Classifying Hedge Fund Strategies with Large Language Models: Systematic vs. Discretionary Performance

Hui-Ching CHUANG[1] Chung-Ming KUAN[2]

[1]National Taipei University
[2]National Taiwan University

European Financial Management Association 2025 Annual Meeting

2025.06.28

We Are Interested in

- Classify hedge funds into systematic or discretionary
  - Model-based "systematic" funds vs. human-based "discretionary" funds (Harvey et al., 2017)
    - Similar classifications: "quantitative" vs. "qualitative"("discretionary" ,"fundamental"), and "machine" vs. "man"(Chincarini,2014; Abis, 2018; Evans et al.,2018)
- Evaluate the performance of classified funds
  - Is fund performance due to authentic skills or sampling luck?
  - Do systematic funds (as a group) outperform discretionary funds?

This Paper

- A new approach to classifying funds
  - Textual analysis is applied to convert text of investment strategies into numeric data and extract "features" from such data.
  - Large Language Model is fine-tuning to classify the investment styles.
  - Our approach captures strategy similarities and avoids subjective judgement or choice of keywords (cf. Harvey et al., 2017; Abis 2018).
- Evaluating fund performance
  - Implementing a statistical test with a false discovery adjustment under two-pass asset pricing models.
  - 10% to 20% of funds exhibit significant positive alphas
  - Funds classified as Systematic yield higher factor-adjusted returns than their Discretionary counterparts, on average.

Related Researches

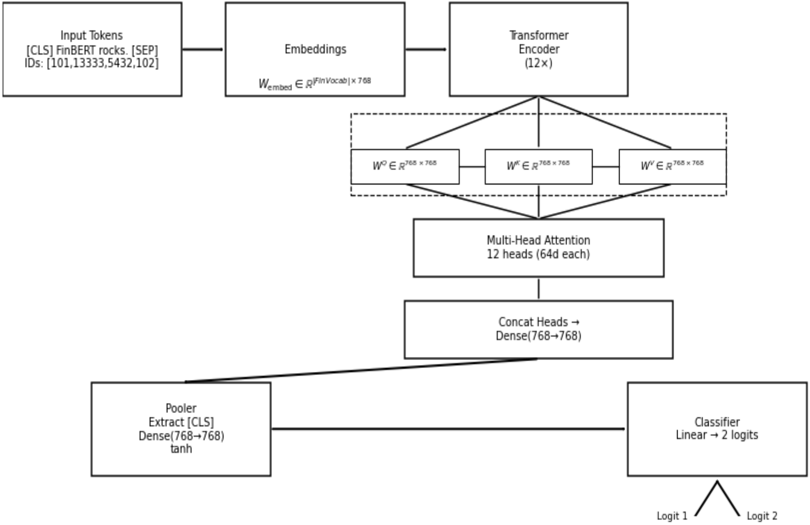
- Chincarini (2014)
  - word count: *algorithm, automate, econometric, mathematical, model, quantitative, statistic*
  - quantitative hedge funds have higher alphas than qualitative ones.
- Harvey et al. (2017)
  - word count approach. *algorithm, approx, computer, model, statistical,and system* are keywords used in their paper.
  - performances are similar.
- Abis (2018)
  - collected 2,607 mutual funds' "Principal Investment Strategies" in prospectuses from SEC.
  - classified manually a sub-sample of 200 prospectuses into two types.
  - apply machine learning (ML) methods to 200 (training sample) to classify the reaming funds (prediction sample).
  - compare stock picking/timing and holding performance and justify her empirical findings by a theoretical model.

Large Language Model: FinBERT

- BERT (Bidirectional Encoder Representations from Transformers), Devlin et al. (2019) pre-trained contextual information from both the left and right sides of a word on Wikipedia and BookCorpus.
- FinBERT, developed by Yang et al. (2020) and Huang et al. (2023), further trains BERT model on
  - SEC corporate filings (10-K and 10-Q),
  - financial analyst reports from Thomson Investext,
  - earnings call transcripts from SeekingAlpha.

  FinBERT captures contextual nuances in financial texts than the original BERT model.
- Domain-specific LLMs: science and biomedicine (Beltagy et al., 2019; Lee et al., 2020), legal studies (Chalkidis et al., 2020), ESG research (Huang et al., 2023; Webersinke et al., 2021), and innovation studies (Lee and Hsiang, 2020; Chuang et al., 2023)

Large Language Model: Fine-Tuning FinBERT



HFR Classification

- As Harvey et al. (2017), we only consider two main strategies (Equity Hedge and Macro) and their six sub-strategies in HFR.

Equity Hedge	Macro
Equity Market Neutral	Active Trading
Quantitative Directional	Commodity: Metals
Fundamental Growth	Commodity: Agriculture
Fundamental Value	Commodity: Energy
Sector: Energy/Basic Materials	Commodity: Multi
Sector: Healthcare	Currency: Discretionary
Sector: Technology	Currency: Systematic
Short Bias	Discretionary Thematic
Multi-Strategy	Systematic Diversified
	Multi-Strategy
Testing	Training

Classification: Training Sample

- Macro funds has natural candidate for training sample.
  - Systematic Diversified Macro funds: "investment processes that typically are functions of mathematical, algorithmic, and technical models, with little or no influence from individuals over the portfolio positioning."
  - Discretionary Thematic Macro funds: "primarily reliant on the evaluation of market data, relationships and influences, as interpreted by an individual or group of individuals who make decisions on portfolio positions."
- Training sample: Binary variable $y_i = 1$ if the i-th fund is a Systematic Diversified Macro fund and $y_i = 0$ if it is a Discretionary Thematic Macro fund; the feature matrix (explanatory variable matrix) of Macro funds as inputs to train classifiers.
- Our approach is free from subjective judgement of investment strategies/keywords.

Classification Performance

- In our training (on Macro funds) process:
  - Hold-out test set 15% (337 funds)
  - The remaining 85%, reserved 15% (286) for validation to monitor model performance
  - Leaving 1,619 funds for training

	Accuracy	AUC	Precision	F1	Recall
Training	93.14%	97.55%	92.38%	94.95%	97.66%
Validation	89.16%	93.47%	89.90%	91.99%	94.18%
Test	92.58%	96.16%	91.91%	94.53%	97.30%

- In our previous studies, we shown that the best performed ML algorithm (Random Forest among LDA, KNN, SVM, Classification Tree, and Gradient Boosting) achieved 86%, 82%, 86%, and 90% for accuracy, AUC, precision, and F1 score in the testing set respectively.

Classification: In sample performance

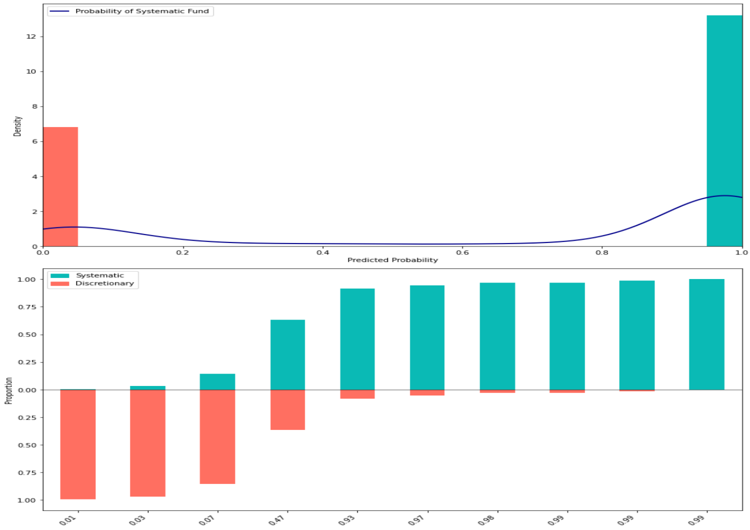


Figure 1: Equity Hedge: Discretionary

Figure 2: Equity Hedge: Systematic

Significant Performance under False Discovery Rate Control

Consider the following factor model:

$$E(r_i) = \alpha_i + \beta_i' \lambda, \quad i = 1, \cdots, N,$$

where r_i: excess return of fund i. α_i pricing error (alpha) of fund. β_i is a vector of risk exposures, and λ is the risk premia. We want to examine:

$$H_{0,i} : \alpha_i \leq 0, \quad i = 1, \cdots, N.$$

Rejecting the null hypothesis $H_{0,i}$ implies that the superior performance (positive alpha) of fund i is statistically significant and cannot merely be attributed to chance.

Significant Performance under False Discovery Rate Control

- We consider the following factor models
  - One factor: MKT
  - Three factors: MKT, SMB, HML
  - Five factors: PTFSBD, PTFSFX, PTFSCOM, PTFSIR, and PTFSSTK (returns from the long position of the lookback straddle of bonds, currencies, commodities, short-term interest rates, and stocks.)
  - Seven factors: MKT, SMB, CS (credit spread), Δ10Y, PTFSBD, PTFSFX, and PTFSCOM
  - Eleven factors: add HML, MOM, PTFSIR, and PTFSSTK to 7 factors

Fund Performance Comparison

Main Strategy	Style	Count	Mean	STD	Median	Mean Diff.
Seven-factor model (F7)						
Equity Hedge	Discretionary	2,728	8.24%	77.80%	15.29%	7.28%
	Systematic	1,177	15.52%	68.54%	15.82%	
Macro	Discretionary	338	10.20%	67.83%	7.52%	26.99%
	Systematic	791	37.19%	69.46%	33.57%	
Eleven-factor model (F11)						
Equity Hedge	Discretionary	2,728	11.35%	84.83%	17.10%	5.02%
	Systematic	1,177	16.37%	76.27%	17.40%	
Macro	Discretionary	338	13.86%	77.85%	15.59%	25.92%
	Systematic	791	39.79%	80.08%	39.39%	

Significant Performance under False Discovery Rate Control

- Giglio et al. (2021)'s test proceeds as follows:
  - First, they use observable risk factors to calculate risk exposures and residuals for each fund through time-series regression
  - Second, they employ matrix completion on the unbalanced residual matrix, Hastie et al. (2015), and use PCA to identify latent risk factors and exposures.
  - Third, they perform a cross-sectional regression of the mean excess return on the concatenated observed false and unobserved exposures to estimate risk premiums and fund alphas.
  - Finally, they apply the (adjusted) Benjamini-Hochberg False-discovery rate test to fund alphas to account for the data-snooping bias.
- We used the test from Giglio et al. (2021) to identify funds with positive alpha in each category without data-snooping bias and control for the false discovery rate at 5% level.

Significant Alphas by style across main strategies

Strategy Level	Style	F7	F11	F3+U4	F5+U2	U7
Panel A: All strategies						
	Discretionary	22.02%	19.57%	22.64%	23.39%	22.11%
	Systematic	18.70%	17.28%	21.39%	20.93%	19.21%
Panel B: Main strategies						
Equity Hedge	Discretionary	22.84%	20.16%	23.17%	24.05%	22.69%
	Systematic	24.55%	22.94%	24.38%	25.66%	23.53%
Macro	Discretionary	15.38%	14.79%	18.34%	18.05%	17.46%
	Systematic	9.99%	8.85%	16.94%	13.91%	12.77%

Conclusion

- This paper fine-tuned the FinBERT, a large language model (LLM) to mitigates the subjective judgment traditionally involved in categorizing investment strategies.
- Our classification performance exceeds most ML based approaches.
- We find that, on average, funds classified as Systematic yield higher factor-adjusted returns than their Discretionary counterparts.
- About 10% and 20% of funds exhibit statistically significant positive alphas in models combining observable and unobservable factors.

Reference

Beltagy, I., Lo, K., Cohan, A., 2019. Scibert: A pretrained language model for scientific text. arXiv preprint arXiv:1903.10676 .

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutsopoulos, I., 2020. LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:2010.02559 .

Chuang, H.C., Hsu, P.H., Lee, Y.N., Walsh, J.P., 2023. What share of patents is commercialized? Technical Report. Working Paper. Manuscript.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), pp. 4171–4186.

Giglio, S., Liao, Y., Xiu, D., 2021. Thousands of alpha tests. The Review of Financial Studies 34, 3456–3496.

Huang, A.H., Wang, H., Yang, Y., 2023. Finbert: A large language model for extracting information from financial text. Contemporary Accounting Research 40, 806–841.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J., 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36, 1234–1240.

Lee, J.S., Hsiang, J., 2020. Patent classification by fine-tuning bert language model. World Patent Information 61, 101965.

Webersinke, N., Kraus, M., Bingler, J.A., Leippold, M., 2021. Climatebert: A pretrained language model for climate-related text. arXiv preprint arXiv:2110.12010 .

Yang, Y., Uy, M.C.S., Huang, A., 2020. Finbert: A pretrained language model for financial communications. arXiv preprint arXiv:2006.08097 .