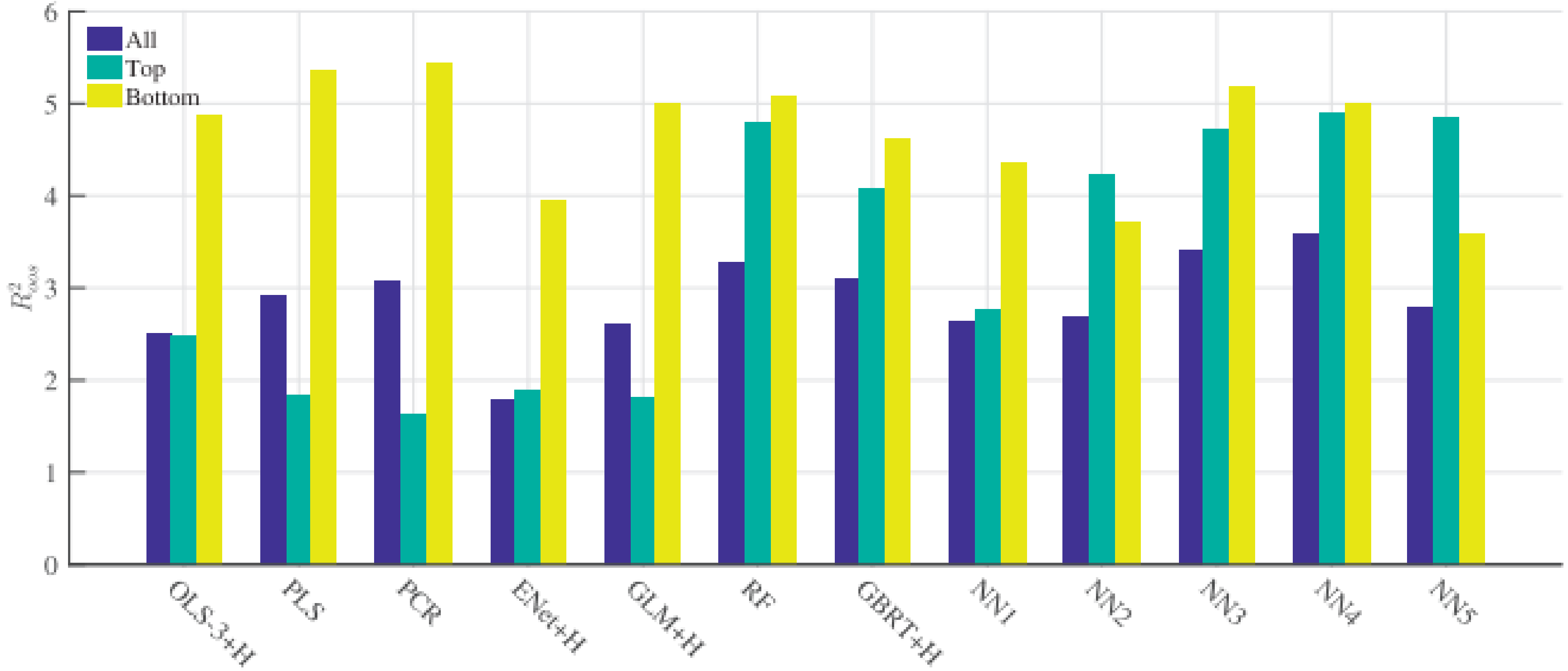
	OLS +H	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
All	−3.46	0.16	0.27	0.26	0.11	0.19	0.33	0.34	0.33	0.39	0.40	0.39	0.36
Top 1,000	−11.28	0.31	−0.14	0.06	0.25	0.14	0.63	0.52	0.49	0.62	0.70	0.67	0.64
Bottom 1,000	−1.30	0.17	0.42	0.34	0.20	0.30	0.35	0.32	0.38	0.46	0.45	0.47	0.42



In this table, we report monthly R^2_{oos} for the entire panel of stocks using OLS with all variables (OLS), OLS using only size, book-to-market, and momentum (OLS-3), PLS, PCR, elastic net (ENet), generalize linear model (GLM), random forest (RF), gradient boosted regression trees (GBRT), and neural networks with 1 to 5 layers (NN1–NN5). "+H" indicates the use of Huber loss instead of the l_2 loss. We also report these R^2_{oos} within subsamples that include only the top-1,000 stocks or bottom-1,000 stocks by market value. The lower panel provides a visual comparison of the R^2_{oos} statistics in the table (omitting OLS because of its large negative values).

Table 2
Annual out-of-sample stock-level prediction performance (percentage R^2_{oos})

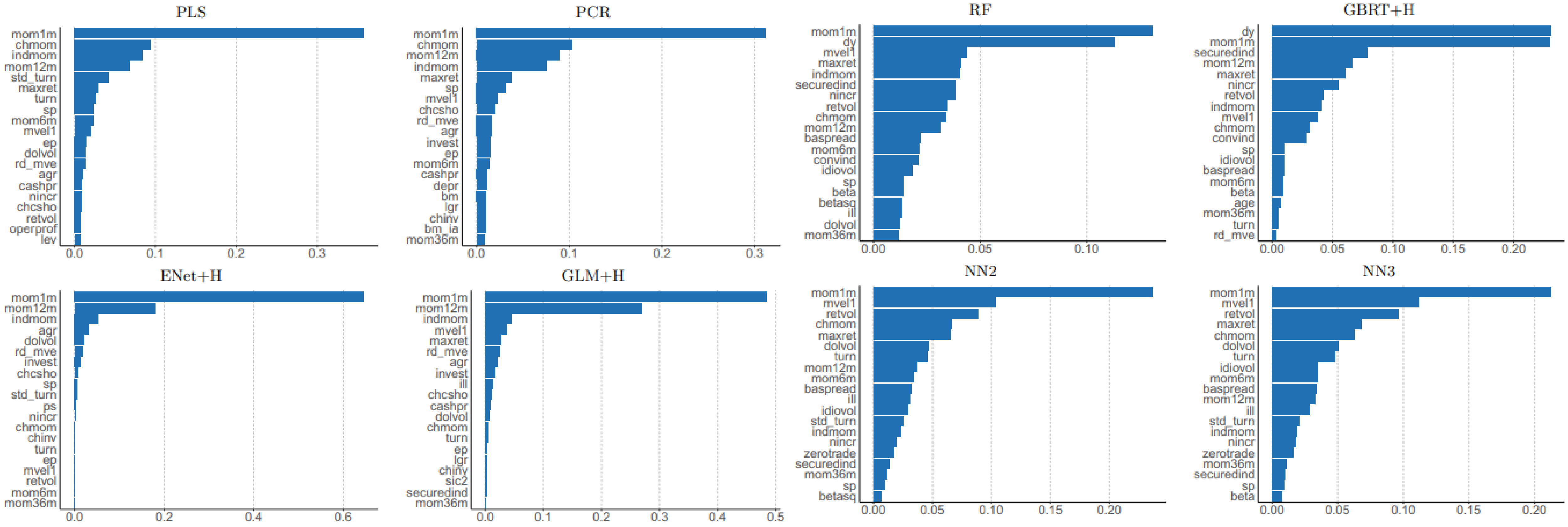
	OLS +H	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
All	–34.86	2.50	2.93	3.08	1.78	2.60	3.28	3.09	2.64	2.70	3.40	3.60	2.79
Top	–54.86	2.48	1.84	1.64	1.90	1.82	4.80	4.07	2.77	4.24	4.73	4.91	4.86
Bottom	–19.22	4.88	5.36	5.44	3.94	5.00	5.08	4.61	4.37	3.72	5.17	5.01	3.58

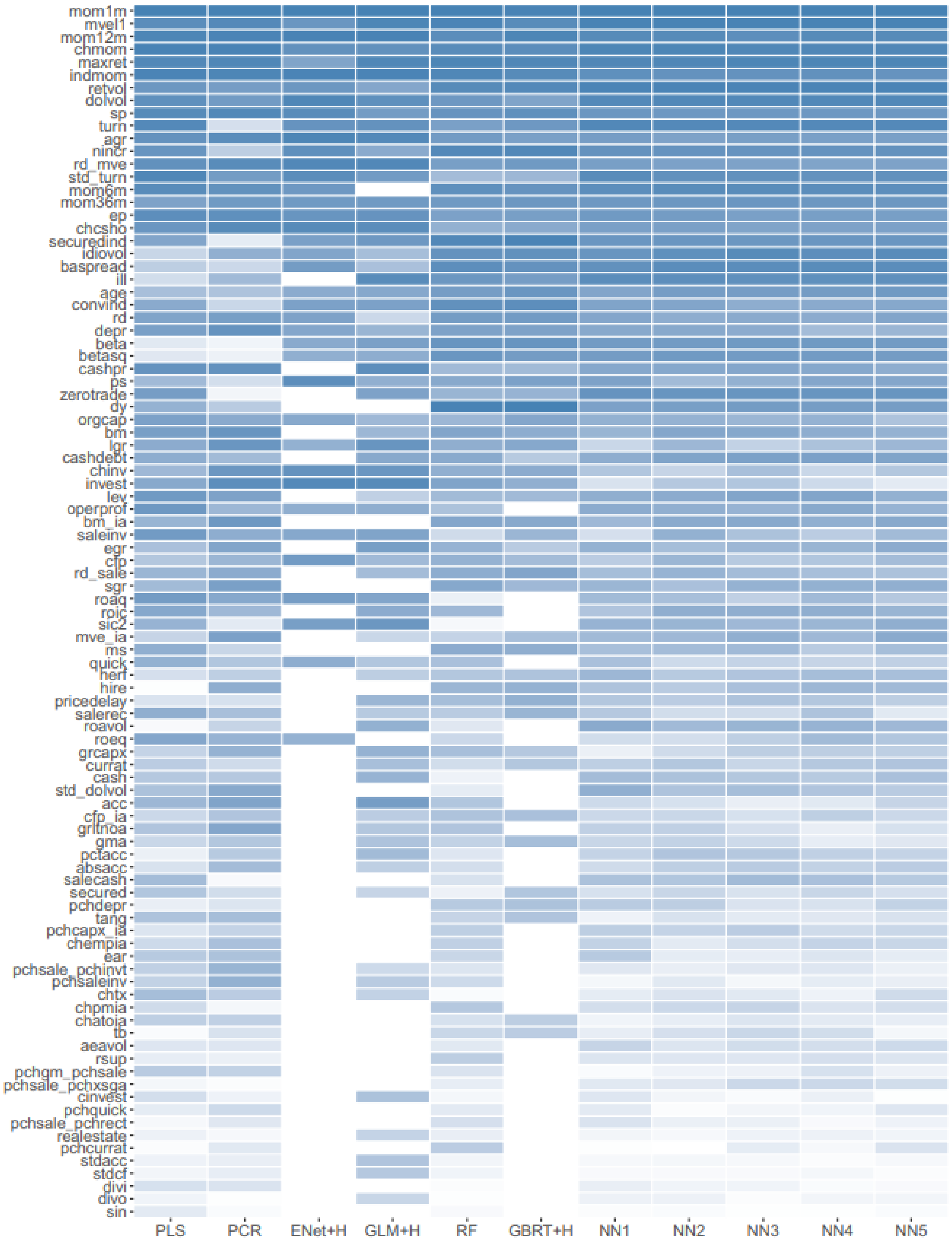


Annual return forecasting R^2_{oos} (see the legend to Table 1).

Which Variable Matter?

Short Term Reversal, Momentum, Momentum Change, Industry Momentum, Amihud illiquidity, Sales-to-price, Log Market Equity



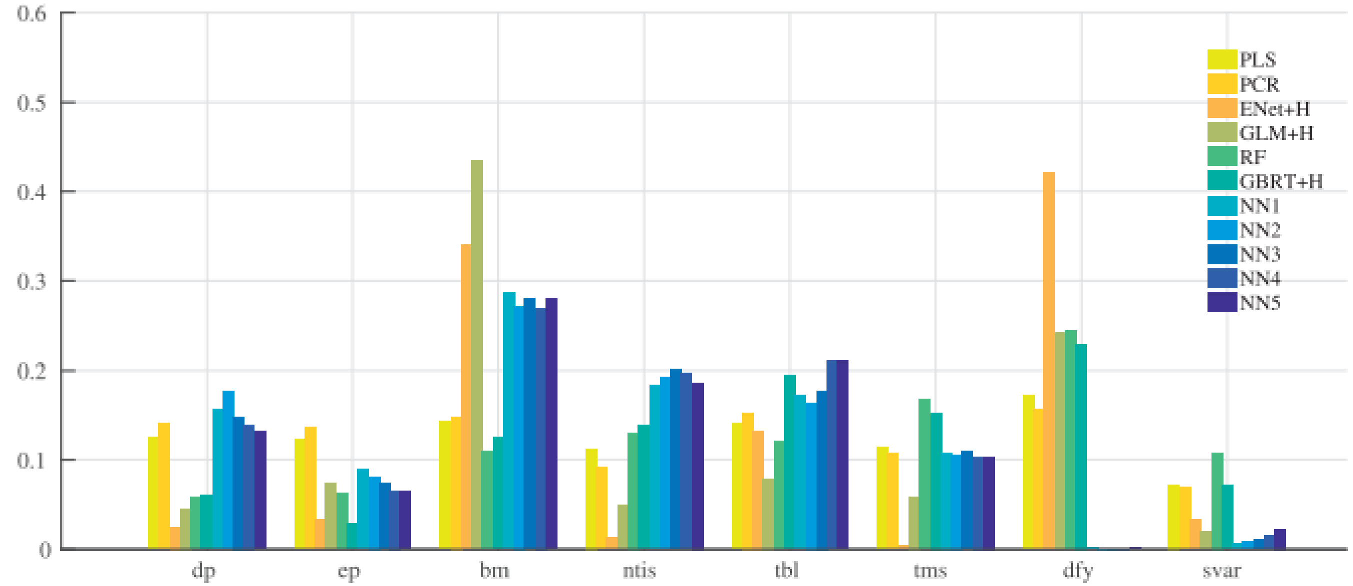

mom1m
mvel1
mom12m
chmom
maxret
indmom
retvol
dolvol
sp
turn
agr
nincr
rd_mve
std_turn
mom6m
mom36m
ep
chcsho
securedind
idiovol
baspread
ill
age
convind
rd
depr
beta
betasq
cashpr
ps
zerotrade
dy
orgcap
bm
lgr
cashdebt
chinv
invest
lev
operprof
bm_ia
saleinv
egr
cfp
rd_sale
sgr
roaq
roic
sic2
mve_ia
ms
quick
herf
hire
pricedelay
salerec
roavol
roeq
grcapx
currat
cash
std_dolvol
acc
cfp_ia
grltnoa
gma
pctacc
absacc
salecash
secured
pchdepr
tang
pchcapx_ia
chempia
ear
pchsale_pchinvt
pchsaleinv
chtx
chpmia
chatoia
tb
aeavol
rsup
pchgm_pchsale
pchsale_pchxsga
cinvest
pchquick
pchsale_pchrect
realestate
pchcurrat
stdacc
stdcf
divi
divo
sin
PLS
PCR
ENet+H
GLM+H
RF
GBRT+H
NN1
NN2
NN3
NN4
NN5

Which Macro Variables Matter?

Dividend/Price
Earnings/Price
Book/Market
Net Equity Expansion
T-bill
Term Spread
Default Spreads
Stock Variance

Table 4
Variable importance for macroeconomic predictors

	PLS	PCR	ENet+H	GLM+H	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
dp	12.52	14.12	2.49	4.54	5.80	6.05	15.57	17.58	14.84	13.95	13.15
ep	12.25	13.52	3.27	7.37	6.27	2.85	8.86	8.09	7.34	6.54	6.47
bm	14.21	14.83	33.95	43.46	10.94	12.49	28.57	27.18	27.92	26.95	27.90
ntis	11.25	9.10	1.30	4.89	13.02	13.79	18.37	19.26	20.15	19.59	18.68
tbl	14.02	15.29	13.29	7.90	11.98	19.49	17.18	16.40	17.76	20.99	21.06
tms	11.35	10.66	0.31	5.87	16.81	15.27	10.79	10.59	10.91	10.38	10.33
dfy	17.17	15.68	42.13	24.10	24.37	22.93	0.09	0.06	0.06	0.04	0.12
svar	7.22	6.80	3.26	1.87	10.82	7.13	0.57	0.85	1.02	1.57	2.29



Variable importance for eight macroeconomic variables in each model. Variable importance is an average over all training samples. Variable importance within each model is normalized to sum to one. The lower panel provides a complementary visual comparison of macroeconomic variable importance.

Table 5
Monthly portfolio-level out-of-sample predictive R^2

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
A. Common factor portfolios												
S&P 500	−0.22	−0.86	−1.55	0.75	0.71	1.37	1.40	1.08	1.13	1.80	1.63	1.17
SMB	0.81	2.09	0.39	1.72	2.36	0.57	0.35	1.40	1.16	1.31	1.20	1.27
HML	0.66	0.50	1.21	0.46	0.84	0.98	0.21	1.22	1.31	1.06	1.25	1.24
RMW	−2.35	1.19	0.41	−1.07	−0.06	−0.54	−0.92	0.68	0.47	0.84	0.53	0.54
CMA	0.80	−0.44	0.03	−1.07	1.24	−0.11	−1.04	1.88	1.60	1.06	1.84	1.31
UMD	−0.90	−1.09	−0.47	0.47	−0.37	1.37	−0.25	−0.56	−0.26	0.19	0.27	0.35
B. Subcomponents of factor portfolios												
Big value	0.10	0.00	−0.33	0.25	0.59	1.31	1.06	0.85	0.87	1.46	1.21	0.99
Big growth	−0.33	−1.26	−1.62	0.70	0.51	1.32	1.19	1.00	1.10	1.50	1.24	1.11
Big neutral	−0.17	−1.09	−1.51	0.80	0.36	1.31	1.28	1.43	1.24	1.70	1.81	1.40
Small value	0.30	1.66	1.05	0.64	0.85	1.24	0.52	1.59	1.37	1.54	1.40	1.30
Small growth	−0.16	0.14	−0.18	−0.33	−0.12	0.71	1.24	0.05	0.42	0.48	0.41	0.50
Small neutral	−0.27	0.60	0.19	0.21	0.28	0.88	0.36	0.58	0.62	0.70	0.58	0.68
Big conservative	−0.57	−0.10	−1.06	1.02	0.46	1.11	0.55	1.15	1.13	1.59	1.37	1.07
Big aggressive	0.20	−0.80	−1.15	0.30	0.67	1.75	2.00	1.33	1.51	1.78	1.55	1.42
Big neutral	−0.29	−1.75	−1.96	0.83	0.48	1.13	0.77	0.85	0.85	1.51	1.45	1.16
Small conservative	−0.05	1.17	0.71	−0.02	0.34	0.96	0.56	0.82	0.87	0.96	0.90	0.83
Small aggressive	−0.10	0.51	0.01	−0.09	0.14	1.00	1.46	0.34	0.64	0.75	0.62	0.71
Small neutral	−0.30	0.45	0.12	0.42	0.35	0.76	−0.01	0.70	0.69	0.83	0.66	0.72
Big robust	−1.02	−1.08	−2.06	0.55	0.35	1.10	0.33	0.74	0.79	1.28	1.03	0.74
Big weak	−0.12	1.42	1.07	0.89	1.10	1.33	1.77	1.79	1.79	2.05	1.66	1.60
Big neutral	0.86	−1.22	−1.26	0.41	0.13	1.10	0.91	0.84	0.94	1.19	1.15	0.99
Small robust	−0.71	0.35	−0.38	−0.04	−0.42	0.70	0.19	0.24	0.50	0.63	0.53	0.55
Small weak	0.05	1.06	0.59	−0.13	0.44	1.05	1.42	0.71	0.92	0.99	0.90	0.89
Small neutral	−0.51	0.07	−0.47	−0.33	−0.32	0.60	−0.08	0.10	0.25	0.38	0.32	0.41
Big up	0.20	−0.25	−1.24	0.66	1.17	1.18	0.90	0.80	0.76	1.13	1.12	0.93
Big down	−1.54	−1.63	−1.55	0.44	−0.33	1.14	0.71	0.36	0.70	1.07	0.90	0.84
Big medium	−0.04	−1.51	−1.94	0.81	−0.08	1.57	1.80	1.29	1.32	1.71	1.55	1.23
Small up	0.07	0.78	0.56	−0.07	0.25	0.62	−0.03	0.06	0.07	0.21	0.19	0.25
Small down	−0.21	0.15	−0.20	0.15	−0.01	1.51	1.38	0.74	0.82	1.02	0.91	0.96
Small medium	0.07	0.82	0.20	0.59	0.37	1.22	1.06	1.09	1.09	1.18	1.00	1.03

Table 6
Market timing Sharpe ratio gains

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
A. Common factor portfolios												
S&P 500	0.07	0.05	−0.06	0.12	0.19	0.18	0.19	0.22	0.20	0.26	0.22	0.19
SMB	0.06	0.17	0.09	0.24	0.26	0.00	−0.07	0.21	0.18	0.15	0.09	0.11
HML	0.00	0.01	0.04	−0.03	−0.02	0.04	0.02	0.04	0.06	0.04	0.02	0.01
RMW	0.00	−0.01	−0.06	−0.19	−0.13	−0.11	−0.01	−0.03	−0.09	0.01	0.01	−0.07
CMA	0.02	0.02	0	−0.09	−0.05	0.08	−0.01	0.00	0.01	0.05	0.04	0.06
UMD	0.01	−0.06	−0.02	−0.02	−0.07	−0.04	−0.07	−0.04	−0.08	−0.04	−0.10	−0.01
B. Subcomponents of factor portfolios												
Big value	−0.01	0.06	−0.03	0.09	0.06	0.09	0.08	0.11	0.11	0.13	0.10	0.11
Big growth	0.08	−0.01	−0.08	0.10	0.17	0.20	0.21	0.22	0.20	0.26	0.22	0.21
Big neutral	0.06	0.03	−0.06	0.11	0.16	0.13	0.17	0.23	0.21	0.23	0.23	0.21
Small value	−0.04	0.15	0.09	0.01	0.08	0.07	0.08	0.11	0.11	0.10	0.11	0.13
Small growth	0.00	0.03	−0.06	−0.03	−0.05	0.04	0.05	0.02	0.03	0.03	0.02	0.02
Small neutral	0.02	0.09	0.05	0.03	0.04	0.11	0.11	0.09	0.08	0.10	0.09	0.11
Big conservative	0.08	0.02	−0.04	0.08	0.15	0.09	0.13	0.17	0.14	0.19	0.16	0.14
Big aggressive	0.08	−0.01	−0.11	0.01	0.13	0.22	0.18	0.21	0.19	0.23	0.20	0.20
Big neutral	0.04	−0.01	−0.08	0.09	0.11	0.09	0.11	0.13	0.12	0.18	0.18	0.16
Small conservative	0.04	0.17	0.12	0.02	0.05	0.17	0.15	0.11	0.11	0.14	0.13	0.15
Small aggressive	0.01	0.05	−0.06	−0.05	−0.03	0.08	0.06	0.02	0.05	0.06	0.04	0.05
Small neutral	0.01	0.06	0.03	0.01	0.04	0.08	0.09	0.07	0.06	0.08	0.07	0.09
Big robust	0.10	0.07	−0.07	0.11	0.18	0.17	0.18	0.18	0.16	0.22	0.19	0.16
Big weak	0.05	0.12	0.05	0.09	0.12	0.21	0.17	0.22	0.20	0.21	0.18	0.19
Big neutral	0.09	0.00	−0.04	0.09	0.20	0.19	0.17	0.22	0.21	0.24	0.21	0.20
Small robust	0.09	0.04	−0.03	0.00	0.00	0.10	0.07	0.04	0.05	0.08	0.08	0.08
Small weak	−0.03	0.09	0.00	−0.03	−0.02	0.07	0.07	0.06	0.06	0.06	0.05	0.06
Small neutral	0.04	0.04	−0.03	0.00	0.01	0.11	0.09	0.04	0.04	0.07	0.07	0.08
Big up	0.10	0.05	−0.06	0.10	0.21	0.16	0.14	0.17	0.14	0.17	0.18	0.17
Big down	−0.02	0.09	−0.08	−0.02	0.02	0.08	0.10	0.10	0.07	0.12	0.11	0.09
Big medium	−0.01	0.04	−0.06	0.14	0.09	0.17	0.20	0.22	0.21	0.25	0.22	0.19
Small up	0.08	0.13	0.10	0.05	0.07	0.16	0.12	0.07	0.06	0.08	0.07	0.10
Small down	−0.14	0.04	−0.05	−0.09	−0.05	0.06	0.04	0.01	0.01	0.02	0.01	0.01
Small medium	0.05	0.11	0.07	0.08	0.09	0.13	0.15	0.13	0.12	0.14	0.13	0.15

This table documents improvement in annualized Sharpe ratio ($SR^* - SR$). We compute the SR^* by weighting the portfolios based on a market timing strategy (see Campbell and Thompson 2008).

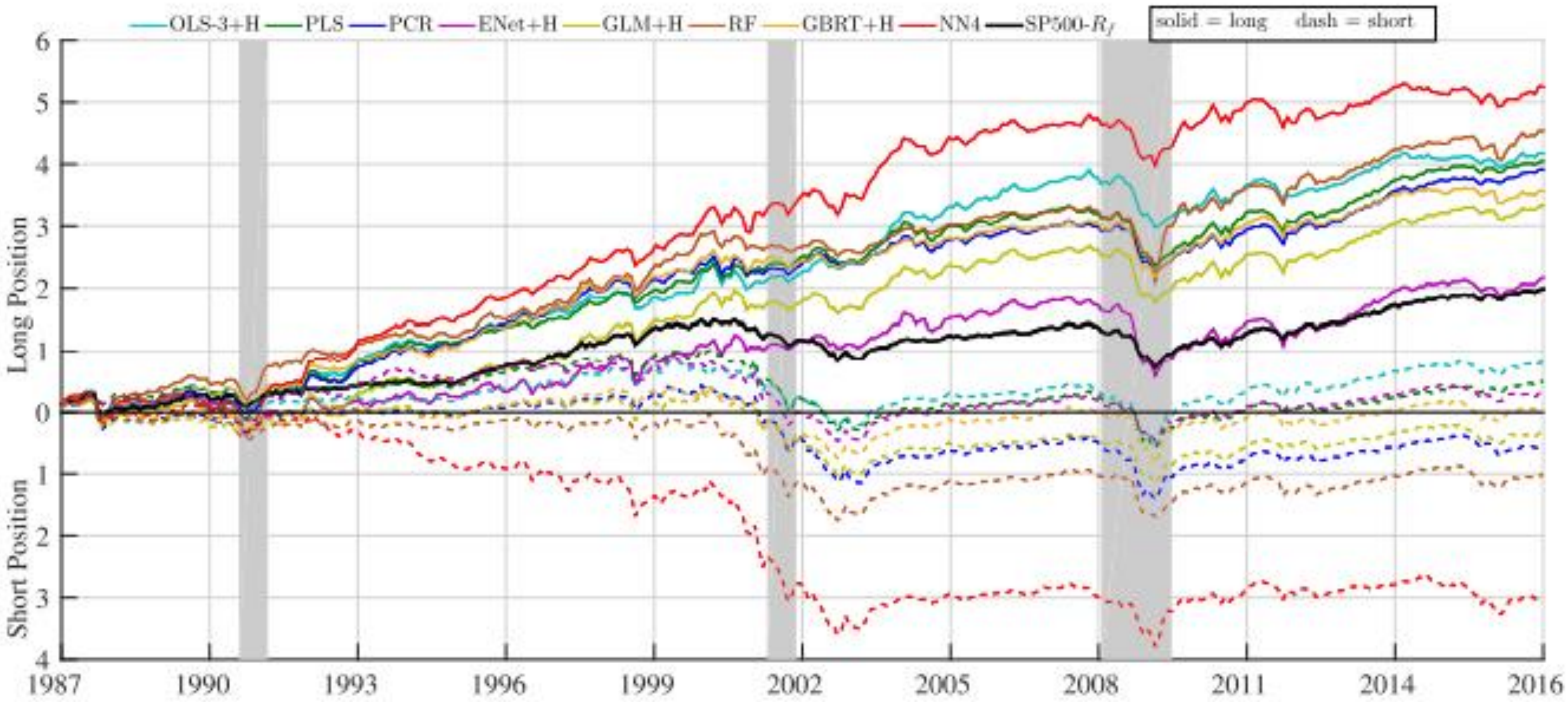


Figure 9

Cumulative return of machine learning portfolios

The figure shows the cumulative log returns of portfolios sorted on out-of-sample machine learning return forecasts. The solid and dashed lines represent long (top decile) and short (bottom decile) positions, respectively. The shaded periods show NBER recession dates. All portfolios are value weighted.

Conclusion

- “Shallow” outperforms “Deep”
- Neural Networks and Trees perform the best
- Generally agree on same set of predictors

How is this valuable from an Academic Perspective?

How is this valuable from an Investor Perspective?