Machine Learning Classification Methods and Portfolio Allocation: An Examination of Market Efficiency

Yang Bai[1]
Kuntara Pukthuanthong[1]

University of Missouri

January, 2021, AFA

(Main Slides: 33 Pages)

[1] Yang Bai: yangbai@mail.missouri.edu; Kuntara Pukthuanthong: pukthuanthongk@missouri.edu.

1. A Note on Manuscript
2. Introduction
3. Empirical Setup
4. Empirical Analysis
5. Conclusion
6. Appendix

Section 1

A Note on Manuscript

A Note on Manuscript

Our manuscript is being updated. The current manuscript does not include all of the findings mentioned in this presentation.

An updated manuscript will be available soon at the following links:

1. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3665051
2. https://kuntara.weebly.com/working-papers.html
3. https://www.yangbai-finance.com/research.html

Section 2

Introduction

Findings in the Literature about Predictability using Historical Information

1. Traditional Methods:

- Goyal and Welch 2008: Popular predictors cannot produce predictability in OOS tests.
- DeMiguel, Garlappi and Uppal 2009: Traditional methods cannot produce excess profit through predictive portfolio allocation with historical information.
- Stambaugh, Yu and Yuan 2015: Prices get corrected slower for the short legs, because limit to arbitrage in short leg is severe.

Findings in the Literature about Predictability using Historical Information

2. New Methods in Finance Machine Learning[2]:

- Rossi 2018: Goyal and Welch predictors produce OOS predictability for market returns with boosted tree models.
- Gu, Kelly and Xiu 2020: Returns are predictable in OOS tests with stock characteristics.
- Chen, Pelgers and Zhu 2020: Deep learning models adapted in the GMM framework can predicatively price stocks in OOS tests with characteristics.
- Cohen, Malloy and Nguyen 2020: Prices are lazy and information may be reflected in the prices with lags.

[2] This is not a comprehensive list of recent developments in finance machine learning literature.

Market Efficiency and Information Economics

- Market Efficiency
  - Market efficiency is defined as information efficiency, i.e., current prices reflect all information and there is no pricing error.
- Grossman and Stiglitz 1980:
  - Information efficiency is conditional.
  - Full information efficiency is impossible.
- Kyle 1985:
  - High noise reduces the information.
- O'Hara 2003:
  - Higher proportion of informed trades induces better information quality reflected by the price.
- Easley and O'Hara 2004:
  - In equilibrium, the quantity and quality of information affect asset prices.

Gaps in the Literature and Our Ideas

1. Gaps

- All the methods in the finance machine learning literature frame the asset pricing problem about risk premium explanation and return predictability as a numeric value prediction problem.
    - The modeling target is numeric value return.
    - The metrics are all error based. We cannot measure accuracy directly and thus we have no explicit measurements on how well the models perform.
    - The modeling uncertainty is hard to measure.
    - There is not enough economic intuition on the source of predictability, i.e., why and how ML models work.

Gaps in the Literature and Our Ideas (Continues)

1. Gaps (Continues)

- Despite of the close relation between predictability and market efficiency, there has not been a formal test on the EMH with the new methods including the features of:
  - OOS test setup
  - ML methods
- With numeric prediction methods, it is hard to formally test the predictability against a benchmark that is implied by the market efficiency.

Gaps in the Literature and Our Ideas (Continues)

2. Our Ideas

- Frame the asset pricing problem about risk premium explanation and return predictability as a machine learning classification problem targeting on the return states and the probabilities of future return states.
- Introduce new testing framework through predictability with binomial test as the tool to evaluate whether return state predictions by the machine learning classification models are statistically meaningful.
- By looking at return state transitions, OOS prediction accuracy and modeling uncertainty, we can better anatomize the ML models and the predictability.

Return States: Our Modeling Target

Specification of Modeling Target	
10 Return States	Criteria
1	Numeric return less than 10 percentile in a month
2	Numeric return less than 20 percentile but greater than or equal to 10 percentile in a month
3	Numeric return less than 30 percentile but greater than or equal to 20 percentile in a month
4	Numeric return less than 40 percentile but greater than or equal to 30 percentile in a month
5	Numeric return less than 50 percentile but greater than or equal to 40 percentile in a month
6	Numeric return less than 60 percentile but greater than or equal to 50 percentile in a month
7	Numeric return less than 70 percentile but greater than or equal to 60 percentile in a month
8	Numeric return less than 80 percentile but greater than or equal to 70 percentile in a month
9	Numeric return less than 90 percentile but greater than or equal to 80 percentile in a month
10	Numeric return greater than or equal to 90 percentile in a month

Section 3

Empirical Setup

Data

Coverage | Statistics

- CRSP, COMPUSTAT, Goyal and Welch Variables, FRED-MD
- Reconstruct Green et al. 2017 for a CRSP Centric Data: 3.34 Million Observations
- 332 lagged predictors
  - 101 firm characteristics
  - 2-digit SIC code
  - 2-digit SIC lagged Industry returns
  - 9 market specific predictors
  - 125 macro indicators
  - 94 anomaly long-short returns based on single sort of 94 numeric firm characteristics
- Security Coverage
  - 196301:201912
  - EXCHCD 1,2,3 and SHRCD 10,11,12

Data

- Note: We ran our models with 2 data setups, one with macroeconomic data components and one with only the characteristics augmented with industry information. Due to the similarity in the performance with or without the macroeconomic data components, we present our results with the data setup including only the characteristics augmented with industry information.

Sample Splits

- OOS tests are in spirit of Martin and Nagel 2020 and Fama and French 2018
  - IS tests can lose economic meanings (Martin and Nagel 2020)

Training and Testing Setup	IS Training	OOS Testing
Time Series Setup 1	196301:199112	199201:201912
Time Series Setup 2	199201:201912	196301:199112
Time Series Main Setup (Combined Cross-Validation)	196301:199112 and 199201:201912	199201:201912 and 196301:199112
Cross-Sectional Setup	Odd Number Months 196301:201911	Even Number Months 196302:201912

Machine Learning Models

Panel A: Architectural Specifications				
Model	Architecture	Specification	Structural Complexity	Structural Capacity
ANN1 128	Neuron Network	Multilayer Perceptron	1 Hidden Layer	# Neurons = 128
ANN1 16	Neuron Network	Multilayer Perceptron	1 Hidden Layer	# Neurons = 16
ANN1 32	Neuron Network	Multilayer Perceptron	1 Hidden Layer	# Neurons = 32
ANN1 64	Neuron Network	Multilayer Perceptron	1 Hidden Layer	# Neurons = 64
ANN2 128	Neuron Network	Multilayer Perceptron	2 Hidden Layers	# Neurons = {128,64}
ANN2 32	Neuron Network	Multilayer Perceptron	2 Hidden Layers	# Neurons = {32,16}
ANN2 64	Neuron Network	Multilayer Perceptron	2 Hidden Layers	# Neurons = {64,32}
ANN3 128	Neuron Network	Multilayer Perceptron	3 Hidden Layers	# Neurons = {128,64,32}
ANN3 64	Neuron Network	Multilayer Perceptron	3 Hidden Layers	# Neurons = {64,32,16}
ANN4 128	Neuron Network	Multilayer Perceptron	4 Hidden Layers	# Neurons = {128,64,32,16}
DART2 100	Tree	Boosting Tree	Maximum Depth = 2	# Trees = 100
DART4 100	Tree	Boosting Tree	Maximum Depth = 4	# Trees = 100
DART6 100	Tree	Boosting Tree	Maximum Depth = 6	# Trees = 100
DART8 100	Tree	Boosting Tree	Maximum Depth = 8	# Trees = 100
DRF2 200	Tree	Forest	Maximum Depth = 2	# Trees = 200
DRF4 200	Tree	Forest	Maximum Depth = 4	# Trees = 200
DRF6 200	Tree	Forest	Maximum Depth = 6	# Trees = 200
DRF8 200	Tree	Forest	Maximum Depth = 8	# Trees = 200
GBM2 100	Tree	Boosting Tree	Maximum Depth = 2	# Trees = 100
GBM4 100	Tree	Boosting Tree	Maximum Depth = 4	# Trees = 100
GBM6 100	Tree	Boosting Tree	Maximum Depth = 6	# Trees = 100
GBM8 100	Tree	Boosting Tree	Maximum Depth = 8	# Trees = 100

Metrics for Performance Evaluation

1. Economic Metrics
   - Monthly Sharpe Ratio (SR)
   - Certainty Equivalent (CEQ)
   - Cumulative Return
   - Maximum Draw-down
   - Turnover
2. Statistical Metrics
   - Overall Metrics
     - Accuracy
     - Cohen's Kappa
   - By-Class Metrics
     - Prevalence
     - Balanced Accuracy
     - Sensitivity(Recall)
     - Specificity
     - Precision
     - F1 Score

Tests

1. Factor Models: FF3F, FF3F+MOM, q4 and q5
2. Binomial Test: A Joint Test for OOS Prediction Accuracy and Market Efficiency

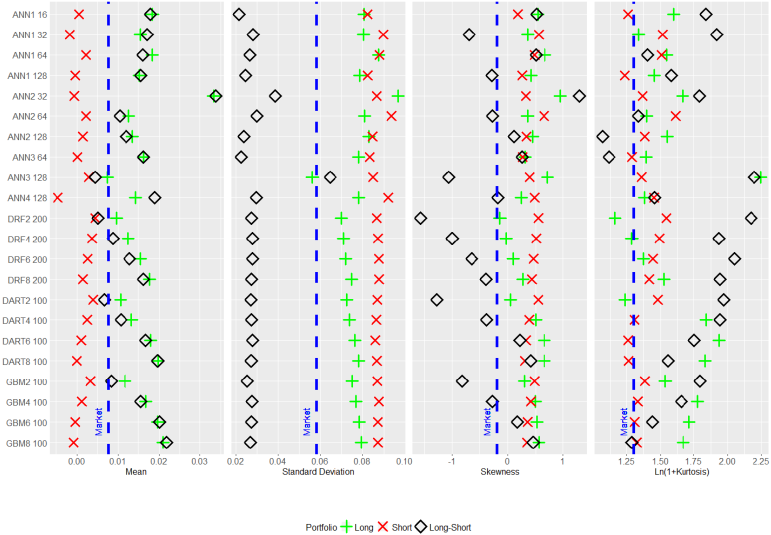
- Benchmarks:
  - Naive Classifier with No Historical Information beyond Return Distribution
  - Martingale Classifier (which predicts the future return state with the current return state.)
- Selection of No Information Accuracy Benchmark: TukeyHSD
- Test with OOS Prediction Accuracy

Section 4

Empirical Analysis

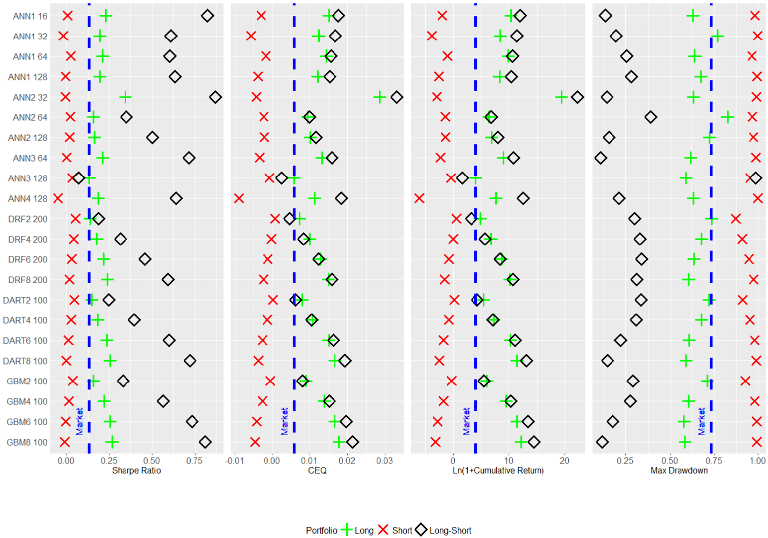
OOS EW Portfolio Performance: Returns (196301:201912)

▸ Return Distribution: An Example
▸ OOS Value Weight Performance: Returns



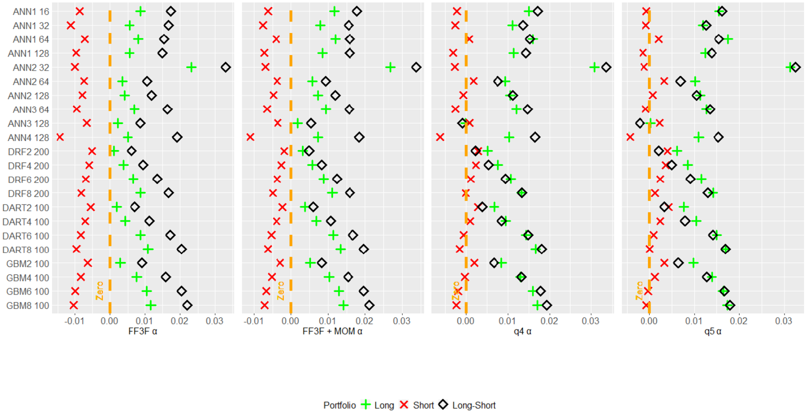
OOS EW Portfolio Performance: Economic Metrics (196301:201912)

Cumulative Return: An Example
OOS Value Weight Performance: Economic Metrics



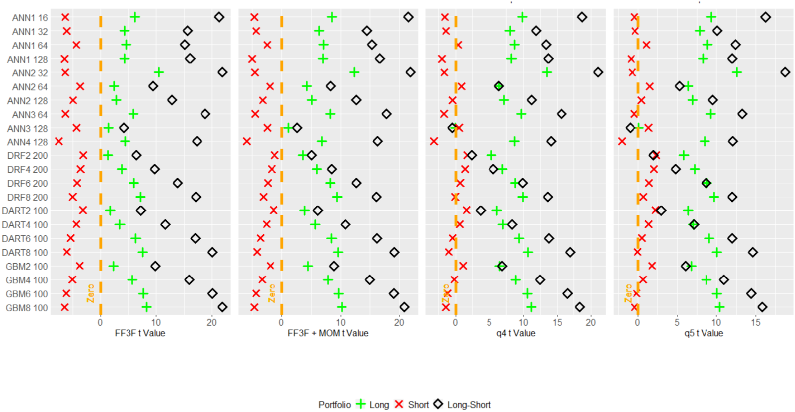
Factor Model Tests on OOS EW Portfolios: α

▸ Factor Model Tests on OOS VW Portfolios: α



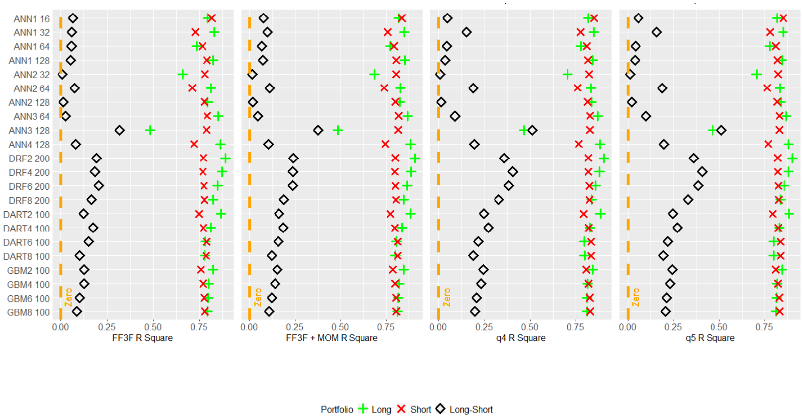
Factor Model Tests on OOS EW Portfolios: t tests on α

▸ Factor Model Tests on OOS VW Portfolios: t tests on α



Factor Model Tests on OOS EW Portfolios: Regression R^2

▸ Factor Model Tests on OOS VW Portfolios: Regression R^2



Binomial Tests

- Accuracy: The correctly predicted portion of return states.
- Correct Prediction = Success in Bernoulli Trial
- Null Hypothesis: The correctly predicted percentage by the model is the same as the correctly predicted percentage by the benchmark classifier.
- Alternative Hypothesis: The correctly predicted percentage by the model is **NOT** the same as the correctly predicted percentage by the benchmark classifier.
- A Joint Test:
  - The historical information
  - The modeling structure
  - The market efficiency/the information efficiency

Binomial Tests: On OOS Predictions

The Selection of No Information Classifier

Table 7 presents the accuracy of each model. The accuracy of a model is the direct evaluation of the correctness of the model predictions. The Kappa statistic measures the level of agreement between the predictions and the actual data and higher Kappa statistic indicates better performance. According to Landis and Koch (1977), a Kappa value greater than 0 but less than 0.2 indicates that the agreement is slight. The confidence interval is the binomial confidence interval based on accuracy. The P values are associated with the hypothesis test on whether the accuracy is different from the 2 benchmark accuracies statistically. We discussed our model specifications and the statistical metrics in Section 2. Table 7 shows that all of our models are better than the no information accuracy which is calculated under the assumption that the historical information is useless in terms of prediction future return states. All of our models are also better than the martingale accuracy, which is calculated under the assumption that the stock returns follow a memoryless process.

Model	Accuracy	Kappa	Lower 99% Bound	Upper 99% Bound	No Info Accuracy	No Info P Value	Martingale Accuracy	Martingale P Value
ANN1 16	0.153	0.059	0.153	0.154	0.102	0.000	0.117	0.000
ANN1 32	0.153	0.058	0.153	0.154	0.102	0.000	0.117	0.000
ANN1 64	0.154	0.059	0.153	0.154	0.102	0.000	0.117	0.000
ANN1 128	0.150	0.056	0.150	0.151	0.102	0.000	0.117	0.000
ANN2 32	0.153	0.058	0.153	0.154	0.102	0.000	0.117	0.000
ANN2 64	0.151	0.056	0.151	0.152	0.102	0.000	0.117	0.000
ANN2 128	0.152	0.057	0.151	0.152	0.102	0.000	0.117	0.000
ANN3 64	0.152	0.058	0.152	0.153	0.102	0.000	0.117	0.000
ANN3 128	0.151	0.056	0.150	0.151	0.102	0.000	0.117	0.000
ANN4 128	0.152	0.058	0.152	0.153	0.102	0.000	0.117	0.000
DART2 100	0.153	0.059	0.153	0.154	0.102	0.000	0.117	0.000
DART4 100	0.155	0.061	0.155	0.156	0.102	0.000	0.117	0.000
DART6 100	0.156	0.062	0.156	0.157	0.102	0.000	0.117	0.000
DART8 100	0.156	0.062	0.155	0.156	0.102	0.000	0.117	0.000
DRF2 200	0.152	0.057	0.152	0.153	0.102	0.000	0.117	0.000
DRF4 200	0.156	0.061	0.155	0.156	0.102	0.000	0.117	0.000
DRF6 200	0.157	0.063	0.157	0.158	0.102	0.000	0.117	0.000
DRF8 200	0.158	0.064	0.158	0.159	0.102	0.000	0.117	0.000
GBM2 100	0.155	0.061	0.155	0.156	0.102	0.000	0.117	0.000
GBM4 100	0.157	0.063	0.157	0.158	0.102	0.000	0.117	0.000
GBM6 100	0.158	0.064	0.157	0.159	0.102	0.000	0.117	0.000
GBM8 100	0.158	0.064	0.157	0.158	0.102	0.000	0.117	0.000

Binomial Tests: Conclusion and Implication

1. With historical information and our modeling structure, the OOS future return states are predictable.
2. All of our models deliver statistically significantly higher prediction accuracies comparing to the benchmarks.
3. There exists information about future return states in historical information.
4. Our models can generate some correct information about future return states with historical information.
5. Combining with the clear OOS economic gains, the information generated with our empirical framework can lead to trading profits and the profits are from the prediction accuracy.

Behind the High OOS Prediction Accuracies

1. Imbalance of Return State Transitions
2. Characteristics, Accuracy, Modeling Uncertainty and Market Return
3. Variable Contribution and Modeling Structures

Return State Transition Probability: the Ground Truth 196301:201912

Panel A: True Return State Transition Probability Matrix 196301:201912

	New 1	New 2	New 3	New 4	New 5	New 6	New 7	New 8	New 9	New 10
Old 1	0.1741	0.1063	0.0816	0.0686	0.0665	0.0660	0.0719	0.0817	0.1052	0.1782
Old 2	0.1137	0.1073	0.0963	0.0891	0.0879	0.0875	0.0918	0.0993	0.1090	0.1180
Old 3	0.0859	0.0987	0.0997	0.1011	0.1007	0.1033	0.1050	0.1054	0.1059	0.0944
Old 4	0.0713	0.0899	0.1007	0.1073	0.1127	0.1134	0.1122	0.1092	0.1014	0.0817
Old 5	0.0696	0.0860	0.0992	0.1094	0.1128	0.1203	0.1167	0.1098	0.0981	0.0779
Old 6	0.0690	0.0868	0.1002	0.1084	0.1138	0.1177	0.1186	0.1116	0.0970	0.0768
Old 7	0.0675	0.0897	0.1025	0.1083	0.1134	0.1163	0.1164	0.1121	0.0980	0.0758
Old 8	0.0753	0.0973	0.1054	0.1067	0.1102	0.1123	0.1092	0.1058	0.0984	0.0794
Old 9	0.0958	0.1103	0.1061	0.1023	0.0976	0.0974	0.0971	0.1009	0.0999	0.0927
Old 10	0.1742	0.1236	0.0966	0.0825	0.0752	0.0736	0.0743	0.0802	0.0912	0.1284

Panel B: Return State Transition Mean Return 196301:201912

	New 1	New 2	New 3	New 4	New 5	New 6	New 7	New 8	New 9	New 10
Old 1	-0.2744	-0.1217	-0.0750	-0.0413	-0.0130	0.0114	0.0394	0.0759	0.1350	0.4003
Old 2	-0.2424	-0.1177	-0.0716	-0.0394	-0.0127	0.0127	0.0404	0.0751	0.1292	0.3169
Old 3	-0.2316	-0.1139	-0.0691	-0.0370	-0.0121	0.0126	0.0388	0.0719	0.1243	0.2961
Old 4	-0.2254	-0.1105	-0.0661	-0.0357	-0.0108	0.0119	0.0375	0.0683	0.1193	0.2871
Old 5	-0.2229	-0.1078	-0.0624	-0.0332	-0.0099	0.0120	0.0357	0.0663	0.1174	0.2940
Old 6	-0.2185	-0.1055	-0.0614	-0.0328	-0.0094	0.0127	0.0358	0.0660	0.1165	0.2959
Old 7	-0.2123	-0.1041	-0.0623	-0.0340	-0.0103	0.0114	0.0348	0.0646	0.1128	0.2830
Old 8	-0.2097	-0.1048	-0.0627	-0.0350	-0.0110	0.0113	0.0370	0.0668	0.1168	0.2884
Old 9	-0.2120	-0.1076	-0.0664	-0.0374	-0.0125	0.0113	0.0375	0.0690	0.1207	0.2983
Old 10	-0.2321	-0.1137	-0.0707	-0.0406	-0.0135	0.0115	0.0378	0.0714	0.1276	0.3506

Return State Transition Probability: Average Model Performance 196301:201912

Table 9 presents our OOS modeling prediction *average* accuracies of return state transitions from the old states to the new states across models. Specifically, we calculate the OOS prediction accuracies of each classification model and form a percentage accuracy table similar to the table below. We then average the numbers across all the models. A stock in return state 1 means that the stock delivers a return that is among the worst performing returns of the trading month. A stock in return state 10 indicates that the stock are among the stocks delivering the best performing returns of the trading month. Details of the return state definition can be found in Table 1.

Combining what is demonstrated in Table 8, Table 9 shows that our models benefit significantly from the most certain return states, i.e., return states 1 and 10. Our models almost give up the most uncertain states, i.e. return states 3, 4, and 9.

	New 1	New 2	New 3	New 4	New 5	New 6	New 7	New 8	New 9	New 10
Old 1	0.5054	0.0228	0.0045	0.0020	0.0142	0.0446	0.0504	0.0476	0.0928	0.4449
Old 2	0.4980	0.0735	0.0122	0.0064	0.0456	0.1478	0.1309	0.1040	0.1516	0.2424
Old 3	0.4403	0.0902	0.0158	0.0125	0.0787	0.2515	0.1773	0.1052	0.1266	0.1753
Old 4	0.4098	0.0874	0.0195	0.0146	0.0995	0.3184	0.2009	0.1023	0.0993	0.1475
Old 5	0.4226	0.0832	0.0177	0.0152	0.1034	0.3350	0.2087	0.0923	0.0874	0.1404
Old 6	0.4244	0.0822	0.0200	0.0153	0.1101	0.3358	0.2028	0.0963	0.0848	0.1390
Old 7	0.4094	0.0942	0.0245	0.0159	0.1028	0.3333	0.2008	0.1025	0.0849	0.1303
Old 8	0.4376	0.1093	0.0295	0.0189	0.1016	0.3133	0.1748	0.0991	0.0876	0.1293
Old 9	0.5011	0.1472	0.0350	0.0221	0.0834	0.2463	0.1214	0.1033	0.0948	0.1270
Old 10	0.7522	0.1173	0.0257	0.0166	0.0446	0.1067	0.0586	0.0613	0.0667	0.0992

Characteristics and Prediction Accuracy at Stock Level: Regression on Prediction Accuracy (Example with GBM8 100)

The table below presents a regression of OOS prediction accuracy on characteristics across individual stocks. The OOS prediction accuracy is produced by GBM8 100 over the entire sample coverage 196301:201912. For each stocks, we calculate the prediction accuracy during its existence in our sample and we include all 95 numeric characteristics augmented with the number of appearance (n) in the sample. The regression presents a R-squared of 0.485.

Characteristics	Estimate	P Value	Characteristics	Estimate	P Value
lag_retvol	0.0685	**0.0000**	lag_chmom	-0.0068	**0.0440**
lag_betasq	0.0365	**0.0000**	lag_herf	0.0012	**0.0463**
lag_mom12m	0.0274	**0.0000**	lag_sue	0.0036	**0.0485**
lag_dolvol	0.0173	**0.0000**	lag_nincr	0.0025	**0.0576**
lag_baspread	0.0119	**0.0000**	lag_sgr	0.0026	**0.0577**
lag_dy	0.0060	**0.0000**	lag_invest	0.0023	**0.0604**
lag_zerotrade	0.0057	**0.0000**	lag_saleinv	0.0014	**0.0739**
lag_lev	0.0042	**0.0000**	lag_agr	0.0022	**0.0823**
lag_indmom	-0.0072	**0.0000**	lag_rd_mve	-0.0021	**0.0825**
lag_sp	-0.0089	**0.0000**	lag_pctacc	-0.0021	**0.0878**
lag_roeq	-0.0091	**0.0000**	lag_lgr	0.0019	**0.0992**
lag_aeavol	-0.0096	**0.0000**	lag_cinvest	-0.0027	**0.0994**
lag_disp	-0.0102	**0.0000**	lag_chnanalyst	-0.0047	0.1131
lag_ill	-0.0114	**0.0000**	lag_ear	0.0036	0.1168
lag_beta	-0.0423	**0.0000**	lag_std_turn	0.0040	0.1275
lag_mom1m	-0.1302	**0.0000**	lag_stdacc	0.0026	0.1310
lag_sich2_ret	0.0127	**0.0000**	lag_pchdepr	0.0016	0.1427
lag_mom36m	0.0105	**0.0000**	lag_currat	-0.0027	0.1503
lag_chtx	0.0097	**0.0000**	lag_pchsaleinv	0.0021	0.1604
lag_roaq	-0.0052	**0.0000**	lag_ms	-0.0013	0.2183
n	0.0000	**0.0000**	lag_pchcapx_ia	0.0014	0.2217
lag_pricedelay	-0.0064	**0.0000**	lag_bm_ia	0.0014	0.2699
lag_mom6m	-0.0154	**0.0000**	lag_chpmia	-0.0014	0.3111
lag_grltnoa	-0.0050	**0.0001**	lag_chatoia	0.0013	0.3510
lag_nanalyst	-0.0053	**0.0001**	lag_cashpr	-0.0008	0.3528
lag_mve_ia	0.0041	**0.0001**	lag_idiovol	0.0010	0.3933
lag_roavol	-0.0036	**0.0002**	lag_absacc	0.0008	0.3945
lag_maxret	-0.0142	**0.0003**	lag_gma	0.0007	0.4578
lag_stdcf	-0.0064	**0.0003**	lag_salecash	-0.0005	0.4786
lag_std_dolvol	0.0039	**0.0004**	lag_cfp	-0.0009	0.4840
lag_secured	-0.0035	**0.0006**	lag_pchquick	0.0022	0.4985
lag_fgr5yr	-0.0049	**0.0007**	lag_tang	-0.0005	0.5089
lag_cashdebt	-0.0026	**0.0009**	lag_chcsho	-0.0006	0.5126
lag_bm	-0.0031	**0.0014**	lag_chempia	0.0009	0.5715
lag_cash	-0.0033	**0.0019**	lag_depr	-0.0003	0.6221
lag_acc	0.0033	**0.0020**	lag_pchsale_pchrect	0.0005	0.6325
lag_rsup	-0.0040	**0.0020**	lag_salerec	0.0003	0.6374
lag_pchsale_pchinvt	-0.0039	**0.0040**	lag_ep	-0.0005	0.6732
lag_turn	-0.0061	**0.0049**	lag_pchgm_pchsale	-0.0004	0.6739
lag_egr	-0.0021	**0.0102**	lag_quick	0.0008	0.6802
lag_ps	0.0031	**0.0106**	lag_realestate	-0.0003	0.7120
lag_roic	-0.0021	**0.0139**	lag_grcapx	-0.0004	0.7130
lag_pchsale_pchxsga	-0.0027	**0.0198**	lag_orgcap	0.0004	0.7181
lag_hire	-0.0034	**0.0248**	lag_sfe	-0.0003	0.7761
lag_chfeps	0.0089	**0.0285**	lag_pchcurrat	-0.0003	0.9153
lag_chinv	-0.0028	**0.0293**	lag_tb	-0.0001	0.9183
lag_cfp_ia	0.0033	**0.0295**	lag_operprof	-0.0001	0.9404
lag_age	0.0019	**0.0420**	lag_rd_sale	0.0000	0.9658
(Intercept)	0.1624	**0.0000**			

Modeling Uncertainty, Accuracy and Market Return: Rolling Window Correlation (Example with GBM8 100)

Table 12 presents the rolling correlation between each pair of value weighted market return, max predicted probability and OOS prediction accuracy. The max predicted probability is the maximum of predicted probabilities across all 10 return states. The OOS prediction accuracy is the accuracy by date based on OOS prediction. Both max predicted probability and OOS prediction accuracy are from GBM8 100. All the time series are normalized across the time period 199201:201912. Note that max predicted probability across time shows the changing model uncertainty prior to the realization of the prediction results, i.e., the max predicted probability can be seen as a measure of pre-realization modeling uncertainty about future returns.

Rolling Window in Months	VW MKT and Max Predicted Probability	OOS Prediction Accuracy and Max Predicted Probability	VW MKT and OOS Prediction Accuracy
1	-0.0632	0.2243	0.1517
6	-0.2725	0.4009	0.0635
12	-0.3999	0.4403	0.0364
18	-0.4403	0.4496	0.0704
24	-0.4308	0.4572	0.1204
30	-0.4197	0.4665	0.1720
36	-0.3986	0.4704	0.2348
42	-0.3819	0.4768	0.2984
48	-0.3662	0.4907	0.3395
54	-0.3562	0.5138	0.3668
60	-0.3488	0.5417	0.3897
66	-0.3414	0.5753	0.4232
72	-0.3320	0.6038	0.4539
78	-0.3224	0.6268	0.4611
84	-0.3219	0.6459	0.4527
90	-0.3292	0.6641	0.4242
96	-0.3336	0.6824	0.3808
102	-0.3205	0.7005	0.3446
108	-0.2935	0.7196	0.3240
114	-0.2584	0.7368	0.3243
120	-0.2181	0.7508	0.3369

Behind the High OOS Prediction Accuracies: Implications

- The efficiency level for different return states may be different.
- Stock characteristics have significant relation with OOS prediction accuracy.
- Modeling uncertainty, accuarcy and market return have meaningful relation.
- We also checked variable contributions with both TS and CS setups.
    - Historical trading information contributes the most to the tree models.
    - Theoretical risk exposure contributes to the models.
    - Corporate announcement related information makes contribution too.
    - Macro variables also contribute to the models but the contribution is very limited.

Section 5

Conclusion

Summary and Contribution: Setup

- We are the first to frame asset pricing problem as classification problem.
  - We focus on probability of future return states and measure accuracy directly.
  - We demonstrate 22 models, 2 time window setups and 2 data setups.
  - We show that the portfolios based on our classification predictions realize significant economic gains in OOS time period.

Summary and Contribution: Tests

- We introduce a new explicit empirical test of market efficiency with new methods, the machine learning classification methods, through the predictability as the bridge.
    - We are the first to introduce the binomial test as a tool to examine market efficiency.
    - The OOS prediction accuracies are statistically significantly higher than the 2 benchmark accuracies, questioning the correctness of prices.
    - Our models can generate correct information about future return states with historical information.

Summary and Contribution: Behind Predictability

- Return state transitions are not uniform. The transition probability implies that the market efficiency level can be different for different return states.
- Our models take the advantage of the imbalance of the return state transition probability.
- Characteristics are significantly associated with OOS prediction accuracy.
- Model uncertainty, OOS prediction accuracy and the market return have a complex correlation.
- Variable importance of our models questions the weak form and semi-strong form EMH. The fact our model can generate useful information from historical information implies the possibility of creating private information with analytical tools.

Section 6

Appendix

Coverage

Go Back

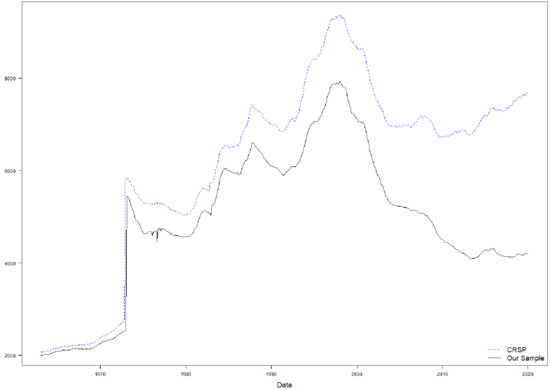


Figure 1 Number of Stocks in CRSP vs Number of Stocks in Our Sample 196301:201912

Figure 1 presents a comparison of the sample coverage between our data set and the CRSP database. The dashed line represents the number of securities included in the CRSP database and the solid line represents the number of stocks included in our sample. Note that CRSP is a general security database. It includes securities other than stocks of the public firms. In our sample, we include only the stocks listed on NYSE, Amex, and NASDAQ. This figure presents the comparison from January 1963 to December 2019. In total, our sample covers distinct 26302 stocks. On average, our sample covers around 4887 stocks for every trading month. The detailed summary statistics of the sample coverage can be found in Table 4.

Summary Statistics

Go Back

Table 4 presents the summary statistics of our data with CRSP database as the reference. Panel A presents number of securities in our sample. Panels B and C present summary statistics and market capitalization in month t-1, respectively.

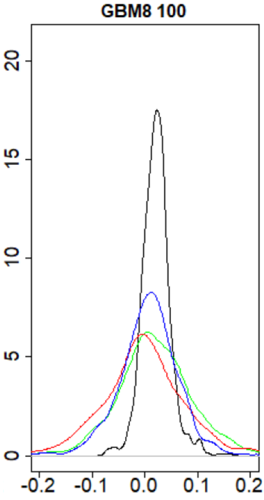
Panel A: Number of Securities Summary					
Sample	Distinct Total	Mean	Min	Max	Filter
CRSP	33004	6146.905	2069	9366	None
Our Sample	26302	4886.6754	1997	7929	No missing return; EXCHCD and SHRCD

Panel B: Summary Statistics of Returns					
Sample	Mean	SD	Skewness	Kurtosis	Filter
CRSP	0.0102	0.176	20.8963	5165.1519	No missing return
Our Sample	0.0109	0.1883	20.6107	4785.8618	No missing return; EXCHCD and SHRCD

Panel C: Summary Statistics of Market Capitalization at t-1					
Sample	Mean	SD	Skewness	Kurtosis	Filter
CRSP	1601233.289	11003218.89	27.6035	1369.1445	No missing t-1 ME
Our Sample	1723086.927	11923239.69	26.0793	1202.901	No missing t-1 ME; EXCHCD and SHRCD

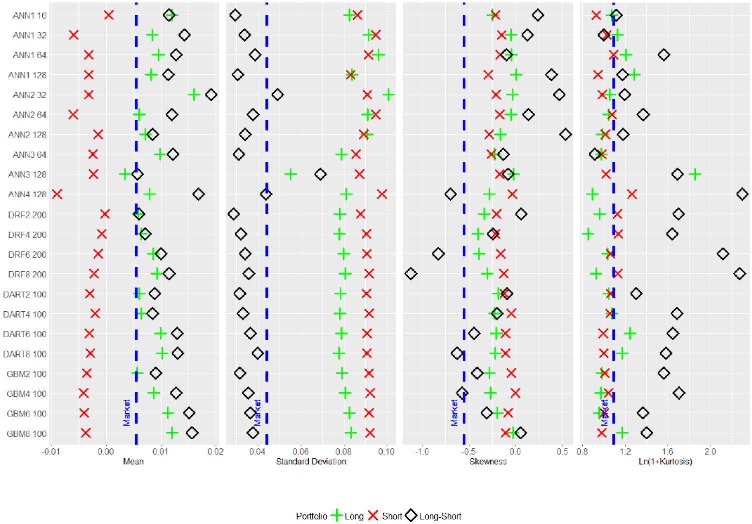
GBM8 100 based OOS EW Portfolio Return Distribution (196301:201912)

Go Back



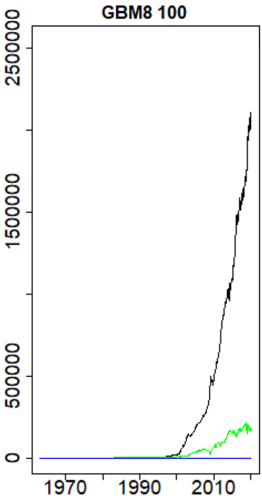
OOS VW Portfolio Performance: Returns (196301:201912)

Go Back



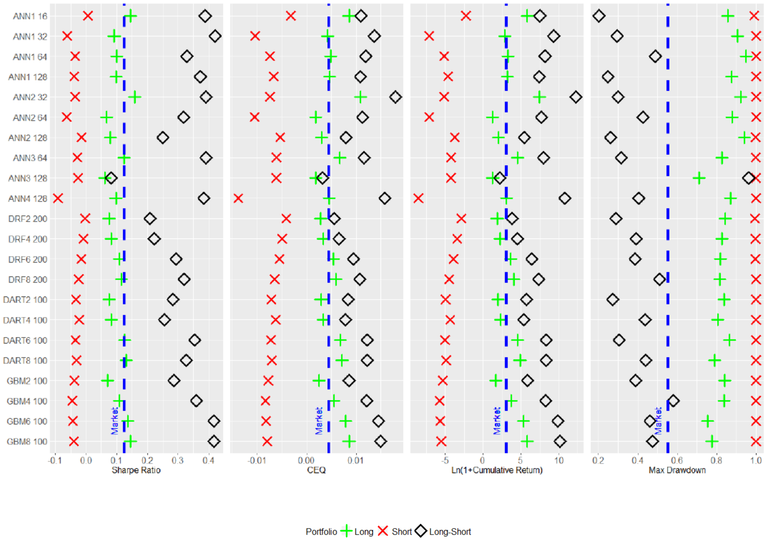
GBM8 100 based OOS EW Portfolio Cumulative Return (in 100%) (196301:201912)

Go Back



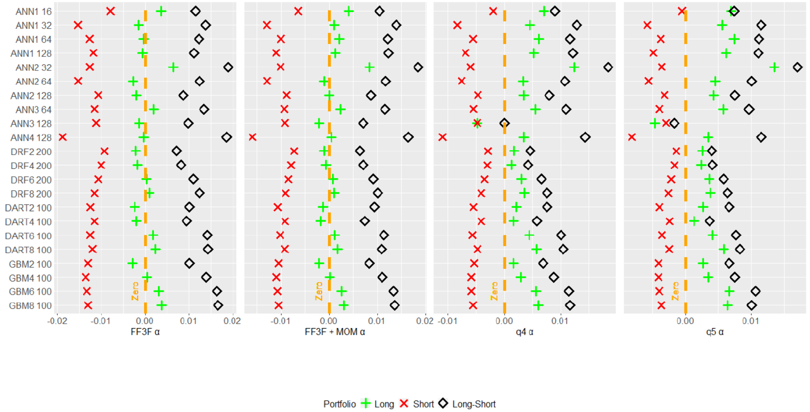
OOS VW Portfolio Performance: Economic Metrics (196301:201912)

Go Back



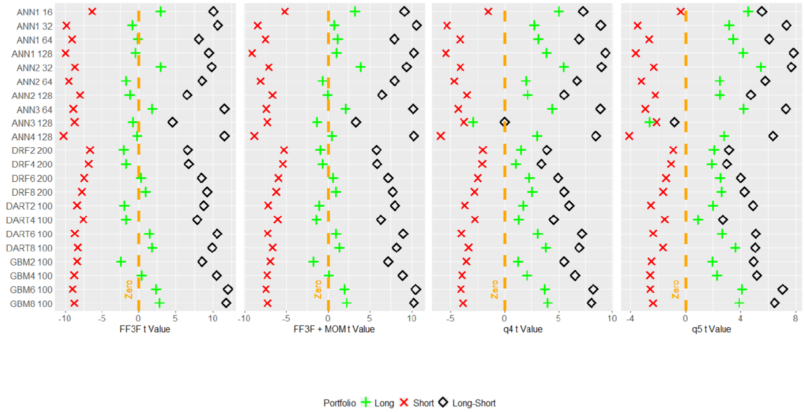
Factor Model Tests on OOS VW Portfolios: α

Go Back



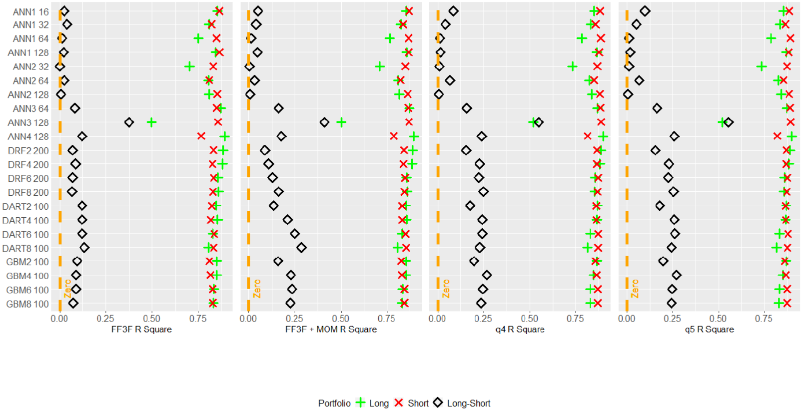
Factor Model Tests on OOS VW Portfolios: t tests on α

Go Back



Factor Model Tests on OOS VW Portfolios: Regression R^2

Go Back



Binomial Tests: On OOS Predictions

Go Back

Table 3 presents the Tukey's HSD multiple comparison test with Monte Carlo simulation. The testing samples are generated with the sample covering 199201:201912. Classifier 1 is the random classifier that assigns return states with equal probability. Classifier 2 is the random classifier that assigns return states with IS sample probability mass function observed in the sample 196301:199112. Classifier 3 is the naïve classifier that assigns the most populated IS return state to all OOS observations. Classifier 4 is the random classifier that assigns return states with OOS sample probability mass function with equal probability. Classifier 5 is the naïve classifier that assigns the most populated OOS returns state to all OOS observations. Note that even with minimum information, the naïve classifier which assigns the most populated IS return state to all OOS observations demonstrates higher accuracy than random classifier that uses no information. To provide a comprehensive evaluation of the proper benchmarks, we also consider the martingale hypothesis about return process, i.e., the best prediction for the future return is today's return and the return process is a memoryless process. We produce Classifier 6 to account for the martingale hypothesis by predicting the future return state with the current return state. Because of introducing historical return state information, Classifier 6 has better overall accuracy in our simulation. We include both Classifier 5 and Classifier 6 as our benchmarks in our binomial tests.

	Difference	Lower 95% Bound	Upper 95% Bound	P Value
1-2	0.0000	-0.0002	0.0002	1.0000
1-3	0.0004	0.0002	0.0006	0.0000
1-4	0.0000	-0.0001	0.0002	0.9671
1-5	0.0005	0.0004	0.0007	0.0000
1-6	0.0201	0.0199	0.0203	0.0000
2-3	0.0004	0.0002	0.0005	0.0000
2-4	0.0000	-0.0001	0.0002	0.9836
2-5	0.0005	0.0004	0.0007	0.0000
2-6	0.0201	0.0199	0.0203	0.0000
3-4	-0.0003	-0.0005	-0.0002	0.0000
3-5	0.0002	0.0000	0.0003	0.1138
3-6	0.0197	0.0196	0.0199	0.0000
4-5	0.0005	0.0003	0.0007	0.0000
4-6	0.0201	0.0199	0.0202	0.0000
5-6	0.0196	0.0194	0.0197	0.0000