Market Signals from Social Media

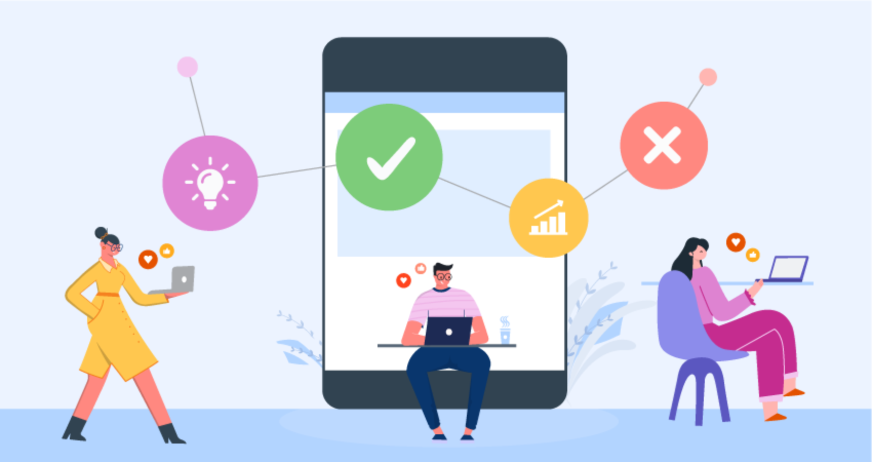
J. Anthony Cookson (*University of Colorado Boulder*)

Runjing Lu (*University of Toronto*)

William Mullins (*UC San Diego*)

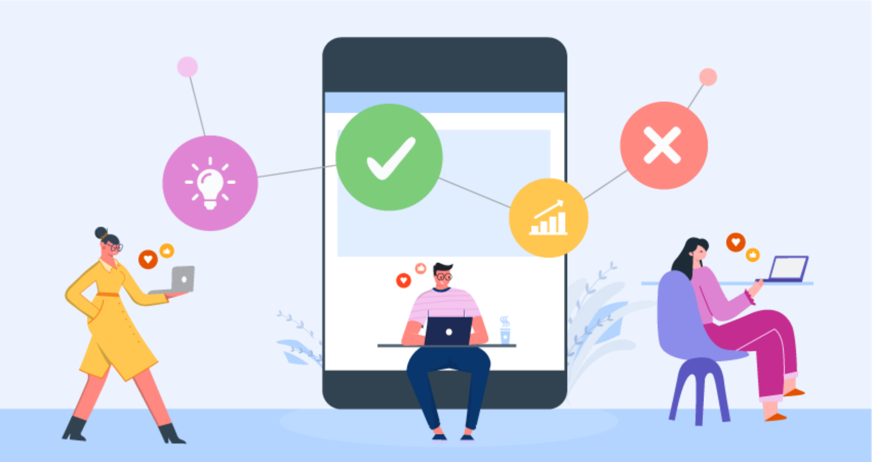
Marina Niessner (*Indiana University*)

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 *mix* 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* 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 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:

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

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

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

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 indexes

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$_t$				
$\text{ARA}_t(z)$	-0.079***	0.021	0.021	0.068**
	(0.030)	(0.032)	(0.026)	(0.027)
$\text{AIA}_t(z)$	0.134***	0.155***	-0.032	-0.009
	(0.032)	(0.030)	(0.025)	(0.025)
$\text{MAI (WSJ)}_t(z)$	-0.051**	-0.098***	-0.022	-0.025
	(0.026)	(0.028)	(0.019)	(0.018)
$\text{MAI (NYT)}_t(z)$	0.047*	0.064***	-0.026	-0.023
	(0.025)	(0.024)	(0.017)	(0.017)
$\text{Twitter EU}_t(z)$	-0.078***	-0.045**	-0.048**	-0.047**
	(0.030)	(0.022)	(0.023)	(0.022)
$\text{RavenPack news}_t(z)$	-0.035	-0.029	0.021	0.019
	(0.030)	(0.028)	(0.020)	(0.019)
$\text{Attention}_t(z)$		-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.

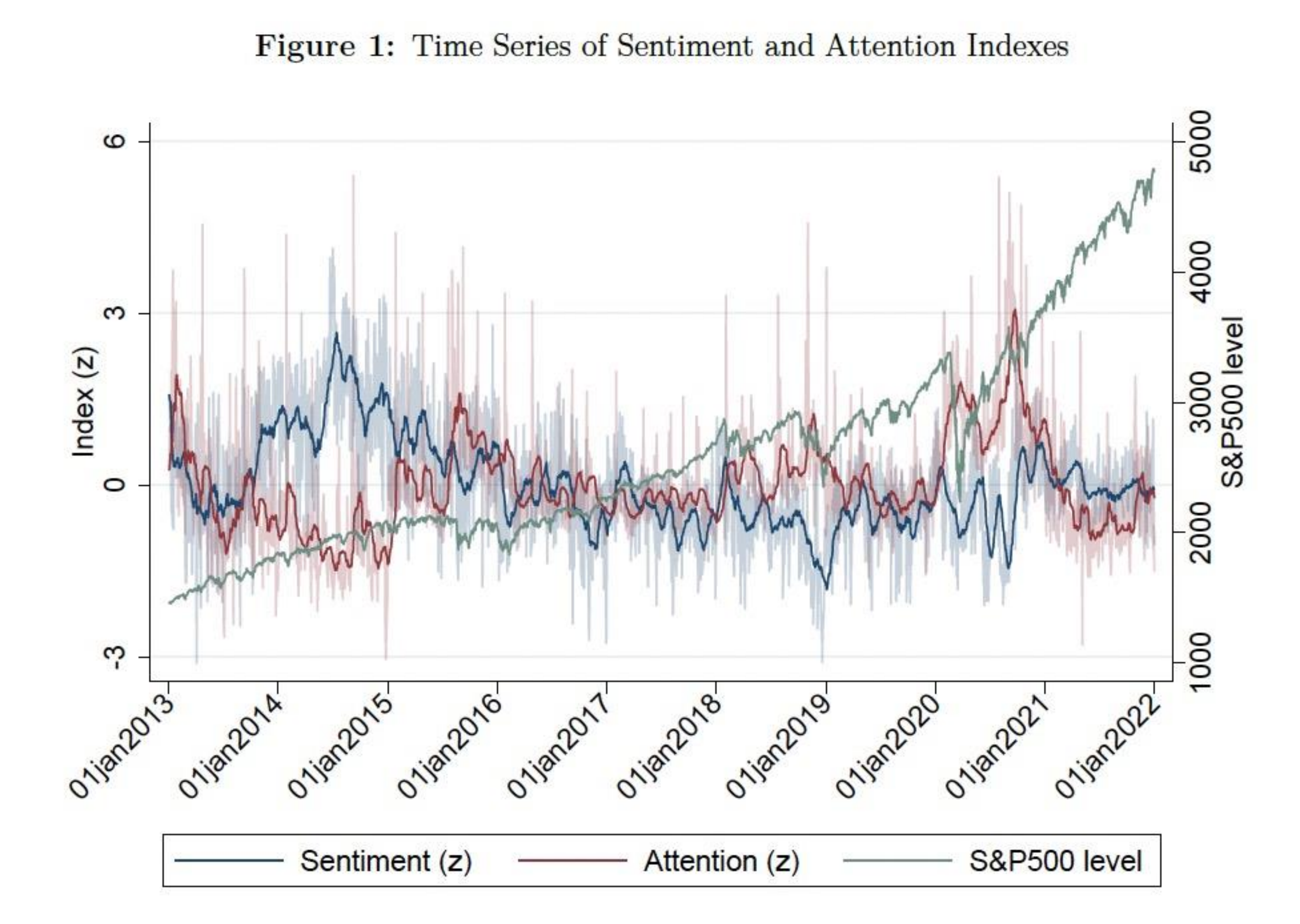
Panel B: Attention$_t$				
$\text{ARA}_t(z)$	0.342***	0.322***	0.315***	0.318***
	(0.059)	(0.056)	(0.043)	(0.042)
$\text{AIA}_t(z)$	0.073**	0.106***	0.162***	0.158***
	(0.032)	(0.031)	(0.026)	(0.026)
$\text{MAI (WSJ)}_t(z)$	-0.161***	-0.173***	-0.017	-0.020
	(0.036)	(0.036)	(0.015)	(0.015)
$\text{MAI (NYT)}_t(z)$	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

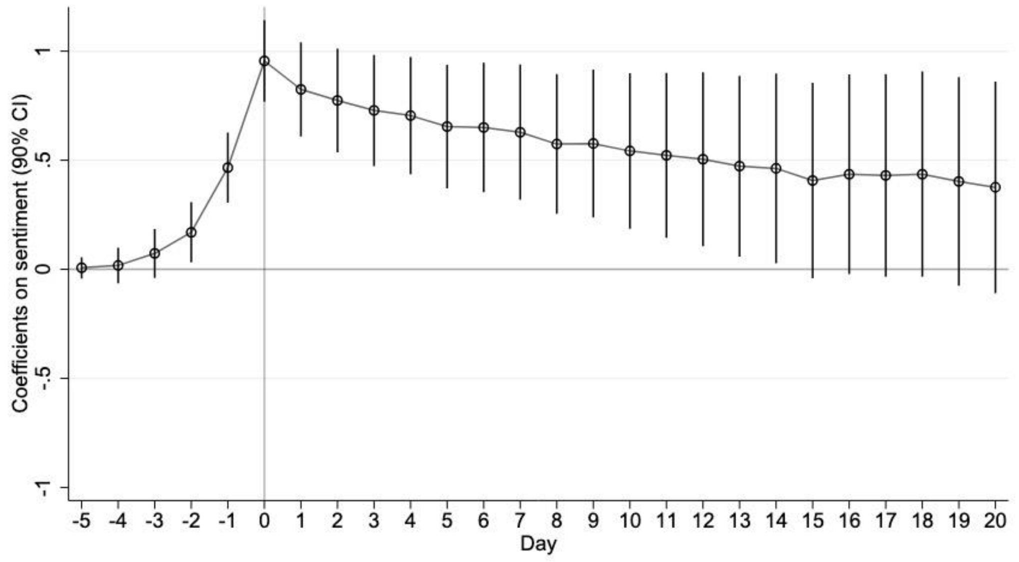
Return Results in One Picture

Both sentiment and attention at date t predict negative returns

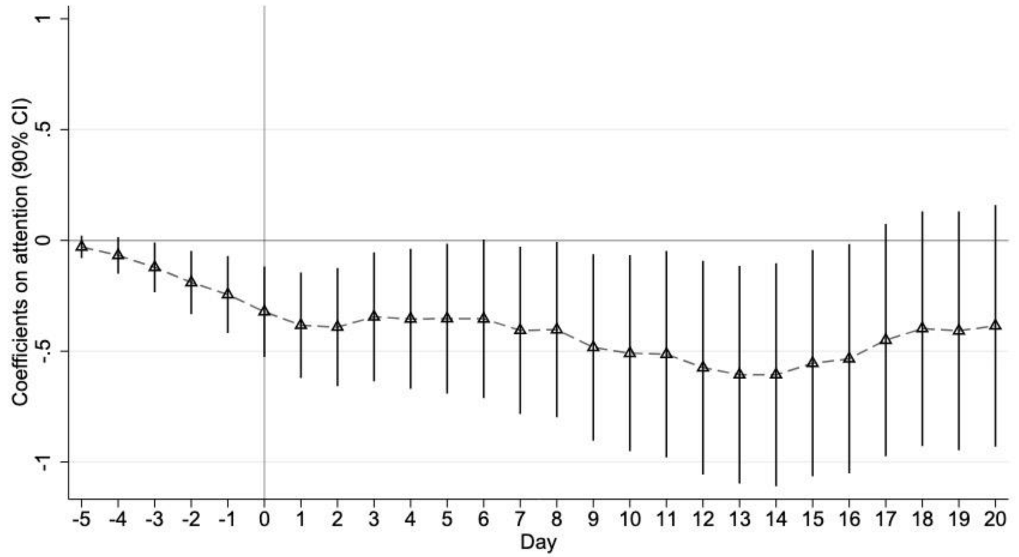
But, for different reasons

- Sentiment reversal (-38bps by t+15)
- Attention continuation (-6.7 bps on day t+1)

Results are similar with year-month FE



(a) Sentiment



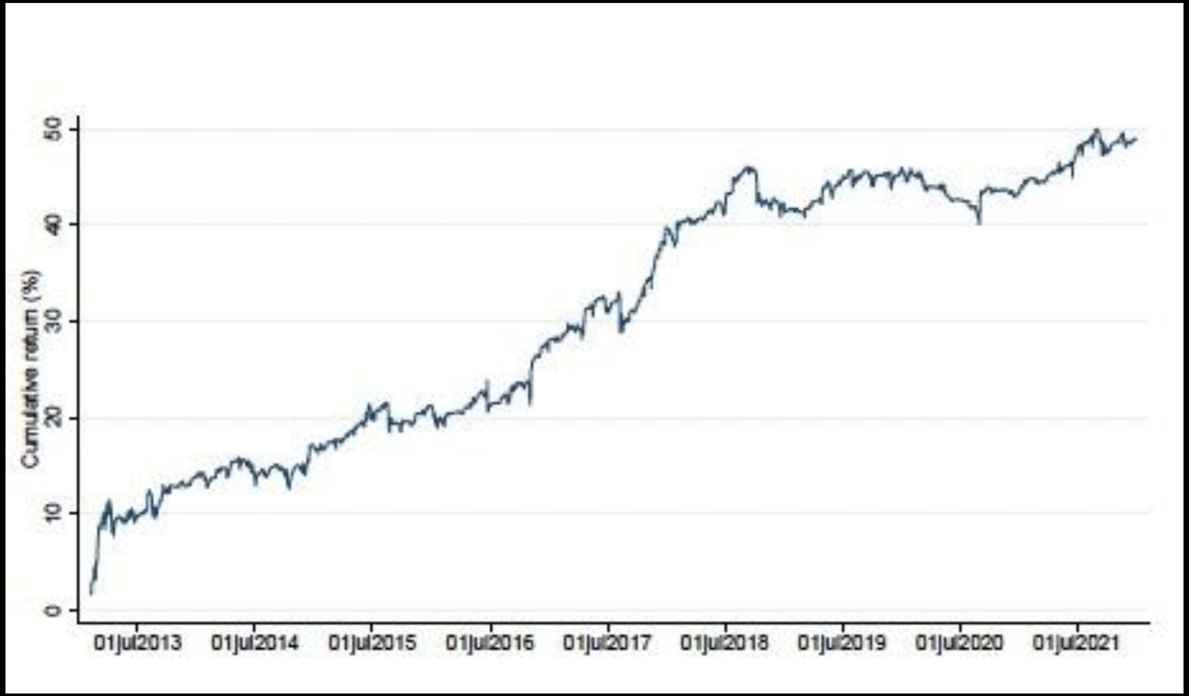
(b) Attention

Not future information?

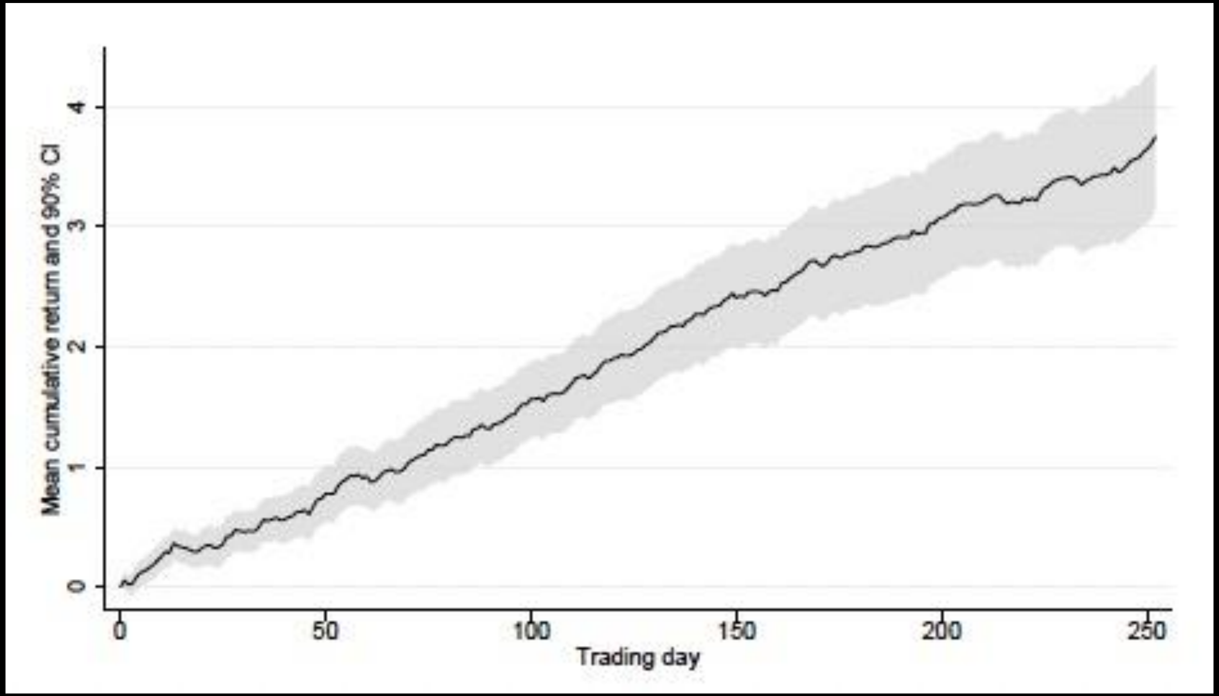
Trading strategy

Strategy delivers 3-4.6% annually (cumulative 50% gain over full sample) with a Sharpe ratio of 1.2, which is not explained by exposure to FF3+momentum.

Cumulative Returns



Representative Year

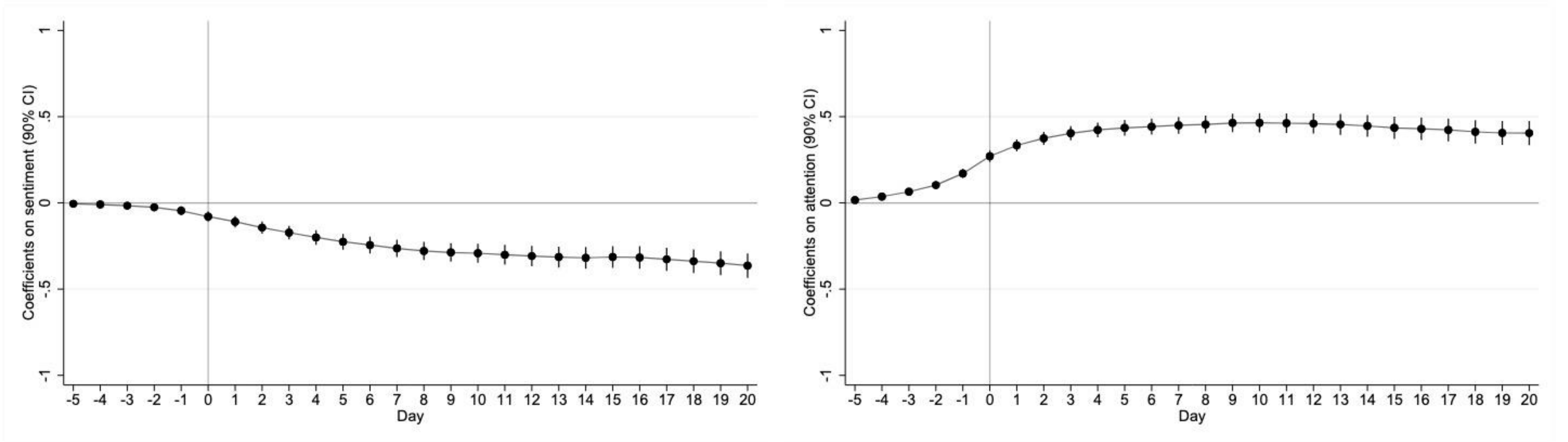


Results on Trading Activity

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?

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)

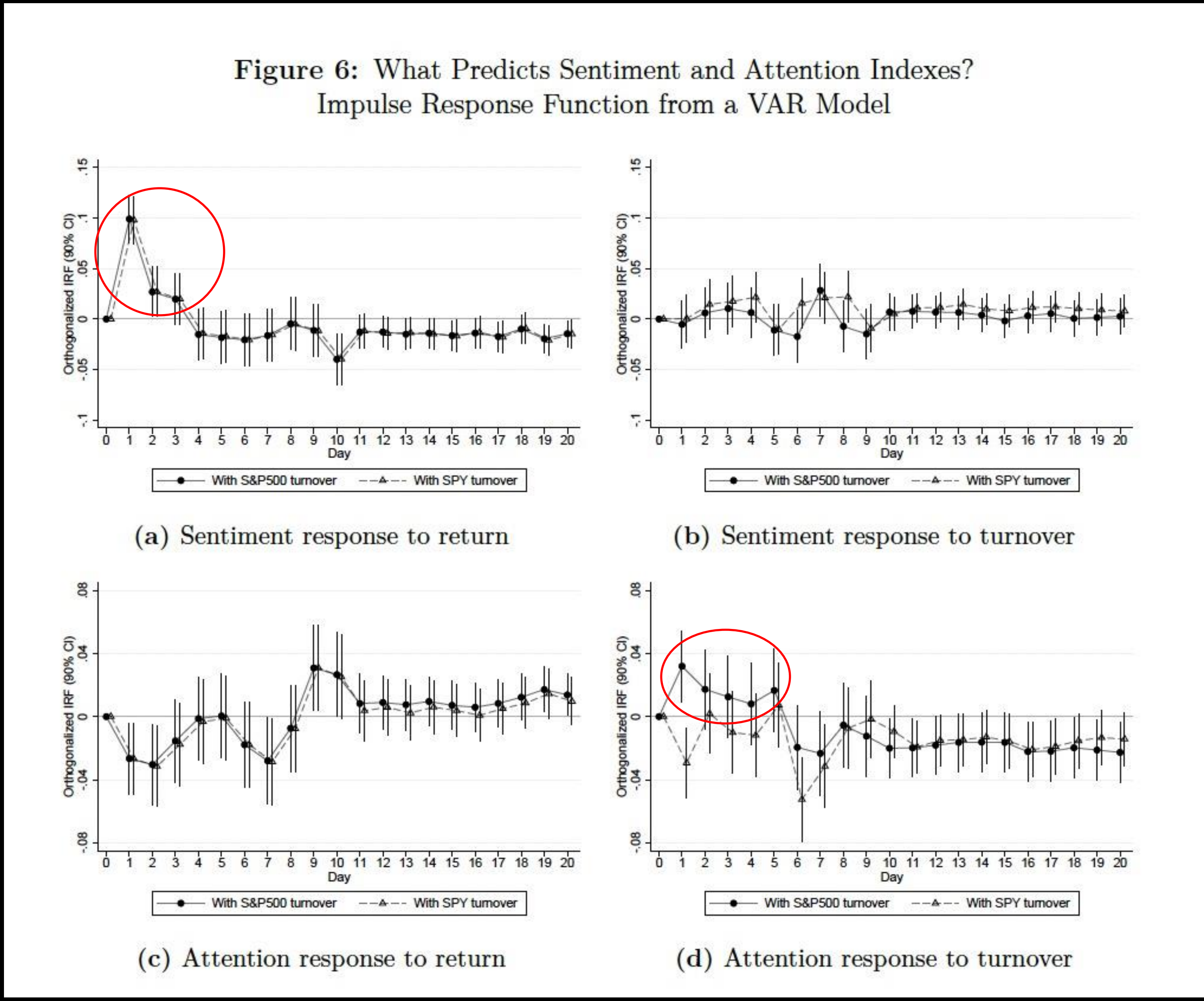


Figure 6: What Predicts Sentiment and Attention Indexes?
Impulse Response Function from a VAR Model

(a) Sentiment response to return

(b) Sentiment response to turnover

(c) Attention response to return

(d) Attention response to turnover

Jumps:

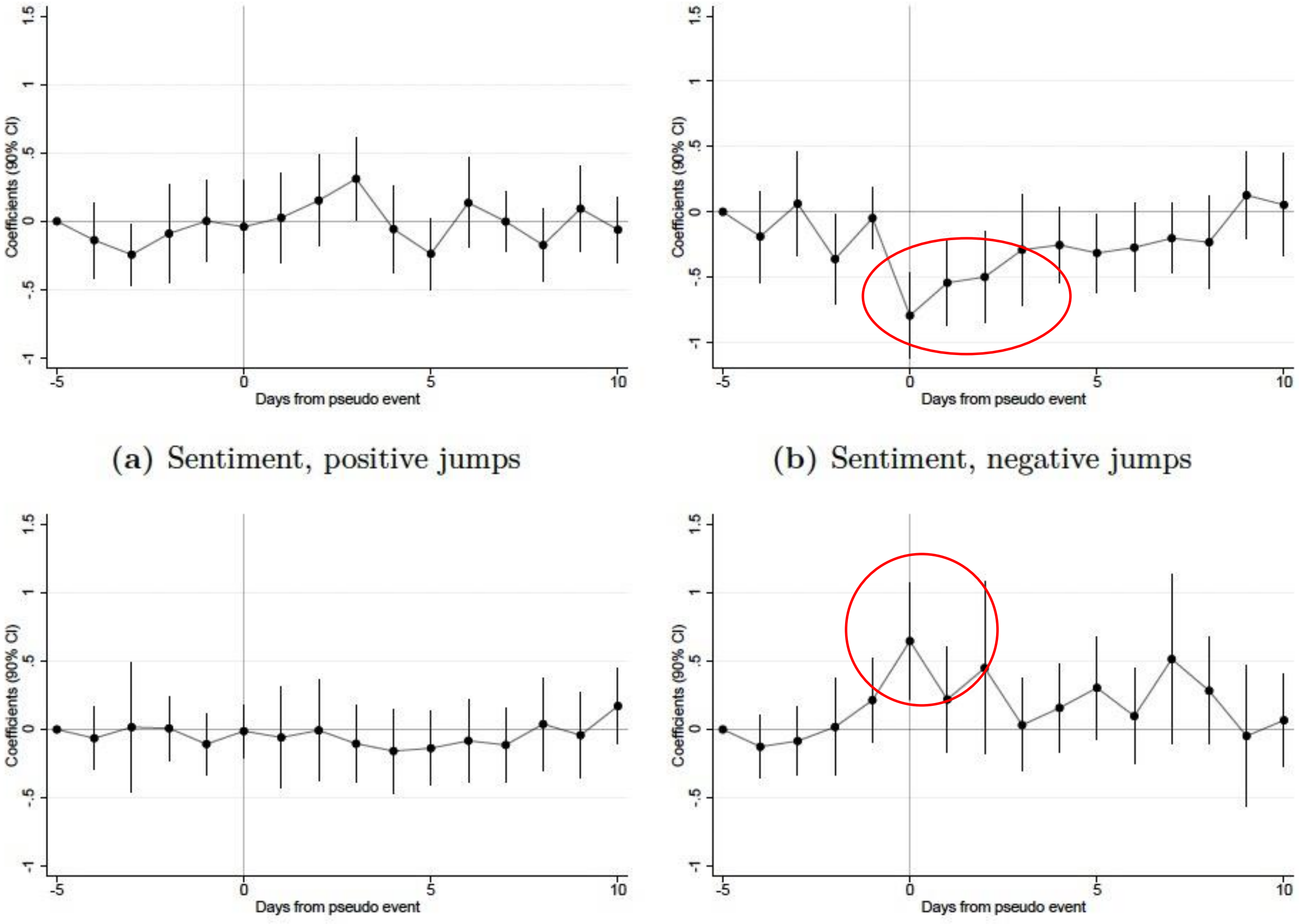
Separate impact of positive returns and negative returns

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	Day t+1	Day t+2	Day t	Day t+1	Day t+2
	(1)	(2)	(3)	(4)	(5)	(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 …

Thanks!

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?