Social media and finance

FIRN Masterclass – Day 3

J. Anthony Cookson
CU Boulder

My plan

1. Social media landscape (today)
   Definitions, growth over time, data sources
   <u>**Some key examples**</u>

2. **Social media as a lens**
   Examples and approach, opportunities with new formats and features

3. Social transmission bias and social media signals
   Examples and approach.

4. Effects of social media
   Subtopic: production, consumption, and distribution of information
   Opportunities and challenges

Social transmission bias

Old roots: much of Gary Becker's research studied social preferences

New interest: Hirshleifer proposes that we study **socially emergent phenomena**.

That is, not just that we as financial actors have social preferences, but that these preferences shape the environment in such a way that… **new things happen**.

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

ARTICLE | Open Access

Presidential Address: Social Transmission Bias in Economics and Finance

DAVID HIRSHLEIFER

First published: 27 May 2020 | https://doi.org/10.1111/jofi.12906 | Citations: 175

CU Full Text

[Correction added on 18 June 2021, after first online publication: Copyright line has been updated in this version.]

SECTIONS PDF TOOLS SHARE

ABSTRACT

I discuss a new intellectual paradigm, *social economics and finance*—the study of the social processes that shape economic thinking and behavior. This emerging field recognizes that people observe and talk to each other. A key, underexploited building block of social economics and finance is *social transmission bias*: systematic directional shift in signals or ideas induced by social transactions. I use five "fables" (models) to illustrate the novelty and scope of the transmission bias approach, and offer several emergent themes. For example, social transmission bias compounds recursively, which can help explain booms, bubbles, return anomalies, and swings in economic sentiment.

Social transmission bias

Figure 2 from Hirshleifer

Six Themes about Social Economics and Finance

I offer six themes about a new intellectual paradigm, *social economics and finance*, and one of its key intellectual building blocks, *social transmission bias*.

- **Compounding:** Social transmission bias compounds recursively, so small bias can have large effects.
- **Idiosyncrasy:** Social transmission bias helps explain why aggregate outcomes are often error-prone and unpredictable.
- **Dynamics:** Social transmission bias offers an endogenous *social* explanation for action booms, price bubbles, and swings in investor sentiment.
- **Emergence:** Socially emergent behavior often looks completely different from individual propensities.
- **Mimicry:** Social emergence can easily create the illusion of a direct individual propensity "for" a behavior when no such propensity exists. So the inferences drawn from empirical tests of behavioral hypotheses are often overstated.
- **Proxies:** This approach suggests *new test variables* for empirical research: (i) general social and network proxies, and (ii) proxies for transmission bias.

This framework is ready-made for social media ecosystems

An example of such a model of bubbles
Chinco (2023MS)

Home > Management Science > Vol. 69, No. 2 >

The Ex Ante Likelihood of Bubbles

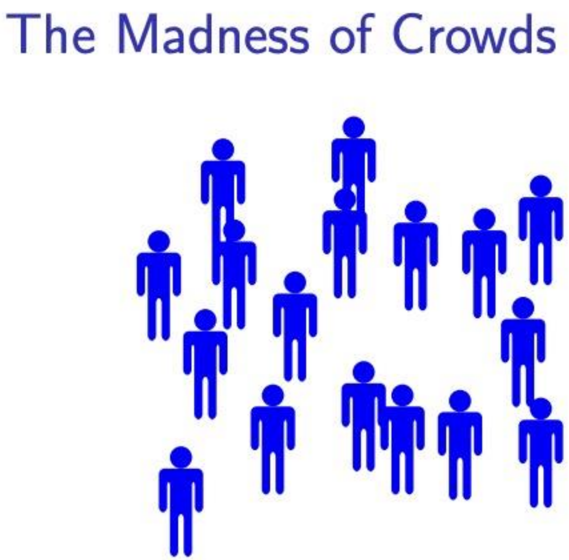
Alex Chinco

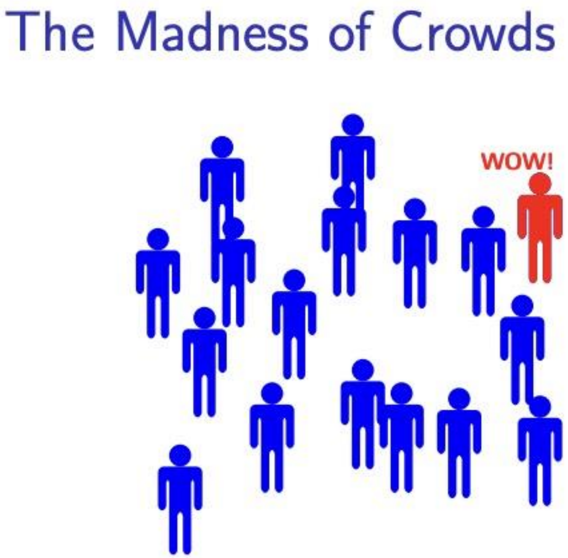
Published Online: 4 May 2022 | https://doi.org/10.1287/mnsc.2022.4351

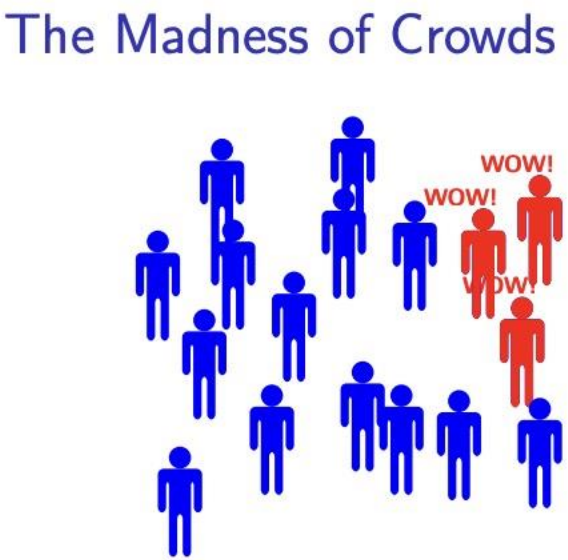
Abstract

The limits of arbitrage explain how a speculative bubble is sustained; they do not explain how likely one is to occur. To do that, you need a theory about the thing that sporadically causes arbitrageur constraints to bind. I propose a first such theory, which is based on social interactions between speculators. The theory says that bubbles should be more likely in assets where increases in past returns make excited-speculators relatively more persuasive to their peers. I empirically verify this ex ante prediction about bubble likelihoods and show that it is robust to some ex post disagreement about bubble definitions.

The model in pictures





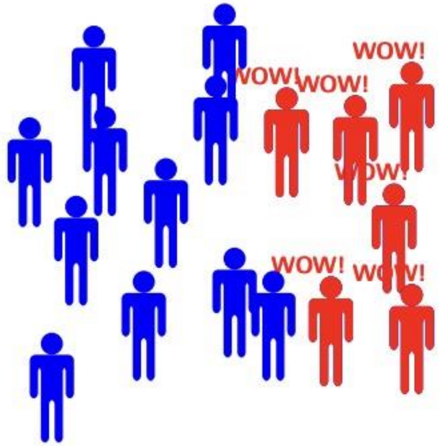


The model in pictures

Intuition: **Excitement** and **apathy** compete in the model & you get bubbles if **`theta`** is big.

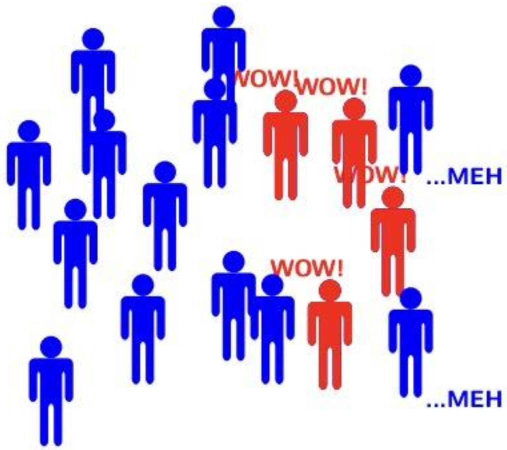
If theta is big, ideas spread explosively (think of SIR models after Covid)

The Madness of Crowds



Excitement rate: $\theta r(1-n)n$

The Madness of Crowds



Excitement rate: $\theta r(1-n)n$

— nonlinear in n
— increases in r (past return)
— increases in θ (persuasiveness)

Apathy rate: n

Empirical evidence in Chinco (2023) for news media

Measure `theta` as Corr(media coverage with past return) by industry

Then, predict industry-level bubble events as in Greenwood, Shleifer and You (2018)

It seems to do well!

But, this was news media… *it would be interesting to measure these constructs in social media*.

I haven't seen someone do this yet, but it would be interesting once someone does..

Dependent Variable: `WillBeBubble`	(1)	(2)	(3)	(4)	(5)
`Intercept`	$9.22^{\star}$ (4.97)	-0.39 (6.68)	0.78 (7.18)	5.34 (8.03)	-5.51 (9.19)
`Theta`	$3.79^{\star\star}$ (1.60)	$3.96^{\star\star}$ (1.57)	$4.17^{\star\star}$ (1.60)	$3.88^{\star\star}$ (1.61)	$4.25^{\star\star\star}$ (1.58)
`%ΔSales`		$1.07^{\star\star}$ (0.51)			$1.08^{\star}$ (0.55)
`Turnover`			5.27 (3.26)		4.80 (3.27)
`%Issuer`				0.12 (0.20)	-0.08 (0.21)
Matched on...					
`PastReturn`	✓	✓	✓	✓	✓
`ReturnVol`	✓	✓	✓	✓	✓
`CAPE`	✓	✓	✓	✓	✓

Social Transmission Bias in Finance

A non-random sampling of papers

Han, Liu and Sui (2025) study Bitcointalk (a forum for Bitcoin discussions) and study how ideas spread.

Chen and Hwang (2022) study Seeking Alpha readership and sharing patterns, finding selective sharing.

Cookson, Engelberg and Mullins (2023 RFS) study selective information sourcing

The Review of Financial Studies

Volume 36, Issue 2
February 2023

JOURNAL ARTICLE

Echo Chambers

Get access ›

J Anthony Cookson ✉, Joseph E Engelberg, William Mullins

The Review of Financial Studies, Volume 36, Issue 2, February 2023, Pages 450–500, https://doi.org/10.1093/rfs/hhac058

Published: 17 August 2022 **Article history** ▾

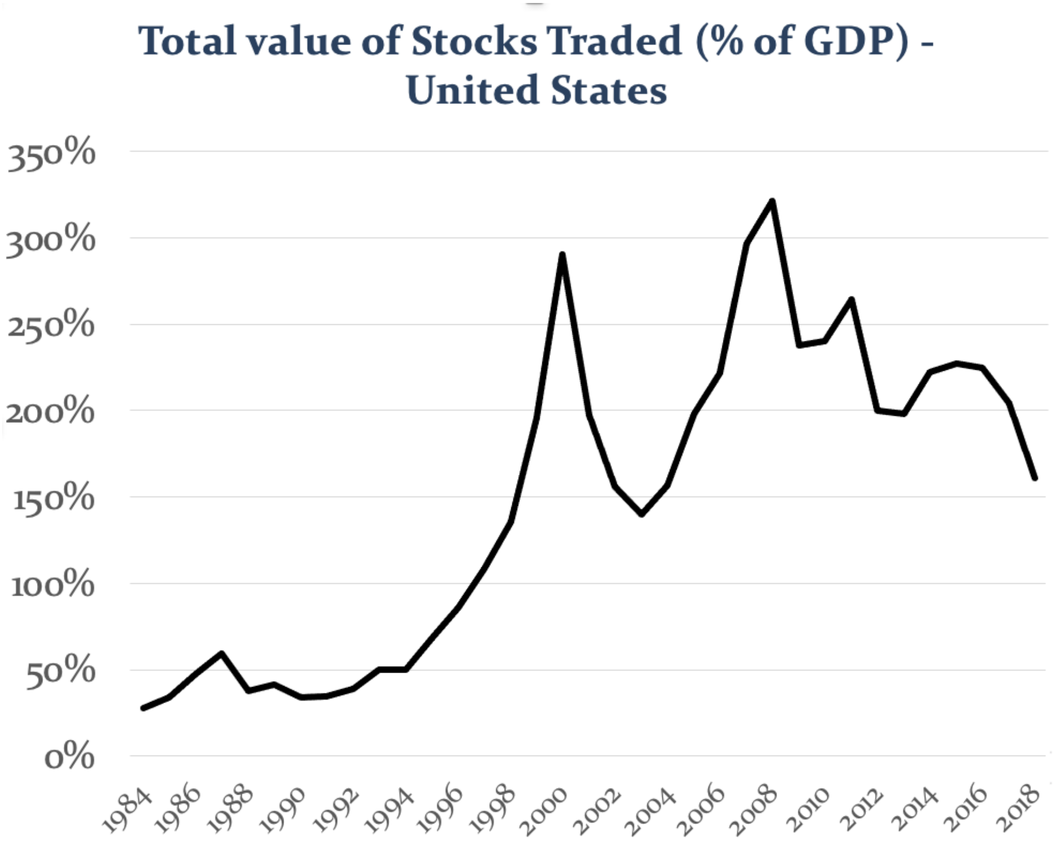
Cite | Permissions | Share ▾

The Volume Puzzle

"The Great Unsolved Problem of Financial Economics" (Cochrane, 2016)

*"The **sheer volume of trading** is the puzzle.*

*… non-information mechanisms — life-cycle, preference shocks, rebalancing among heterogeneous agents, preference shifts, generate trading volume. But **they do not generate the astronomical magnitude and concentration of volume that we see**."*



The Volume Puzzle

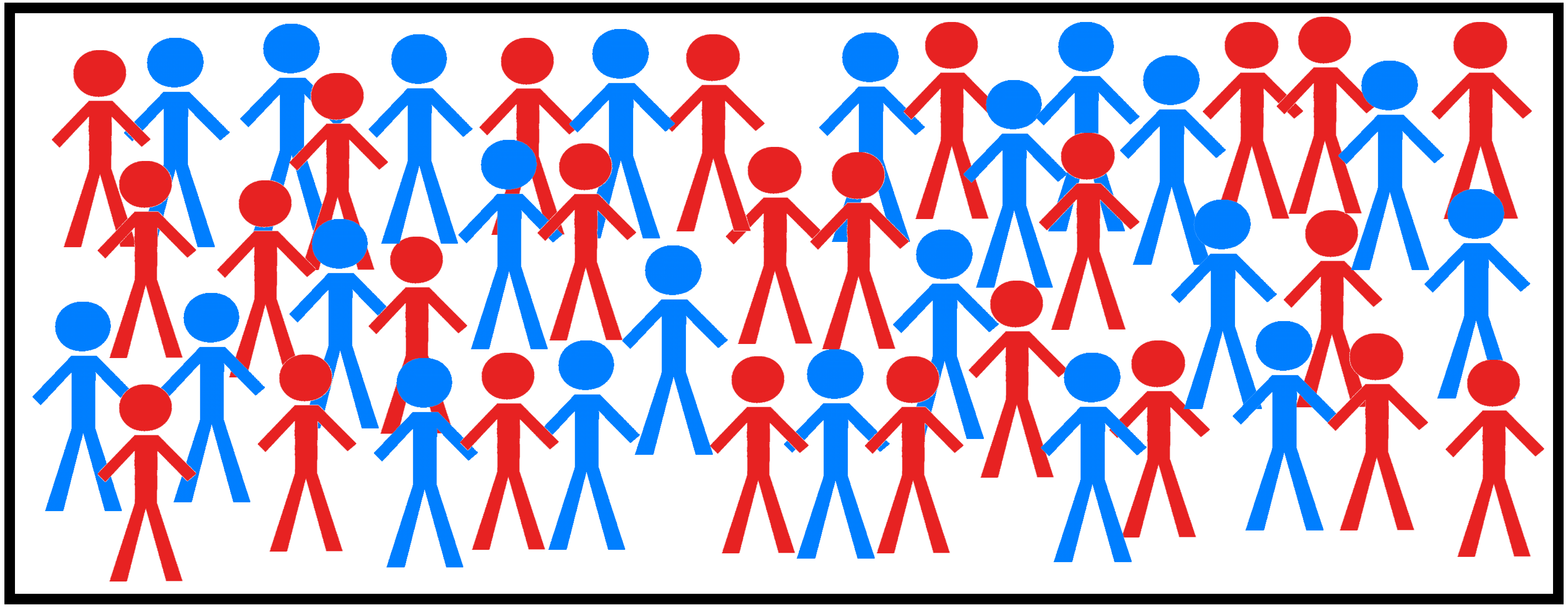
- An important explanation for high trading volume is investor **disagreement** (Harris & Raviv, 1993)

  - Disagreement due to investors:
    - (i) using different models (e.g., Kandel & Pearson 1995)
    - (ii) using different information (e.g., Hong & Stein, 1999)

- But, “disagreement” as an explanation for volume begs the question: *why* do investors persistently disagree?

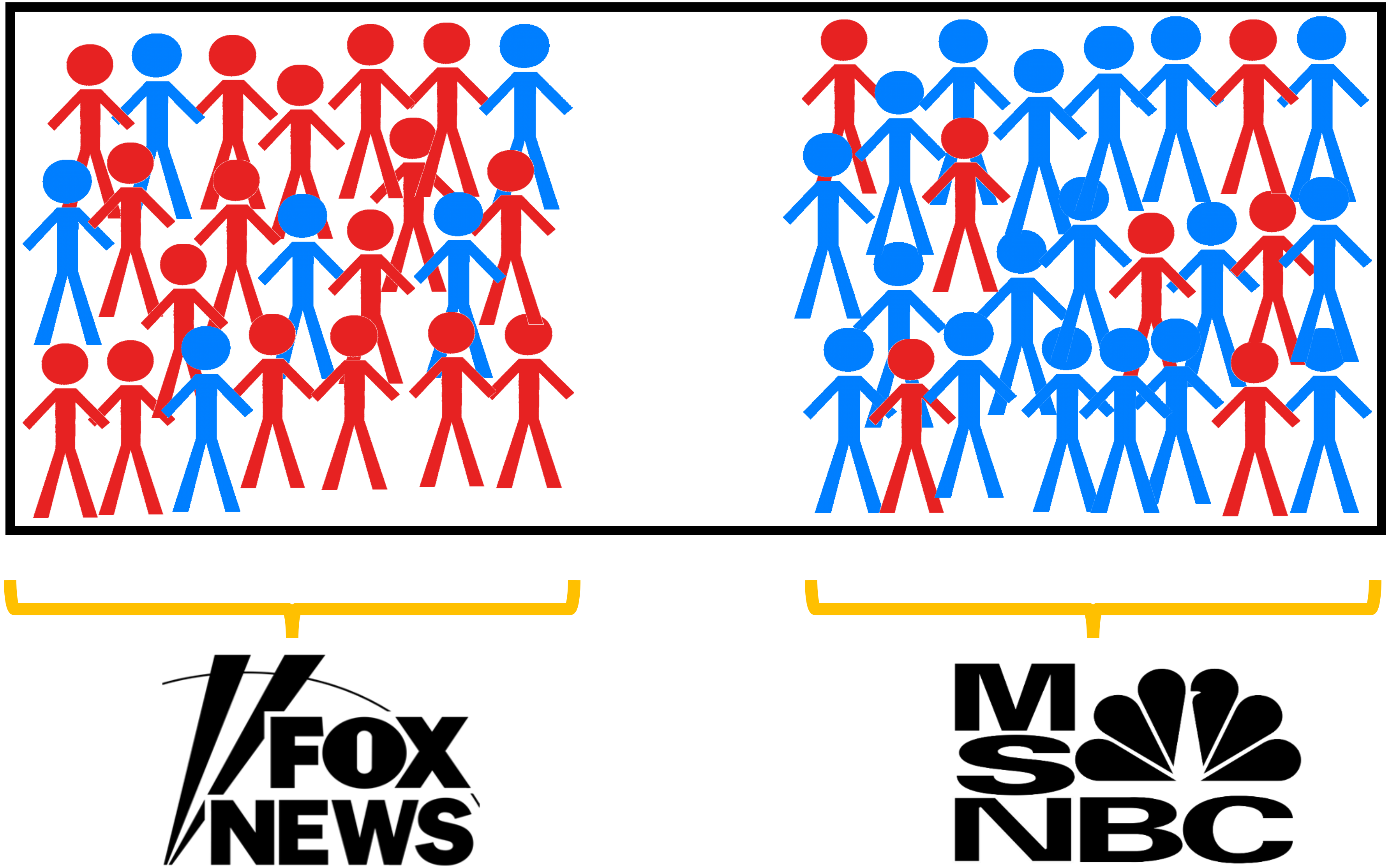
- This paper proposes a new-to-finance mechanism that causes persistent disagreement in other settings (politics, religion, etc.): **Echo Chambers**

Echo Chambers in Politics

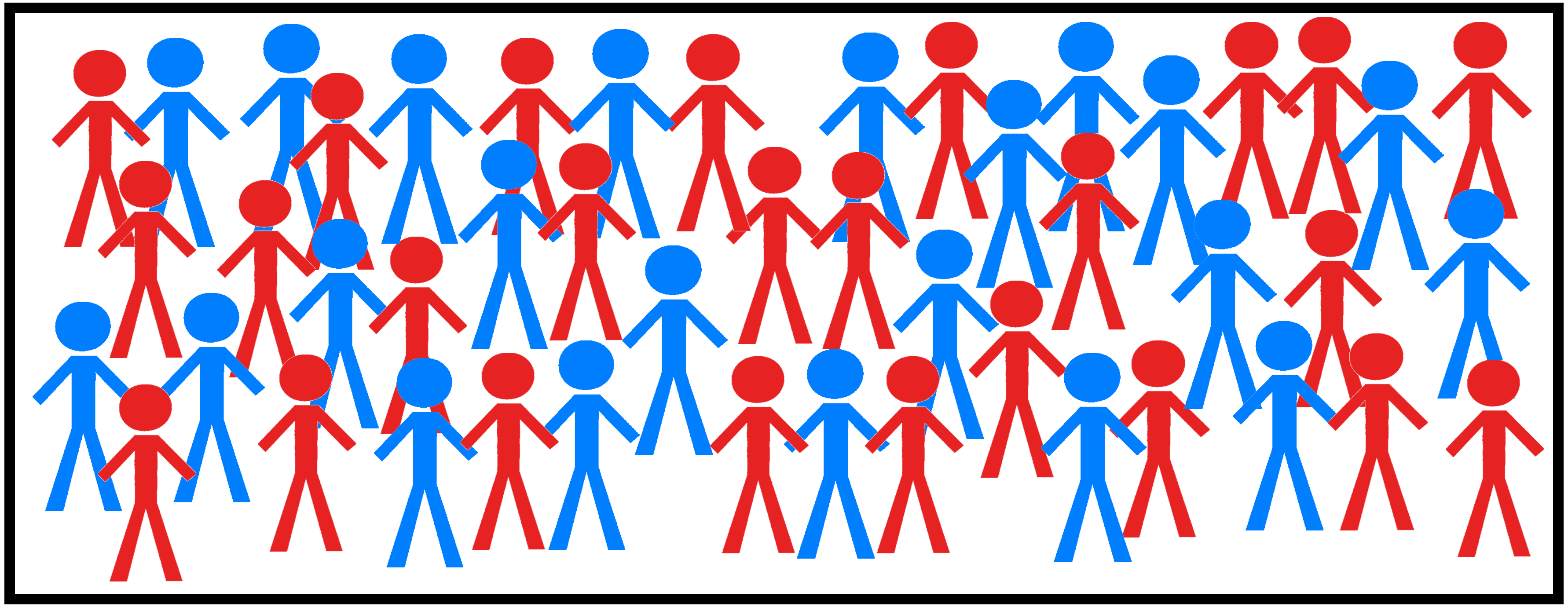


- Imagine a population of Republicans (red) and Democrats (blue)
- Do they choose to see the same information?

Echo Chambers in Politics

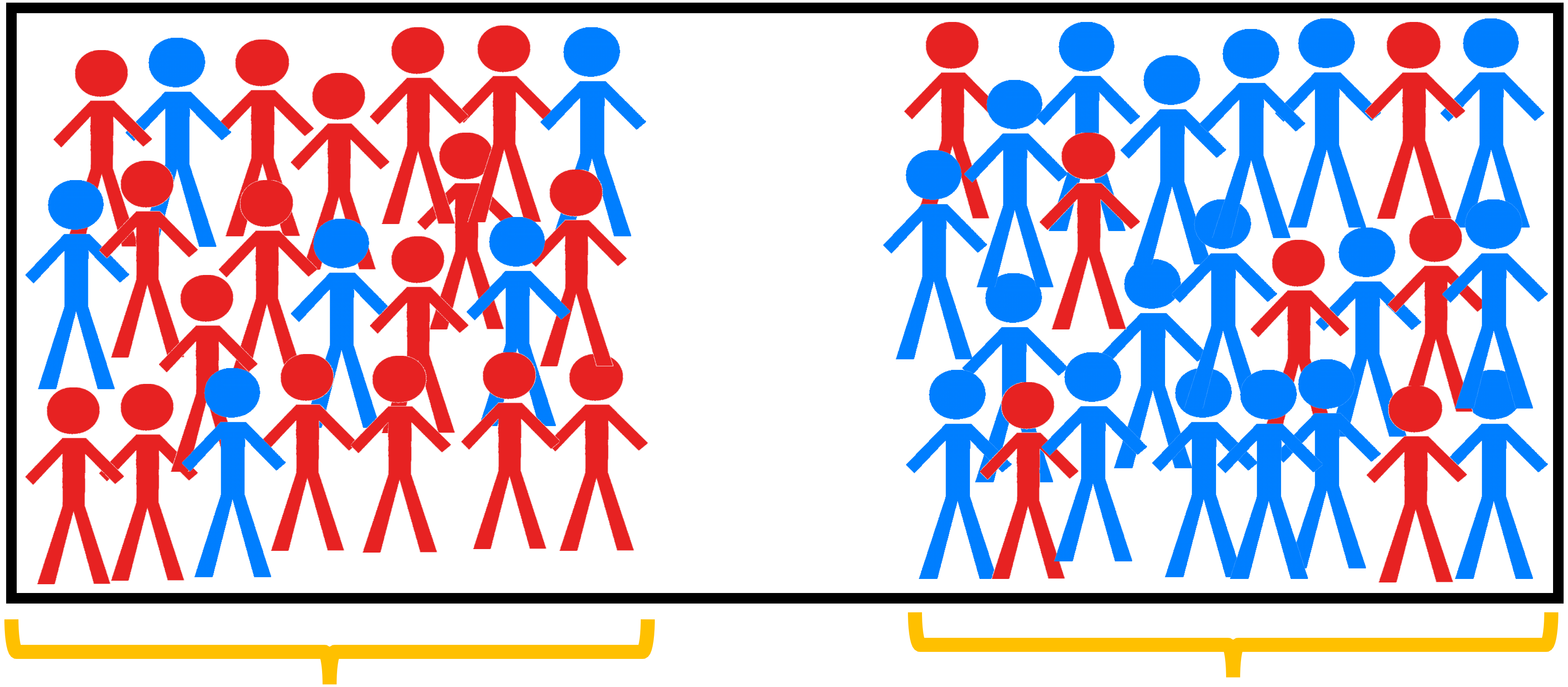


Echo Chambers among Investors



- Now imagine a population of Tesla **Bulls** (red) and Tesla **Bears** (blue)

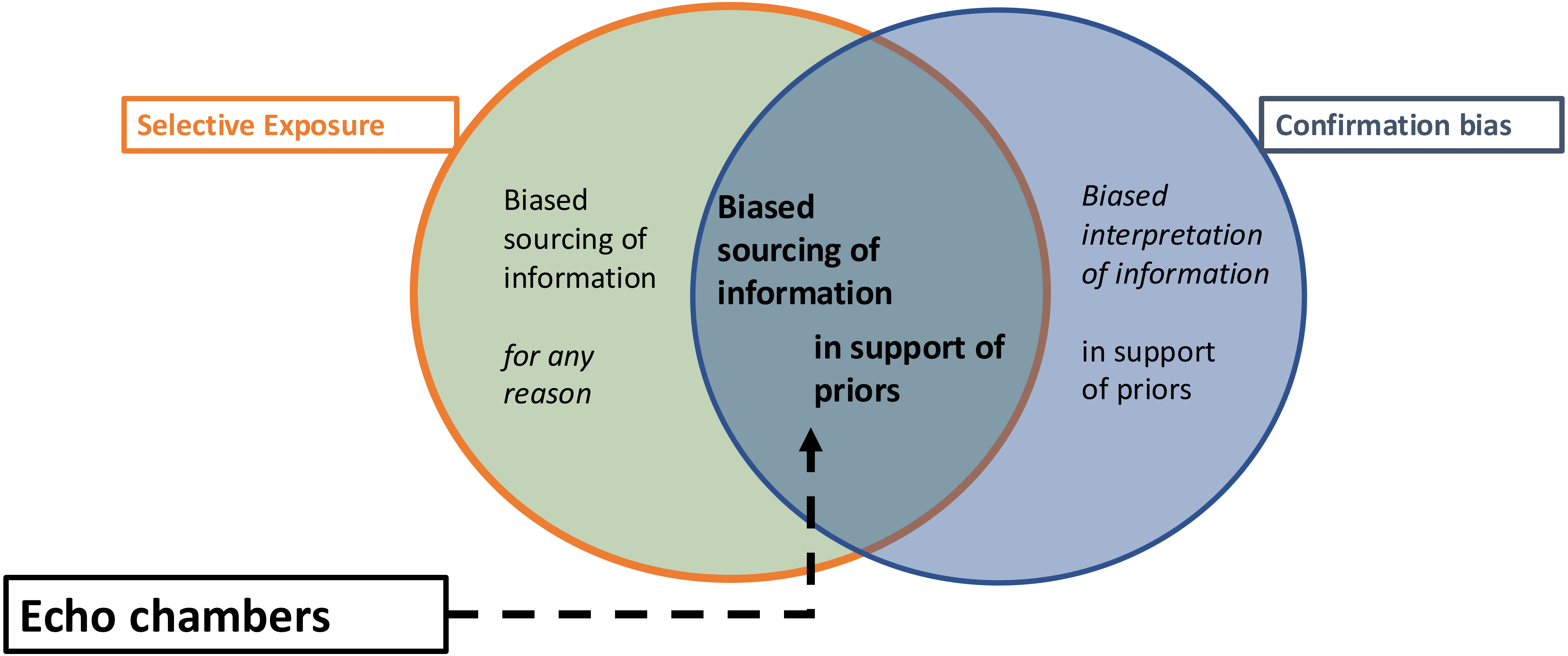
Echo Chambers among Investors (*sentiment*-based)



Bullish Information

Bearish Information

What is an Echo Chamber?



The Null Hypothesis

Is it surprising to see Echo Chambers in financial markets?

- Yes!

- In markets, forming correct beliefs about prices is valuable
  - Should want to consume value-relevant information *irrespective of prior beliefs*

- Not true for other settings in which echo chambers have been found
  - No *immediate* financial incentive to be correct in politics, beliefs about science, religion

What we do

- **StockTwits**, **but some different features versus** **Cookson and Niessner (2020)**
- We use the StockTwits universe of ~400,000 users that post >33 million self-labeled "***bearish***" and "***bullish***" posts
- Key: We examine *who* each user chooses to "**follow**"
  - *A follow is a decision to place a user's future posts into your newsfeed.*
  - ⇒ we can see if user ***chooses to build an echo chamber***, thus distorting the info they will receive in future via their newsfeed

What we find

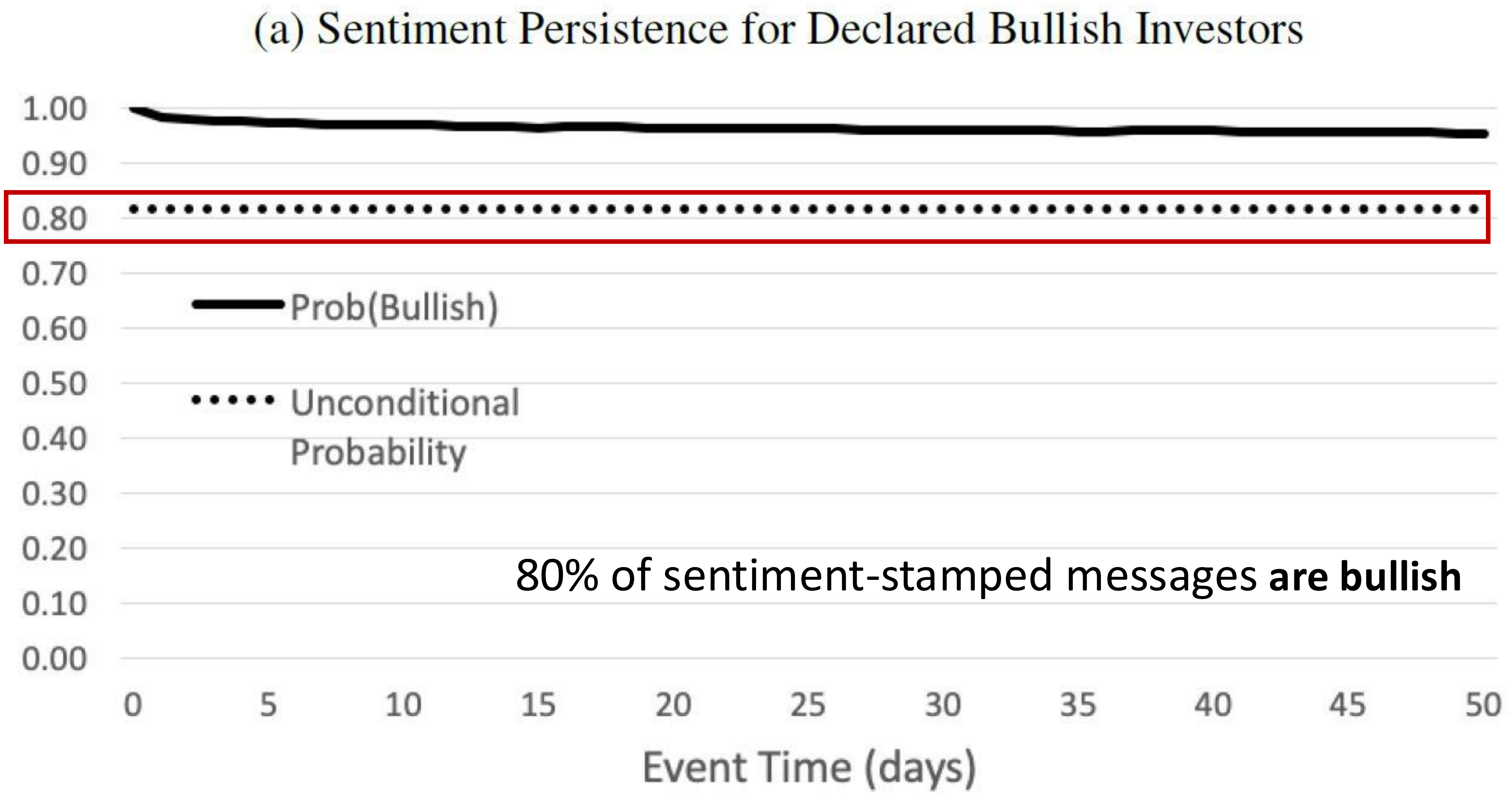
- Strong sentiment echo-chambers: Self-described Bulls are **5 times** more likely to *follow* other users with a bullish view of the same stock than Bears
    - True for **professional investors**
    - Twice as strong for those who **actually trade**
    - Stronger on **news days**
    - Counter-attitudinal returns reduce echo-chamber behavior

- Effect on newsfeed: Bulls see **62 more bullish messages** and **24 fewer bearish messages** over the next 50 days than Bears

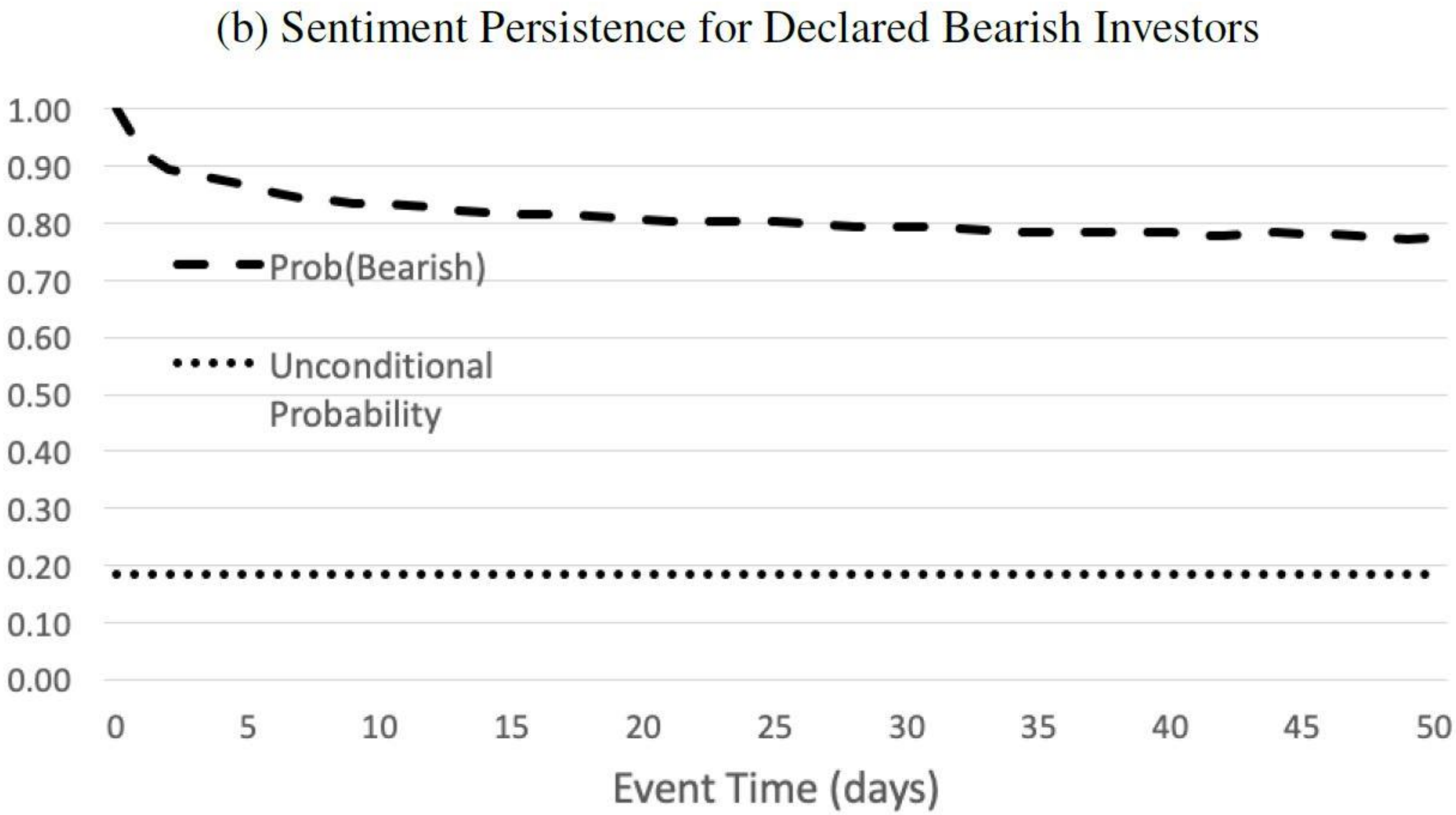
- Stock picks made in Echo Chambers → **worse returns**

- Echo Chambers are strongly associated with **more trading volume**

Sentiment is persistent within-user: Bulls



Sentiment is persistent within-user: Bears



Echo Chamber Main Result: Choice of who to **Follow**

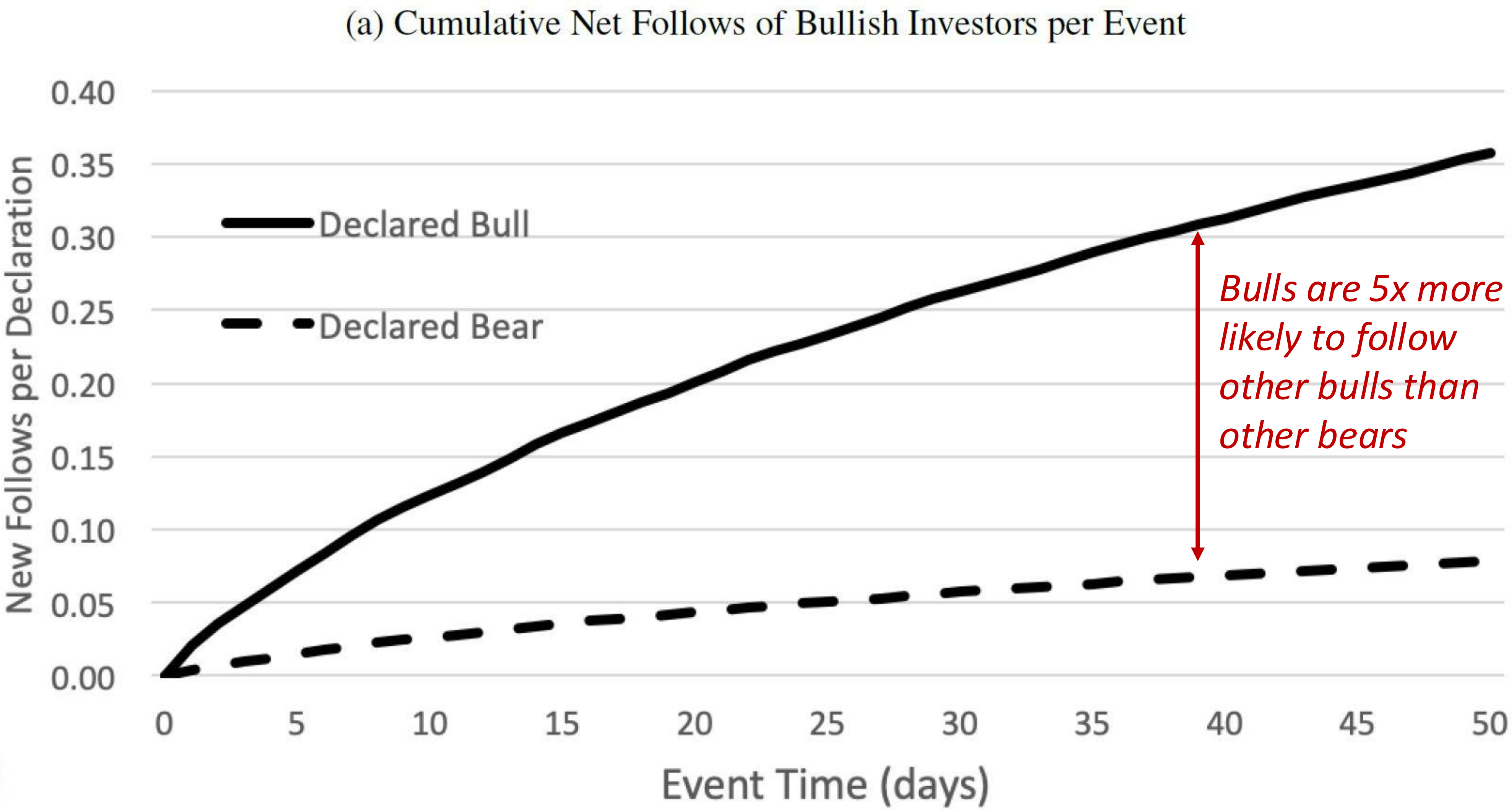
- To follow another user is to sign up to their information feed (future posts)
  - From StockTwits, we have a timestamp for when each UserID chooses to follow another

- In the following figures:
  1. We identify users **declaring as bulls (bears)** about a stock on day *t=0* (the event)
  2. We **count (cumulative) # of follows of bulls (bears)** in event time (up to 50 days)

- A followed user is a bear (bull) on date t if the user "stamps" message as a bullish (bearish) about stock *s* on date *t*

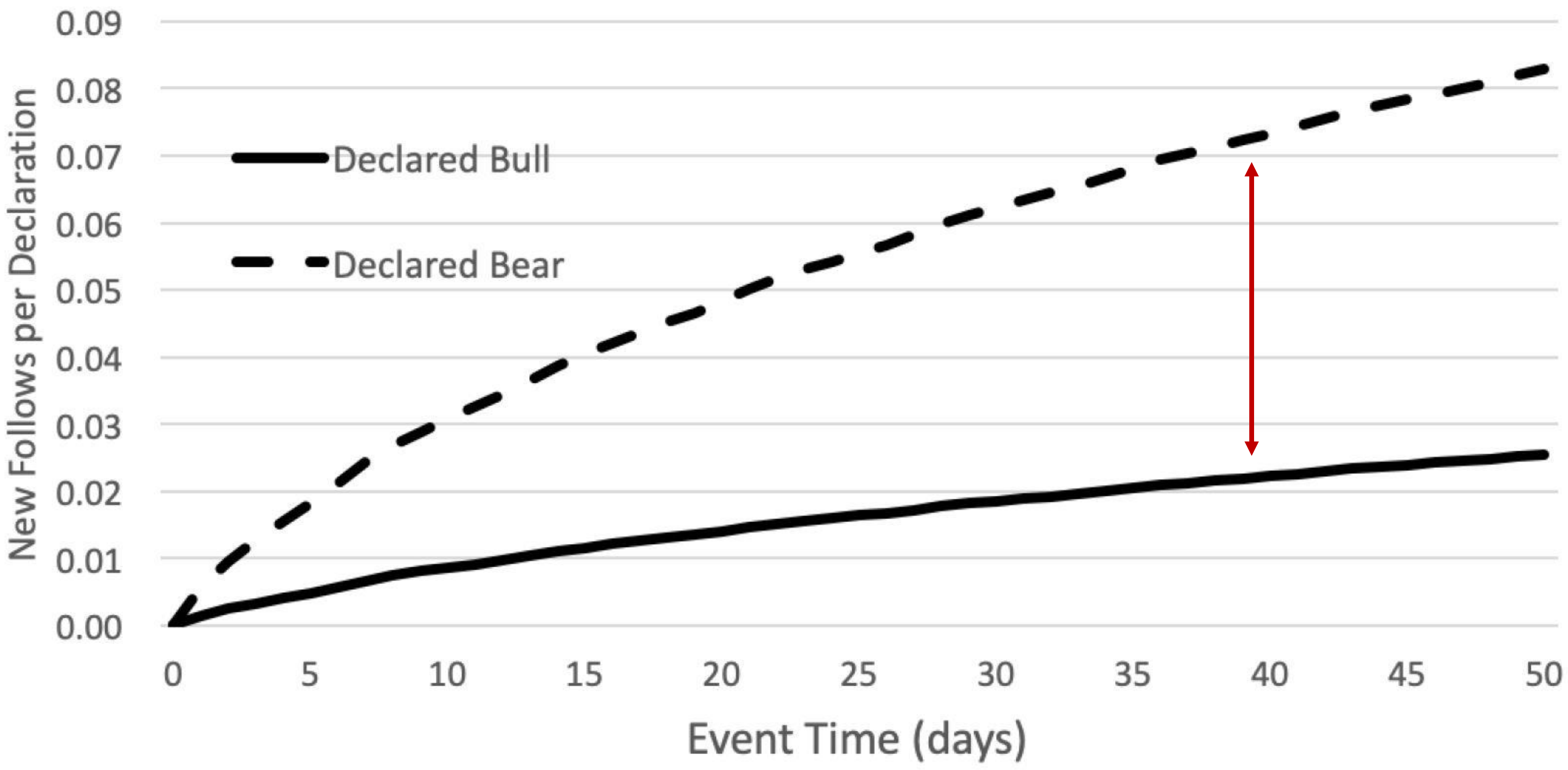
- This setting is akin to:
  - *seeing someone declare themselves a Republican or Democrat today and then*
  - *seeing if they set their DVR to record Hannity or Maddow tomorrow*

Selective Exposure: choosing to follow Bulls



Selective Exposure: choosing to follow Bears

(b) Cumulative Net Follows of Bearish Investors per Event



Regression Analogue for **Following** decision

	Dep. var.: $\mathbb{1}$ x100 if new follows$_{i,s,t+x}$ are more Bull than Bear				
	(1)	(2)	(3)	(4)	(5)
		+ User FE		+ User FE	with Bull bins
	t+1	t+1	t+1 $\rightarrow$ t+5	t+1 $\rightarrow$ t+5	t+1 $\rightarrow$ t+5
$\mathbb{1}$Declare Bull$_{i,s,t}$	0.88***	0.68***	2.41***	1.78***	
	[0.02]	[0.01]	[0.04]	[0.03]	
$\mathbb{1}$ if net # Bull declarations$_{i,s,t} \leq$ -1					-0.81***
					[0.03]
$\mathbb{1}$ if net # Bull declarations$_{i,s,t}$ = 2					1.30***
					[0.02]
$\mathbb{1}$ if net # Bull declarations$_{i,s,t}$ = 3					2.28***
					[0.04]
$\mathbb{1}$ if net # Bull declarations$_{i,s,t} \geq$ 4					4.92***
					[0.06]
# observations	13,893,332	13,893,332	13,893,332	13,893,332	13,893,332
# clusters (users)	305,967	305,967	305,967	305,967	305,967
R^2	0.05	0.12	0.09	0.19	0.20
Mean of dependent var.(%)	1.64	1.64	4.54	4.54	4.54
Effect size (% of mean)	54	41	53	39	
User FE	-	Y	-	Y	Y
User x symbol FE	-	-	-	-	-
Day x symbol FE	Y	Y	Y	Y	Y

Selective exposure gets stronger with more bullish declarations on a day.

Regression Analogue for **Following** decision

	Dep. var.: $\mathbb{1}$ x100 if new follows$_{i,s,t+x}$ are more Bull than Bear									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		+ User FE		+ User FE	with Bull bins	+ User-Symbol FE	Conditional on new follows	Conditional on new follows	Conditional on new follows	Conditional on new follows
	t+1	t+1	t+1 $\rightarrow$ t+5	t+1 $\rightarrow$ t+5	t+1 $\rightarrow$ t+5	t+1 $\rightarrow$ t+5	t+1 $\rightarrow$ t+5	t+2 $\rightarrow$ t+10	t+11 $\rightarrow$ t+30	t+31 $\rightarrow$ t+50
$\mathbb{1}$Declare Bull$_{i,s,t}$	0.88***	0.68***	2.41***	1.78***		1.08***	12.21***	7.70***	3.34***	0.27
	[0.02]	[0.01]	[0.04]	[0.03]		[0.04]	[0.57]	[0.47]	[0.38]	[0.31]
$\mathbb{1}$ if net # Bull declarations$_{i,s,t} \leq$ -1					-0.81***					
					[0.03]					
$\mathbb{1}$ if net # Bull declarations$_{i,s,t}$ = 2					1.30***					
					[0.02]					
$\mathbb{1}$ if net # Bull declarations$_{i,s,t}$ = 3					2.28***					
					[0.04]					
$\mathbb{1}$ if net # Bull declarations$_{i,s,t} \geq$ 4					4.92***					
					[0.06]					
# observations	13,893,332	13,893,332	13,893,332	13,893,332	13,893,332	12,262,524	596,518	919,306	1,302,544	1,308,130
# clusters (users)	305,967	305,967	305,967	305,967	305,967	259,476	63,594	77,761	84,419	82,568
R^2	0.05	0.12	0.09	0.19	0.20	0.34	0.72	0.78	0.83	0.87
Mean of dependent var.(%)	1.64	1.64	4.54	4.54	4.54	4.92	77.03	70.13	60.39	40.85
Effect size (% of mean)	54	41	53	39		22	16	11	6	1
User FE	-	Y	-	Y	Y	-	-	-	-	-
User x symbol FE	-	-	-	-	-	Y	Y	Y	Y	Y
Day x symbol FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

The magnitude dissipates by 50 days after declaration.
Unlike content of newsfeed

Echo Chamber Heterogeneity

1. Investment sophistication

   Professionals also place themselves in echo chambers (but less so)

2. Skin-in-the game

   Buyers/sellers are *more* likely in echo chambers.

3. Information arrival (earnings announcements)

   Echo chambers are **stronger** on EA days.

4. Experienced returns

   Bad returns take **bulls** out of echo chambers (and vice versa).

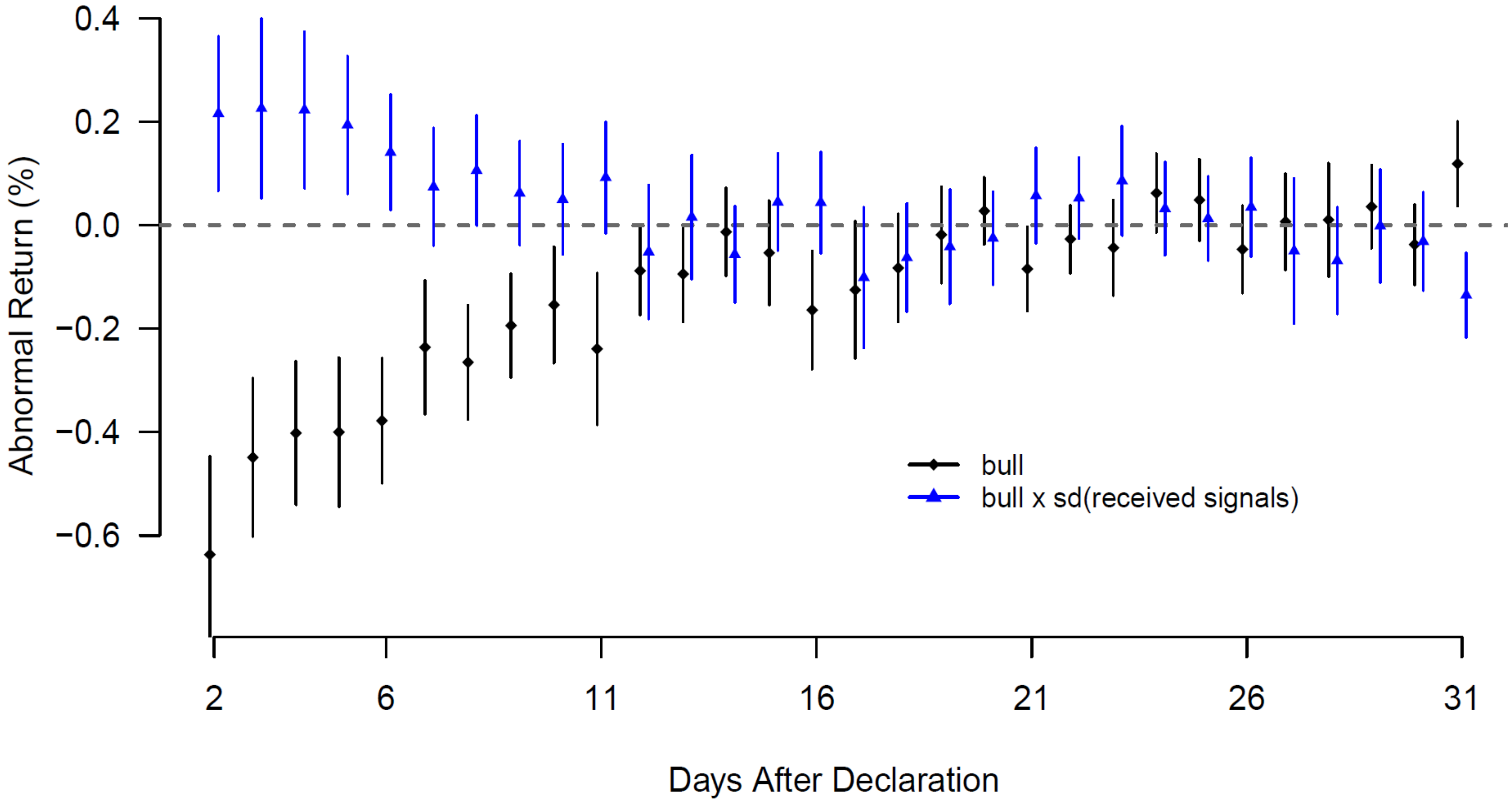
Echo Chambers and Returns: worse stock picks?

Do users in echo chambers make **worse stock picks**?

- We follow stock "pick" performance (buy bullish, sell bearish) for 30 days for users inside vs outside echo chambers.
- Echo chambers == no "signal diversity" (low or no difference of opinion in newsfeed)
- **Idea**: Users who see all the same signals (i.e., no signal diversity) are in an echo chamber

Bad performance, especially in echo chamber

Black diamonds == complete echo chamber.



What is the effect of siloed information on trading?

- Key implication: Echo chambers ➔ Information silos

- In information silos, people's newsfeeds become:
  - More different from one another (↑ ***received disagreement***)
  - More internally consistent (↓ ***received uncertainty***)

- We saw received uncertainty ➔ underperformance
  - Does it also affect trading on US stock markets?
  - Yes! About the same magnitude as disagreement itself!

Note: As with all the results, the paper makes this point with regressions that control for confounding explanations.

Analysis now at the stock-day level

Relation to Stock Turnover (on US markets)

Both siloing measures show ***more trading when information is more siloed***

Magnitude similar to disagreement itself

	Dep. var.: Abnormal Log Turnover$_{s,t}$				
	(1)	(2)	(3)	(4)	(5)
Sender Disagreement$_{s,t}$	0.017***	0.015***	0.027***	0.024***	0.025***
	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]
Received Disagreement$_{s,t}$		0.004**		$\boxed{0.008^{***}}$	$\boxed{0.011^{***}}$
		[0.002]		$\boxed{[0.002]}$	$\boxed{[0.002]}$
Received Uncertainty$_{s,t}$			-0.014***	$\boxed{-0.016^{***}}$	$\boxed{-0.016^{***}}$
			[0.002]	$\boxed{[0.001]}$	$\boxed{[0.001]}$
Abnormal Log Turnover$_{s,t-1}$	0.188***	0.188***	0.189***	0.188***	0.205***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.005]
$\mathbb{1}$Media Article$_{s,t}$	0.128***	0.128***	0.128***	0.128***	
	[0.006]	[0.006]	[0.006]	[0.006]	
Log GoogleASVI$_{s,t}$	0.347***	0.346***	0.347***	0.346***	
	[0.015]	[0.015]	[0.015]	[0.015]	
Volatility$_{s,(t-5 \text{ to } t-1)}$	0.117***	0.116***	0.116***	0.115***	
	[0.035]	[0.035]	[0.035]	[0.035]	
Cum. Abnormal Returns$_{s,(t-5 \text{ to } t-1)}$	0.010	0.010	0.010	0.010	
	[0.012]	[0.012]	[0.012]	[0.012]	
Cum. Abnormal Returns$_{s,(t-30 \text{ to } t-6)}$	-0.051***	-0.051***	-0.051***	-0.051***	
	[0.009]	[0.009]	[0.009]	[0.009]	
# obs.	421,915	421,915	421,915	421,915	421,915
# clusters (stock)	1,075	1,075	1,075	1,075	1,075
# clusters (day)	1,886	1,886	1,886	1,886	1,886
R^2	0.83	0.83	0.83	0.83	0.82
Mean of dependent var.	0.43	0.43	0.43	0.43	0.43
Day FE	Y	Y	Y	Y	Y
Month x stock FE	Y	Y	Y	Y	Y
Message number FE	Y	Y	Y	Y	Y

Conclusion

- High trading volume is a puzzle for Finance; main explanation is **disagreement**

- But what sustains disagreement?
  - This paper proposes and finds evidence for a new mechanism that generates sustained disagreement and trading: echo chambers

Social media signals

Given the role of social transmission bias and "social finance" movement in general, there is increasing interest in signals drawn from social media:

- **Biases** (echo chambers, misinformation) versus **Informative signals** (see the Chen et al "wisdom of crowds" paper from Day 1).
- How does this reflect changes in the information environment? Does social media induce changes in the information environment?

Journal of Financial Economics

Volume 158, August 2024, 103870

The social signal ☆

J. Anthony Cookson [a], Runjing Lu [b], William Mullins [c], Marina Niessner [d]

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.jfineco.2024.103870

Motivation: Investor social media

Literature uses each platform as "**social media**"

- ❖Different user populations
- ❖Different contributor incentives & formats
- ❖Different timing

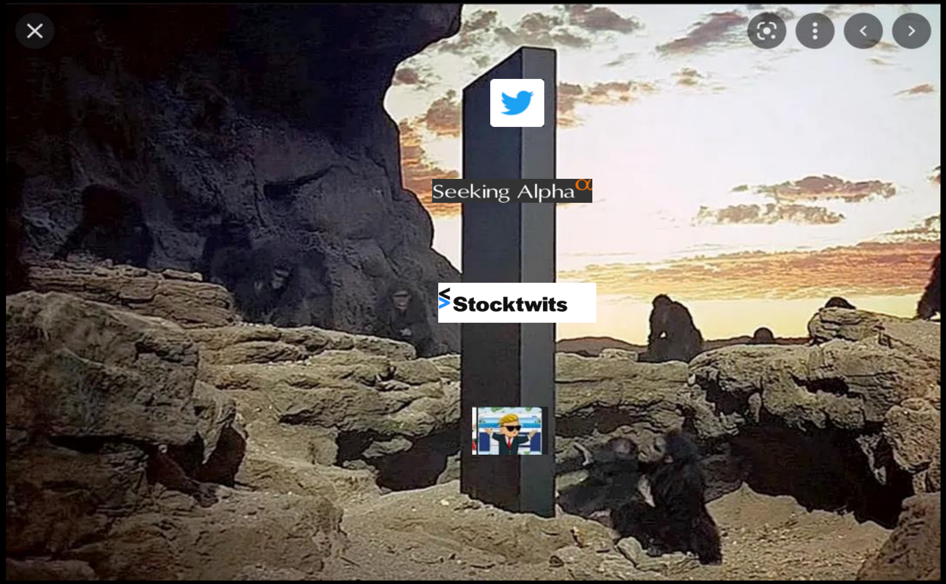
But how different is the information across platforms?

- ❖Important for connecting research on StockTwits, Twitter & Seeking Alpha

Two views of investor social media

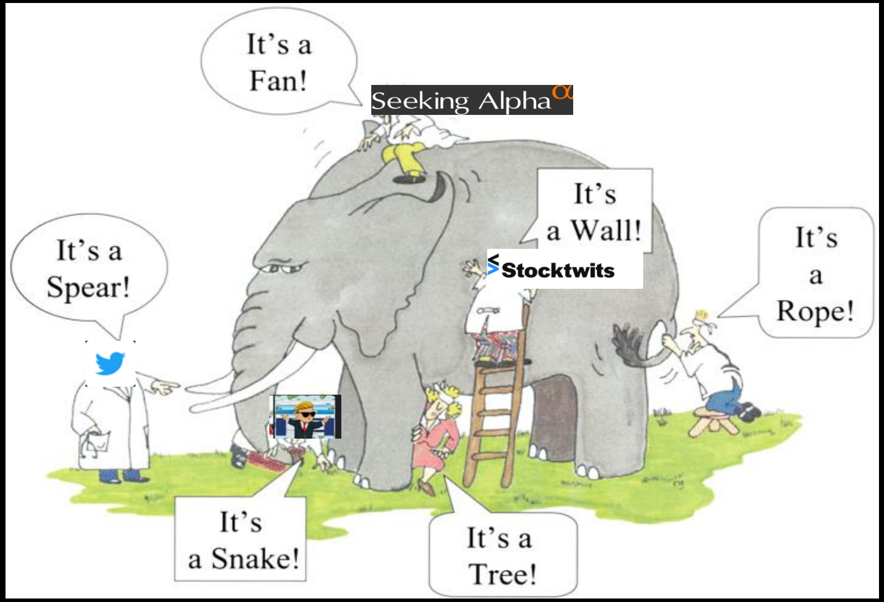
Monolith view

Platforms are **substitutes**



Idiosyncratic view

Platforms convey different info: **complements**



Example posts

Stocktwits

TheCoolGuy Bullish 01:34 AM

$AAPL iPhone 14 pro model demand is so strong that even with 10% capacity increase the lead time still remains at 3-4 weeks to get a phone, showing very strong demand in the face of a budget constrained consumer. Apple literally can't make them fast enough and may lose on revenue if they can't close the supply and demand gap.

1

Seeking Alpha

The World Is Ending - Somebody Tell Apple Stock

Sep. 26, 2022 11:49 AM ET | **Apple Inc. (AAPL)** | 107 Comments | 16 Likes

Summary

- As you know, we are all doomed. Capitalism is ending, the Fed has ruined everything by being first too soft and now too tough.
- It's all going to zero.
- There's just one problem. The largest constituent of the S&P 500 and the Nasdaq 100 is only 17% below its all-time highs.
- So who is wrong - the doomsayers, or Apple shareholders?
- We investigate below.
- This idea was discussed in more depth with members of my private investing community, Growth Investor Pro. Learn More »

App Economy Insights @EconomyApp · Sep 24

DCF models are broken and flawed.

Why?

Imagine building a DCF for $AAPL in 2006.

Analysts assumed a linear deceleration of iPod sales.

❌ No iPhone.
❌ No Watch.
❌ No AirPods.
❌ No App Store.

You can't put innovation in a spreadsheet.

Findings

- We distinguish *attention* from *sentiment*
  - Attention is ***monolithic*** across platforms, sentiment is ***idiosyncratic***
    - *Not driven by traditional news, holds across users on the same platform*

Findings

- We distinguish *attention* from *sentiment*
  - Attention is ***monolithic*** *across platforms*, sentiment is ***idiosyncratic***
- Events show how **platform features** and different user bases affect signal:
  - **StockTwits character limit increase** → improved informativeness of StockTwits, but not Twitter or SeekingAlpha
  - **Jan 2021 GME phenomenon** → reduced informativeness across platforms
    - Driven by new users

Findings

- We distinguish *attention* from *sentiment*
  - Attention is ***monolithic*** *across platforms*, sentiment is ***idiosyncratic***
- Events show how platform features and different user bases affect signal:
  - **StockTwits character limit increase** → improved informativeness of StockTwits, but not Twitter or SeekingAlpha
  - **Jan 2021 GME phenomenon** → reduced informativeness across platforms
    - Driven by new users
- Both sentiment and attention relate strongly & positively to retail trading
- Different return implications:
  - Sentiment predicts positive & quick return to baseline
  - Attention predicts negative & persistent decline
  - Both conditional on each other

What ***is*** the social signal?

Data and Attention measures

Data:

- Stocktwits
  - Split by groups (top 1%, professionals, novices, intermediate, self-labeled posts)
- Twitter (*from a company called Social Market Analytics*)
- Seeking Alpha (*from Ravenpack 1.0*)

Attention measure:

- Start with N of posted messages
- Within platform-day we compute share of messages about each ticker
  - E.g. 100 messages in total on platform, 30 about TSLA → TSLA share = 30%
  - Removes time variation in N of messages

Sentiment measure:

- Each platform provides either message-level sentiment (StockTwits, SeekingAlpha) or firm-day sentiment (Twitter)
- Aggregate to firm-day(-group)
  - Average sentiment by firm-day is *Sentiment*

Sentiment and attention are distinct

Omnibus Principal Components Analysis

Put six signals (attention, sentiment) x (StockTwits, Twitter, SA) into a PCA.

First two principal components are (mostly):

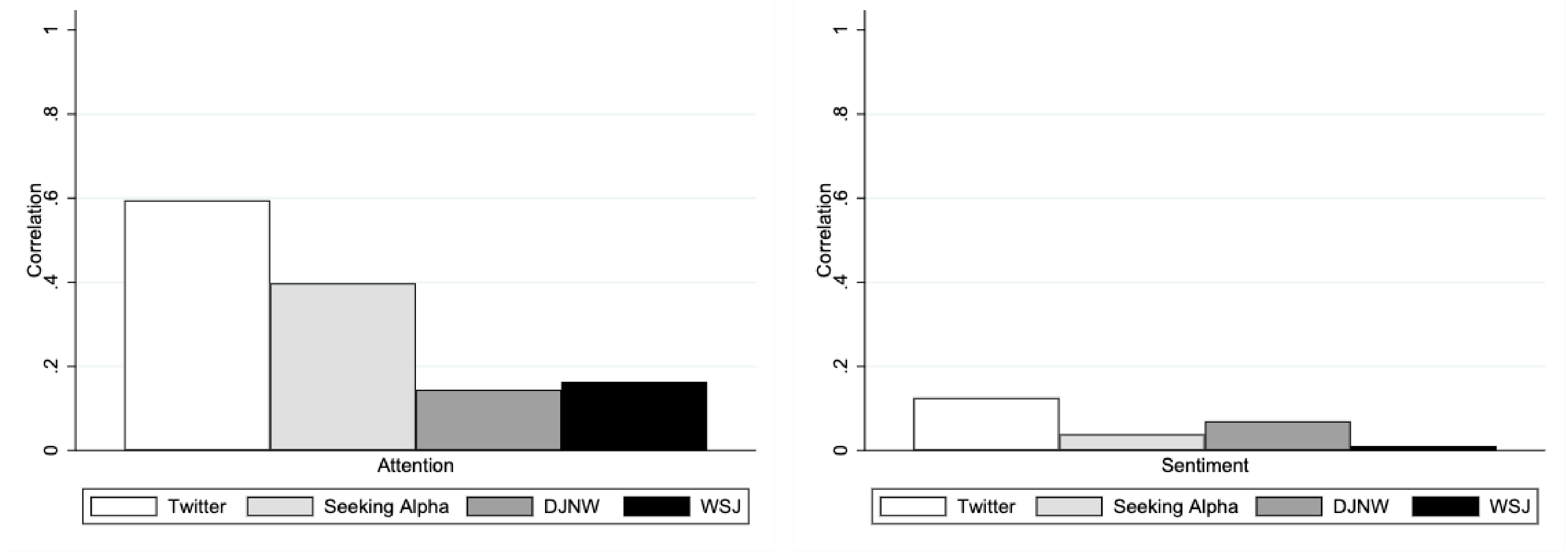
- Attention (35.6% of the overall variation)
- Sentiment (19.3% of the variation)

Motivates our approach, which treats attention and sentiment as separate analyses.

Note: Attention explains ~2x the sentiment variation.

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6
Sentiment:						
StockTwits	-0.031	0.644	-0.345	0.682	0.017	-0.014
Twitter	0.082	0.647	-0.225	-0.720	-0.008	0.071
Seeking Alpha	0.160	0.384	0.885	0.087	0.190	0.008
Attention:						
StockTwits	0.548	-0.130	-0.188	0.047	0.638	0.488
Twitter	0.605	-0.033	-0.098	-0.009	0.048	-0.788
Seeking Alpha	0.548	-0.007	0.052	0.084	-0.745	0.368
Fraction of variation	35.6%	19.3%	15.9%	14.5%	9.2%	5.5%

Attention is highly correlated across platforms
Sentiment is not



(a) Attention **(b)** Sentiment

Could be due to:

- News
- Sentiment classification (ML)
- Platform features
- Different users

Attention and sentiment

Principal Component Analyses

1st PC of attention explains 70% of attention variation

→ Lots of common information in attention

1st PC of sentiment explains *only* 39% of sentiment variation

Null if all three were orthogonal: 1/3 of variation in each PC

Panel B: PCA of Platform-Level Attention Signals

	Comp1	Comp2	Comp3
StockTwits	0.565	-0.665	0.489
Twitter	0.614	-0.057	-0.787
Seeking Alpha	0.551	0.745	0.376
Fraction of variation	$\boxed{70.0\%}$	18.9%	11.1%

Panel C: PCA of Platform-Level Sentiment Signals

	Comp1	Comp2	Comp3
StockTwits	0.611	-0.464	0.642
Twitter	0.662	-0.147	-0.735
Seeking Alpha	0.435	0.874	0.217
Fraction of variation	$\boxed{38.8\%}$	32.3%	29.0%

Robustness to controlling for news and firm FE
Conditional PCAs

PCA is robust to first residualizing by news & public events
(EAs and 8-Ks, lagged up to 7 days)

Similarly, PCA results are not due to correlation with time-invariant firm characteristics
(Panel C residualizes by firm FEs before PCA).

Table 1.1: How Common is the Social Signal across Platforms? Residual Signals Controlling for News and Firm FEs

Panel A: Residualized Attention Signals

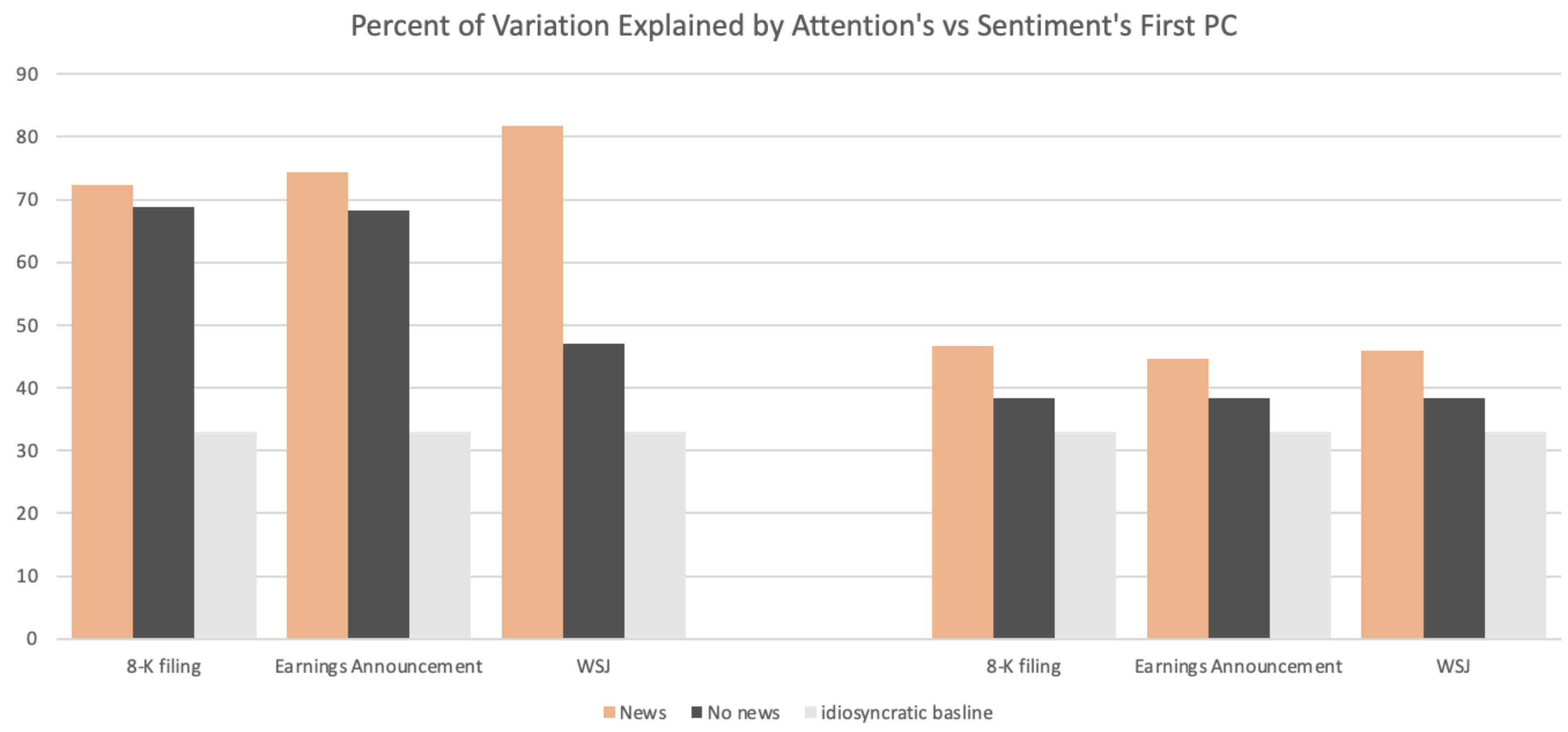
	Comp1	Comp2	Comp3
StockTwits	0.606	-0.428	0.671
Twitter	0.626	-0.264	-0.734
Seeking Alpha	0.491	0.864	0.108
Fraction of variation	$\boxed{63.9\%}$	23.8%	12.3%

Panel B: Residualized Sentiment Signals

	Comp1	Comp2	Comp3
StockTwits	0.660	-0.288	0.694
Twitter	0.675	-0.178	-0.716
Seeking Alpha	0.330	0.941	0.077
Fraction of variation	$\boxed{38.2\%}$	32.8%	29.0%

Heterogeneity by news

PCAs on news versus non-news days



Sentiment and Attention across user groups

Same pattern within StockTwits

Panel A: Correlations with the StockTwits Signal

	Top 1%	Professional	Intermediate	Novice	No experience	Self-labeled
StockTwits attention	0.819	0.884	0.966	0.929	0.987	0.931
StockTwits sentiment	0.232	0.313	0.348	0.222	0.783	0.387

Attention is highly correlated *within platform*

Sentiment is not

(see similar results in the PCA)

*This comparison holds constant

- sentiment classification algorithm → not driving low sentiment correlation
- platform features

Information Events

Information events

Regression of next day returns on social signals

- StockTwits increased # of characters from 140 to 1,000 (May 8, 2019)

- Goal to isolate a platform feature in driving informational differences.

Stocktwits

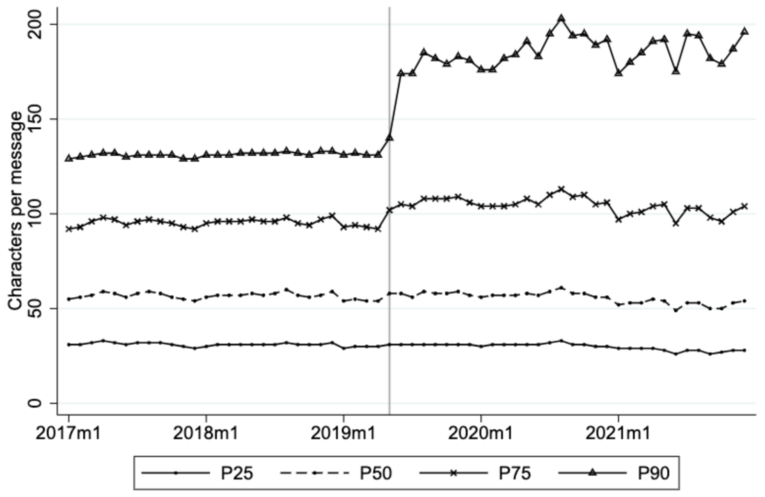
StockTwits Character Limit Increase

Platform-by-platform, run regressions of CAR and RT imbalance on single platform social signal

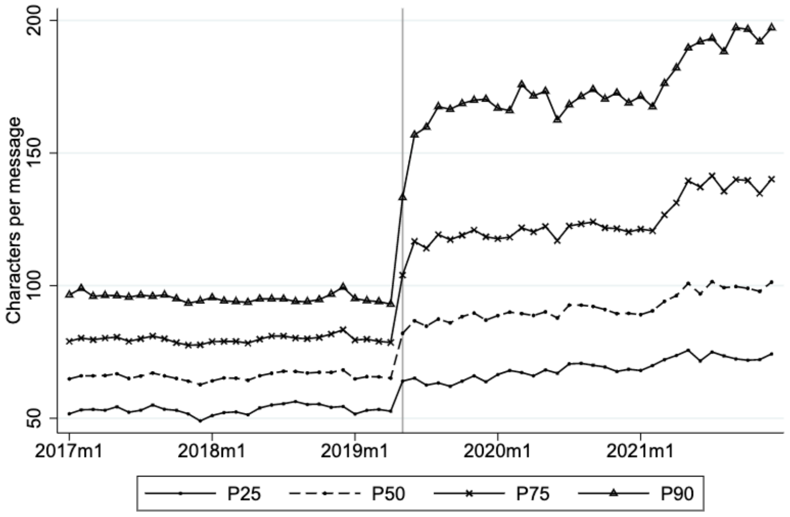
Prediction: The character limit increase should only affect the informativeness of StockTwits

Separately, build longStockTwits signals and shortStockTwits signals

Prediction: Long messages are the most affected



(a) Number of characters per message



(b) Firm-day level average number of characters per message

StockTwits sentiment became more informative for next-day returns, particularly long messages

	Dep. variable: CAR t+1			
	(1) StockTwits	(2) StockTwits top quartile	(3) Twitter	(4) Seeking Alpha
Post × Sentiment (z)	0.066*	0.137**	-0.010	-0.004
	(0.034)	(0.055)	(0.043)	(0.034)
Post × Attention (z)	0.153*	-0.325	-0.007	-0.016
	(0.085)	(0.222)	(0.027)	(0.032)
Sentiment (z)	0.030	0.002	-0.003	0.079***
	(0.023)	(0.037)	(0.018)	(0.024)
Attention (z)	-0.355***	-0.249	-0.029	-0.016
	(0.110)	(0.185)	(0.027)	(0.023)
DJNW sentiment (z)	0.097***	0.092***	0.101***	0.086***
	(0.016)	(0.026)	(0.016)	(0.016)
DJNW attention (z)	0.026	-0.011	-0.003	-0.009
	(0.027)	(0.072)	(0.029)	(0.026)
Controls	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
Outcome Mean	-0.093	0.004	-0.093	-0.093
Outcome SD	7.823	6.451	7.823	7.823
Observations	214,260	53,447	214,260	214,260
R^2	0.027	0.066	0.026	0.026

Information events

Regression of next day returns on social signals

- GME euphoria changed the composition and attitudes of retail investors (Jan 28, 2021)

- Goal to understand impact of new users and ideas

The GME Event

Investors on StockTwits mention "short squeezes" much more frequently after Jan 2021

- Bradley et al. (2022) on Reddit "*DD reports*"

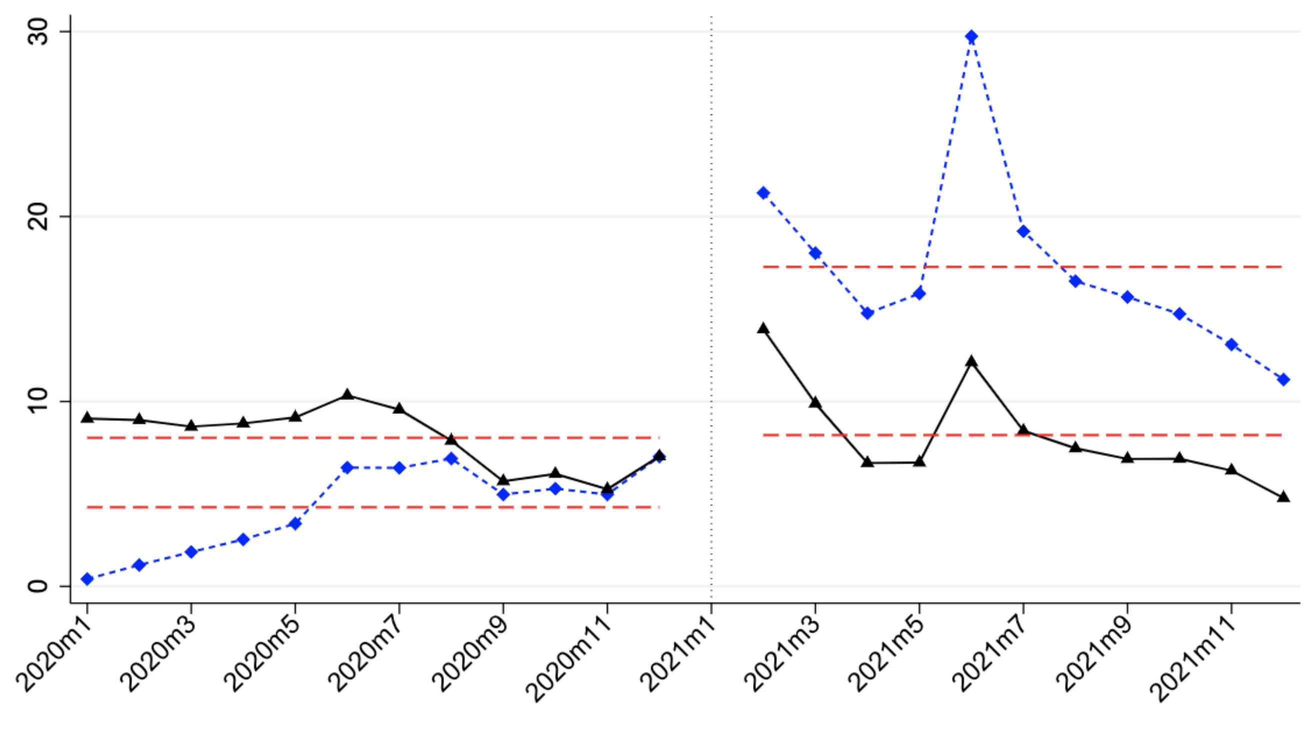


Figure 9: StockTwits Mentions of "Short Squeeze" Around the GameStop Event
Old versus New Users

Biggest increase from "new" users

Pre/post GME

PC Sentiment (cols 1 and 2)

Sentiment's (PC1) ability to predict CAR *falls to zero* after GME

- PC2 and PC3 provide no marginal value

New vs old (cols 3-4, StockTwits)

- No change in informativeness of existing users (joined before 2020)
- Effect concentrated among new users

No change in attention's impacts pre vs post

	Dep. variable: CAR t+1			
	(1) PC signal	(2) PC signal	(3) StockTwits old	(4) StockTwits new
Post × Sentiment (z)	-0.098**	-0.098**	0.001	-0.102**
	(0.045)	(0.045)	(0.034)	(0.043)
Post × Attention (z)	0.006	0.006	-0.017	0.016
	(0.097)	(0.097)	(0.091)	(0.109)
Sentiment (z)	0.095**	0.095**	0.035	0.095**
	(0.039)	(0.039)	(0.028)	(0.039)
Attention (z)	-0.077	-0.078	-0.070	-0.069
	(0.058)	(0.058)	(0.056)	(0.067)
Post × Sentiment PC2 (z)		-0.009		
		(0.031)		
Post × Sentiment PC3 (z)		0.022		
		(0.030)		
Sentiment PC2 (z)		0.011		
		(0.025)		
Sentiment PC3 (z)		0.016		
		(0.026)		
DJNW sentiment (z)	0.086***	0.082***	0.087***	0.088***
	(0.016)	(0.016)	(0.016)	(0.017)
DJNW attention (z)	-0.065**	-0.066**	-0.066**	-0.070**
	(0.031)	(0.032)	(0.031)	(0.031)
Controls	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
Outcome Mean	-0.004	-0.004	-0.004	-0.004
Outcome SD	7.867	7.867	7.867	7.867
Observations	287,833	287,833	287,833	287,833
R^2	0.049	0.049	0.049	0.049

How the social signal relates to

Retail trading & Returns

Does "social induced" retail trading imbalance predict $\boldsymbol{CAR_{t+1}}$?

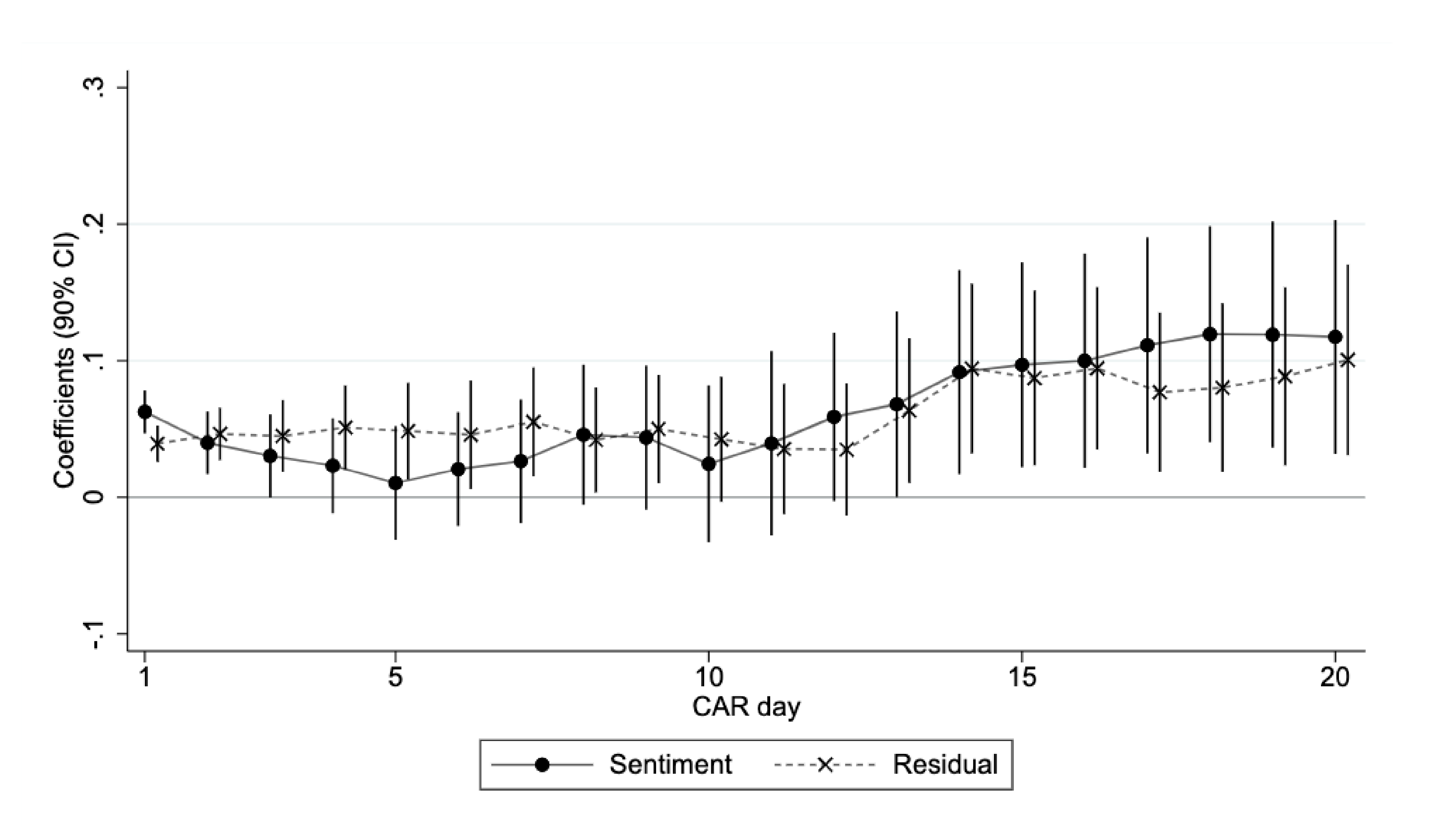
To answer this question:

1. Project retail trading imbalance (RT imbalance) onto social signal variables (all attention and sentiment variables from each platform)

2. Use fitted values and residuals as explanatory variables for next day CAR (and longer term CARs)

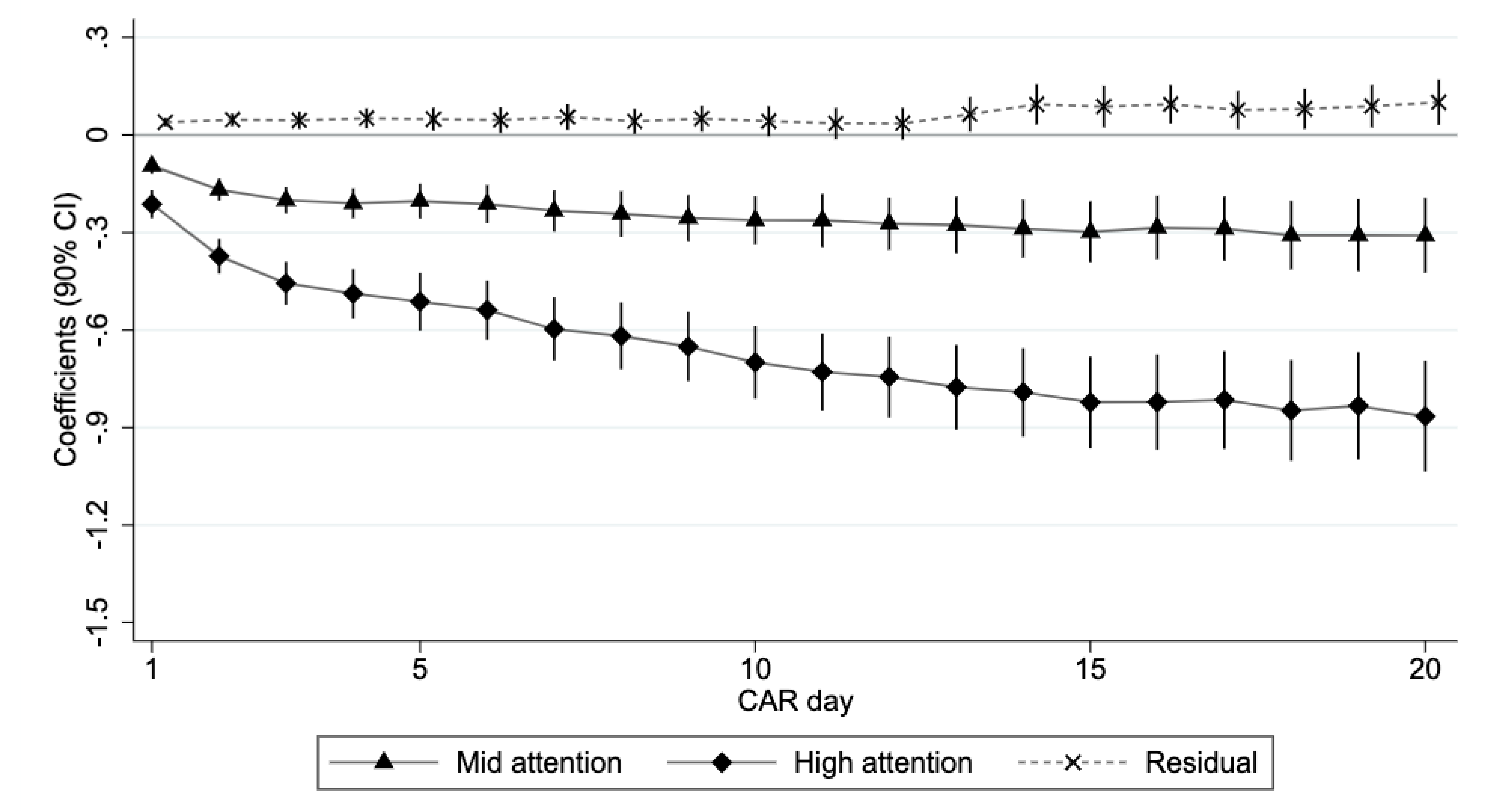
Sentiment-driven return plateaus quickly

… then returns to the residual baseline



Attention-driven return is negative

… does not return to baseline (~0.9 percentage pts by day 20)



Conclusion

Cross-platform analysis via PCAs shows

- Sentiment from the social signal is idiosyncratic
- Attention is more monolithic across platforms

- StockTwits & GME experiments:
  - Provide more-identified evidence for why investor social media has information
    - Shows platform *features* affect power of social signal → Cross-platform differences matter
    - Shows *user bases* of platform also affect information in social signal

Retail trading and CAR respond to attention and sentiment *in different ways*:

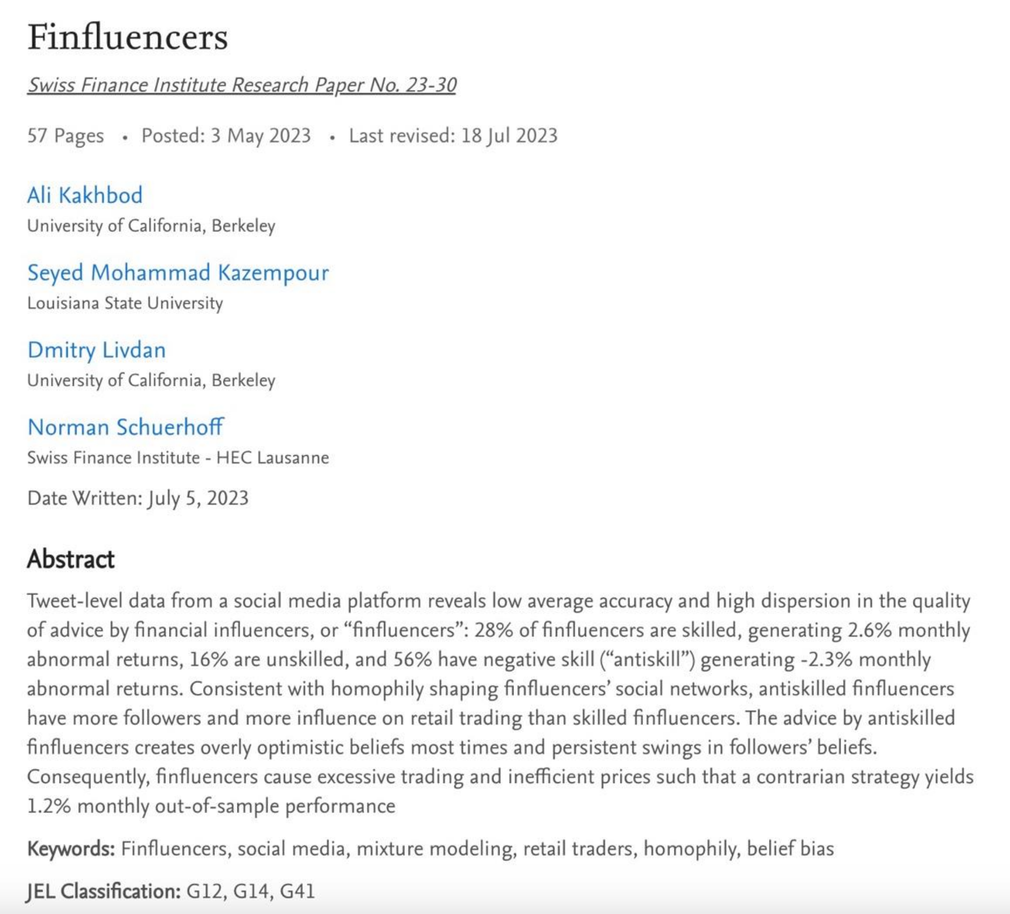
- Sentiment leads to trading and positive next-day returns that revert
- Attention leads to trading and negative & persistent future returns

Other work on social media signals

Heterogeneity in skill and influence

- Finfluencers – Kakhbod et al (2025WP)
- Dim (2025JFQA)
- Martineau et al (2025 WP)
- Wisdom or Whims? Chen et al (2025 WP)

Finfluencers (Kakhbod et al WP)



Finfluencers

Swiss Finance Institute Research Paper No. 23-30

57 Pages • Posted: 3 May 2023 • Last revised: 18 Jul 2023

Ali Kakhbod
University of California, Berkeley

Seyed Mohammad Kazempour
Louisiana State University

Dmitry Livdan
University of California, Berkeley

Norman Schuerhoff
Swiss Finance Institute - HEC Lausanne

Date Written: July 5, 2023

Abstract

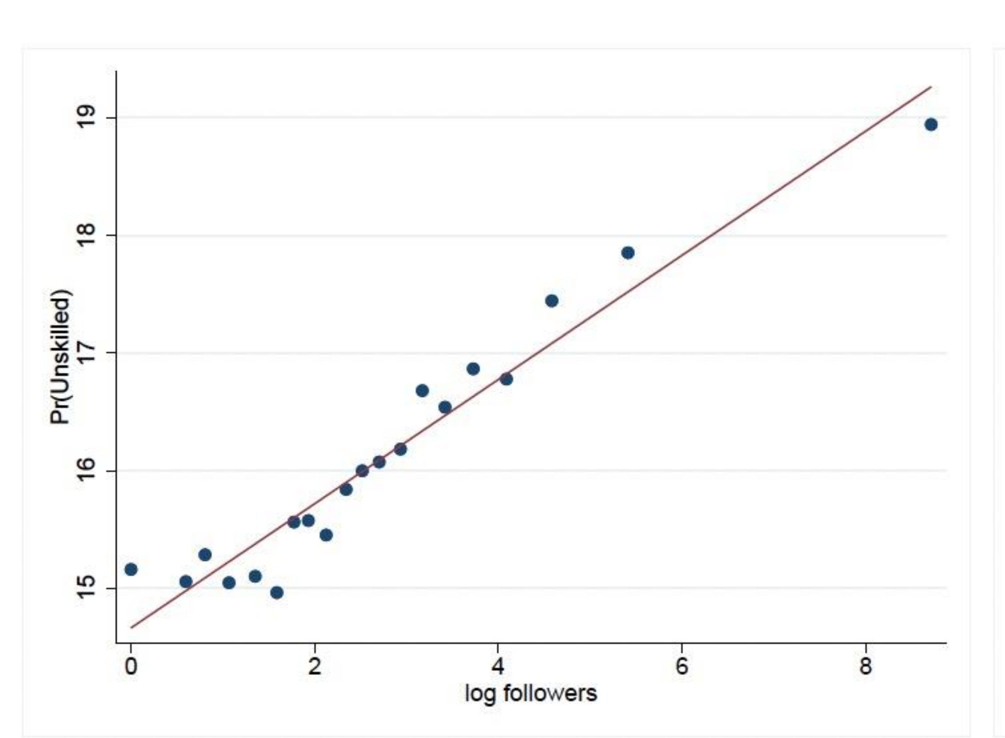
Tweet-level data from a social media platform reveals low average accuracy and high dispersion in the quality of advice by financial influencers, or "finfluencers": 28% of finfluencers are skilled, generating 2.6% monthly abnormal returns, 16% are unskilled, and 56% have negative skill ("antiskill") generating -2.3% monthly abnormal returns. Consistent with homophily shaping finfluencers' social networks, antiskilled finfluencers have more followers and more influence on retail trading than skilled finfluencers. The advice by antiskilled finfluencers creates overly optimistic beliefs most times and persistent swings in followers' beliefs. Consequently, finfluencers cause excessive trading and inefficient prices such that a contrarian strategy yields 1.2% monthly out-of-sample performance

Keywords: Finfluencers, social media, mixture modeling, retail traders, homophily, belief bias

JEL Classification: G12, G14, G41

Estimate skill & alpha in a mixture model (also in Crane and Crotty and in Dim)

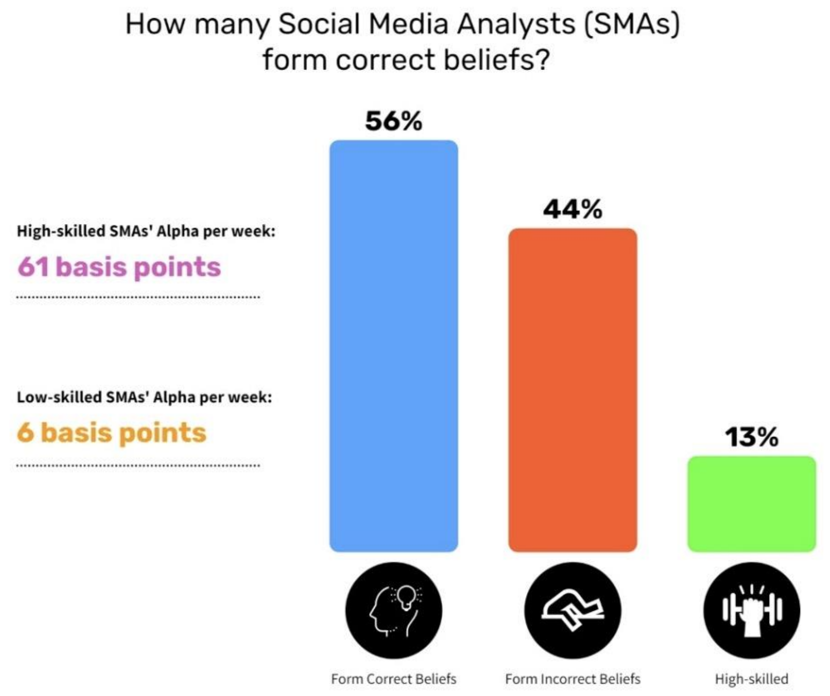
Main finding: skill is **negatively related** to following on StockTwits



Social media analysts' skill (Dim 2025)

Also estimates skill & alpha in a mixture model

Main finding: only 13% of Seeking Alpha authors are skilled



Social Media Analysts' Skill: Evidence from Text-Implied Beliefs

- *Journal of Financial and Quantitative Analysis, forthcoming*
- 2022 Crowell Third Prize
- Brattle Group PhD Candidate Award for Outstanding Research, 2022 WFA Conference
- Best PhD Paper Award, 2021 FMCG Conference
- Nominated for the 2021 Hillsdale Investment Management – CFA Society Toronto Research Award
- Conferences: WFA 2022, Finance Down Under 2022, CICF 2022, SGF 2022, NFA 2021, FIRS 2021, 15th Behavioural Finance Working Group Annual Conference, SoFiE Seminar 2021, 2nd LTI/Bank of Italy Research Workshop, FMCG 2021, AFFI 2021, European Retail Investing Conference 2021, Dauphine PhD Workshop 2021

This paper documents that 56% of nonprofessional social media investment analysts (SMAs) are skilled and declare beliefs that generate positive abnormal returns, while 44% produce negative abnormal returns. 13% of all SMAs are high-skill type and produce a one-week three-factor alpha of 61 bps, while the remaining 87% generate only 6 bps. The distinctive features of high-skill SMAs are primarily firm and industry specializations. Although SMAs tend to extrapolate and herd, their expectations are not systematically wrong. For higher-skilled SMAs compared to the less-skilled ones, extrapolation fades more quickly, and herding is lower, consistent with theory.

Social media and distortion of price revelation

Social Media and the Distortion of Price Revelation

66 Pages • Posted: 16 May 2023 • Last revised: 19 Mar 2025

Edna Lopez Avila
University of Toronto - Rotman School of Management

Charles Martineau
University of Toronto - Rotman School of Management and UTSC Management

Jordi Mondria
University of Toronto - Department of Economics

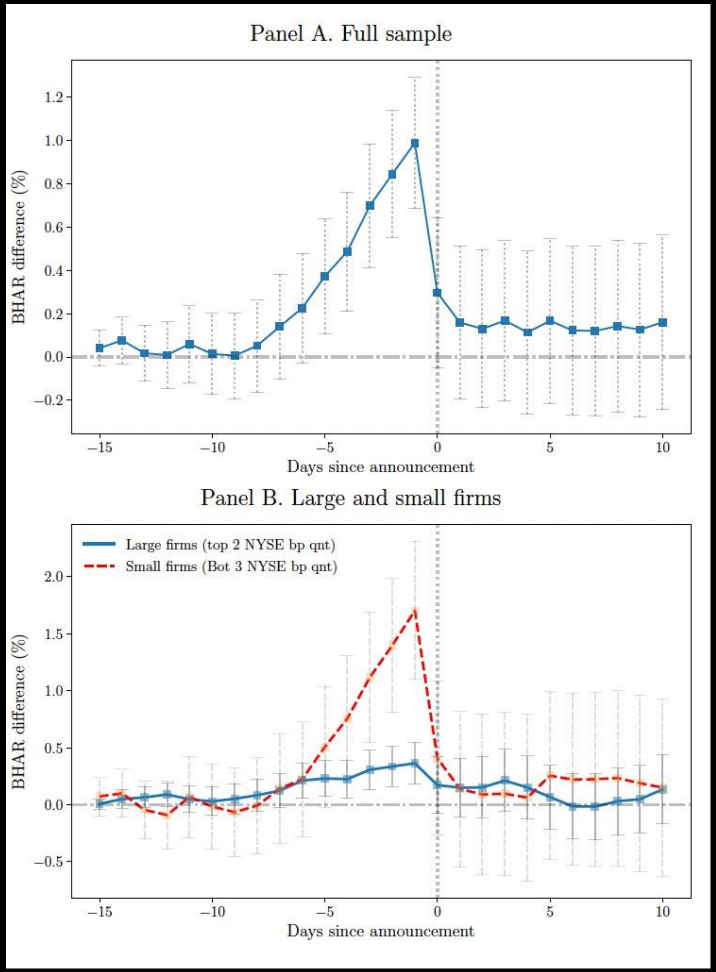
Date Written: May 16, 2024

Abstract

Social media attention before earnings announcements is excessively optimistic, fails to predict fundamentals, and generates buying pressure, leading to a 58 bps stock return as intermediaries seek higher returns to mitigate inventory risk. A return reversal occurs immediately on announcement dates as markets correct the mispricing. The social-media induced buying pressure predicts the reversal and the magnitude of the reversal is amplified by the uncertainty of the earnings news. How stock prices respond to earning news is endogenous to the effect of social media in the pre-announcement price formation. Social media worsens price revelation ahead of earnings announcements.

Keywords: JEL Classification: G12, G14, G40 earnings announcements, investor attention, price efficiency, return reversal, social media, StockTwits, price formation, inventory risk

JEL Classification: E50, G12, G14



Studies social media attention before earnings announcements.

EAs with high attention are **mispriced, no predictability** for earnings surprises

Other work on social media signals

Heterogeneity in skill and influence

- Finfluencers – Kakhbod et al (WP)
- Dim (2025JFQA)
- Martineau et al (2025 WP)
- Wisdom or Whims? Chen et al (2025 WP)

An under-explored angle to this literature – **market level** signals.

- One exception is Gifluence (working paper), which examines the sentiment of GIFs over time.

Market Signals from Social Media

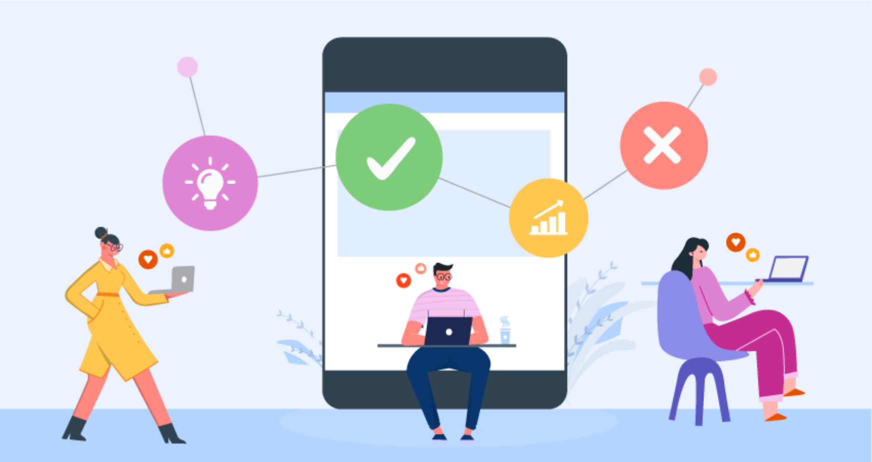
J. Anthony Cookson *(University of Colorado Boulder)*

Runjing Lu *(University of Toronto)*

William Mullins *(UC San Diego)*

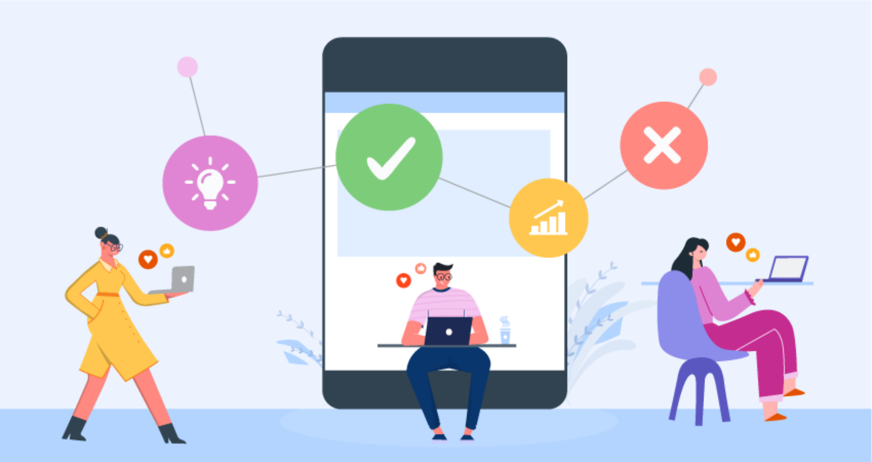
Marina Niessner *(Indiana University)*

Motivation



- Market Signals
  - Sentiment of major interest (especially since Baker and Wurgler 2006)
  - A testing ground for updating models – extrapolation, diagnostic expectations, memory (Bordalo et al. 2018, 2020)
  - Yet, many market signals are low frequency (despite frequent updating) & "sentiment" is sometimes a *mix* of sentiment and attention

Motivation



- Market Signals
  - Sentiment of major interest (especially since Baker and Wurgler 2006)
  - A testing ground for updating models – extrapolation, diagnostic expectations, memory (Bordalo et al. 2018, 2020)
  - Yet, many market signals are low frequency (despite frequent updating) & "sentiment" is sometimes a <u>*mix*</u> of sentiment and attention

- Social Media
  - Increasingly a ***primary*** news source (Pew, 2021)
  - A natural place to look for animal spirits (Gamestop, SVB, etc.)
  - Allows us to separate sentiment from attention

Findings

- High attention and sentiment each independently predict *negative* market returns
  - Return dynamics are distinct:
    - Sentiment: a within-month ***reversal*** after a run-up
    - Attention: a ***continuation*** of negative returns
  - *Economic content: a dynamic trading strategy yields 1.2 Sharpe Ratio*

Findings

- High attention and sentiment each independently predict *negative* market returns
  - Return dynamics are distinct:
    - Sentiment: a within-month ***reversal*** after a run-up
    - Attention: a ***continuation*** of negative returns
  - *Economic content: a dynamic trading strategy yields 1.2 Sharpe Ratio*

- Sentiment and attention have *opposite* relation to aggregate trading
  - S&P500 turnover increases after
    - ***low*** sentiment
    - ***high*** attention

Findings

- High attention and sentiment each independently predict *negative* returns
  - Return dynamics are distinct:
    - Sentiment: a within-month ***reversal*** after a run-up
    - Attention: a ***continuation*** of negative returns
  - *Economic content: a dynamic trading strategy yields 1.2 Sharpe Ratio*

- Sentiment and attention have *opposite* relation to aggregate trading
  - S&P500 turnover increases after ***low*** sentiment and ***high*** attention

- What ***drives*** market-wide sentiment and attention?
  - VAR: strong connection between ***lagged trading::attention*** and ***lagged returns::sentiment***.
  - Market Price **Jumps**:
    - negative jumps ➔ sentiment ↓ and attention ↑
      - Similar when using spikes in VIX
    - positive jumps do not matter

Contributions

- Daily measures of market-wide sentiment and attention
  - Distinct patterns for sentiment versus attention should be of interest to macro updating literature
  - *A high-frequency measure. All results hold with year-month FE*
- Thinking about extrapolation in market sentiment
  - Sentiment is extrapolative with respect to lagged returns
  - … but this relationship is driven by *negative* market jumps
- Social media contribution
  - Aggregate focus (vs. firm-level) is novel relative to this growing literature

Constructing Market Signals from Social Media

Step 1: Data and measures

Firm-Day Data:

- StockTwits
- Twitter (*from a company called Context Analytics*)
- Seeking Alpha (*from Ravenpack 1.0*)

Step 1: Data and measures

Firm-Day Data:

- StockTwits
- Twitter (*from a company called Context Analytics*)
- Seeking Alpha (*from Ravenpack 1.0*)

Attention measure:

- Each source gives # of posted messages per firm-day

Sentiment measure:

- Firm-day sentiment (Twitter)
- Message-level sentiment (StockTwits, SeekingAlpha) => average by firm-day

Sample Restriction: at least 10 StockTwits messages to include firm-day

Step 2: Purge and Aggregate

Firm-day signals S_{it} could be driven by idiosyncratic reactions to news and differences across firms.

Purge: for each platform, we run auxiliary regressions for each signal:

$$S_{it}^{st} = \Gamma^{st} X_{it} + \beta_{st} \overline{S_{i,-y}^{st}} + \epsilon_{it}^{st}$$

$$S_{it}^{tw} = \Gamma^{tw} X_{it} + \beta_{tw} \overline{S_{i,-y}^{tw}} + \epsilon_{it}^{tw}$$

$$S_{it}^{SA} = \Gamma^{SA} X_{it} + \beta_{SA} \overline{S_{i,-y}^{SA}} + \epsilon_{it}^{SA}$$

Where X_{it} includes indicators for traditional news, 8-K filings, and earnings announcements on days $t-7$ to t. The regressions also control for firm’s average signal in the prior year

Step 2: Purge and Aggregate

Firm-day signals S_{it} could be driven by idiosyncratic reactions to news and differences across firms.

Purge: for each platform, we run auxiliary regressions for each signal:

$$S_{it}^{st} = \Gamma^{st} X_{it} + \beta_{st}\, \overline{S_{i,-y}^{st}} + \epsilon_{it}^{st}$$

$$S_{it}^{tw} = \Gamma^{\mathrm{tw}} X_{it} + \beta_{tw}\, \overline{S_{i,-y}^{tw}} + \epsilon_{it}^{tw}$$

$$S_{it}^{SA} = \Gamma^{SA} X_{it} + \beta_{SA}\, \overline{S_{i,-y}^{SA}} + \epsilon_{it}^{SA}$$

Where X_{it} includes indicators for traditional news, 8-K filings, and earnings announcements on days $t-7$ to t. The regressions also control for firm's average signal in the prior year

Aggregate the residuals (ϵ_{it}^{st}, ϵ_{it}^{tw} and ϵ_{it}^{SA}**)** into daily attention and sentiment series by platform via market cap-weighted average.

Step 3: Combine into separate Sentiment and Attention Series

- After the purge-and-aggregate step, we have six signals: 3 platforms x {sentiment, attention}
- Perform two PCAs: one on sentiment signals, another on attention signals
- The first PCs of each = daily sentiment and attention indexes

How do sentiment and attention aggregate?

Purging explains more variation in attention than sentiment

Sentiment puts twice the weight on StockTwits and Twitter as on Seeking Alpha

Attention puts most weight on StockTwits and Twitter

Note: Aggregation puts more weight on large firms

Table 2: Sentiment and Attention Index Construction

Panel A: Residualizing regressions for platform-day signal

	Dep. var.: Sentiment$_{i,t}$ (z)			Dep. var.: Attention$_{i,t}$ (z)		
	ST	TW	SA	ST	TW	SA
Firm annual avg$_{i,y(t)-1}$	0.373***	0.569***	0.295***	0.834***	0.789***	0.531***
	(0.018)	(0.015)	(0.026)	(0.084)	(0.043)	(0.046)
Firm news controls	Y	Y	Y	Y	Y	Y
Observations	738,438	738,438	738,438	738,438	738,438	738,438
R^2	0.0349	0.1093	0.0665	0.0811	0.4612	0.4031

Panel B: PCA of platform-day signal

	Sentiment PC1	Attention PC1
StockTwits	0.649	0.707
	(0.020)	(0.014)
Twitter	0.675	0.706
	(0.013)	(0.016)
Seeking Alpha	0.352	0.040
	(0.091)	(0.099)
Fraction(%)	46.876	53.696
	(1.207)	(2.525)

How do sentiment and attention validate?

Sentiment is negatively related to Twitter EPU and to attention.

Some other relationships, but not much robust to calendar patterns.

Even these correlations are modest R2 ~ 10% without FE.

Panel A: Sentiment$_t$				
ARA$_t(z)$	-0.079***	0.021	0.021	0.068**
	(0.030)	(0.032)	(0.026)	(0.027)
AIA$_t(z)$	0.134***	0.155***	-0.032	-0.009
	(0.032)	(0.030)	(0.025)	(0.025)
MAI (WSJ)$_t(z)$	-0.051**	-0.098***	-0.022	-0.025
	(0.026)	(0.028)	(0.019)	(0.018)
MAI (NYT)$_t(z)$	0.047*	0.064***	-0.026	-0.023
	(0.025)	(0.024)	(0.017)	(0.017)
Twitter EU$_t(z)$	-0.078***	-0.045**	-0.048**	-0.047**
	(0.030)	(0.022)	(0.023)	(0.022)
RavenPack news$_t(z)$	-0.035	-0.029	0.021	0.019
	(0.030)	(0.028)	(0.020)	(0.019)
Attention$_t(z)$		-0.294***		-0.147***
		(0.049)		(0.030)
Observations	2,267	2,267	2,267	2,267
R^2	0.028	0.099	0.509	0.518
DOW FE	N	N	Y	Y
MOY FE	N	N	Y	Y
YQ FE	N	N	Y	Y

How do sentiment and attention validate?

Attention is positively related to retail (ARA) and institutional attention (AIA).

Negatively related to attention, strongest connection to retail attention

Some other relationships, but not much robust to calendar patterns.

Even these correlations are modest R2 ~ 24% without FE.

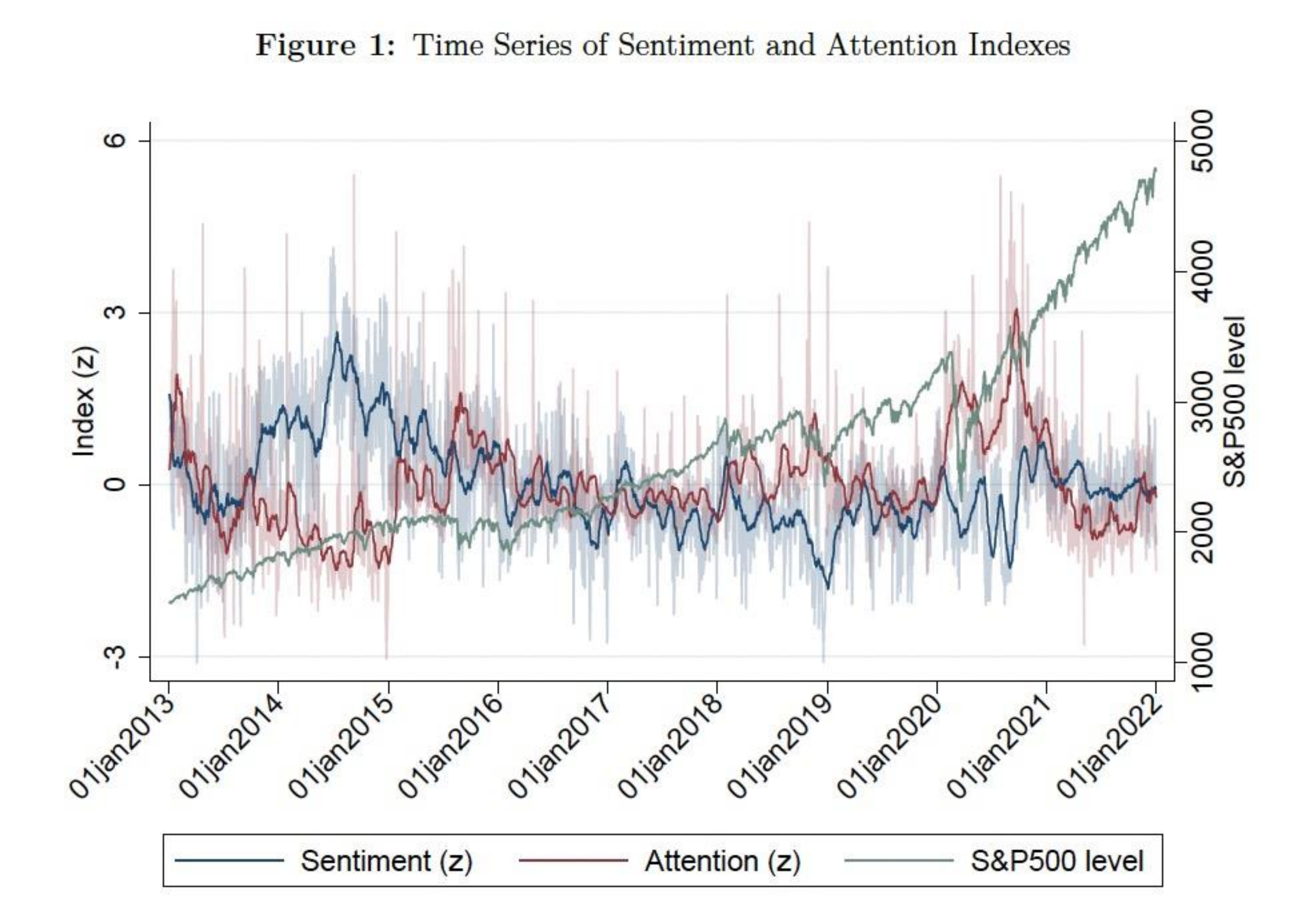
Panel B: Attention$_t$				
$\text{ARA}_t(z)$	0.342***	0.322***	0.315***	0.318***
	(0.059)	(0.056)	(0.043)	(0.042)
$\text{AIA}_t(z)$	0.073**	0.106***	0.162***	0.158***
	(0.032)	(0.031)	(0.026)	(0.026)
$\text{MAI (WSJ)}_t(z)$	-0.161***	-0.173***	-0.017	-0.020
	(0.036)	(0.036)	(0.015)	(0.015)
$\text{MAI (NYT)}_t(z)$	0.059**	0.070***	0.023	0.020
	(0.023)	(0.022)	(0.016)	(0.016)
$\text{Twitter EU}_t(z)$	0.110**	0.090*	0.004	-0.002
	(0.054)	(0.050)	(0.017)	(0.016)
$\text{RavenPack news}_t(z)$	0.021	0.012	-0.016	-0.014
	(0.030)	(0.028)	(0.019)	(0.018)
$\text{Sentiment}_t(z)$		-0.248***		-0.123***
		(0.030)		(0.023)
Observations	2,267	2,267	2,267	2,267
R^2	0.182	0.242	0.589	0.597
DOW FE	N	N	Y	Y
MOY FE	N	N	Y	Y
YQ FE	N	N	Y	Y

Time series variation in social media indexes

Sentiment and attention have distinct variation.

Low frequency variation highlights important episodes
- Bull run in 2014
- China trade war in 2018
- Pandemic onset in 2020

Lots of *high frequency variation* (the focus of our paper)



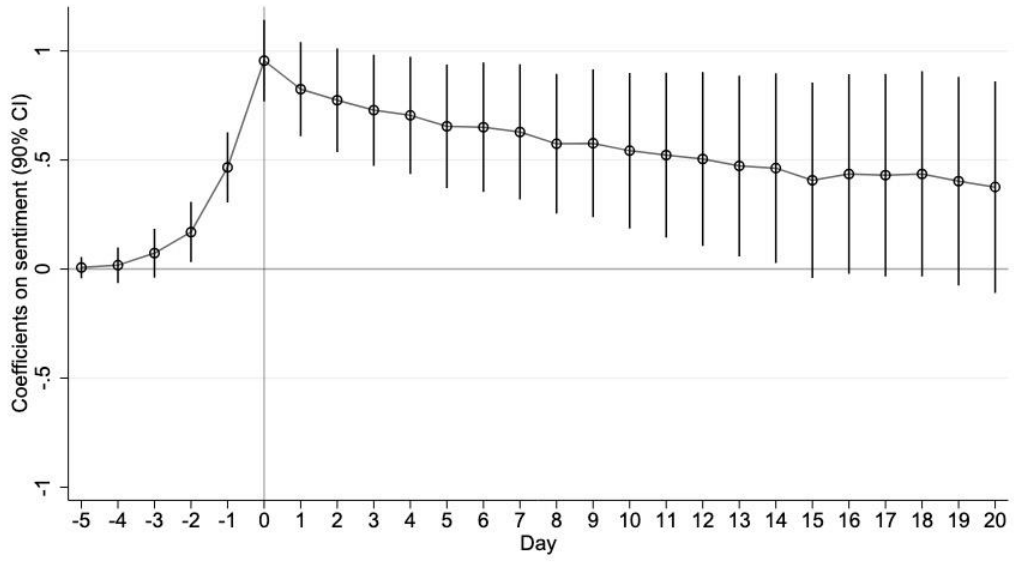
Return Results

Return Results in One Picture

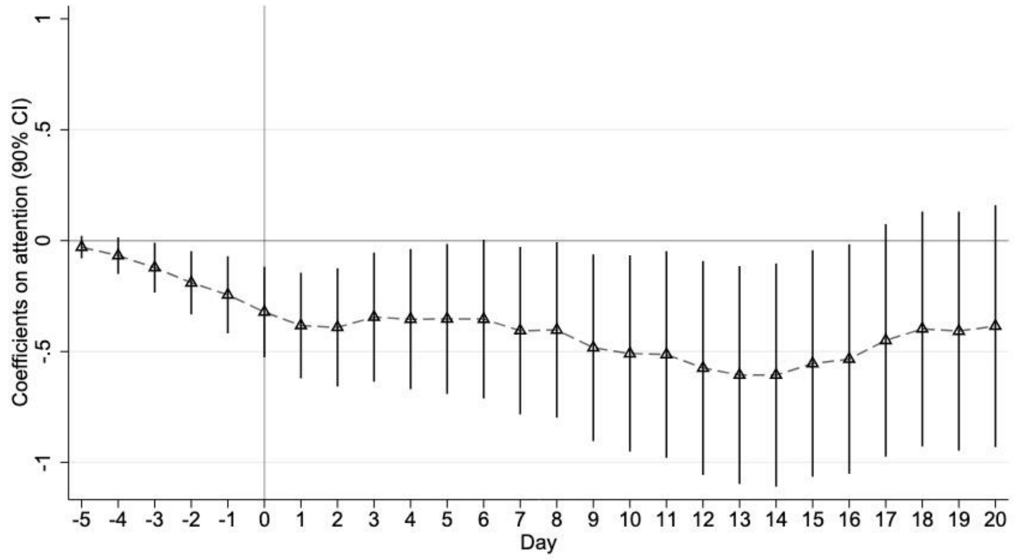
Both sentiment and attention at date t predict negative returns

But, for different reasons

- Sentiment reversal
- Attention continuation



(a) Sentiment



(b) Attention

Market Signals Predict Future Negative Returns

Table 4: Do Sentiment and Attention Indexes Predict Returns?

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+1∼t+15	(6) Day t+1∼t+15
$\text{Sentiment}_t(z)$	0.524***	0.544***	-0.106***	-0.108***	-0.383**	-0.382**
	(0.041)	(0.042)	(0.035)	(0.037)	(0.174)	(0.183)
$\text{Attention}_t(z)$	-0.095***	-0.097***	-0.068**	-0.067**	-0.335	-0.335
	(0.029)	(0.030)	(0.033)	(0.033)	(0.206)	(0.207)
$\text{Sentiment} \times \text{Attention}_t(z)$		0.158***		-0.022		0.008
		(0.038)		(0.031)		(0.191)
Controls	Y	Y	Y	Y	Y	Y
DOW FE	Y	Y	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y	Y	Y
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.173	0.192	0.035	0.036	0.326	0.326

1 sd higher sentiment predicts -38 basis points through day 15 (reversal)

1 sd higher attention predicts -6.7 basis points on day $t+1$ (continuation)

Results are similar with year-month FE

Quantifying Economic Importance

Trading strategy

Follow a dynamic trading strategy to allocate to S&P500

- Fit day $t+1$ returns on day t attention and sentiment + interaction from 2013 through prior month $(m-1)$.

$$\hat{r}_{t+1} = \hat{\beta}_{1,m-1} Sentiment_t + \hat{\beta}_{2,m-1} Attention_t + \hat{\beta}_{3,m-1} (Sentiment \times Attention)_t + \hat{\gamma}_{m-1} \Omega_t$$

- For month m, use fitted model & day t realizations to construct prediction for day $t+1$ returns, and form weight (allocation) on S&P500

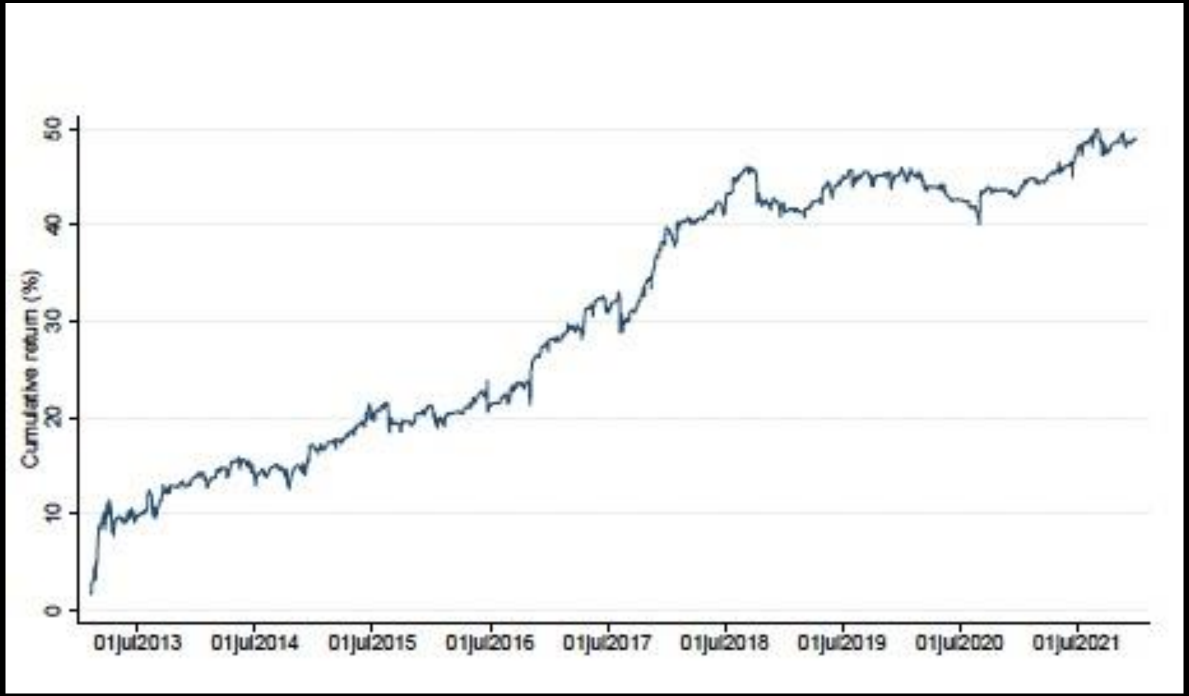
$$w_t^{social} \equiv \frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}$$

Quantifying Economic Importance

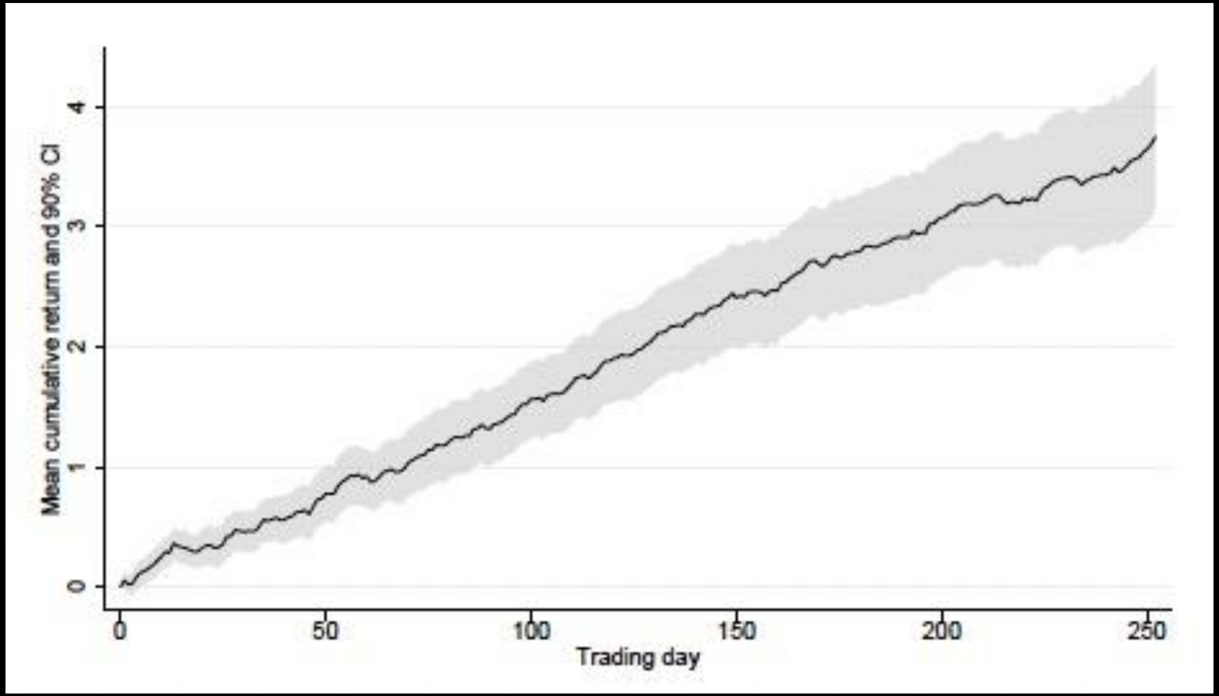
Trading strategy

Strategy delivers 3-4.6% annually (cumulative 50% gain over full sample) with a Sharpe ratio of 1.2.

Cumulative Returns



Representative Year



Quantifying Economic Importance

Trading strategy

Descriptively, there is (apparently) significant alpha in excess of FF3+Momentum

The strategy is exposed to market risk and momentum.

- But portfolio returns are mostly idiosyncratic (R^2 of 15.7%)

Table 7: Dynamic Strategy: Abnormal Returns and Factor Decomposition

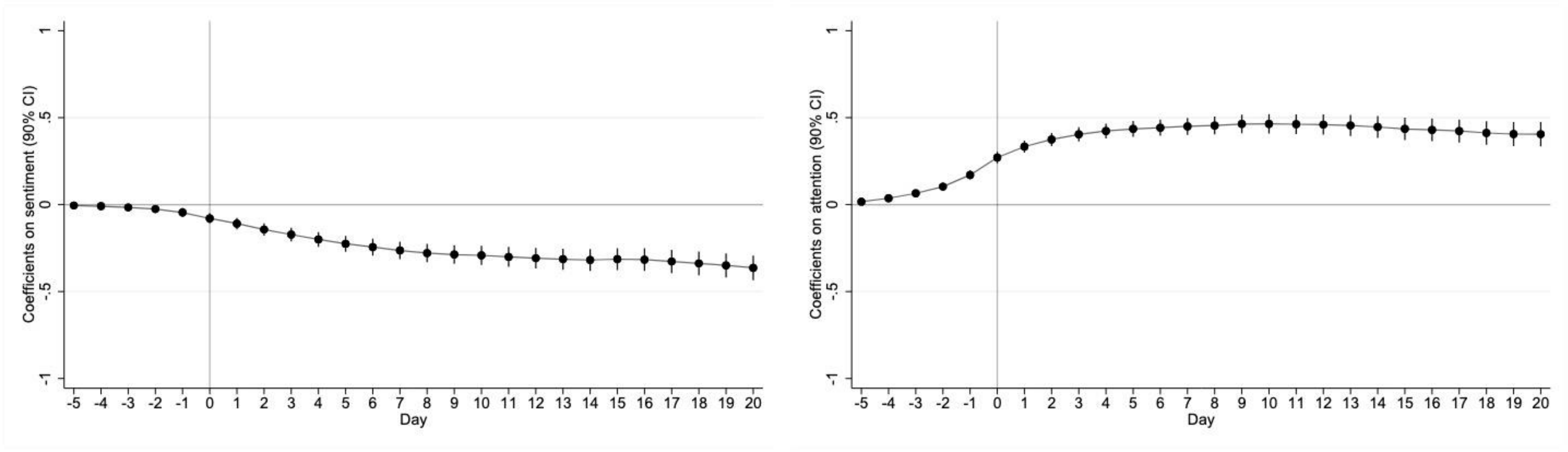
	Dependent var.: Portfolio excess return$_{t+1}$(%)			
	(1)	(2)	(3)	(4)
Alpha	0.018***	0.012**	0.012**	0.012**
	(0.005)	(0.005)	(0.005)	(0.005)
Market excess return$_{t+1}$		0.096***	0.098***	0.097***
		(0.016)	(0.017)	(0.017)
SMB$_{t+1}$			-0.011	-0.010
			(0.009)	(0.009)
HML$_{t+1}$			-0.011	0.002
			(0.009)	(0.010)
MOM$_{t+1}$				0.018**
				(0.008)
Observations	2,246	2,246	2,246	2,246
R^2	–	0.152	0.154	0.157
Alpha (annualized)	4.564***	3.051**	2.984**	3.020**
	(1.249)	(1.226)	(1.234)	(1.232)
Information ratio (annualized)	1.224	0.833	0.810	0.821

Results on Trading Activity

Trading Activity Results in One Picture

Distinct relation to aggregate trading: more trading after

- Low sentiment
- High attention



(a) Sentiment and S&P 500 turnover

(b) Attention and S&P 500 turnover

Indexes predict aggregate trading

Trading activity results are significant too.
Result is robust to controlling for year-month FE and other aggregate attention indexes (Da et al. 2024, Fisher et al. 2022)

Table 5: Do Sentiment and Attention Indexes Predict Turnover?

	(1) Day t	(2) Day t	(3) Day t+1	(4) Day t+1	(5) Day t+1∼t+15	(6) Day t+1∼t+15
Panel A: S&P turnover						
$\text{Sentiment}_t(z)$	-0.020***	-0.021***	-0.018***	-0.019***	-0.173***	-0.178***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.033)	(0.033)
$\text{Attention}_t(z)$	0.071***	0.071***	0.042***	0.042***	0.120***	0.121***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.033)	(0.033)
$\text{Sentiment} \times \text{Attention}_t(z)$		-0.007		-0.008		-0.043
		(0.005)		(0.005)		(0.028)
Observations	2,267	2,267	2,267	2,267	2,267	2,267
R^2	0.596	0.597	0.482	0.483	0.724	0.724

What drives sentiment and attention?

OLS Regressions: *Predictors of Sentiment and Attention*

Sentiment is higher after high recent returns and low market turnover.

Attention is higher after high market turnover, especially S&P500 constituents.

But, OLS doesn't account for feedback dynamics →estimate a VAR

Table 9: What Predicts Social Media Sentiment and Attention Indexes?

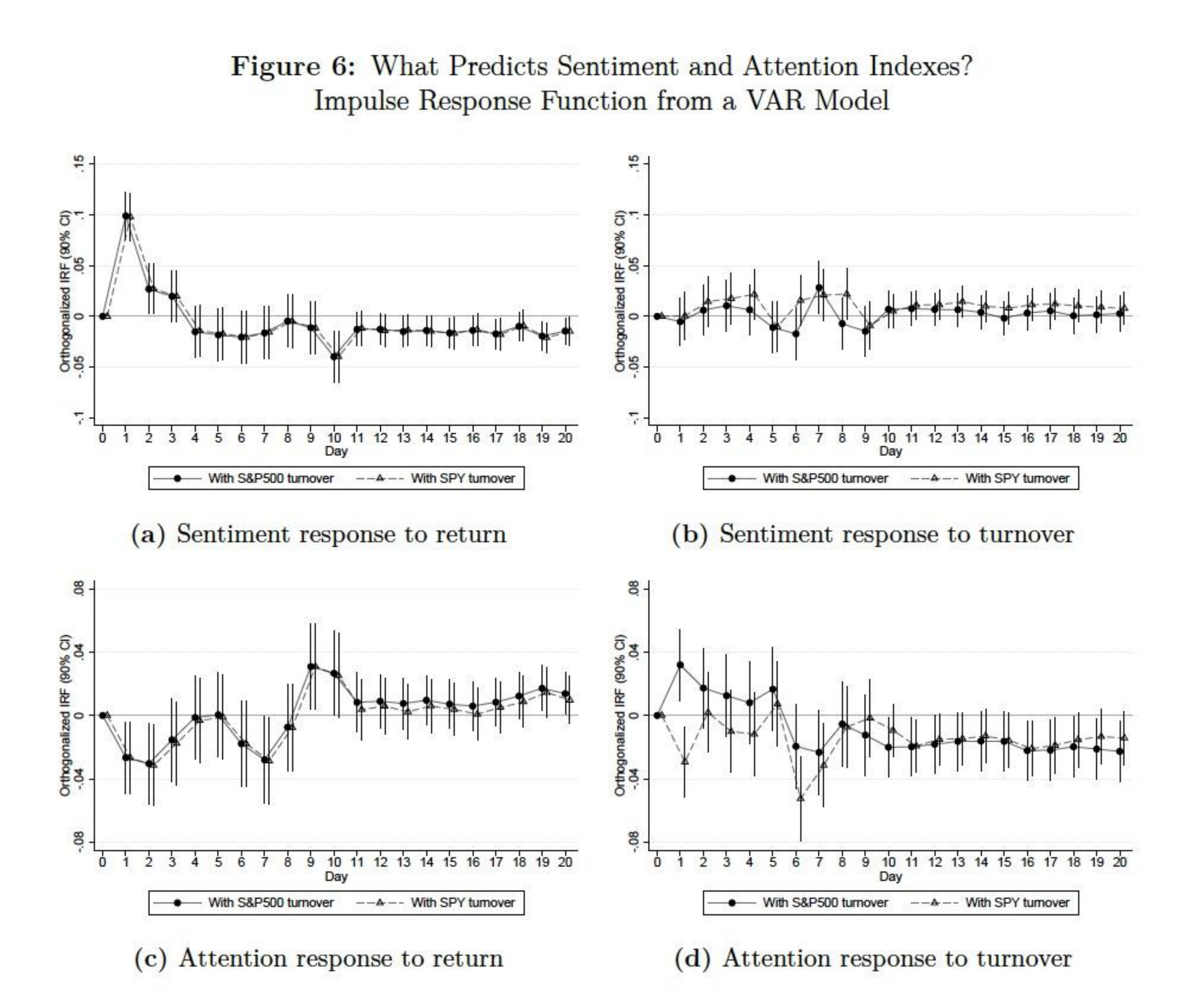
	Dependent var.: $\text{Sentiment}_t(z)$		Dependent var.: $\text{Attention}_t(z)$	
	(1) S&P turnover	(2) SPY turnover	(3) S&P turnover	(4) SPY turnover
Return_{t-1}	0.144***	0.139***	-0.024*	-0.035**
	(0.027)	(0.028)	(0.013)	(0.014)
Return_{t-2}	0.074***	0.073***	-0.024	-0.031*
	(0.017)	(0.019)	(0.015)	(0.018)
Return_{t-3}	0.020	0.019	-0.011	-0.020
	(0.015)	(0.015)	(0.014)	(0.015)
Return_{t-4}	0.003	0.005	-0.007	-0.022
	(0.016)	(0.016)	(0.013)	(0.015)
Return_{t-5}	0.008	0.007	-0.000	-0.019
	(0.014)	(0.014)	(0.017)	(0.018)
Ab. $\log(\text{turnover})_{t-1}$	-0.208**	-0.136**	0.897***	0.167***
	(0.096)	(0.057)	(0.091)	(0.055)
Ab. $\log(\text{turnover})_{t-2}$	0.024	0.039	0.016	0.058
	(0.091)	(0.055)	(0.078)	(0.050)
Ab. $\log(\text{turnover})_{t-3}$	-0.035	-0.021	0.028	-0.025
	(0.085)	(0.055)	(0.073)	(0.053)
Ab. $\log(\text{turnover})_{t-4}$	-0.035	0.015	-0.025	-0.089
	(0.105)	(0.058)	(0.077)	(0.058)
Ab. $\log(\text{turnover})_{t-5}$	-0.196**	-0.090*	0.115	-0.005
	(0.090)	(0.052)	(0.078)	(0.055)
DOW FE	Y	Y	Y	Y
MOY FE	Y	Y	Y	Y
YQ FE	Y	Y	Y	Y
Observations	2267	2267	2267	2267
R^2	0.535	0.533	0.533	0.505

VAR: Impulse Response

10 daily lags of attention, sentiment, returns and turnover

1 SD increase in Return:
significant spike in sentiment, drop in attention.

1 SD increase in Turnover:
No response in sentiment, but a persistent increase in attention (for S&P500 turnover)



Jumps:

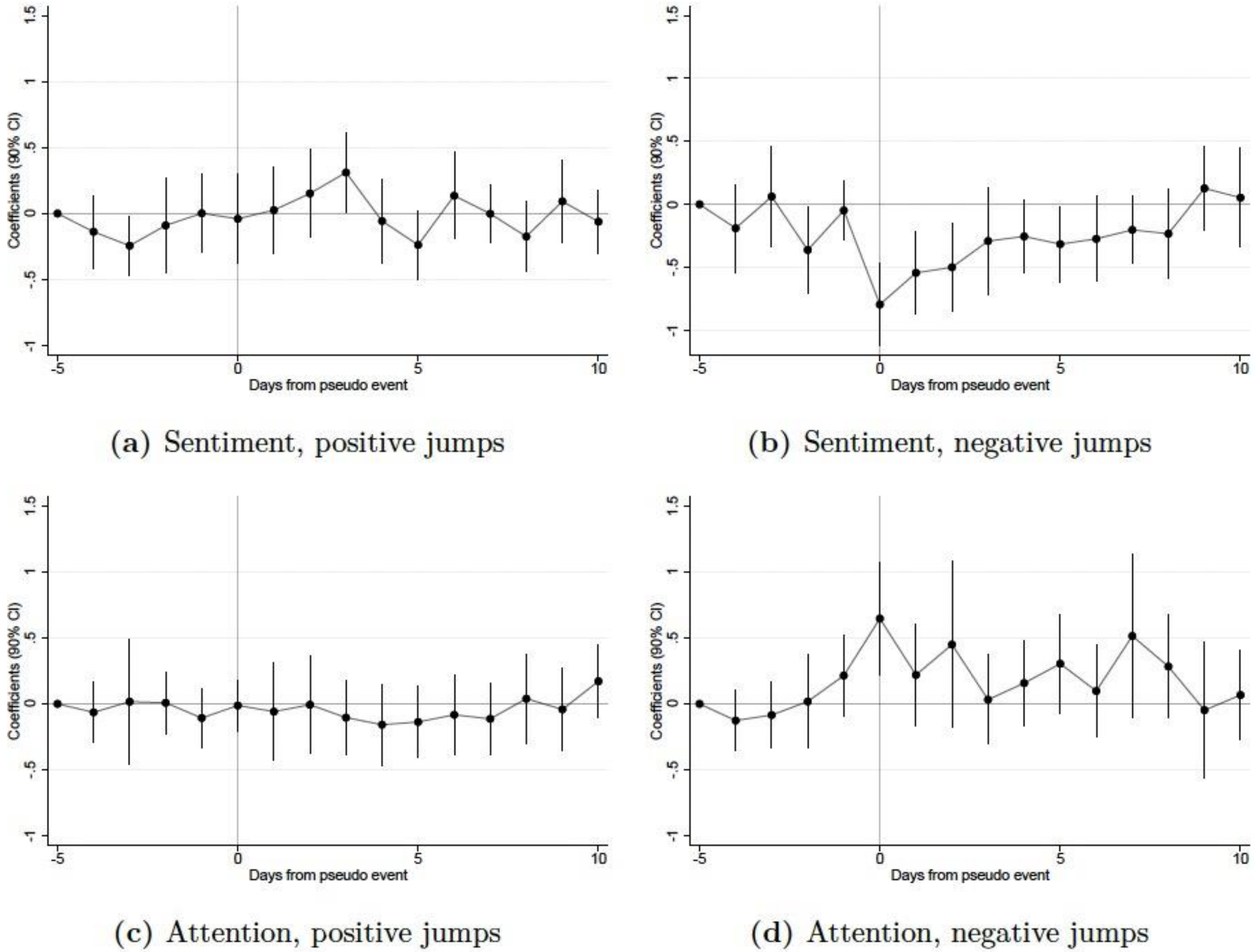
Separate impact of positive returns and negative returns

Positive Jumps (left column):
No shift in sentiment/attention!

Negative Jumps (right column):
Sentiment drops sharply, attention rises

- Similar impacts for positive spikes in VIX.
- Not driven by FOMC days

Figure 7: How Do Sentiment and Attention Indexes Change around Return Jumps



(a) Sentiment, positive jumps

(b) Sentiment, negative jumps

(c) Attention, positive jumps

(d) Attention, negative jumps

Spillovers from Central Firms

Bian, Huang, Li and Tang (2025)

- Measure centrality in the data economy.
- Propose a shock the Apple Tracking Transparency (ATT) policy.

Take their ~40 most central firms, compute central firm sentiment and attention.

Then, we estimate

$$Sent_t = \beta_1 Sent_t^{CF} + \beta_2 Sent_t^{CF} \times post_t + \beta_3 post_t + \epsilon_t$$

Spillovers from Central Firms

Central firms' connection to overall sentiment dampens after ATT.

Not much happens for connection of central attention to overall attention

	Dep. var.: Overall index (z)			Dep. var.: Non-central firm index (z)		
	Day t (1)	Day t+1 (2)	Day t+2 (3)	Day t (4)	Day t+1 (5)	Day t+2 (6)
Panel A: Sentiment						
Post ATT × Central sentiment$_t(z)$	-0.375***	-0.496***	-0.459***	-0.514***	-0.541***	-0.545***
	(0.087)	(0.105)	(0.125)	(0.114)	(0.114)	(0.129)
Central sentiment$_t(z)$	0.910***	0.485***	0.437***	0.687***	0.445***	0.436***
	(0.070)	(0.095)	(0.106)	(0.095)	(0.102)	(0.106)
Observations	422	421	420	422	421	420
R^2	0.709	0.308	0.280	0.472	0.301	0.294
Panel B: Attention						
Post ATT × Central attention$_t(z)$	-0.093	0.030	0.145	-0.036	0.009	0.168
	(0.079)	(0.107)	(0.126)	(0.181)	(0.178)	(0.217)
Central attention$_t(z)$	0.954***	0.462***	0.154	0.330***	0.220*	0.143
	(0.050)	(0.076)	(0.105)	(0.112)	(0.115)	(0.113)
Observations	422	421	420	422	421	420
R^2	0.898	0.670	0.601	0.647	0.628	0.619
DOW FE	Y	Y	Y	Y	Y	Y
Event quarter FE	Y	Y	Y	Y	Y	Y

Summarizing …

We develop new ***market*** indexes of sentiment and attention

- Predictive of returns within month (new relative to vast sentiment literature)
- Distinct dynamics for attention versus sentiment
- Sentiment and attention indexes have distinct predictions for aggregate turnover

What drives sentiment and attention indexes?

- Sentiment → extrapolative of returns but driven by the downside.
- Attention → attention increases after rises in S&P500 turnover

Much more to investigate

- Can sentiment predict other outcomes/portfolios?
- What implications do these dynamics have for how to think about macro updating?

Summarizing Day 3

Social media in Social Finance

- Social media is a venue for studying selective sharing, updating and information sourcing, which can lead to **socially emergent phenomena**.
- As a forum, it is a high-data environment where social finance is happening; this has promise.

Understanding social media signals remains an active area

- Platform differences, heterogeneity across people, interactions between retail and sophisticated investors.
- *What is not as well understood? New forums, new features and formats (videos, GIFs, Emojis), connections across forums and to traditional media.*
  - ***Ripe for new research***