Machine-Learning the Skill of Mutual Fund Managers

Ron Kaniel [1], Zihan Lin [2], Markus Pelger [2], Stijn Van Nieuwerburgh [3]

[1]Rochester University, [2]Stanford University, [3]Columbia University

Research question

Key questions:

- Do mutual fund managers have skills and outperform their benchmarks?
- If yes, can we predict which and when mutual fund managers have skills?

Challenges:

- Big Data: Information about mutual funds is high-dimensional.
- Non-parametric: Skill can depend in a complex way on the information set.
- Time variation: Skill can depend on time-varying economic conditions.

Our solution: A machine learning approach

- Machine learning methods are very flexible and deal with big data.
- We predict risk-adjusted fund performance with neural networks
- We use large set of fund and holding-based stock characteristics and macroeconomic information

Literature (Partial List)

Machine learning in finance: Gu et al. (2020), Freyberger et al. (2020), Chen et al. (2019), Bryzgalova et al. (2019).

ML for funds: Independent work included as subset of our analysis:

- Li and Rossi (2021): predict returns (orthogonal objective:we predict abnormal returns), only stock characteristics (we find fund characteristics important for identifying skill), no macro information as predictor (we uncover interaction effects with sentiment)
- DeMiguel et al. (2022): abnormal returns, only fund charact., no macro

Fund return predictability: Gruber (1996), Zheng (1999), Sapp and Tiwari (2004), Carhart (1997), Bollen and Jeffrey (2005), Kacperczyk et al. (2005)

Performance and macroeconomic conditions: Stambaugh et al. (2012), Moskowitz (2000), Kosowski (2011), Kacperczyk et al. (2014, 2016)

Fund performance: Berk and Green (2004), Fama and French (2010).

Skill of fund managers: Risk-adjusted returns

Object of interest: Abnormal returns:

$$\begin{aligned} R_{i,t}^{abn} &= R_{i,t} - F_t\hat{\beta}_{i,t-1}, \\ R_{i,t-36:t-1} &= \alpha_i + F_{t-36:t-1}\hat{\beta}_{i,t-1} + \eta_{i,t-36:t-1} \end{aligned}$$

- Monthly abnormal returns measure the skill of fund managers
- Abnormal returns are return residuals after subtracting Carhart (1997) 4-factor exposures
    - Results robust to addition of other risk factors.
- Rolling window regression to capture time-variation

Information set

Actively managed equity mutual funds:

- 407,158 observations for 3,275 funds from 1980/01 to 2019/01.

Holding based stock characteristics

- 46 stock characteristics, weighted by mutual fund holdings.
- Six groups: Past returns, Investment, Profitability, Intangibles, Value, and Trading Frictions.

13 Fund-specific characteristics

- Fund Residual Momentum: F_ST_Rev, F_r2_1, F_r12_2
- Fund Characteristics: age, tna, flow, exp_ratio, turnover ratio
- Fund Family Characteristics: Family_tna, fund_no, Family_r12_2, Family_age, Family_flow

Macroeconomic state variables

- Investor sentiment as in Baker and Wurgler (2006)
- Chicago Fed National Activity Index (CFNAI) for real activity

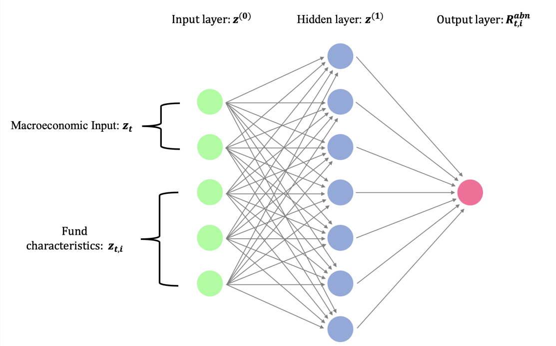
Machine Learning Solution

- Predict fund abnormal returns with a neural network of lagged predictors:

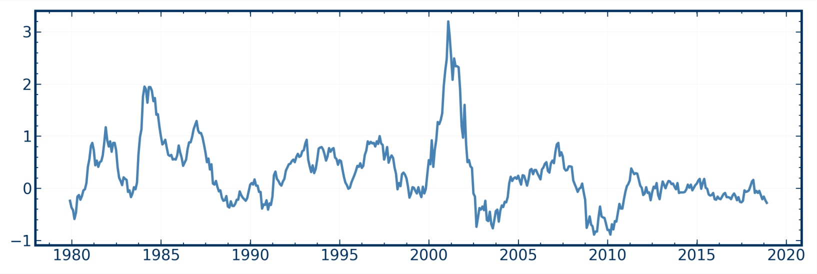
$$R^{abn}_{i,t+1} = g(z_{i,t}, z_t) + \epsilon_{i,t+1}$$

$\Rightarrow$ Estimate skill (abnormal return) conditional on fund specific information $z_{i,t}$ and macro states z_t

- Neural networks can reliably estimate a complex functional relationship among a large set of variables.
- Illustration of Feedforward Network with Single Hidden Layer



Time series of macroeconomic states



Sentiment

Problem: All economic conditions must be represented in all subsamples.

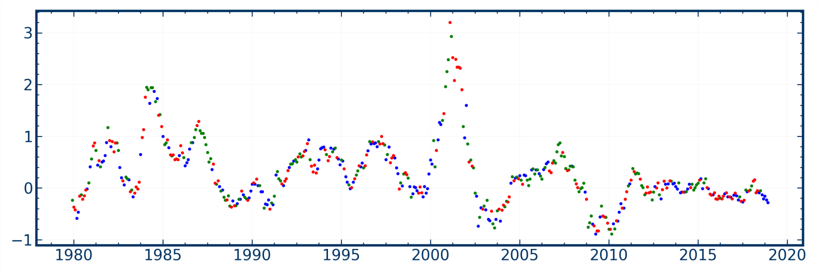
Our solution: Cross-out-of-sample evaluation

- Randomly split full sample into three periods. In each fold, use one period for out-of-sample evaluation and combine data from the other two for training and validation.
- Combine models estimated on each of the three folds.

⇒ Each and every data point is evaluated out-of-sample.

⇒ High and low macro states in all training and evaluation samples.

Time series of macroeconomic states



Sentiment for cross-out-of-sample folds

Problem: All economic conditions must be represented in all subsamples.

Our solution: Cross-out-of-sample evaluation

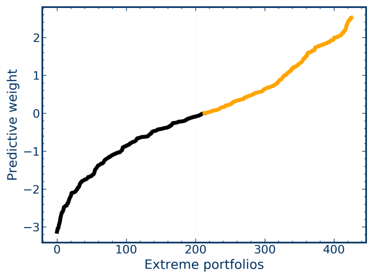
- Randomly split full sample into three periods. In each fold, use one period for out-of-sample evaluation and combine data from the other two for training and validation.
- Combine models estimated on each of the three folds.

⇒ Each and every data point is evaluated out-of-sample.

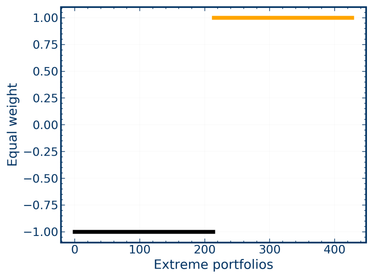
⇒ High and low macro states in all training and evaluation samples.

Prediction-sorted portfolios

- Prediction sorted portfolios measure the predictability of skill:
  Sort funds into deciles based on neural network prediction
- Equally-weighted portfolios only use the ranking of the prediction signal.
- Prediction-weighted portfolios use both the ranking and relative magnitude.



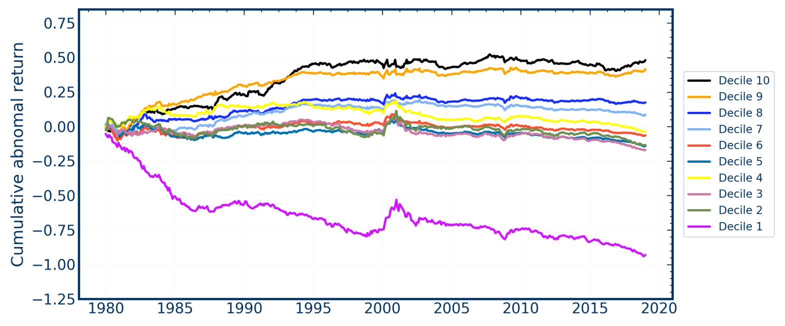
(a) Prediction weights



(b) Equal weights

Portfolio weights for top and bottom deciles

All characteristics + sentiment

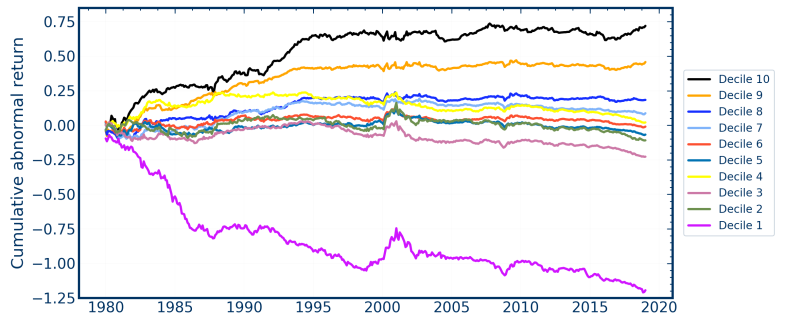


Equally-weighted deciles

Abnormal returns are predictable

- 10% best funds: cumulative abnormal return of 48% equally-weighted.
- 10% worst fund: cumulative abnormal return of -93% equally-weighted.

All characteristics + sentiment



Prediction-weighted deciles

Abnormal returns are predictable

- 10% best funds: cumulative abnormal return of 48% equally-weighted.
- 10% worst fund: cumulative abnormal return of -93% equally-weighted.

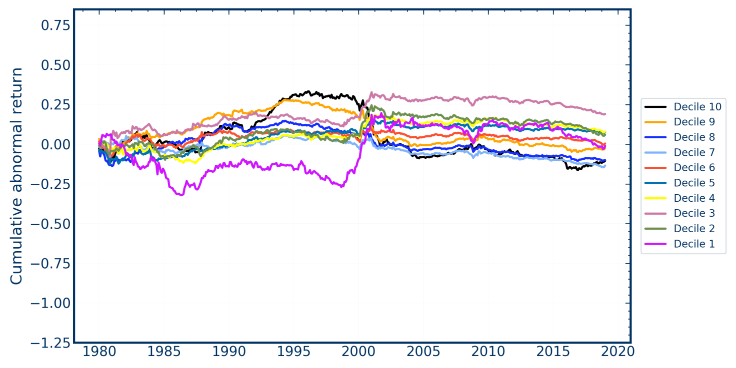
Prediction-weights better capture economic benefits:

- 10% best funds: cumulative abnormal return of 72% prediction-weighted
- 10% worst fund: cumulative abnormal return of -119% prediction-weighted

⇒ Avoiding the worst mutual funds more valuable than investing in the best.

⇒ After fees: Same spread, cumulative abnornal return of 37% for top 10%.

Information set comparison



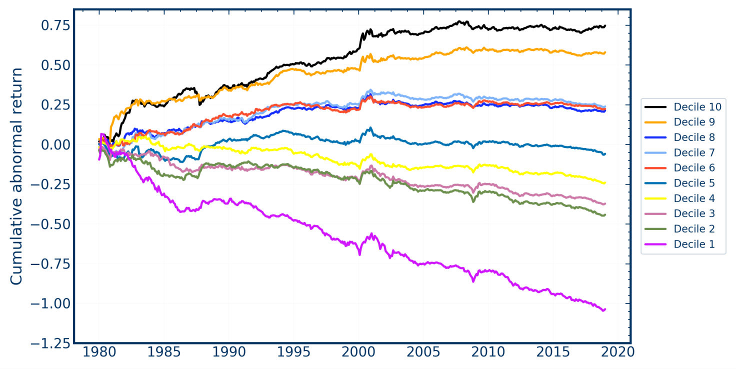
Stock-specific characteristics only

- Compare information sets for the same flexible machine learning method: Stock-specific or fund-specific characteristics

⇒ Holding-based stock characteristics not predictive for abnormal returns

⇒ Fund-specific characteristics predict abnormal returns

Information set comparison



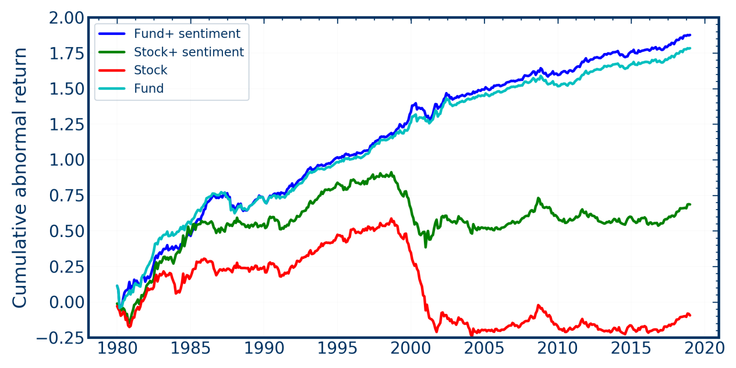
Fund-specific characteristics

- Compare information sets for the same flexible machine learning method: Stock-specific or fund-specific characteristics

$\Rightarrow$ Holding-based stock characteristics not predictive for abnormal returns

$\Rightarrow$ Fund-specific characteristics predict abnormal returns

Information set comparison



Cumulative abnormal returns of long-short prediction decile portfolios

- Spread in skill: Long-short prediction portfolio of top and bottom deciles
- Economic measure of skill - not necessarily tradable investment strategy
- Stock characteristics are not predictive.
- Fund characteristics and sentiment are extremely useful for prediction.

Information set comparison: A refinement

Information set	mean (%)	t-stat	SR	R^2_F (%)
Stock	-0.02	-0.2	-0.01	-1.60
Stock+ sentiment	0.15	1.6	0.07	1.27
Stock+ fund	0.28	3.3***	0.15	2.30
Stock+ fund+ sentiment	0.41	4.5***	0.21	5.00
Fund	0.38	5.5***	0.25	0.19
Fund+ sentiment	0.40	5.4***	0.25	2.73
Fund momentum + Flow + sentiment	0.48	5.2***	0.24	0.92

Statistics of long-short prediction decile portfolios

Rank vs level prediction:

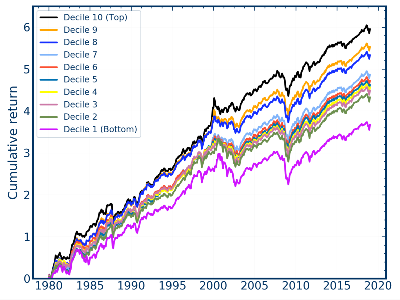
- R^2_F measures how well the realized long-short portfolio return is predicted.

⇒ Sentiment improves level prediction without changing relative ranking

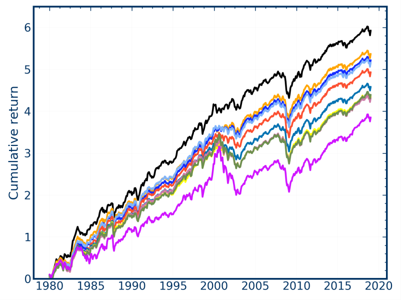
Which information matters?

- Stock-specific information: Low Sharpe ratios and insignificant spread
- Fund-specific characteristics necessary for high Sharpe ratios and significant spreads
- Fund momentum, flow + sentiment capture most of the relative ranking

Predicting returns vs. abnormal returns



Fund-specific characteristics + sentiment



Stock-specific characteristics + sentiment

Return prediction different from abnormal return prediction

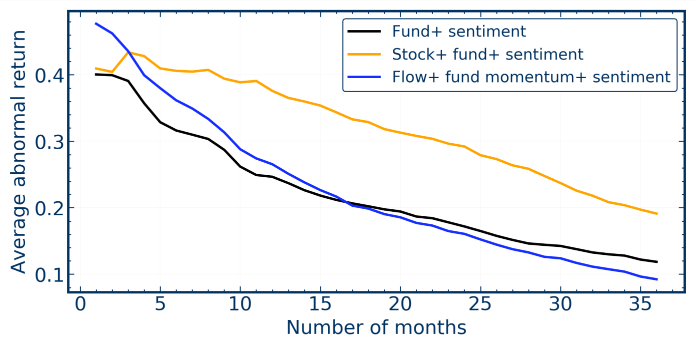
- Returns have a strong market component (level effect)
- Returns are predictable by stock and fund characteristics

⇒ Stock characteristics predict systematic factor component

Abnormal return prediction is relative objective and preferable:

- Higher Sharpe ratio and R_F^2 than for abnormal returns than for returns
- Return prediction: SR=0.14, R_F^2=-26.54
- Abnormal return prediction SR=0.21, R_F^2=5.00

Holding Periods



Mean of abnormal returns of long-short portfolios for different holding periods

Longer holding periods:

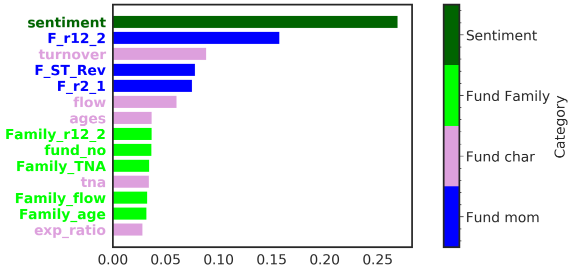
- Investments held for longer periods (based on 1 month prediction)
- Same results for Sharpe ratios and t-statistics

$\Rightarrow$ Predictability lasts over long time horizons.

Better performance when directly predicting for longer horizons:

- Monthly rebalancing not crucial for high abnormal returns
- Annual rebalancing: Mean 0.31 and $SR = 0.27$

Which variables are important?



Top variable importance

Variable importance measure:

$$\text{Sensitivity}(z_k) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{\partial \hat{R}_{i,t}^{abn}}{\partial z_{i,k,t}}\right)^2}$$

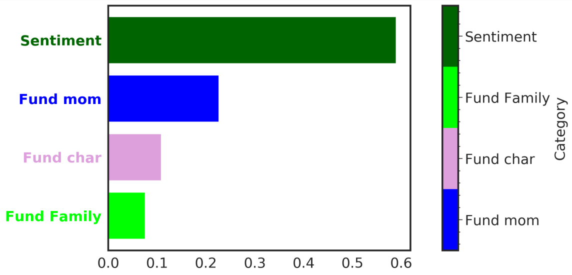
- Generalizes slopes in linear regression models

⇒ Sentiment, fund momentum, turnover and flow most important

Formal statistical significance tests for neural networks:

⇒ Leading variables are statistically significant!

Which variables are important?



Top variable group importance

Variable importance measure:

$$\text{Sensitivity}(z_k) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{\partial \hat{R}_{i,t}^{abn}}{\partial z_{i,k,t}}\right)^2}$$

- Generalizes slopes in linear regression models

$\Rightarrow$ Sentiment, fund momentum, turnover and flow most important

Formal statistical significance tests for neural networks:

$\Rightarrow$ Leading variables are statistically significant!

Results in different sentiment terciles

	T^L				T^M				T^H			
Portfolio	SR	mean	t-stat	R_F^2	SR	mean	t-stat	R_F^2	SR	mean	t-stat	R_F^2
D10-D1	0.12	0.23	1.6	0.50	0.37	0.42	4.6***	3.39	0.32	0.55	4.0***	4.83
D1	-0.11	-0.18	-1.4	0.71	-0.25	-0.23	-3.1***	3.65	-0.23	-0.29	-2.9***	1.35
D10	0.05	0.04	0.6	-0.86	0.22	0.19	2.7***	1.00	0.21	0.27	2.6**	2.68

Prediction deciles conditional on sentiment terciles

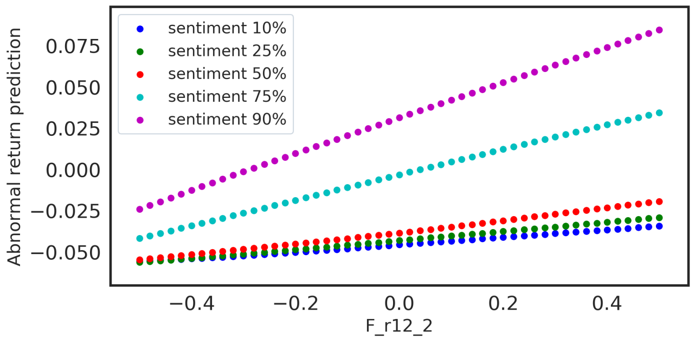
Conditional mean depends on sentiment:

- Highest Sharpe ratio earned in medium and high sentiment periods.
- During high sentiment state, long-short portfolio earns more than twice the expected return compared to low sentiment state.

Predictability depends on sentiment:

- Abnormal returns most predictable in medium and high sentiment periods.
- Market timing strategy: Investing into top funds during high sentiment: Earns an average monthly abnormal return of 0.27%.

Interactions with sentiment



Conditional abnormal returns $g(z_{i,t}, z_t)$

- Conditional abnormal fund returns as function of momentum for different sentiment quantiles
- 2-dimensional function by keeping other covariates at their mean.

⇒ Nonlinear interaction effects with sentiment.

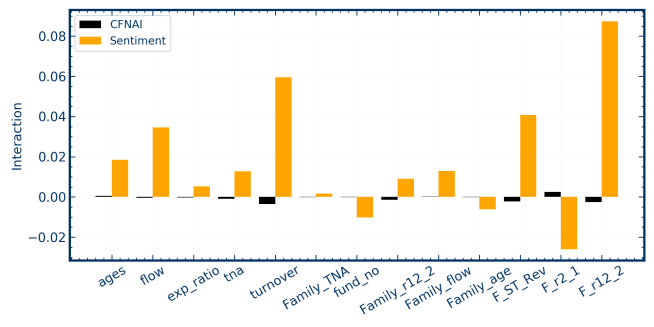
⇒ Similar results for other fund characteristics.

Interaction measure

New interaction measure:

- Differences in slopes for high and low macroeconomic states:

$$\text{Interaction}(z, \text{macro}) = \left(\hat{R}^{abn}(\text{high } z, \text{ high macro}) - \hat{R}^{abn}(\text{low } z, \text{ high macro})\right) - \left(\hat{R}^{abn}(\text{high } z, \text{ low macro}) - \hat{R}^{abn}(\text{low } z, \text{ low macro})\right).$$



- Momentum, reversal, turnover and flow strongly interact with sentiment
  - We show statistical significance for interaction effects with sentiment
- CFNAI does not interact with fund characteristics

Economic Mechanism

Flow predicts performance

- Investors detect skill and reallocate investment.

Fund momentum predicts performance

- Reallocation slower than in frictionless model Berk and Green (2004)
- Skill leaves trail through gradual flows

Interaction with sentiment:

- Funds attract flows through marketing
  $\Rightarrow$ buying pressure for the stocks funds hold
- Downward-sloping demand curve
  $\Rightarrow$ raises prices and lifts fund returns
- Creates more inflows next period
  $\Rightarrow$ stronger in high sentiment periods.

More results

Chronological cross-out-of-sample analysis: chronological

- Predictability and performance robust to sampling
- Economic model depends on which sentiment states are observed

Fund fees: fee

- Spread in abnormal returns not explained by fees
- Prediction results hold net-of-fees

Decomposition of abnormal returns: decomposition

- Decomposition of abnormal returns into between-disclosure (between quarter) and within-disclosure component (within quarter)
- 50% comes from active trading within quarter, 50% with fixed holdings

Spanning tests: spanning

- Outperformance is not compensation for standard risk factors
- Important: Time-series regression ex post on prediction portfolios different from ex ante local regressions to obtain abnormal returns

Robustness to fund size: size

- Results robust to excluding small funds or value-weighting

Tuning parameters: implementation

- Results are very robust to network structure and tuning parameters

Conclusion

Empirical results

1. Strong predictability: Predictability of fund performance (i.e., risk-adjusted returns) is out-of-sample, long-lived and economically meaningful.
2. Variable selection: Identify fund flow and fund residual momentum as key predictors. Characteristics of stock holdings are not predictive.
3. Macro interaction: Fund flow and residual momentum matter more when sentiment is high. No interaction effects with CFNAI.

Methodology Contributions

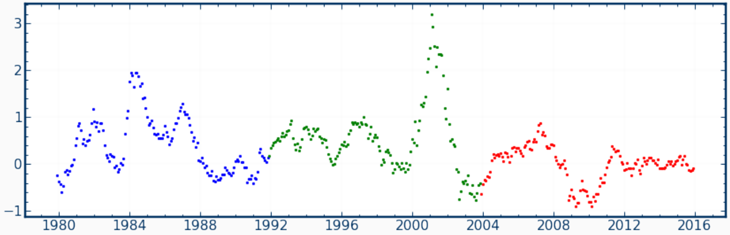
1. Prediction methods:
    - Abnormal returns economically motivated and statistically better target.
    - Prediction-weights better reflect economic benefits.
2. Model evaluation:
    - New method for out-of-sample evaluation with macroeconomic states.
    - Novel measure for interaction, including statistical significance test.
3. Protocol: Vary information set to compare economic benefits.

Appendix

Additional Results

- Chronological cross-out-of-sample analysis chronological
- Information set comparison set
- Predicting returns vs. abnormal returns return
- Statistical significance tests test
- After fee performance fee
- Longer holding periods holding
- Persistence of fund characteristics and classification persistence
- Separate results for deciles deciles
- Decomposition of abnormal returns decomposition
- Which macroeconomic variable? macro
- Spanning tests spanning
- Simplified model simple
- Robustness to fund size size
- Data and implementation summary
- Missing data missing

Chronological cross-out-of-sample

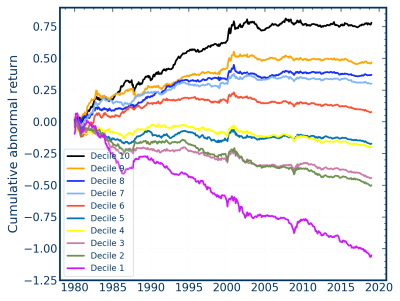


- Split full sample into three periods, two for training and validation and third for out-of-sample evaluation.
- Combine models estimated on each of the three folds.

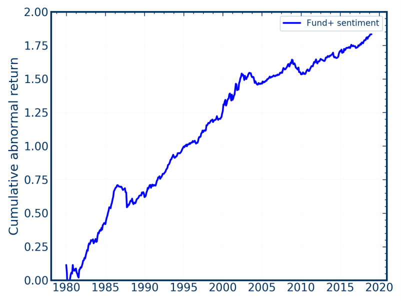
⇒ Each and every data point is evaluated out-of-sample.

⇒ High and low sentiment periods are not represented in all folds.

Chronological cross-out-of-sample: Prediction-sorted portfolios



(a) Prediction-weighted deciles



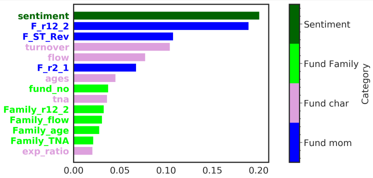
(b) Long-short prediction decile

Sampling	mean(%)	t-stat	SR	R_F^2 (%)
Chronological folds	0.39	5.0***	0.23	1.47
Random folds	0.40	5.4***	0.25	2.73

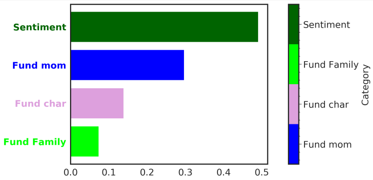
- Fund-specific characteristics + sentiment as input.

⇒ Predictability and economic significance robust to sampling

Chronological cross-out-of-sample: Variable importance



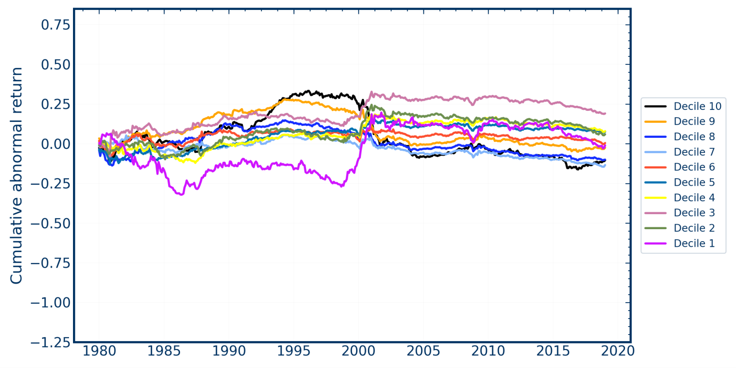
(a) Top variable importance.



(b) Top variable group importance.

- High sentiment not present in the third fold
- As expected lower importance and interaction effects under chronological sampling

Information set comparison



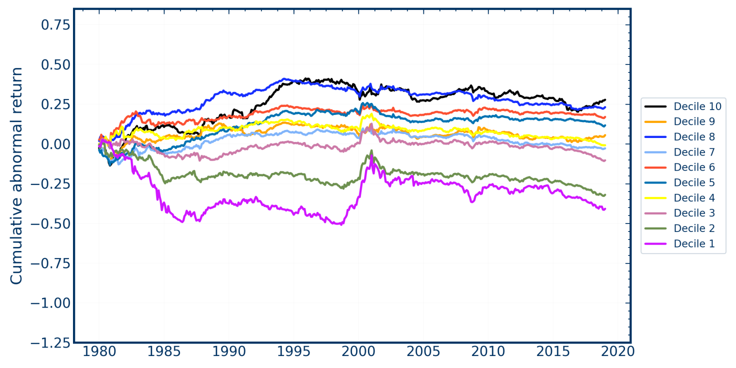
Stock-specific characteristics only

- Compare 4 information sets for the same flexible machine learning method: Stock-specific or fund-specific characteristics with/without sentiment

⇒ Holding-based stock characteristics not predictive for abnormal returns

⇒ Fund-specific characteristics and sentiment predict abnormal returns

Information set comparison



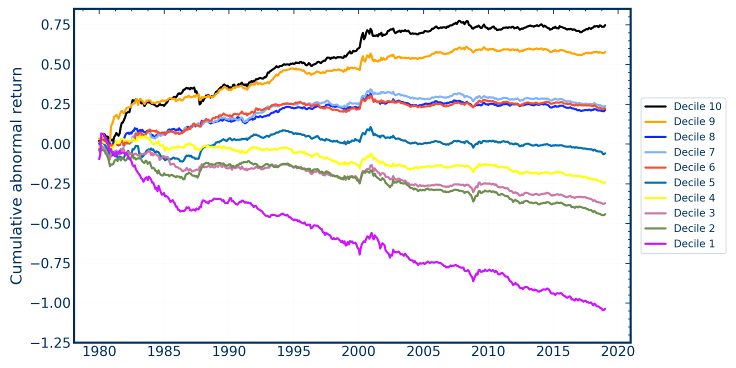
Stocks-specific characteristics + sentiment

- Compare 4 information sets for the same flexible machine learning method: Stock-specific or fund-specific characteristics with/without sentiment

⇒ Holding-based stock characteristics not predictive for abnormal returns

⇒ Fund-specific characteristics and sentiment predict abnormal returns

Information set comparison



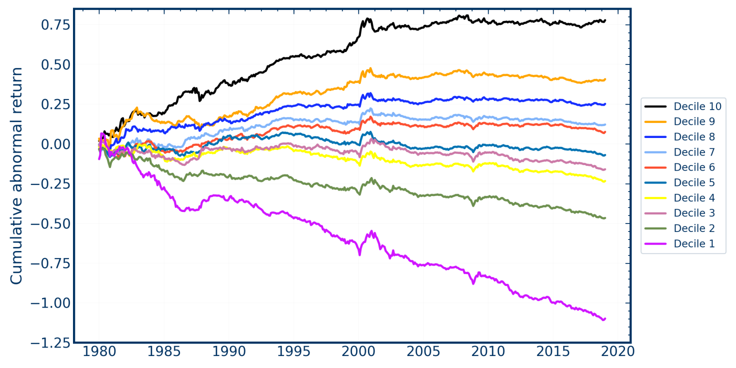
Fund-specific characteristics

- Compare 4 information sets for the same flexible machine learning method: Stock-specific or fund-specific characteristics with/without sentiment

⇒ Holding-based stock characteristics not predictive for abnormal returns

⇒ Fund-specific characteristics and sentiment predict abnormal returns

Information set comparison



Fund-specific characteristics + sentiment

- Compare 4 information sets for the same flexible machine learning method: Stock-specific or fund-specific characteristics with/without sentiment

⇒ Holding-based stock characteristics not predictive for abnormal returns

⇒ Fund-specific characteristics and sentiment predict abnormal returns

Predicting returns vs. abnormal returns

Data	mean(%)	t-stat	SR	R^2_F (%)
Fund+ sentiment	0.49	3.0***	0.14	0.97
Fund	0.53	3.5***	0.16	0.97
Stock+ sentiment	0.44	3.1***	0.14	-20.03
Stock	0.11	1.1	0.05	-53.21
Stock+ fund+ sentiment	0.45	3.1***	0.14	-26.54

Statistics of long-short prediction decile portfolios based on returns

- Returns are predictable by stock and fund characteristics
- Stock characteristics predict returns but not abnormal returns.
- Sharpe ratio of abnormal return long-short portfolio is higher
- The level of fund returns hard to predict

$\Rightarrow$ Abnormal return prediction is relative objective and preferable: higher Sharpe ratio and R^2_F than for returns

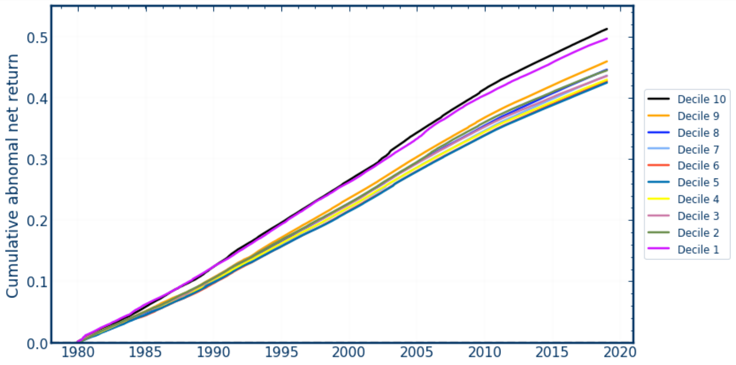
Significance tests

Fund char	Sensitivity	Interaction w. sentiment
sentiment	0.14***	
F_r12_2	0.08***	0.09***
turnover	0.05***	0.06***
F_ST_Rev	0.04***	0.04***
F_r2_1	0.04***	-0.03***
flow	0.03***	0.03***
ages	0.02***	0.02***
fund_no	0.02***	-0.01**
tna	0.02***	0.01**
Family_r12_2	0.02***	0.01
Family_flow	0.02**	0.01***
Family_TNA	0.02**	0.00
Family_age	0.02*	-0.01
exp_ratio	0.01	0.01

Formal statistical significance tests for neural networks:

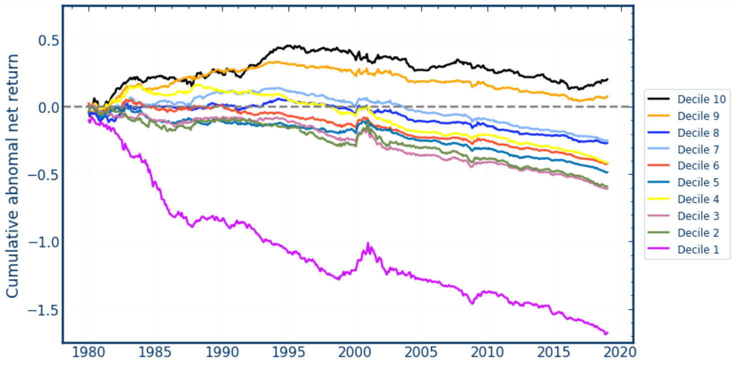
- Apply and extend the significance tests of Horel and Giesecke (2020).
- Most important fund characteristics and sentiment interactions are statistically significant.

After fee performance



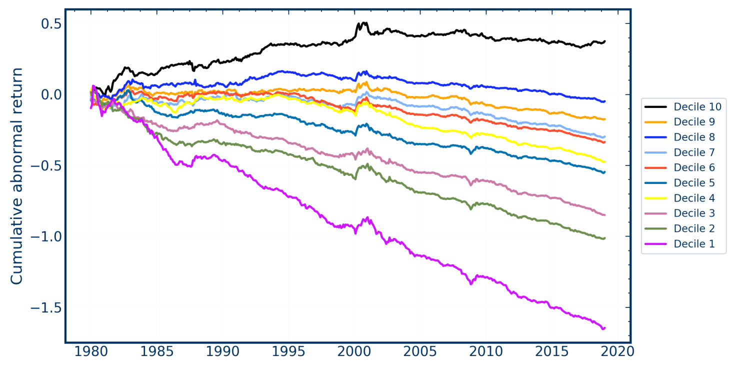
- Best 10% (predicted and realized) funds charge higher fees
- But so do the 10% worst funds; both 50% cumulative expense ratio

After fee performance



⇒ Spread the same after fees.

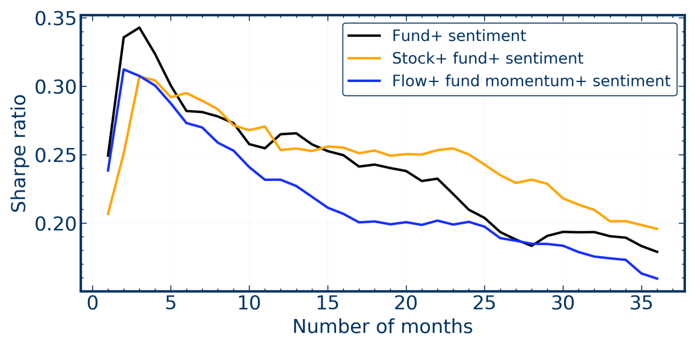
After fee performance



Predicted After-fee abnormal returns

- We directly predict after-fee returns
- Results in 207% cumulative return spread (vs. 191% baseline)
- Cumulative abnormal after-fee return of 37% for top 10%.

Holding Periods

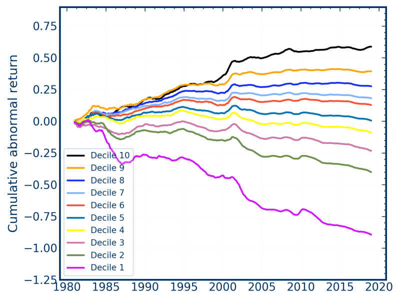


Sharpe ratio over long holding horizons

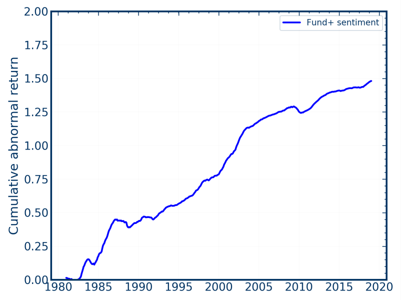
- Investments held for longer periods

⇒ Predictability lasts over long time horizons.

Prediction of one-year abnormal returns



Cumulative predictions deciles



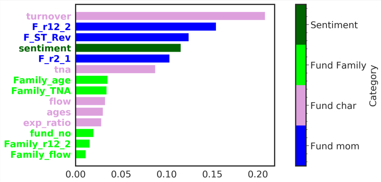
Cumulative long-short portfolio

Information set	SR	mean(%)	t-stat
Fund+ sentiment	0.27	0.31	6.6***

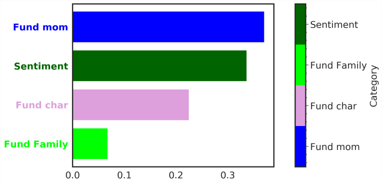
- Cumulative abnormal returns with annual abnormal return prediction
- Information set: fund-specific characteristics and sentiment

⇒ Predictability lasts over long time horizons.

Prediction of one-year abnormal returns: Variable importance



Top variable importance.

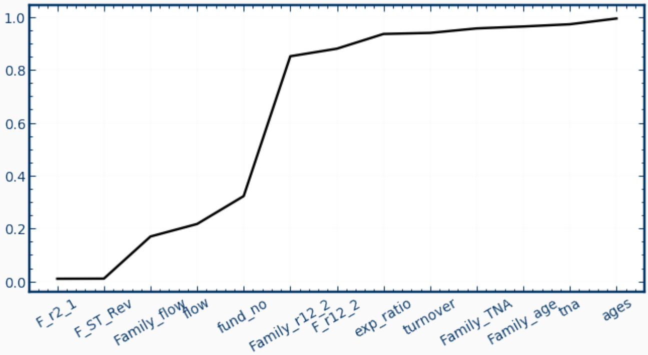


Top variable group importance.

Top variable importance for explaining annual overlapping abnormal returns.

- Variable importance shifts to more persistent variables

Persistence of fund characteristics

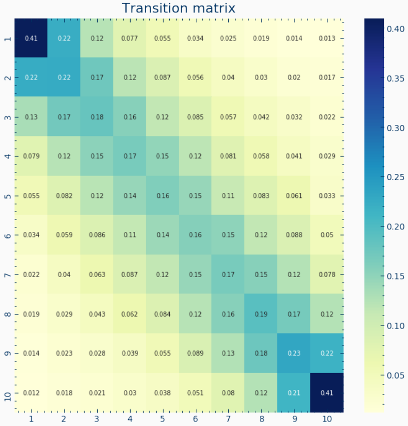


Autocorelation of fund characteristics

- Many fund characteristics are persistent

⇒ Predictability over longer time horizons

Persistence of classification



Transition between prediction deciles

- Extreme deciles are persistent

⇒ Predictability over longer time horizons

Results for top deciles

Information set	mean(%)	t-stat	SR	R_F^2 (%)
Stock+ fund	0.10	1.7*	0.08	-0.52
Stock+ fund+ sentiment	0.15	2.9***	0.13	1.87
Stock	-0.02	-0.4	-0.02	-2.52
Stock+ sentiment	0.06	1.2	0.06	0.61
Fund	0.16	3.7***	0.17	-1.20
Fund+ sentiment	0.17	3.5***	0.16	1.46
Flow+ fund momentum+ sentiment	0.19	3.2***	0.15	-0.15
Fund exclude momentum and flow	-0.01	-0.2	-0.01	-0.17
F_r12_2+ sentiment	0.12	2.0**	0.09	-0.58

Results for bottom deciles

Information set	mean(%)	t-stat	SR	R^2_F (%)
Stock+ fund	-0.19	-2.6***	-0.15	1.33
Stock+ fund+ sentiment	-0.25	-3.5***	-0.22	1.99
Stock	-0.00	-0.0	-0.00	-0.82
Stock+ sentiment	-0.09	-1.2	-0.08	-0.03
Fund	-0.22	-3.7***	-0.23	0.74
Fund+ sentiment	-0.23	-3.8***	-0.23	1.38
Flow+ fund momentum+ sentiment	-0.29	-4.2***	-0.23	1.05
Fund exclude momentum and flow	-0.07	-1.8*	-0.09	-0.32
F_r12_2+ sentiment	-0.23	-3.8***	-0.18	0.88

Inspecting the mechanism

- Is the abnormal return mostly trading within disclosure dates or trading within a disclosure period?
- A decomposition:

$$R^{abn}_{i,t} = \underbrace{\tilde{R}_{i,t} - f_t\tilde{\beta}_i}_{\text{Between disclosure abnormal return}} + \underbrace{R_{i,t} - f_t\beta_i - (\tilde{R}_{i,t} - f_t\tilde{\beta}_i)}_{\text{Within disclosure abnormal return}} \quad (1)$$

$$= \underbrace{\tilde{R}_{i,t} - f_t\tilde{\beta}_i}_{\text{Between disclosure abnormal return}} + \underbrace{R_{i,t} - \tilde{R}_{i,t}}_{\text{Return gap}} + \underbrace{f_t(\tilde{\beta}_i - \beta_i)}_{\text{Risk exposure difference}} \quad (2)$$

Inspecting the mechanism: machine learning

	Total		Between-disclosure		Within-disclosure		Risk difference		Return gap	
	SR	mean	SR	mean	SR	mean	SR	mean	SR	mean
Stock+ fund	0.15	0.28***	0.05	0.13	0.14	0.15***	0.06	0.06	0.11	0.09***
Stock+ fund+ sentiment	0.21	0.41***	0.10	0.28**	0.13	0.13***	0.07	0.06	0.09	0.06**
Stock	-0.01	-0.02	-0.01	-0.03	0.01	0.01	-0.01	-0.01	0.03	0.02
Stock+ sentiment	0.07	0.15	0.04	0.12	0.02	0.02	0.00	0.00	0.02	0.02
Fund	0.25	0.38***	0.15	0.20***	0.17	0.18***	0.15	0.12***	0.08	0.06**
Fund+ sentiment	0.25	0.40***	0.15	0.24***	0.16	0.16***	0.16	0.13***	0.03	0.03

Decomposition of deep learning prediction long-short deciles

Which macroeconomic variable?

Information set	mean (%)	t-stat	SR	R_F^2 (%)
Fund+sentiment	0.40	5.4***	0.25	2.73
Fund+CFNAI	0.39	6.0***	0.28	0.72
Fund+sentiment+CFNAI	0.42	6.3***	0.29	2.48
Fund+sentiment_orth	0.43	6.4***	0.29	1.22
Fund+CFNAI_orth	0.38	5.4***	0.25	0.92
Fund	0.38	5.5***	0.25	0.19

- Models with sentiment predict the abnormal return factor better.
- Macro variables affect the mean of long-short portfolio:
  Up to 0.05% increase
- Largest effect of macroeconomic variables in the level

Spanning of ML long-short portfolios with different factor models.

	FF 4 factors		FF 5 factors		FF 6 factors		FF 8 factors		
	α	R^2	α	R^2	α	R^2	α	R^2	mean μ
Stock+ fund	0.07	0.14	0.08*	0.13	0.06	0.15	0.04	0.19	0.04
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
Stock+ fund+ sentiment	0.13***	0.29	0.10**	0.33	0.08**	0.36	0.07*	0.37	0.13***
	(0.04)		(0.04)		(0.04)		(0.04)		(0.05)
Stock	0.05	0.15	0.04	0.16	0.03	0.17	0.01	0.22	0.01
	(0.04)		(0.04)		(0.04)		(0.04)		(0.05)
Stock+ sentiment	0.09**	0.31	0.04	0.39	0.03	0.40	0.02	0.41	0.08*
	(0.04)		(0.04)		(0.04)		(0.04)		(0.05)
Fund	0.14***	0.17	0.20***	0.04	0.16***	0.18	0.16***	0.18	0.18***
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
Fund+ sentiment	0.17***	0.16	0.22***	0.04	0.18***	0.16	0.19***	0.18	0.20***
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
Flow+ fund momentum+ sentiment	0.11***	0.28	0.22***	0.13	0.16***	0.33	0.18***	0.37	0.15***
	(0.04)		(0.04)		(0.04)		(0.04)		(0.05)
F_r12_2+ sentiment	0.13***	0.30	0.25***	0.11	0.19***	0.34	0.19***	0.34	0.19***
	(0.04)		(0.04)		(0.04)		(0.04)		(0.05)

Machine learning prediction long-short portfolios

- Time-series regressions of long-short portfolios on factors: Fama-French 5, momentum, long-term and short-term reversal factors
- Long-short portfolio returns are not compensation for risk. outline

Spanning of univariate sorted portfolios with different factor models.

	FF 4 factors		FF 5 factors		FF 6 factors		FF 8 factors		
	α	R^2	α	R^2	α	R^2	α	R^2	mean μ
F_ST_Rev	0.18***	0.04	0.19***	0.02	0.18***	0.04	0.23***	0.31	0.20***
	(0.04)		(0.04)		(0.04)		(0.04)		(0.05)
F_r2_1	0.01	0.11	0.03	0.03	-0.01	0.12	0.01	0.17	0.08
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
F_r12_2	0.17***	0.23	0.22***	0.06	0.16***	0.23	0.16***	0.23	0.28***
	(0.04)		(0.04)		(0.04)		(0.04)		(0.05)
flow	0.10**	0.02	0.08	0.03	0.08	0.03	0.07	0.03	0.12**
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
turnover	0.02	0.01	-0.05	0.08	-0.05	0.08	-0.07	0.12	0.03
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
fund_no	0.09*	0.03	0.12**	0.03	0.11**	0.04	0.13***	0.07	0.13***
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
Family_r12_2	0.14***	0.04	0.17***	0.01	0.15***	0.05	0.14***	0.05	0.19***
	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)

Univariate long-short portfolios

- Time-series regressions of univariate long-short portfolios on asset pricing factors: Fama-French 5, momentum, long-term and short-term reversal factors
- Results for seven most important fund-specific characteristics.
- All R^2s are small and alphas are highly significant.

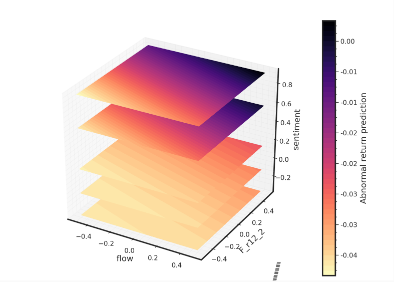
Spanning of univariate sorted portfolios with Carhart four factor model.

	Mkr	SMB	HML	Mom	α	Factor mean	R^2
F_r12_2	0.29***	0.04	0.11**	0.44***	0.17***	0.28***	0.23
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	
flow	0.12**	-0.10**	0.03	0.03	0.10**	0.12**	0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
F_ST_Rev	-0.09*	0.07	0.08*	0.13***	0.18***	0.20***	0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
F_r2_1	0.10**	-0.05	0.05	0.34***	0.01	0.08	0.11
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
turnover	-0.00	0.04	0.03	0.06	0.02	0.03	0.01
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
fund_no	0.17***	0.02	0.03	0.08	0.09*	0.13***	0.03
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
Family_r12_2	0.10*	0.04	0.10**	0.21***	0.14***	0.19***	0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	

Univariate long-short portfolios

- Time-series regressions on 4 Fama-French-Carhart factors.
  Results for seven most important fund-specific characteristics.
- All R^2s are small and alphas are highly significant.
  The mean return and mean intercept are similar in magnitude. outline

Simplified Model: Only flow, momentum, sentiment



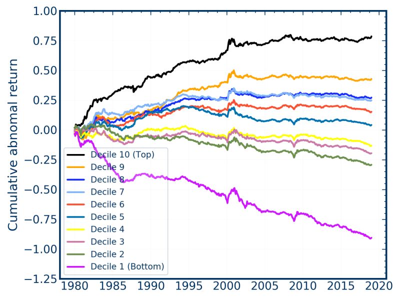
Conditional abnormal return for flow, momentum and sentiment

Decile	mean(%)	t-stat	SR	R^2_F (%)
Long-short	0.40	5.4***	0.25	0.70
Long	0.17	3.4***	0.16	-0.73
Short	-0.23	-3.6***	-0.21	0.82

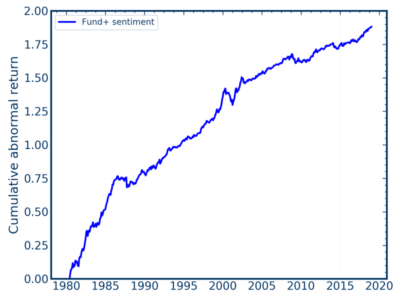
- Large performance with only three variables
- Visualization of complete function possible

$\Rightarrow$ Strong interaction effects

Removing small funds



Cumulative predictions deciles



Cumulative long-short portfolio

Decile	mean(%)	t-stat	SR	R^2_F (%)
Long-short	0.40	6.5***	0.30	4.07
Long	0.17	3.9***	0.18	2.03
Short	-0.23	-4.5***	-0.23	2.05

- Exclude mutual funds smaller than 15 millions AUM

⇒ Results are robust to excluding small funds

- Further robustness results for value-weighted funds

Summary statistics

Table 1: Summary statistics of the fund characteristics data

Statistic	N	Mean	St. Dev.	Median
turnover	358,303	0.826	1.015	0.620
ages	407,139	13.669	10.200	11.000
flow (%)	406,661	1.601	419.975	−0.392
r12_2	407,158	0.108	0.173	0.107
LME	407,158	−0.385	0.108	−0.424
BEME	407,158	−0.153	0.376	−0.161
abnormal return (%)	407,158	−0.028	2.000	−0.028
exp_ratio (%)	407,043	0.097	0.086	0.095
TNA	406,802	1,153.180	4,833.920	214.700

Model tuning

- Select tuning parameters on validation data.

Notation	Hyperparameters	Candidates
HL	Number of layers in Neural Network	1, 2, 3
HU	Number of hidden units in each layer	2^{6-i} or 2^{7-i}, $i = 1$ to HL
DR	Dropout	0.90, 0.95
LR	Learning rate	0.001 , 0.1
L2	l2 regularization	0.0, 1e-3, 1e-2

- The optimal network structure is one hidden layer with 64 hidden states.

model

HL	HU	DR	l2	LR
1	64	0.95	0.001	0.01

Weighting schemes

- Prediction-weighted: $\mu_{i,t}$ model prediction,

$$\text{For top portfolio: } \tilde{\mu}_{i,t} = \mu_{i,t} - \min_{i \in \text{Top}} (\mu_{i,t}) \tag{3}$$

$$\text{For bottom portfolio: } \tilde{\mu}_{i,t} = \mu_{i,t} - \max_{i \in \text{Bottom}} (\mu_{i,t}) \tag{4}$$

$$w_{i,t}^{\text{pred}} = \frac{\tilde{\mu}_{i,t}}{\sum_{i=1}^{N} \tilde{\mu}_{i,t}} \tag{5}$$

- Construct long-short factor as difference between top and bottom deciles.

prediction

Data source

- Mutual fund holding data: TFN/CDA S12.
  - Quarterly frequency - used the last observed fund holding.
- Mutual fund characteristics data: CRSP.
  - Monthly frequency.
- Stock characteristics data: CRSP and CompuStat. data
  - Monthly frequency - characteristics shown to have predictive power for the cross-section of expected returns.

Macro Definitions

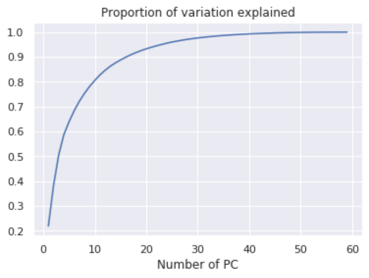
- Sentiment is the Principal Component (with twists) of
    - The closed-end fund discount
    - NYSE share turnover
    - The number of IPOs
    - The average first-day returns on IPOs
    - The equity share in new shares
    - Dividend premium
- CFNAI is the first principal component of 85 economic indicators from four broad categories: production and income; employment, unemployment, and hours; personal consumption and housing; and sales, orders, and inventories. data

Impute Missing Data

- Method in Pelger and Xiong 2022 data :

$$C_t = \underbrace{\Lambda_t}_{L \times K} \underbrace{V_t}_{K \times N_t}$$

- Intuition $\Rightarrow$ Correlation between fund characteristics.
- Characteristics are cross-sectionally normalized to [-0.5, 0.5].
- Few PCs explain most variation in characteristics space (illustration for 1992/01):



Solving the model

- Estimate a "characteristics covariance matrix" with observed entries:

$$\Sigma_{t,l,r} = \frac{1}{|Q_{l,r}|} \sum_{i \in Q_{l,r}} C_{t,l,i} C_{t,r,i}$$

  Apply Λ_t as the normalized eigenvectors of Σ_t.

- Estimate the "characteristics factor" with a "regression":

$$\underbrace{V_{t,i}}_{K \times 1} = \left(\sum_{l \in Q_{i,t}} \Lambda_{tl}^2 \right)^{-1} \sum_{l \in Q_{i,t}} \Lambda_{tl} C_{t,l,i}$$

- Given $V_{t,i}$ and Λ_t we estimate the missing entries as

$$\tilde{C}_{t,i} = \Lambda_t V_{t,i}$$

- Normalize $\tilde{C}_{t,l}$ with a second round of quantile ranking. data

References

M. Baker and J. Wurgler. Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4): 1645–1680, 2006.

J. B. Berk and R. C. Green. Mutual fund flows and performance in rational markets. *Journal of political economy*, 112(6):1269–1295, 2004.

N. Bollen and B. Jeffrey. Short-term persistence in mutual fund performance. *Review of Financial Studies*, 18(2): 569–597, 2005.

S. Bryzgalova, M. Pelger, and J. Zhu. Forest through the trees: Building cross-sections of stock returns. *Working paper*, 2019.

M. M. Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82, 1997.

L. Chen, M. Pelger, and J. Zhu. Deep learning in asset pricing. *Working Paper*, 2019.

V. DeMiguel, J. Gil-Bazo, F. J. Nogales, and A. A. P. Santos. Machine learning and fund characteristics help to select mutual funds with positive alpha. *Working paper*, 2022.

E. F. Fama and K. R. French. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, 65 (5):1915–1947, 2010.

J. Freyberger, A. Neuhierl, and M. Weber. Dissecting characteristics nonparametrically. *Review of Financial Studies*, 2020. forthcoming.

M. Gruber. Another puzzle: The growth of actively managed mutual funds. *Journal of Finance*, 51(3):783–810, 1996.

S. Gu, B. Kelly, and D. Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5): 2223–2273, 2020.

E. Horel and K. Giesecke. Significance tests for neural networks. *Journal of Machine Learning Research*, 21(227): 1–29, 2020.

M. Kacperczyk, C. Sialm, and L. Zheng. On the industry concentration of actively managed equity mutual funds. *Journal of Finance*, 60(4):1983–2011, 2005.

M. Kacperczyk, S. Van Nieuwerburgh, and L. Veldkamp. Time-varying fund manager skill. *The Journal of Finance*, 69(4):1455–1484, 2014.

M. Kacperczyk, S. Van Nieuwerburgh, and L. Veldkamp. A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626, 2016.

R. Kosowski. Do mutual funds perform when it matters most to investors? us mutual fund performance and risk in recessions and expansions. *Quarterly Journal of Finance*, 1(3):607–664, 2011.

B. Li and A. G. Rossi. Selecting mutual funds from the stocks they hold: A machine learning approach. *Working Paper*, 2021.

T. J. Moskowitz. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses: Discussion. *The Journal of Finance*, 55(4):1695–1703, 2000.

R. F. Stambaugh, J. Yu, and Y. Yuan. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302, 2012.

L. Zheng. Is money smart? a study of mutual fund investors' fund selection ability. *Journal of Finance*, 54: 901–933, 1999.