What to Believe on Social Media Platforms? The Importance of Consistency Between Words and Actions

Zhuo Chen, Honghai Yu, Yunmiao Zhang, Guofu Zhou

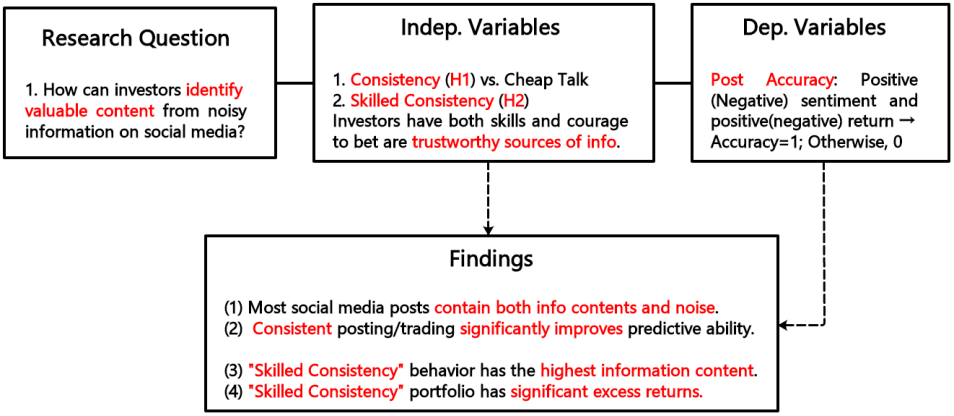
CICF, 2025

Interpreter: Weihao Zhao

School of Economics and Management WHU

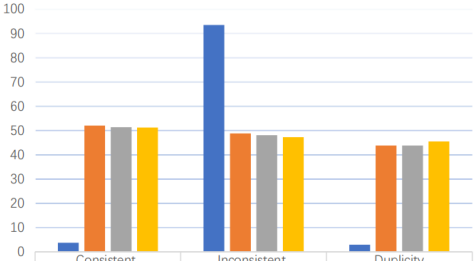
2025-06-27

Framework


Research Question
1. How can investors identify valuable content from noisy information on social media?
Indep. Variables
1. Consistency (H1) vs. Cheap Talk
2. Skilled Consistency (H2)
Investors have both skills and courage to bet are trustworthy sources of info.
Dep. Variables
Post Accuracy: Positive (Negative) sentiment and positive(negative) return → Accuracy=1; Otherwise, 0
Findings
(1) Most social media posts contain both info contents and noise.
(2) Consistent posting/trading significantly improves predictive ability.
(3) "Skilled Consistency" behavior has the highest information content.
(4) "Skilled Consistency" portfolio has significant excess returns.

Motivation

The vast majority of posts **cannot accurately predict** future prices(b/c inconsisitency).



	Consistent	Inconsistent	Duplicity
Proportion (%)	3.66	93.48	2.86
Pos_Acc_1week (%)	52	48.79	43.81
Pos_Acc_1month (%)	51.36	48.02	43.81
Pos_Acc_3months (%)	51.19	47.25	45.45

Social media generally suffers from information noise and serious bias.

Motivation

1. **Tons of investors make investment decisions based on opinions of others.**
   - Investment info on social media is wide spreading and important.
   - These 'opinions' contain both **valuable content and a lot of noise.**
   - The poster may post untrue opinions due to **impression management**.
2. **No effective method to distinguish between "true and false statements":**
   - Relying solely on text content to determine the authenticity of info may be insufficient;
   - Lack of attention to the real trading behavior of the poster;
3. **Whether viewpoint of 'skilled investors' is more credible has not been explored.**

Contributions

1. Literature on the Impact of Social Media on Financial Markets

Prior: There is a correlation between social media content and market indicators(stock volatility, price trends, etc).(Farrell et al. 2022)

- Openness and anonymity of social media brought info quality issues.(Jia et al. 2020)

Extend: Using 'behavioral consistency' as a new variable for identifying information content.

- Unifying previous conflicting conclusions about "effective/ineffective" info on social media.

2. Literature on Heterogeneity of Investor Behavior and Ability Recognition

Prior: Info itself has noise, there is a distinction between "skilled/non-skilled" investors.

Extend: Introduce 'real transaction data' to identify 'skills' and 'confidence' of poster.

Hypothesis

H1: 'Consistent Words and Actions'

If an investor actually makes a matching transaction after publishing a certain viewpoint, it indicates that his/her viewpoint is more credible and has predictive ability.

H2: 'Skilled Consistency'

If an investor has already made a profit before expressing opinion and subsequent trading behavior is consistent with opinion, then this "skills + confidence" statement is more likely to contain valuable information.

Samples

Sample platform: Snowball; **Sample time**: 01/2016 ~ 12/2023

Original post data: 2449427; **Posts with stock tags**: 1364506

Real-time transaction records: 1306334; **Real-time users**: 11246

Variable	Def	Calc. Method
Consistency	发帖情绪与交易方向是否一致	若情绪为正且买入，或情绪为负且卖出，记为 1；反向为-1；无交易为 0
Profitable	发帖前是否在该股票上盈利	基于前一次买入与当前价格的差值，正为 1，否则为 0
Skilled Consistency	技能 + 一致性复合指标	若发帖者之前盈利且之后交易与观点一致，记为 1，否则为 0

Model Design and Analysis Methods

(1) Consistency and Info Content Model:

$$\textbf{Posting Accuracy}_{ijt} = \alpha + \beta_1 \cdot \textbf{Consistency}_{ijt} + \gamma' X_{ijt} + \epsilon_{ijt}$$

(2) Investor skill testing model:

$$\textbf{Posting Accuracy}_{ijt} = \alpha + \beta_1 \cdot \textbf{Profitable}_{ijt} + \gamma' X_{ijt} + \epsilon_{ijt}$$

(3) Skilled Consistency Model:

$$\textbf{Posting Accuracy}_{ijt} = \alpha + \beta_1 \cdot \textbf{SkilledConsistency}_{ijt} + \gamma' X_{ijt} + \epsilon_{ijt}$$

H1: Consistency of behavior after posting enhances credibility of info

	One week			One month	Three months
	(1)	(2)	(3)	(4)	(5)
Consistency	0.040***	0.039***	0.041***	0.031***	0.018***
	(5.74)	(5.55)	(5.10)	(3.82)	(2.12)
Volatility		-0.010***	-0.018***	-0.020***	-0.032***
		(-3.08)	(-3.25)	(-2.73)	(-4.02)
Illiq		0.003	0.008***	0.003	0.009**
		(1.41)	(2.85)	(0.83)	(2.41)
Social Engagement		-0.001	0.001	0.001	0.004*
		(-0.75)	(0.25)	(0.62)	(1.71)
Length		-0.006	-0.008	-0.016***	-0.005
		(-0.96)	(-1.45)	(-3.42)	(-1.31)
Article		-0.005	-0.009	-0.003	-0.015
		(-0.45)	(-1.00)	(-0.27)	(-1.62)
Analyst Coverage		-0.003	-0.003	-0.009	-0.015**
		(-0.75)	(-0.64)	(-1.49)	(-2.40)
News Coverage		0.000	-0.007	-0.012*	-0.016**
		(0.04)	(-1.36)	(-1.87)	(-2.22)
Investor Fixed Effect	No	No	Yes	Yes	Yes
Date Fixed Effect	No	No	Yes	Yes	Yes
Stock Fixed Effect	No	No	Yes	Yes	Yes
Observations	81,388	81,388	79,593	79,593	79,593
Within_Adjusted.R^2	0.000	0.001	0.001	0.002	0.003

	One week			One month	Three months
	(1)	(2)	(3)	(4)	(5)
Profitable	-0.019***	-0.019***	-0.017***	-0.039***	-0.051***
	(-3.67)	(-3.41)	(-3.18)	(-5.80)	(-5.98)
Volatility		-0.011***	-0.018***	-0.020***	-0.032***
		(-3.17)	(-3.32)	(-2.83)	(-4.14)
Illiq		0.003	0.008**	0.001	0.007*
		(1.13)	(2.57)	(0.36)	(1.85)
Social Engagement		-0.002	0.000	0.001	0.003
		(-0.96)	(0.15)	(0.39)	(1.47)
Length		-0.006	-0.008	-0.016***	-0.005
		(-0.88)	(-1.45)	(-3.42)	(-1.27)
Article		-0.005	-0.009	-0.003	-0.015
		(-0.43)	(-1.03)	(-0.27)	(-1.63)
Analyst Coverage		-0.003	-0.003	-0.008	-0.013**
		(-0.68)	(-0.51)	(-1.26)	(-2.14)
News Coverage		-0.000	-0.007	-0.012*	-0.016**
		(-0.08)	(-1.37)	(-1.86)	(-2.23)
Investor Fixed Effect	No	No	Yes	Yes	Yes
Date Fixed Effect	No	No	Yes	Yes	Yes
Stock Fixed Effect	No	No	Yes	Yes	Yes
Observations	81,388	81,388	79,593	79,593	79,593
Within_Adjusted.R^2	0.000	0.000	0.001	0.003	0.004

Skills alone are not enough to prove the credibility of a post.

H2: "Skilled + Confidence" has the most predictive ability

		One week		One month	Three months
	(1)	(2)	(3)	(4)	(5)
Skilled Consistency	0.037***	0.036***	0.032**	0.033**	0.042***
	(3.12)	(3.06)	(2.25)	(2.06)	(2.64)
Volatility		-0.011***	-0.018***	-0.020***	-0.032***
		(-3.19)	(-3.28)	(-2.74)	(-4.03)
Illiq		0.003	0.009***	0.003	0.009**
		(1.46)	(2.86)	(0.84)	(2.44)
Social Engagement		-0.001	0.001	0.001	0.004*
		(-0.73)	(0.24)	(0.62)	(1.72)
Length		-0.006	-0.008	-0.016***	-0.005
		(-0.95)	(-1.46)	(-3.42)	(-1.31)
Article		-0.005	-0.009	-0.003	-0.015
		(-0.44)	(-1.02)	(-0.28)	(-1.63)
Analyst Coverage		-0.003	-0.003	-0.009	-0.015**
		(-0.75)	(-0.64)	(-1.49)	(-2.41)
News Coverage		0.000	-0.007	-0.012*	-0.016**
		(0.07)	(-1.37)	(-1.88)	(-2.22)
Investor Fixed Effect	No	No	Yes	Yes	Yes
Date Fixed Effect	No	No	Yes	Yes	Yes
Stock Fixed Effect	No	No	Yes	Yes	Yes
Observations	81,388	81,388	79,593	79,593	79,593
Within_Adjusted.R^2	0.000	0.001	0.001	0.002	0.003

Posts with consistent behavior can bring returns

	Raw return	CAPM α	FF3 α	FF5 α
Panel A: Skilled and Consistency				
Equal-Weighted				
Long	0.886	0.588	0.297	0.227
	(0.89)	(0.78)	(0.39)	(0.27)
Short	-1.915	-2.293*	-2.345**	-2.714**
	(-1.52)	(-1.95)	(-2.03)	(-2.40)
Long-Short	2.801*	2.755*	2.516*	2.815**
	(1.92)	(1.93)	(1.77)	(1.99)
Value-Weighted				
Long	0.720	0.418	0.147	0.176
	(0.70)	(0.54)	(0.19)	(0.21)
Short	-1.548	-1.910**	-1.938**	-2.407***
	(-1.38)	(-2.10)	(-2.09)	(-2.57)
Long-Short	2.268*	2.202*	1.959*	2.457**
	(1.68)	(1.94)	(1.71)	(2.22)

Panel B: Non Skilled and Consistency				
Equal-Weighted				
Long	0.014	-0.394*	-0.355*	-0.341*
	(0.03)	(-1.78)	(-1.70)	(-1.73)
Short	-0.994*	-1.405***	-1.380**	-1.345***
	(-1.71)	(-4.19)	(-4.81)	(-4.67)
Long-Short	1.009***	0.885***	0.900***	0.878***
	(3.84)	(3.07)	(3.43)	(3.10)
Value-Weighted				
Long	0.896	0.507	0.530**	0.404
	(1.52)	(1.45)	(2.36)	(1.62)
Short	-0.237	-0.618**	-0.638***	-0.616**
	(-0.43)	(-2.40)	(-2.65)	(-2.51)
Long-Short	1.132***	0.999***	1.042**	0.894***
	(2.86)	(2.80)	(3.42)	(2.73)

The posts not only has predictive power, also can be used to build profit strategies.

研究局限性与未来展望

1. 在其他金融市场（如美国的 Seeking Alpha、StockTwits）开展类似研究

- 检验“行为一致性”在不同市场制度与文化下是否同样有效；
- 探索是否存在地域性偏差、监管因素对发帖/交易行为的影响。

2. 多维行为一致性框架：交易行为之外的信息一致性是否同样重要?

- 将“行为一致性”从发帖 vs 交易，拓展为，发帖 vs 其他行为轨迹的一致性，例如：
- 发帖观点 vs 投票态度（如点赞哪些股票）
- 发帖内容 vs 评论内容（是否反复自我矛盾）
- 同一人前后发帖是否保持一致性，还是存在“反复横跳”行为。

Appendix– 数据筛选

利用基于 Python 的网络爬虫工具，成功获取了雪球投资者的详细交易数据

- 数据涵盖了交易日期、时间、股票名称、交易价格和持仓比例。
- 投资者基本信息：用户 ID、地理位置、性别、发帖数量、关注者数量和粉丝数量；
- 帖子的相关信息：帖子 ID、发布日期、发布时间、点赞数、收藏数、评论数、分享数、提示数和发布设备。

样本筛选标准（数据清洗）

- 剔除非投资相关帖子（基于股票标签和情感词过滤）
- 剔除异常交易记录（如价格为 0、仓位比例异常）
- 剔除非 A 股标的（如 ETF、港股、美股等）
- 时间归属规则：按交易日下午 3 点为界划分帖子所属交易日

Appendix –控制变量

变量	含义	系数含义
Volatility	个股历史波动率	负系数：波动性越大，预测越困难
Illiq	越高表示越不流动	多数不显著，仅个别显著为正
Social Engagement	点赞 + 评论 + 转发	多数不显著 → “互动热度”不能提升帖子准确性
Length	帖子长度	在 (4) 显著为负 → 长帖未必更准，可能“长而空”
Article	是否为文章型长贴	不显著，文章格式对预测无影响
Analyst Coverage	被分析师覆盖数量	说明“热门股”预测更难
News Coverage	新闻曝光数量	信息越多越混杂，反而难预测

组合构建与噪声定义

为验证实证结果的经济意义，文章设计了动态多空组合策略：

- 每日筛选发帖者过去盈利且言行一致的帖，若情绪为正，股票纳入多头；否则纳入空头；
- 每周动态调仓，计算多空收益，并估算其在 CAPM、FF3、FF5 模型下的 alpha

社交媒体中信息噪声与误导性问题

- 自我提升偏误（self-enhancing transmission bias）：投资者倾向于分享自己盈利的经历，而非失败案例
- 印象管理动机（impression management）：投资者为展示“专业形象”而发布并非真实看法的内容
- “回音室效应（echo chamber）”：投资者只关注与自身观点一致的声音，信息趋同导致放大偏差