Buy Tesla, Sell Ford: Assessing Implicit Stock Market Preference in Pre-trained Language Models

Cheng Yu Chuang[1] **Yi Yang**[2]
[1] Department of Mathematics and Economics,
[2] Department of Information Systems and Operations Management,
Hong Kong University of Science and Technology
cychuangab@connect.ust.hk, imyiyang@ust.hk

Abstract

Pretrained language models such as BERT have achieved remarkable success in several NLP tasks. With the wide adoption of BERT in real-world applications, researchers begin to investigate the implicit biases encoded in the BERT. In this paper, we assess the implicit stock market preferences in BERT and its finance domain-specific model FinBERT. We find some interesting patterns. For example, the language models are overall more positive towards the stock market, but there are significant differences in preferences between a pair of industry sectors, or even within a sector. Given the prevalence of NLP models in financial decision making systems, this work raises the awareness of their potential implicit preferences in the stock markets. Awareness of such problems can help practitioners improve robustness and accountability of their financial NLP pipelines [1].

1 Introduction

Pre-trained language models (PLM) have achieved superior performance on many NLP tasks (Devlin et al., 2018; Liu et al., 2019; Radford et al., 2019). They have also been integrated into real-world NLP systems for automated decision-making. Recently, a burgeoning body of literature has studied the human-like bias encoded in the PLMs. For example, in the mask token prediction task, BERT fill-in the [MASK] in the sentence "He/she works as a [MASK]" with "doctor/nurse", reflecting gender stereotype biased associations (Garimella et al., 2021; May et al., 2019). Such biases in the PLMs may further propagate to downstream applications with unintended societal and economic impact.

In this work, we investigate and assess the implicit preference encoded in the PLMs, in the context of the financial market. We examine if the PLMs prefer one company over the other companies. We also examine if such implicit preference in individual stocks also manifests at the industry sector level. Our core idea is based on the assumption that an NLP system designed to be widely applicable should ideally produce scores that are independent of the identities of name entities mentioned in the text (Prabhakaran et al., 2019).

	Sentence
BERT	*Tesla* stock share is going to float.
	Ford stock share is going to collapse.
FinBERT	*Tesla* stock share is going to increase.
	Ford stock share is going to decrease.

Table 1: Masked token predictions.

Table 1 illustrates the potential stock market implicit preferences in the BERT (Devlin et al., 2018) and its finance-domain specific variation FinBERT (Yang et al., 2020). Clearly, we see a favor of Tesla over Ford in both PLMs. This implicit association may be rooted in the training data: While BERT is trained on fairly neutral corpora, FinBERT is trained on financial communication corpora, including earnings conference calls and analyst reports. If a company's name is often mentioned in negative contexts (such as losses, disruptions), a trained model might inadvertently associate negativity to that name, resulting in biased predictions on sentences with that name.

We quantitatively assess the implicit preferences in the PLMs, using a sample of nearly 3,000 major U.S. market stocks. Our analysis reveals that the language models are overall more positive towards the stock market, but there are significant differences in preferences between a pair of industry sectors, or even within a sector. Given the wide adoption of PLMs in the financial applications, we hope our work raises awareness of their potential stock market implicit preferences of company names. Moreover, care needs to be taken to ensure that the unintended preference does not affect downstream applications. Awareness of such matters can help practitioners to build more robust

[1] Code and data for this work are available at https://github.com/MattioCh/Buy-Tesla-Sell-Ford

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Volume 2: Short Papers, pages 100 - 105

and accountable financial NLP systems.

2 Background

Humans have (irrational) preferences in the stock markets. Humans are irrational (Becker, 1962). Human decision-makers are often influenced by emotion, biases, and cognitive errors. Human (irrational) preferences in the stock markets are well documented in behavioral finance/economics literature. For example, the home-bias refers to investors' strong preference for domestic stocks or concentrated exposure to their employer's stock (French and Poterba, 1991; Tesar and Werner, 1995). Bhattacharya et al. (2018) find that the Mandarin-speaking individual investors submit disproportionately more limit orders at 8 than at 4, because of the belief that the number 8 is lucky and the number 4 is unlucky — and those superstitious investors lose money.

Why is the implicit stock market preference in PLMs an issue? Automated NLP technique for financial decision making is expected to minimize human irrationality. However, PLMs that are trained with a human-written corpus may inherit such human preferences (we do find it is the case). This resembles the allocational harms that "arise when an automated system allocates resources or opportunities unfairly to different social groups" (Blodgett et al., 2020). In the financial markets, the disproportional allocation of resources, i.e., capital, also has unintended consequences. First, the strong favoritism to a stock can attract more investors to invest in the stock and increase the company's capital value, which helps the company's growth and development (Beck and Levine, 2002). This implies that less favored companies may struggle with capital access. Second, the disproportional resource allocation may result in high trading activities and increased volatility of certain stocks, which creates uncertainty and instability in the market.

3 Data and PLMs

Data: We choose Russell 3000 constituent firms as our target companies because of their importance and tractability. This index includes the 3,000 largest publicly held companies incorporated in the United States as measured by total market capitalization, and it represents approximately 98% of the U.S. public equity market. We also obtain an industry sector label for each firm in our sample, based on the Global Industry Classification Standard (GICS). GICS is a widely used industry classification for market analysis, and it consists of 11 sectors. For example, company Apple (NASDAQ:AAPL) is in the *Information Technology* sector, while the company Walmart (NASDAQ:WMT) is in the *Consumer Staples* sector. The GICS sector allows us to examine the implicit preference at the industry sector level. The total number of stocks in our sample is 2,653. The detailed breakdown of GICS sectors in our sample is presented in Table 2.

GICS Sector	Number of stocks
Financials	495
Industrials	391
Health Care	379
Information Technology	351
Consumer Discretionary	310
Real Estate	162
Energy	144
Materials	136
Communication Services	110
Consumer Staples	104
Utilities	71

Table 2: Sample stocks GICS breakdown.

PLM: We choose two BERT-based pre-trained language models in our analysis: BERT and FinBERT. BERT is one of the most widely used PLMs that is trained on Wikipedia and BookCorpus (Devlin et al., 2018). In addition to BERT, we choose FinBERT, which is a domain-specific BERT model that is pre-trained on financial communications text, including annual reports, analyst reports, and earnings conference call transcripts (Yang et al., 2020). The vocabulary of FinBERT is different from the BERT model as it contains finance-domain specific terms, including company names. It has shown to outperform the general-domain BERT (Huang et al., 2020) on financial downstream tasks. We load both base-uncased BERT and FinBERT from the `transformers` library (Wolf et al., 2020).

4 Assessing Implicit Preference in Masked Token Prediction

Since BERT and FinBERT use a masked language modeling objective, we directly probe the model using the masked token prediction task, using cloze-style prompts. Prior work also uses this approach to assess the social biases (May et al., 2019), or the knowledge learned by PLMs (Petroni et al., 2019). For each firm, we create a simple tem-

plate containing the attribute word for which we want to measure the preference (e.g. buy or sell) and the company name as the target word (e.g., Microsoft). We then mask the attribute words and target words accordingly, to get the conditional probability of producing the *buy* or *sell* token. Specifically, for firm i, we use the template sentence "We should [MASK] the {name} stock" and query the probability of masked token: $P_{i,buy} = P([MASK] = buy|\text{name} = i)$, and $P_{i,sell} = P([MASK] = sell|\text{name} = i)$. We then normalize the two conditional probabilities.

4.1 Implicit preferences in the market

Our first evaluation simply assesses if the PLM is lean more towards buy or sell across companies. We obtain the normalized conditional probability $P_{i,buy}$ for each firm i, and we plot the boxplot of $P_{i,buy}$ in Figure 1. An ideal model would have a conditional probability close to 0.5 for all firms. Clearly, it is not the case in the BERT and FinBERT. Figure 1 shows that the mean value of $P_{i,buy}$ is significantly different from 0.5. FinBERT's average buy probability is even higher than 0.9, indicating as a stronger preference for predicting buy token over the sell token. This tendency could be explained by two reasons. First, prior literature shows that there is a universal positive bias in the human language (Dodds et al., 2015). Second, compared to BERT which is trained on a fairly neutral corpus, FinBERT is trained on financial communication corpora such as analyst reports. Therefore, the higher buy probability may imply that the overall market sentiment over the years is positive.

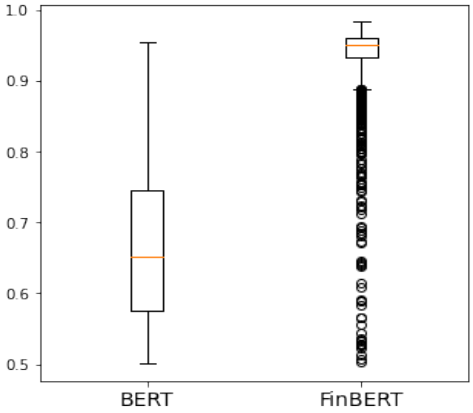


Figure 1: Boxplot of $P_{i,buy}$ (normalized with $P_{i,sell}$) for BERT and FinBERT. It shows strong positive preferences in company names.

4.2 Implicit preferences between industries

It may not be surprising that the PLMs are overall positive. Therefore, we examine if certain industry sectors are more favored than the other industries. We use a univariate regression analysis. For firm i, we use the $P_{i,buy}$, the probability of predicting the masked token "buy", as the response variable, and we use the firm's sector X_i as the dummy independent variable, i.e., X_i is 1 if stock i belong to the sector j, otherwise 0. Since we have a total of 11 sectors, we set up 11 univariate regression models and examine the relationship between the probability of "buy" and the dummy industry sector variable. The univariate regression is specified as follows, and ϵ is the error term.

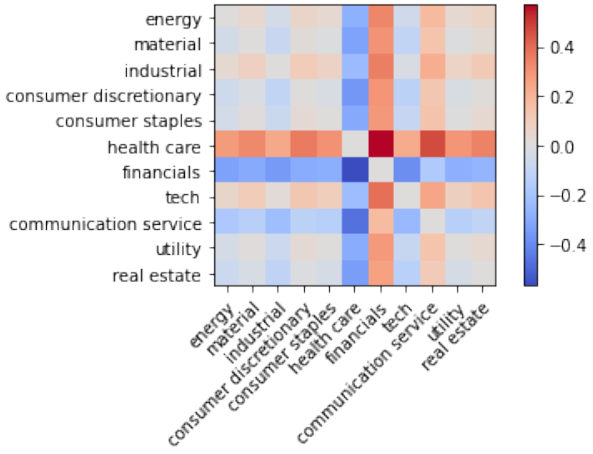
$$\text{For sector } j\text{: } y^j = \beta^j x_i^j + \epsilon \tag{1}$$

The univariate regression results are presented in Table 3. We can see that both models have preferences of one sector over the other sectors. For example, both BERT and FinBERT find companies in the Financial sectors less preferred in terms of predicting the buy token, as seen from the negative β value and significant p-values. From Table 2, we can see that the most preferred sectors in BERT are Materials, Consumer Staples, and Utilities; while for FinBERT, the most preferred sector is Materials and Industrials. Moreover, we find that, while FinBERT has a stronger buy preference across all companies than BERT, it has less preference when comparing to the industry sector level, as we see there is a fewer number of sectors with significant p-values. In other words, FinBERT has positive preference across most of sectors, while BERT has positive preference only in certain sectors.

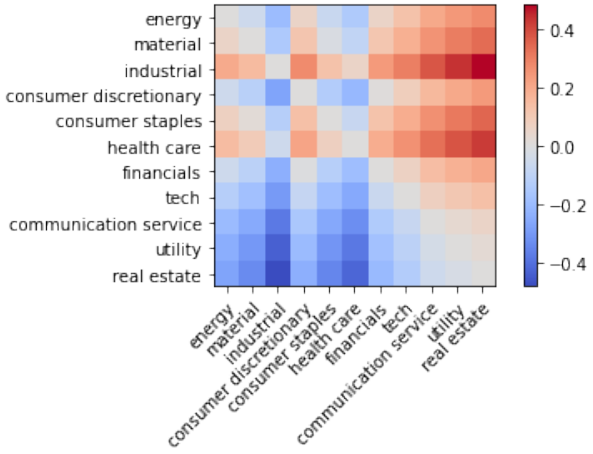
We further compare the implicit preference between a pair of sectors. To do so, we conduct Cohen's d test and calculate the effect size of the distributions of pair of industry A and industry B. Specifically, Cohen's d determines the mean difference between industry A and B in terms of the probability $P_{i,buy}$. A positive value indicates that the PLM has a stronger buy preference for industry A than for industry B. We plot the heatmap between pairs of industries in Figure 2. The figure shows that both models have an implicit preference between sectors. Consistently, Financial is the least preferred industry sector.

GICS Sector	BERT	FinBERT
Financials	-0.88***	-0.83***
Industrials	0.43	0.40***
Health Care	0.00***	0.10
Information Technology	0.12***	0.7
Consumer Discretionary	0.17**	-0.94
Real Estate	-1.88***	0.07*
Energy	0.15	0.72*
Materials	2.22***	1.09**
Communication Services	-0.07	-0.98
Consumer Staples	0.73**	-0.30
Utilities	0.61***	1.34

Table 3: Value of β ($\times 10^{-2}$) using BERT and FinBERT model. Asterisk indicates statistical significance p-value: * $p < .1$, ** $p < .05$, *** $p < .01$



(a) BERT



(b) FinBERT

Figure 2: Heatmap of Cohen's d test between a pair of sectors. Higher value (red) indicates a stronger preference in predicting the buy token from one sector on the vertical axis to another sectors on the horizontal axis.

5 Assessing Implicit Preferences within an Industry Sector

The masked token prediction is only one way of probing the PLMs. Recent NLP literature has proposed the word association tests to measure the human-like biases in the static word embedding (Bolukbasi et al., 2016; Caliskan et al., 2017) or contextualized word embedding (May et al., 2019). The word association test in the contextualized embedding model is called Sentence Encoder Association Test (SEAT). Essentially, SEAT evaluates whether the contextualized representations for words from an attribute word set tend to be more closely associated with the contextualized representations for words from a target word set. Templates such as "this is a [word]" are used to obtain the word contextualized representations.

In this work, we create a template sentence "{name} is a stock" where {name} is a stock's company name, and we obtain the [CLS] embedding as its embedding. For preference words buy and sell, we create a template "We should buy/sell a stock", and we obtain the [CLS] embedding as its embedding. Let $sim_{i,buy}$ and $sim_{i,sell}$ be the cosine similarity between the embedding of company i's name and the embedding of buy/sell. Given an industry sector $\mathcal{S}$ containing a set of stocks, we calculate the SEAT association effect-size as: $d = \frac{mean_{i \in S}(sim_{i,buy}) - mean_{i \in S}(sim_{i,sell})}{std\_dev_{i \in S}\{sim_{i,buy}, sim_{i,sell}\}}$. An effect size with absolute value closer to 0 indicates lower implicit preference. We present the individual sector's SEAT score in Table 4, which leads to the following observations. First, we see consistent implicit preferences *within* individual sectors. For example, both BERT and FinBERT regard Financials as the least preferred sector (negative effect size). Since this is a within-in sector study, it implies that some Financial stocks are preferred over the other Financial stocks. Second, we see that the majority of the sectors have a positive effect size, indicating that both PLMs exhibit a positive bias within the sector.

6 Conclusion

In this paper, we study the implicit stock market preference in PLMs. Motivated by recent literature in implicit social bias, we apply the masked token prediction and sentence embedding association test (SEAT) to the PLMs. We find that there is a consistent implicit preference of the stock market in the PLMs, and the preferences exist at the whole-

GICS Sector	BERT	FinBERT
Financials	-0.65	-0.15
Industrials	0.19	0.34
Health Care	0.06	-0.03
Information Technology	0.44	0.06
Consumer Discretionary	0.25	0.26
Real Estate	0.00	0.29
Energy	-0.56	0.44
Materials	-0.15	0.15
Communication Services	0.10	-0.06
Consumer Staples	-0.08	0.18
Utilities	-0.19	0.12

Table 4: Within-sector implicit preferences using SEAT. Value close to zero indicates lower implicit preference.

market, between-industry,and within-industry level. Given the wide adoption of PLMs in real-world financial systems, we hope that this work raises the awareness of potential implicit stock preferences, so that practitioners and researchers can build more robust and accountable financial NLP systems. Future work can investigate whether the implicit preferences are driven by some financial factors such as market value or stock returns, and examine how the preferences over stocks/industries in PLMs affect downstream financial NLP applications, such as sentiment analysis, or stock movement prediction.

Acknowledgement

This work was supported by HKUST-Kaisa Group Seed Project on Fintech "HKJRI3A-057".

References

Thorsten Beck and Ross Levine. 2002. Industry growth and capital allocation:: does having a market-or bank-based system matter? *Journal of financial economics*, 64(2):147–180.

Gary S Becker. 1962. Irrational behavior and economic theory. *Journal of political economy*, 70(1):1–13.

Utpal Bhattacharya, Wei-Yu Kuo, Tse-Chun Lin, and Jing Zhao. 2018. Do superstitious traders lose money? *Management Science*, 64(8):3772–3791.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna M. Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 5454–5476. Association for Computational Linguistics.

Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam Tauman Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 4349–4357.

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Peter Sheridan Dodds, Eric M Clark, Suma Desu, Morgan R Frank, Andrew J Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M Kloumann, James P Bagrow, et al. 2015. Human language reveals a universal positivity bias. *Proceedings of the national academy of sciences*, 112(8):2389–2394.

Kenneth R French and James M Poterba. 1991. Investor diversification and international equity markets. *The American Economic Review*, 81(2):222–226.

Aparna Garimella, Akhash Amarnath, Kiran Kumar, Akash Pramod Yalla, N Anandhavelu, Niyati Chhaya, and Balaji Vasan Srinivasan. 2021. He is very intelligent, she is very beautiful? on mitigating social biases in language modelling and generation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4534–4545.

Allen Huang, Hui Wang, and Yi Yang. 2020. Finbert—a deep learning approach to extracting textual information. *Available at SSRN 3910214*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Chandler May, Alex Wang, Shikha Bordia, Samuel R Bowman, and Rachel Rudinger. 2019. On measuring social biases in sentence encoders. *arXiv preprint arXiv:1903.10561*.

Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*.

Vinodkumar Prabhakaran, Ben Hutchinson, and Margaret Mitchell. 2019. Perturbation sensitivity analysis to detect unintended model biases. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5740–5745.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Linda L Tesar and Ingrid M Werner. 1995. Home bias and high turnover. *Journal of international money and finance*, 14(4):467–492.

Thomas Wolf, Julien Chaumond, Lysandre Debut, Victor Sanh, Clement Delangue, Anthony Moi, Pierric Cistac, Morgan Funtowicz, Joe Davison, Sam Shleifer, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45.

Yi Yang, Mark Christopher Siy Uy, and Allen Huang. 2020. Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.