

Discussion: Machine Learning Mutual Fund Flows

J. Fausch, M. Frigg, S. Ruenzi, F. Weigert

Emanuele Luzzi

USI Lugano and Swiss Finance Institute

December 17, 2025



Summary

Background: Understanding flows allows fund managers to better manage liquidity, and researchers to better differentiate skill.

Question: Can ML improve on mutual fund flow predictions?

Answer: Yes. R^2 : 22.8% (RF) vs. 18.2% (OLS) OOS (+25%)

Insight:

- Morningstar ratings dominating factor alphas.

Overall Assessment

- Solid methodology with REFORMS checklist (Kapoor et al., 2024)
- Interpretable ML: SHAP values reveal interesting information.
- First systematic ML application to flow prediction (vs. alpha prediction)
- Fun and intriguing read!
- My comments will focus on interpretation and mechanism.

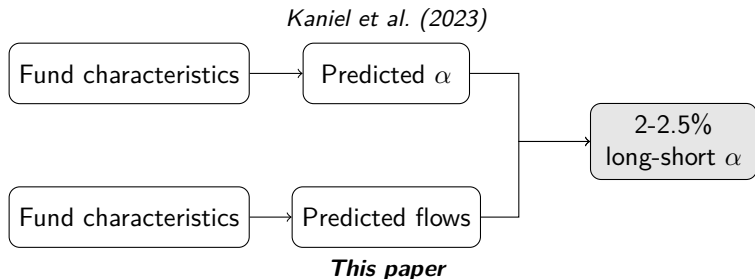
Comment 1: Morningstar puzzle

- Puzzle: Morningstar ratings predict more than performance.
- Ben-David et al. (2022) investors chase stars.
- **This paper:** Rating is $2\times$ more important than best α (SHAP).

What's driving this?

- ① **Information:** Nonlinear risk-adjustment captures something alphas miss.
→ *Add higher moments of returns; does rating importance drop?*
- ② **Behavioral:** Investors see stars, not factor regressions.
→ *Rating \times alpha interaction: ignored when stars are high?*
- ③ **Mechanical:** Categorical variable more ML-friendly than continuous alpha.
→ *Discretize alpha into quintiles; does importance rise?*

Comment 2: Flows vs α



Do flows add information beyond direct α prediction?

- Horse race: double sort on predicted flows \times predicted α .
- Mechanism: temporary (price pressure) vs. permanent (smart money)?

Comment 3: Prediction and performance

Finding: ML improvement over historical mean \uparrow Performance \downarrow

Interpretation: Poor prediction \rightarrow Poor liquidity management \rightarrow Poor performance.

Alternative explanations:

- Reverse: Poor performance \rightarrow Erratic behavior \rightarrow Poor prediction.
- Benchmark: historical mean. Is it fund's most likely used method?
- Strong assumption: funds not using ML in the sample period.

Comment 4: Is the OLS benchmark the right comparison?

Standard flow regressions in the literature include:

- Fund fixed effects → capture time-invariant characteristics.
- Time fixed effects → capture aggregate shocks.

This paper instead uses:

- Lagged flows as proxy for fund FE.
- 9 macro variables as proxy for time FE.

Suggestion:

- How does OLS *with* fund + time FE perform as a benchmark?

Minor comments

Data and variables:

- Text data (prospectuses, shareholder letters) + NLP could capture soft information.
- Holdings data: portfolio-level characteristics (concentration, active share).

Robustness and extensions:

- Crisis subsamples (2008, 2020): Does predictability collapse or strengthen?
- Heterogeneity: Does ML help more for small/young/high-turnover funds?
- International evidence: Do results generalize outside US?
- Time-varying expense ratio deserves more prominence.
- Imputing holdings for more granular prediction comparison (data if available).

Conclusion

- Fun and suggestive paper!
- Mixes methodological novelties with an important question.
- Relevant both for academics and practitioners.

Thank You!