Distant Investments: Decoding Mutual Fund Skill through Fund-Firm Semantic Alignment

Xiyuan Ma, Matthew Spiegel, Hong Zhang, Yijun Zhou

Discussion by Hongye Guo

June 2026

This paper

- Mutual fund positions in "distant" firms $\Longrightarrow$ important information
  - Distance: textual embeddings of **firm** 10K and **fund** perspectus
- Activeness $\times$ fund positions distance (SDI) ...
  - ... predicts fund performance
  - Tables 3 & 4
- For a given stock, SDI weighted trades across funds...
  - ... predicts stock performance
  - Tables 6

This paper

- We know texts contain useful info for both stocks and funds
- This paper:
  - **Stock**-level texts contain info about its holding **funds**
  - **Fund**-level texts contain info about **stocks** it holds
- **Cross-entity type** learning enabled by fund-stock pair level info
- A new type of channel that opens the door for a new branch of studies

Comment 1: equal weight

- An equally-weighted portfolio on a wide universe is hard to trade
  - Owing to large positions in small stocks
  - Overall, it is good to report them, but I view them as auxiliary
- Yet it is easy to find anomalies on small illiquid stocks
  - Unreasonably high t-stats & alphas & Sharpe ratios & ICs
  - Raised threshold $\Longrightarrow$ adverse selection
- FMB regressions use an equal weight
  - But your study is much less exposed to this problem!
  - Because mutual funds tend to trade large, liquid stocks
  - Large benchmark and active weights tend to go hand in hand
  - Traded stocks are not hard to trade

Comment 1: equal weight

- It may be a good idea to emphasize this point more
- One way to do this is to do equal weight on a narrow universe
  - e.g. S&P 500 universe
- Alternatively, cap-weight the FMB regressions
- Not going to be a problem for you
  - e.g. Table 3 already cap-weight

Comment 2: a closer look at the predictive power

- In constructing SDI, the authors remove the industry component
- Industry level information is not hard to process ...
- ...its removal produces a sharper signal for fund skill
- Common theme in quantitative signals
  - e.g. STR adjusts for some versions of peer group

Comment 2: a closer look at the predictive power

- How about the stock return prediction exercise?
- Signal: weighted fund trades across funds
  - The weight is SDI
- The fund had high value-added in the past, is it a stronger signal?
- The stock has low mutual fund ownership, is it a weaker signal?
- Alternative measures of fund trades
  - Currently 1, 0, -1 (right?)
  - The fund is buying a large number stocks (b/c, say, inflow)
  - Would you want to adjust this trade measure?
  - Fund-level adjustment before aggregating to the stock level
  - Analogous to the industry-level adjustment you do for distance

Other minor comments

- What is the right standard error for stock-fund pair level regressions?
  - Time should at least be one dimension
- Do high SDI funds invest in DUI firms overall?
- While we are at it: do we really want the acronym DUI?

Summary

- A paper with a good idea and good execution
- Also very comfortable to read
  - e.g., I like it very much that all tables are self-contained
- Promising results which may become even stronger