Machine-Learning the Skill of Mutual Fund Managers

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

¹Rochester University, ²Stanford University, ³Columbia University
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.

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 character., no macro


Object of interest: Abnormal returns:

\[ R_{i,t}^{abn} = R_{i,t} - F_t \hat{\beta}_{i,t-1}, \]
\[ R_{i,t-36:t-1} = \alpha_i + F_{t-36:t-1} \hat{\beta}_{i,t-1} + \eta_{i,t-36:t-1} \]

- 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.

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:
  \[ R_{i,t+1}^{abn} = g(z_{i,t}, z_t) + \epsilon_{i,t+1} \]

- 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
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**.

![Prediction weights](image1)

![Equal weights](image2)

(a) Prediction weights  
(b) Equal weights

**Portfolio weights for top and bottom deciles**
All characteristics + sentiment

Abnormal returns are predictable

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

- 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

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

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

### 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

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
  ⇒ 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?

Variable importance measure:

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

- 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!
Results in different sentiment terciles

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>SR</th>
<th>mean</th>
<th>t-stat</th>
<th>$R_F^2$</th>
<th>SR</th>
<th>mean</th>
<th>t-stat</th>
<th>$R_F^2$</th>
<th>SR</th>
<th>mean</th>
<th>t-stat</th>
<th>$R_F^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D10-D1</td>
<td>0.12</td>
<td>0.23</td>
<td>1.6</td>
<td>0.50</td>
<td>0.37</td>
<td>0.42</td>
<td>4.6***</td>
<td>3.39</td>
<td>0.32</td>
<td>0.55</td>
<td>4.0***</td>
<td>4.83</td>
</tr>
<tr>
<td>D1</td>
<td>-0.11</td>
<td>-0.18</td>
<td>-1.4</td>
<td>0.71</td>
<td>-0.25</td>
<td>-0.23</td>
<td>-3.1***</td>
<td>3.65</td>
<td>-0.23</td>
<td>-0.29</td>
<td>-2.9***</td>
<td>1.35</td>
</tr>
<tr>
<td>D10</td>
<td>0.05</td>
<td>0.04</td>
<td>0.6</td>
<td>-0.86</td>
<td>0.22</td>
<td>0.19</td>
<td>2.7***</td>
<td>1.00</td>
<td>0.21</td>
<td>0.27</td>
<td>2.6**</td>
<td>2.68</td>
</tr>
</tbody>
</table>

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.
New interaction measure:

- Differences in slopes for high and low macroeconomic states:
  \[
  \text{Interaction}(z, \text{macro}) = \left( \hat{R}_{\text{abn}}^{\text{high} \ z, \text{high} \ \text{macro}} - \hat{R}_{\text{abn}}^{\text{low} \ z, \text{high} \ \text{macro}} \right) \\
  \quad - \left( \hat{R}_{\text{abn}}^{\text{high} \ z, \text{low} \ \text{macro}} - \hat{R}_{\text{abn}}^{\text{low} \ z, \text{low} \ \text{macro}} \right) .
  \]

- 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
  ⇒ buying pressure for the stocks funds hold
- Downward-sloping demand curve
  ⇒ raises prices and lifts fund returns
- Creates more inflows next period
  ⇒ stronger in high sentiment periods.
More results

**Chronological cross-out-of-sample analysis:**
- Predictability and performance robust to sampling
- Economic model depends on which sentiment states are observed

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

**Decomposition of abnormal returns:**
- 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:**
- 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:**
- Results robust to excluding small funds or value-weighting

**Tuning parameters:**
- 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**:
   - 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
- Information set comparison
- Predicting returns vs. abnormal returns
- Statistical significance tests
- After fee performance
- Longer holding periods
- Persistence of fund characteristics and classification
- Separate results for deciles
- Decomposition of abnormal returns
- Which macroeconomic variable?
- Spanning tests
- Simplified model
- Robustness to fund size
- Data and implementation
- Missing data
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

<table>
<thead>
<tr>
<th>Sampling</th>
<th>mean(%)</th>
<th>t-stat</th>
<th>SR</th>
<th>$R_F^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological folds</td>
<td>0.39</td>
<td>5.0***</td>
<td>0.23</td>
<td>1.47</td>
</tr>
<tr>
<td>Random folds</td>
<td>0.40</td>
<td>5.4***</td>
<td>0.25</td>
<td>2.73</td>
</tr>
</tbody>
</table>

- Fund-specific characteristics + sentiment as input.

⇒ Predictability and economic significance robust to sampling
Chronological cross-out-of-sample: Variable importance

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

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

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

$\Rightarrow$ Fund-specific characteristics and sentiment predict abnormal returns
Information set comparison

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

- 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

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

<table>
<thead>
<tr>
<th>Data</th>
<th>mean(%)</th>
<th>t-stat</th>
<th>SR</th>
<th>$R^2_F$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fund+ sentiment</td>
<td>0.49</td>
<td>3.0***</td>
<td>0.14</td>
<td>0.97</td>
</tr>
<tr>
<td>Fund</td>
<td>0.53</td>
<td>3.5***</td>
<td>0.16</td>
<td>0.97</td>
</tr>
<tr>
<td>Stock+ sentiment</td>
<td>0.44</td>
<td>3.1***</td>
<td>0.14</td>
<td>-20.03</td>
</tr>
<tr>
<td>Stock</td>
<td>0.11</td>
<td>1.1</td>
<td>0.05</td>
<td>-53.21</td>
</tr>
<tr>
<td>Stock+ fund+ sentiment</td>
<td>0.45</td>
<td>3.1***</td>
<td>0.14</td>
<td>-26.54</td>
</tr>
</tbody>
</table>

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
  
⇒ Abnormal return prediction is relative objective and preferable: higher Sharpe ratio and $R^2_F$ than for returns
### Significance tests

<table>
<thead>
<tr>
<th>Fund char</th>
<th>Sensitivity</th>
<th>Interaction w. sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>sentiment</td>
<td>0.14***</td>
<td></td>
</tr>
<tr>
<td>F_r12_2</td>
<td>0.08***</td>
<td>0.09***</td>
</tr>
<tr>
<td>turnover</td>
<td>0.05***</td>
<td>0.06***</td>
</tr>
<tr>
<td>F_ST_Rev</td>
<td>0.04***</td>
<td>0.04***</td>
</tr>
<tr>
<td>F_r2_1</td>
<td>0.04***</td>
<td>-0.03***</td>
</tr>
<tr>
<td>flow</td>
<td>0.03***</td>
<td>0.03***</td>
</tr>
<tr>
<td>ages</td>
<td>0.02***</td>
<td>0.02***</td>
</tr>
<tr>
<td>fund_no</td>
<td>0.02***</td>
<td>-0.01**</td>
</tr>
<tr>
<td>tna</td>
<td>0.02***</td>
<td>0.01**</td>
</tr>
<tr>
<td>Family_r12_2</td>
<td>0.02***</td>
<td>0.01</td>
</tr>
<tr>
<td>Family_flow</td>
<td>0.02**</td>
<td>0.01***</td>
</tr>
<tr>
<td>Family_TNA</td>
<td>0.02**</td>
<td>0.00</td>
</tr>
<tr>
<td>Family_age</td>
<td>0.02*</td>
<td>-0.01</td>
</tr>
<tr>
<td>exp_ratio</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

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.
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.
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

<table>
<thead>
<tr>
<th>Information set</th>
<th>SR</th>
<th>mean(%)</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fund+ sentiment</td>
<td>0.27</td>
<td>0.31</td>
<td>6.6***</td>
</tr>
</tbody>
</table>

- 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

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

<table>
<thead>
<tr>
<th>Information set</th>
<th>mean(%)</th>
<th>t-stat</th>
<th>SR</th>
<th>$R_F^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock+ fund</td>
<td>0.10</td>
<td>1.7*</td>
<td>0.08</td>
<td>-0.52</td>
</tr>
<tr>
<td>Stock+ fund+ sentiment</td>
<td>0.15</td>
<td>2.9***</td>
<td>0.13</td>
<td>1.87</td>
</tr>
<tr>
<td>Stock</td>
<td>-0.02</td>
<td>-0.4</td>
<td>-0.02</td>
<td>-2.52</td>
</tr>
<tr>
<td>Stock+ sentiment</td>
<td>0.06</td>
<td>1.2</td>
<td>0.06</td>
<td>0.61</td>
</tr>
<tr>
<td>Fund</td>
<td>0.16</td>
<td>3.7***</td>
<td>0.17</td>
<td>-1.20</td>
</tr>
<tr>
<td>Fund+ sentiment</td>
<td>0.17</td>
<td>3.5***</td>
<td>0.16</td>
<td>1.46</td>
</tr>
<tr>
<td>Flow+ fund momentum+ sentiment</td>
<td>0.19</td>
<td>3.2***</td>
<td>0.15</td>
<td>-0.15</td>
</tr>
<tr>
<td>Fund exclude momentum and flow</td>
<td>-0.01</td>
<td>-0.2</td>
<td>-0.01</td>
<td>-0.17</td>
</tr>
<tr>
<td>$F_{r12.2}$+ sentiment</td>
<td>0.12</td>
<td>2.0**</td>
<td>0.09</td>
<td>-0.58</td>
</tr>
</tbody>
</table>
### Results for bottom deciles

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

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

\[
R_{i,t}^{abn} = \underbrace{\tilde{R}_{i,t} - f_t \tilde{\beta}_i}_{	ext{Between disclosure abnormal return}} + \underbrace{R_{i,t} - f_t \beta_i - (\tilde{R}_{i,t} - f_t \tilde{\beta}_i)}_{	ext{Within disclosure abnormal return}}
\]

\[
= \underbrace{\tilde{R}_{i,t} - f_t \tilde{\beta}_i}_{	ext{Between disclosure abnormal return}} + \underbrace{R_{i,t} - \tilde{R}_{i,t}}_{	ext{Return gap}} + f_t (\tilde{\beta}_i - \beta_i)
\]

Risk exposure difference
## Inspecting the mechanism: machine learning

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

Decomposition of deep learning prediction long-short deciles
### Which macroeconomic variable?

<table>
<thead>
<tr>
<th>Information set</th>
<th>mean (%)</th>
<th>t-stat</th>
<th>SR</th>
<th>$R_F^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fund+sentiment</td>
<td>0.40</td>
<td>5.4***</td>
<td>0.25</td>
<td>2.73</td>
</tr>
<tr>
<td>Fund+CFNAI</td>
<td>0.39</td>
<td>6.0***</td>
<td>0.28</td>
<td>0.72</td>
</tr>
<tr>
<td>Fund+sentiment+CFNAI</td>
<td>0.42</td>
<td>6.3***</td>
<td>0.29</td>
<td>2.48</td>
</tr>
<tr>
<td>Fund+sentiment_orth</td>
<td>0.43</td>
<td>6.4***</td>
<td>0.29</td>
<td>1.22</td>
</tr>
<tr>
<td>Fund+CFNAI_orth</td>
<td>0.38</td>
<td>5.4***</td>
<td>0.25</td>
<td>0.92</td>
</tr>
<tr>
<td>Fund</td>
<td>0.38</td>
<td>5.5***</td>
<td>0.25</td>
<td>0.19</td>
</tr>
</tbody>
</table>

- 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.

<table>
<thead>
<tr>
<th>Model Description</th>
<th>FF 4 factors</th>
<th></th>
<th></th>
<th>FF 5 factors</th>
<th></th>
<th></th>
<th>FF 6 factors</th>
<th></th>
<th></th>
<th>FF 8 factors</th>
<th></th>
<th></th>
<th>mean μ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>R²</td>
<td>α</td>
<td>R²</td>
<td>α</td>
<td>R²</td>
<td>α</td>
<td>R²</td>
<td>α</td>
<td>R²</td>
<td>α</td>
<td>R²</td>
<td></td>
</tr>
<tr>
<td>Stock+ fund</td>
<td>0.07</td>
<td>0.14</td>
<td>0.08*</td>
<td>0.13</td>
<td>0.06</td>
<td>0.15</td>
<td>0.04</td>
<td>0.19</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock+ fund+ sentiment</td>
<td>0.13***</td>
<td>0.29</td>
<td>0.10**</td>
<td>0.33</td>
<td>0.08**</td>
<td>0.36</td>
<td>0.07*</td>
<td>0.37</td>
<td>0.13***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock</td>
<td>0.05</td>
<td>0.15</td>
<td>0.04</td>
<td>0.16</td>
<td>0.03</td>
<td>0.17</td>
<td>0.01</td>
<td>0.22</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock+ sentiment</td>
<td>0.09**</td>
<td>0.31</td>
<td>0.04</td>
<td>0.39</td>
<td>0.03</td>
<td>0.40</td>
<td>0.02</td>
<td>0.41</td>
<td>0.08*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fund</td>
<td>0.14***</td>
<td>0.17</td>
<td>0.20***</td>
<td>0.04</td>
<td>0.16***</td>
<td>0.18</td>
<td>0.16***</td>
<td>0.18</td>
<td>0.18***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fund+ sentiment</td>
<td>0.17***</td>
<td>0.16</td>
<td>0.22***</td>
<td>0.04</td>
<td>0.18***</td>
<td>0.16</td>
<td>0.19***</td>
<td>0.18</td>
<td>0.20***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flow+ fund momentum+ sentiment</td>
<td>0.11***</td>
<td>0.28</td>
<td>0.22***</td>
<td>0.13</td>
<td>0.16***</td>
<td>0.33</td>
<td>0.18***</td>
<td>0.37</td>
<td>0.15***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_r12.2+ sentiment</td>
<td>0.13***</td>
<td>0.30</td>
<td>0.25***</td>
<td>0.11</td>
<td>0.19***</td>
<td>0.34</td>
<td>0.19***</td>
<td>0.34</td>
<td>0.19***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
Spanning of univariate sorted portfolios with different factor models.

<table>
<thead>
<tr>
<th></th>
<th>FF 4 factors</th>
<th>FF 5 factors</th>
<th>FF 6 factors</th>
<th>FF 8 factors</th>
<th>mean µ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>R²</td>
<td>α</td>
<td>R²</td>
<td>α</td>
</tr>
<tr>
<td>F_ST_Rev</td>
<td>0.18*** (0.04)</td>
<td>0.19*** (0.04)</td>
<td>0.18*** (0.04)</td>
<td>0.23*** (0.04)</td>
<td>0.20***</td>
</tr>
<tr>
<td>F_r2.1</td>
<td>0.01 (0.05)</td>
<td>0.03 (0.05)</td>
<td>-0.01 (0.05)</td>
<td>0.12 (0.05)</td>
<td>0.17</td>
</tr>
<tr>
<td>F_r12.2</td>
<td>0.17*** (0.04)</td>
<td>0.22*** (0.04)</td>
<td>0.16*** (0.04)</td>
<td>0.16*** (0.04)</td>
<td>0.28***</td>
</tr>
<tr>
<td>flow</td>
<td>0.10** (0.05)</td>
<td>0.08 (0.05)</td>
<td>0.08 (0.05)</td>
<td>0.07 (0.05)</td>
<td>0.12***</td>
</tr>
<tr>
<td>turnover</td>
<td>0.02 (0.05)</td>
<td>-0.05 (0.05)</td>
<td>-0.05 (0.05)</td>
<td>-0.07 (0.05)</td>
<td>0.12</td>
</tr>
<tr>
<td>fund_no</td>
<td>0.09* (0.05)</td>
<td>0.12** (0.05)</td>
<td>0.11** (0.05)</td>
<td>0.13*** (0.05)</td>
<td>0.13***</td>
</tr>
<tr>
<td>Family_r12.2</td>
<td>0.14*** (0.05)</td>
<td>0.17*** (0.05)</td>
<td>0.15*** (0.05)</td>
<td>0.14*** (0.05)</td>
<td>0.19***</td>
</tr>
</tbody>
</table>

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 $R^2$'s are small and alphas are highly significant.
Spanning of univariate sorted portfolios with Carhart four factor model.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mkr</th>
<th>SMB</th>
<th>HML</th>
<th>Mom</th>
<th>(\alpha)</th>
<th>Factor mean</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_r12_2</td>
<td>0.29***</td>
<td>0.04</td>
<td>0.11**</td>
<td>0.44***</td>
<td>0.17***</td>
<td>0.28***</td>
<td>0.23</td>
</tr>
<tr>
<td>flow</td>
<td>0.12**</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10**</td>
<td>0.12**</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>F_ST_Rev</td>
<td>-0.09*</td>
<td>0.07</td>
<td>0.08*</td>
<td>0.13***</td>
<td>0.18***</td>
<td>0.20***</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>F_r2_1</td>
<td>0.10**</td>
<td>-0.05</td>
<td>0.05</td>
<td>0.34***</td>
<td>0.01</td>
<td>0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>turnover</td>
<td>-0.00</td>
<td>0.04</td>
<td>0.03</td>
<td>0.06</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>fund_no</td>
<td>0.17***</td>
<td>0.02</td>
<td>0.03</td>
<td>0.08</td>
<td>0.09*</td>
<td>0.13***</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>Family_r12_2</td>
<td>0.10*</td>
<td>0.04</td>
<td>0.10**</td>
<td>0.21***</td>
<td>0.14***</td>
<td>0.19***</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
</tr>
</tbody>
</table>

Univariate long-short portfolios

- Time-series regressions on 4 Fama-French-Carhart factors.
  Results for seven most important fund-specific characteristics.
- All \(R^2\)s are small and alphas are highly significant.
  The mean return and mean intercept are similar in magnitude.
Simplified Model: Only flow, momentum, sentiment

Conditional abnormal return for flow, momentum and sentiment

<table>
<thead>
<tr>
<th>Decile</th>
<th>mean(%)</th>
<th>t-stat</th>
<th>SR</th>
<th>$R^2_F$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-short</td>
<td>0.40</td>
<td>5.4***</td>
<td>0.25</td>
<td>0.70</td>
</tr>
<tr>
<td>Long</td>
<td>0.17</td>
<td>3.4***</td>
<td>0.16</td>
<td>-0.73</td>
</tr>
<tr>
<td>Short</td>
<td>-0.23</td>
<td>-3.6***</td>
<td>-0.21</td>
<td>0.82</td>
</tr>
</tbody>
</table>

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

⇒ Strong interaction effects
Removing small funds

<table>
<thead>
<tr>
<th>Decile</th>
<th>mean(%)</th>
<th>t-stat</th>
<th>SR</th>
<th>$R_F^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-short</td>
<td>0.40</td>
<td>6.5***</td>
<td>0.30</td>
<td>4.07</td>
</tr>
<tr>
<td>Long</td>
<td>0.17</td>
<td>3.9***</td>
<td>0.18</td>
<td>2.03</td>
</tr>
<tr>
<td>Short</td>
<td>-0.23</td>
<td>-4.5***</td>
<td>-0.23</td>
<td>2.05</td>
</tr>
</tbody>
</table>

- Exclude mutual funds smaller than 15 millions AUM
  ⇒ Results are robust to excluding small funds
- Further robustness results for value-weighted funds
### Table 1: Summary statistics of the fund characteristics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>turnover</td>
<td>358,303</td>
<td>0.826</td>
<td>1.015</td>
<td>0.620</td>
</tr>
<tr>
<td>ages</td>
<td>407,139</td>
<td>13.669</td>
<td>10.200</td>
<td>11.000</td>
</tr>
<tr>
<td>flow (%)</td>
<td>406,661</td>
<td>1.601</td>
<td>419.975</td>
<td>-0.392</td>
</tr>
<tr>
<td>r12.2</td>
<td>407,158</td>
<td>0.108</td>
<td>0.173</td>
<td>0.107</td>
</tr>
<tr>
<td>LME</td>
<td>407,158</td>
<td>-0.385</td>
<td>0.108</td>
<td>-0.424</td>
</tr>
<tr>
<td>BEME</td>
<td>407,158</td>
<td>-0.153</td>
<td>0.376</td>
<td>-0.161</td>
</tr>
<tr>
<td>abnormal return (%)</td>
<td>407,158</td>
<td>-0.028</td>
<td>2.000</td>
<td>-0.028</td>
</tr>
<tr>
<td>exp_ratio (%)</td>
<td>407,043</td>
<td>0.097</td>
<td>0.086</td>
<td>0.095</td>
</tr>
<tr>
<td>TNA</td>
<td>406,802</td>
<td>1,153.180</td>
<td>4,833.920</td>
<td>214.700</td>
</tr>
</tbody>
</table>
Model tuning

- Select tuning parameters on validation data.

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

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

```
<table>
<thead>
<tr>
<th>HL</th>
<th>HU</th>
<th>DR</th>
<th>L2</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64</td>
<td>0.95</td>
<td>0.001</td>
<td>0.01</td>
</tr>
</tbody>
</table>
```
Weighting schemes

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

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

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

  $$w^{\text{pred}}_{i,t} = \frac{\tilde{\mu}_{i,t}}{\sum_{i=1}^{N} \tilde{\mu}_{i,t}}$$ (5)

- Construct long-short factor as difference between top and bottom deciles.
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.
  - 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.
Method in Pelger and Xiong 2022: 

\[ C_t = \Lambda_t V_t \]

- 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 \( \Lambda_t \) as the normalized eigenvectors of \( \Sigma_t \).

- Estimate the “characteristics factor” with a “regression”:
  \[ V_{t,i} = \left( \sum_{l \in Q_{i,t}} \Lambda_{tl}^2 \right)^{-1} \sum_{l \in Q_{i,t}} \Lambda_{tl} C_{t,l,i} \]

- Given \( V_{t,i} \) and \( \Lambda_t \) 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.
References


