Social media and finance

FIRN Masterclass – Day 1

J. Anthony Cookson
CU Boulder

Social media has become pervasive

People use it more for news and for financial advice.

- Americans 3.6 hours per day on social media (any kind) from Forbes
- 53% of users on Twitter (X) use it for financial advice (Pew 2023)

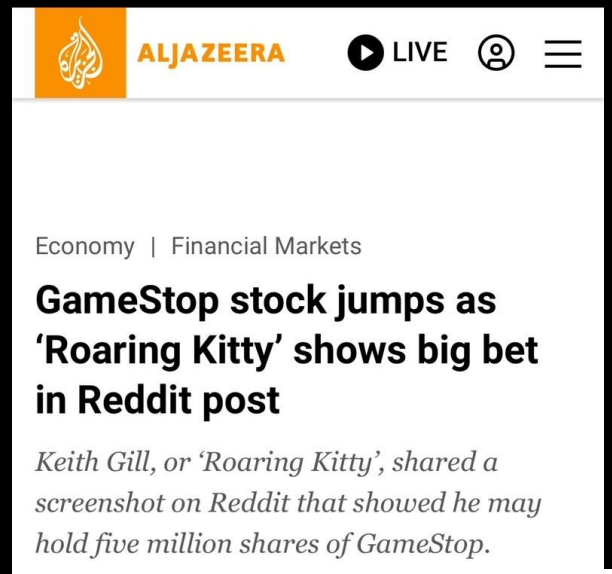
Major events have been linked to social media

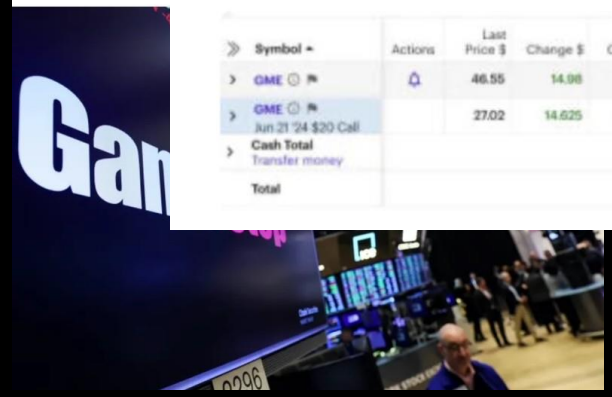
- GameStop short squeeze, highlighting coordination of retail investors (Pedersen 2022)
- Silicon Valley Bank, highlighting speed of communication.

Part of the democratization of finance

An engaging technology, prone to taking advantage of our behavioral biases

In the news (on repeat…)





In context of the GME short squeeze

Research approaches and questions

Social media research in finance is aligned across two major axes:

Axis 1: How social media is used by the researcher:

social media as a lens or social media's effects

Cookson, Mullins and Niessner (2024) draws this distinction, primarily.

Research approaches and questions

Social media research in finance is aligned across two major axes:

Axis 1: How social media is used by the researcher:

social media as a lens or social media's effects

Cookson, Mullins and Niessner (2024) draws this distinction, primarily.

Axis 2: Financial functions of social media (within social media's effects):

production, consumption or distribution of information.

Subset of the approach taken in Aridor et al ("The economics of social media").

My plan

1. Social media landscape (today)
   Definitions, growth over time, data sources
   Some key examples

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

3. Social transmission bias
   Examples and approach, opportunities with new formats and features

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

Social media landscape

It has grown over time, changed format

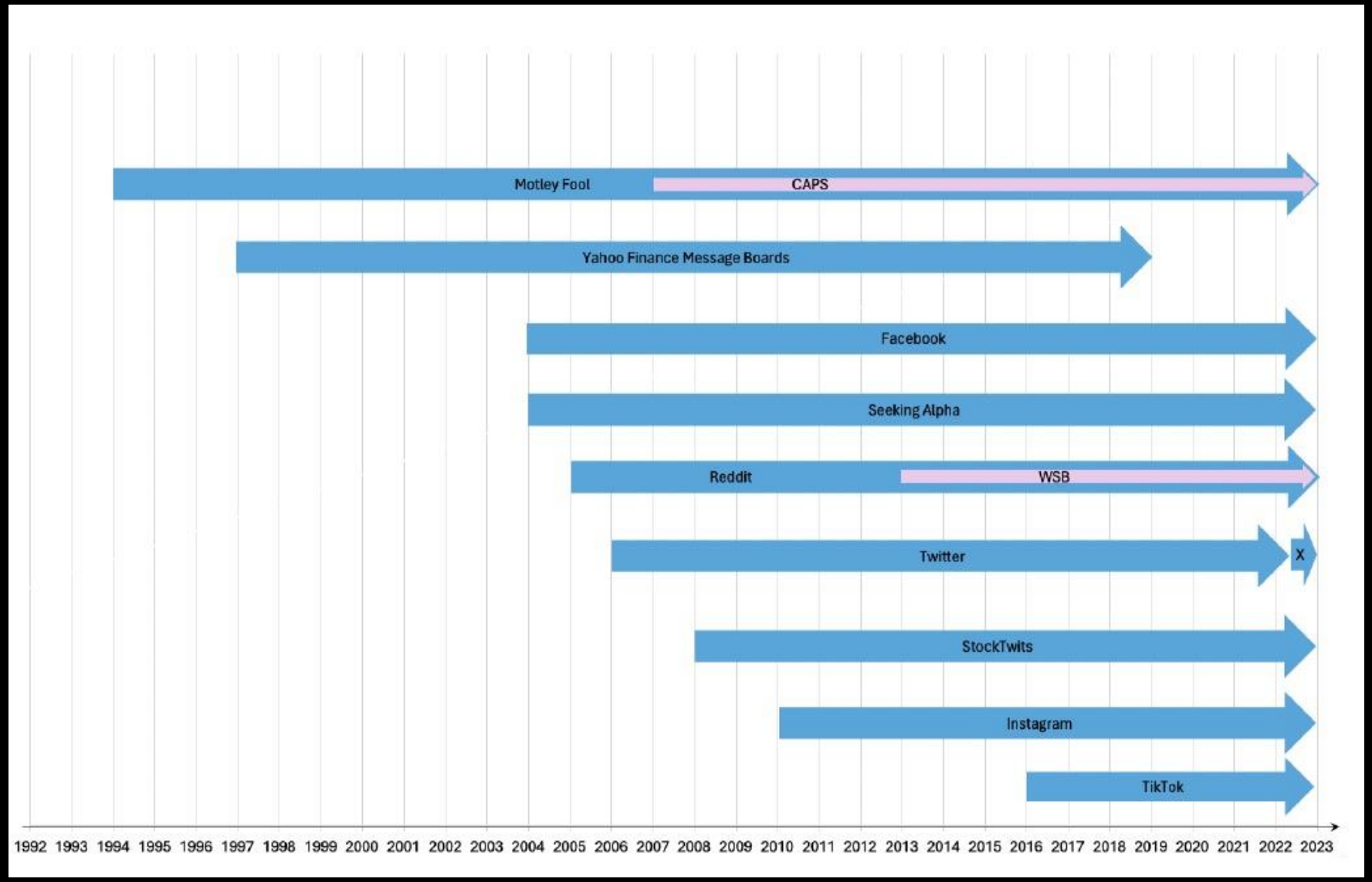
What is social media?

Financial social media evolved out of internet discussion boards (e.g., see Antweiler and Frank 2004, who study internet message boards).

Facebook, MySpace, Twitter (X), Instagram, TikTok, Discord, WhatsApp, Reddit, Bitcointalk, StockTwits,…

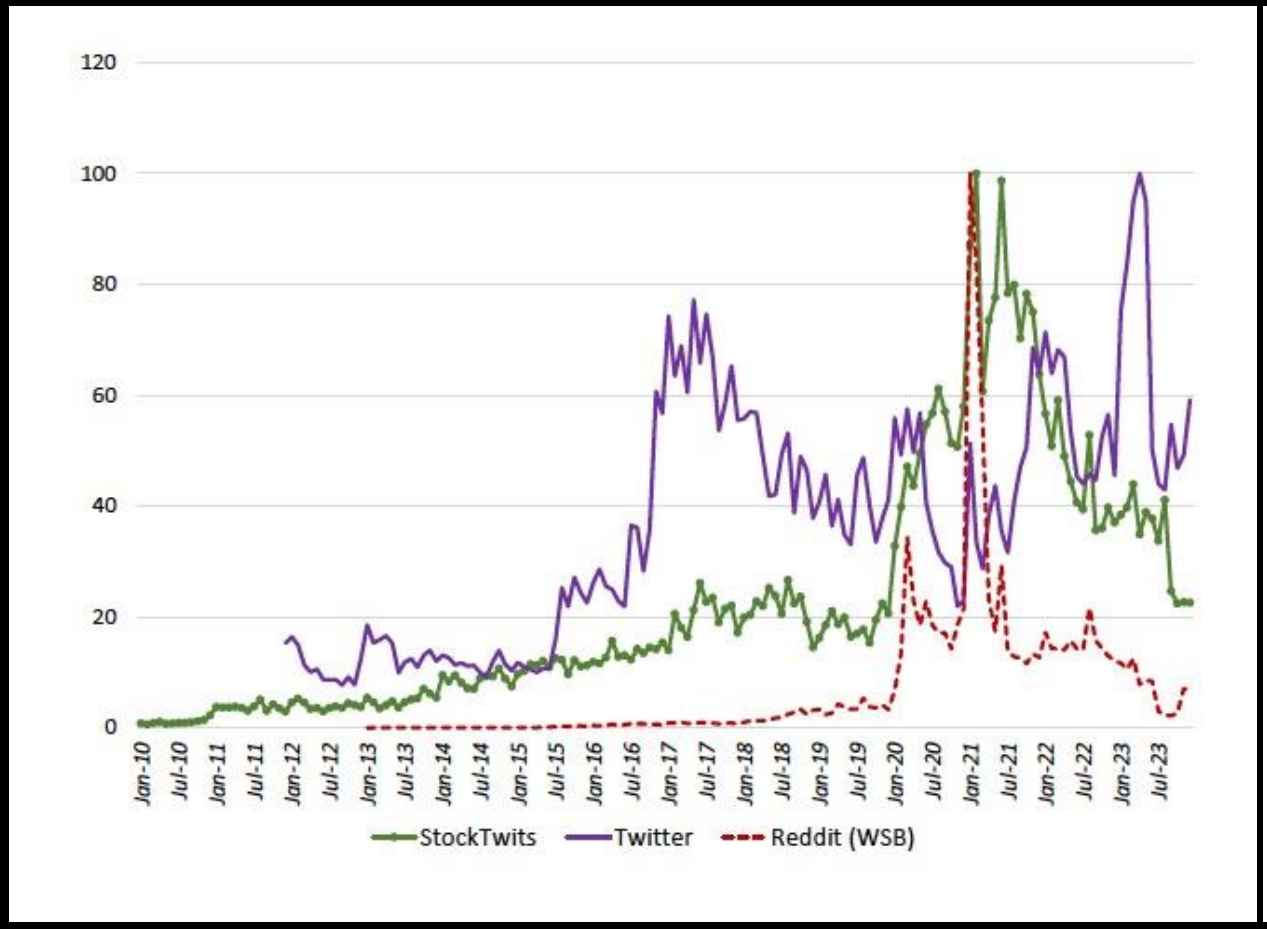
- There's a lot of different media that are called social media (e.g., is Wikipedia "social media"?)
  - Different platform features are understudied: virality, "friends of friends," etc.
- Different social media are not all the same, but also not all different (see "The Social Signal" JFE 2024)

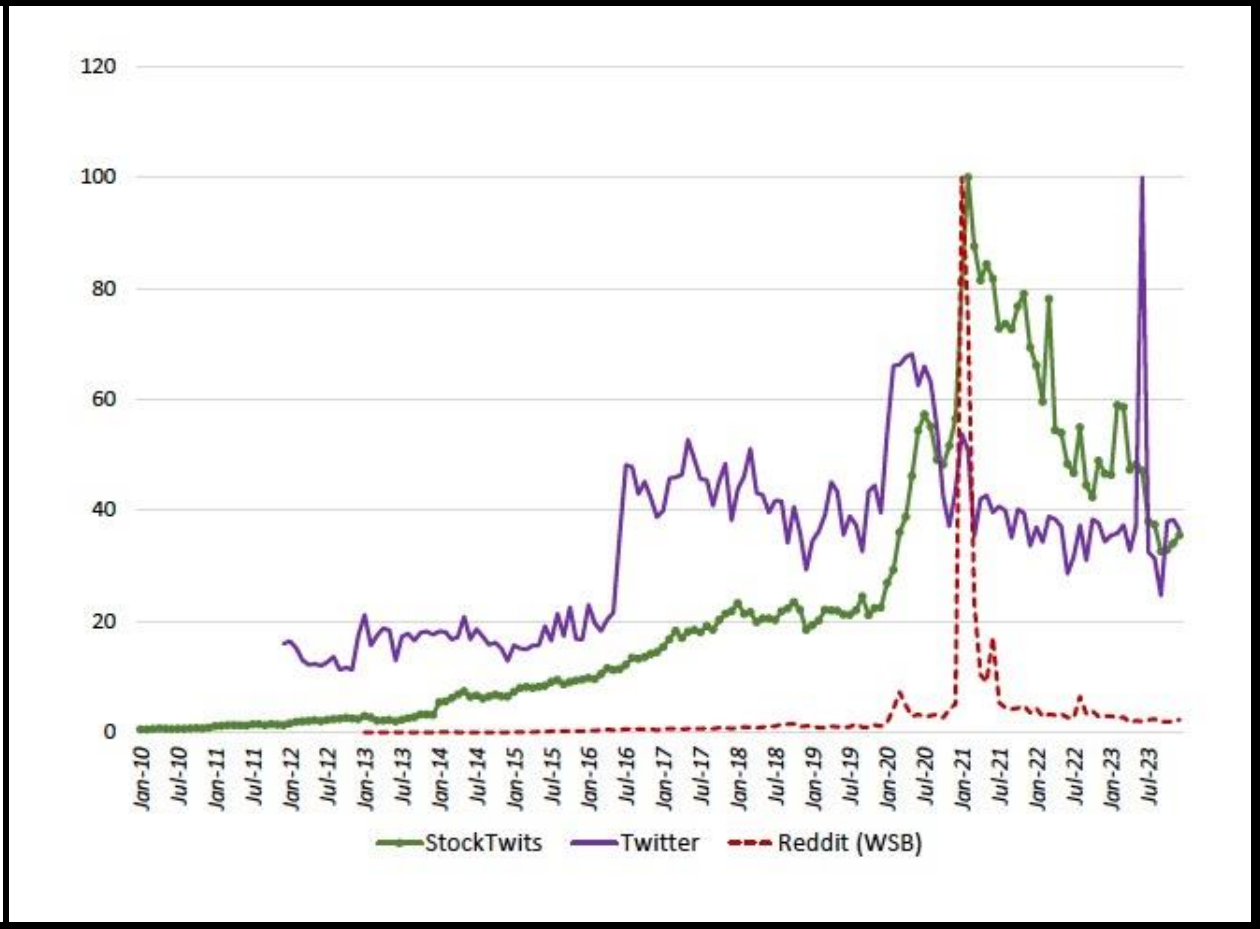
Growth: More platforms with investing content



More posts and users per platform

Finance-specific social media platforms





Social media data

How do I enter this area?

Early work involved collaborations with the (investor) social platform.

- My disagreement work with StockTwits was based on free access to an API because we asked for it.
- Nowadays, platforms are less cooperative (generative AI has made platforms more wary)
  - … but (i) StockTwits has started to sell its data, (ii) scraping, collecting information from online is easier

Social media data

How do I enter this area?

Early work involved collaborations with the (investor) social platform.

- My disagreement work with StockTwits was based on free access to an API because we asked for it.
- Nowadays, platforms are less cooperative (generative AI has made platforms more wary)
  - … but (i) StockTwits has started to sell its data, (ii) scraping, collecting information from online is easier

Commercial products have become more common

- Context Analytics gives away its firm-day Twitter signal
- Refinitiv bought MarketPsych (both firm-day Twitter and StockTwits)
- Ravenpack has sentiment and article information on Seeking Alpha
- Bloomberg has integrated social media tracking products into its platform

Social media data

Most can be found here: http://socialfinance.site/data.jsp?type=ds

Data from websites:

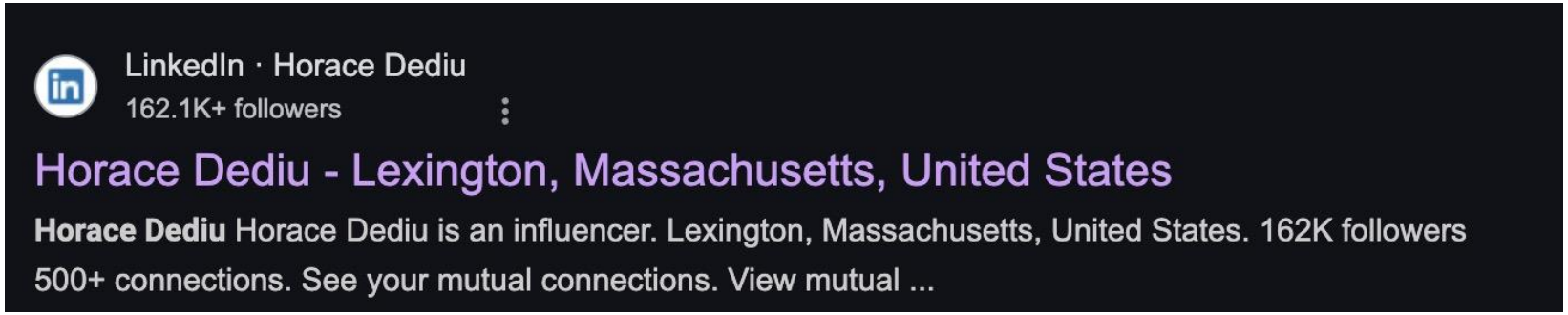
- Disagreement (2010-2021) at the firm-day level (Cookson and Niessner 2023)
- Sentiment and attention (2012-2021) at firm-day level (Cookson, Lu, Mullins and Niessner 2024) from first PC across major platforms.
- Facebook connection data (Stroebel and coauthors), Social capital data (Chetty et al); cross-sectional connectedness metrics.
- Müller and Schwarz have posted their county-level exposure to South-by-Southwest attendance (an instrument for Twitter usage).
- Byoung Hwang has posted firm-day Twitter attention data on his website.

Key examples of social media research

Social media and wisdom of crowds

"*The issue for the pros is that the institution of [financial] analysis risks becoming de-professionalized. In the same way many jobs … became commoditized by the use of new tools or access to information, the era of DIY [do-it-yourself] financial analysis is dawning.*"

-- Horace Dediu, former analyst, now *influencer*



Social media: Seeking Alpha

Paper shows the content of Seeking Alpha predicts returns and earnings surprises.

Not just articles, but comments

JOURNAL ARTICLE

Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media

Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, Byoung-Hyoun Hwang

The Review of Financial Studies, Volume 27, Issue 5, May 2014, Pages 1367–1403,
https://doi.org/10.1093/rfs/hhu001
Published: 04 February 2014

Abstract

Social media has become a popular venue for individuals to share the results of their own analysis on financial securities. This paper investigates the extent to which investor opinions transmitted through social media predict future stock returns and earnings surprises. We conduct textual analysis of articles published on one of the most popular social media platforms for investors in the United States. We also consider the readers' perspective as inferred via commentaries written in response to these articles. We find that the views expressed in both articles and commentaries predict future stock returns and earnings surprises.

Main Contribution: to show social media is informative

Empirical strategy is to count the fraction of **negative** words in Seeking Alpha articles (and comments on those articles until day t+2)

Then, ask whether this fraction predicts **returns t+3 through t+60**, controlling for the usual suspects: *lagged returns, news sentiment, upgrades/downgrades, positive and negative earnings surprises, etc.*

Digging into details: What is a negative word?

Loughran and McDonald (2011) curated lists of positive and negative words, producing a set of dictionaries that proved to be the standard for the next decade+

Also, lists of complexity, modality, etc.
For more, see their data website:
https://sraf.nd.edu/loughranmcdonald-master-dictionary/

Chen et al (2014) uses the negative word list

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

Full Access

When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks

TIM LOUGHRAN, BILL MCDONALD

First published: 06 January 2011 | https://doi.org/10.1111/j.1540-6261.2010.01625.x | Citations: 3,620

CU Full Text

SECTIONS PDF TOOLS SHARE

ABSTRACT

Previous research uses negative word counts to measure the tone of a text. We show that word lists developed for other disciplines misclassify common words in financial text. In a large sample of 10-Ks during 1994 to 2008, almost three-fourths of the words identified as negative by the widely used Harvard Dictionary are words typically not considered negative in financial contexts. We develop an alternative negative word list, along with five other word lists, that better reflect tone in financial text. We link the word lists to 10-K filing returns, trading volume, return volatility, fraud, material weakness, and unexpected earnings.

Digging into details: What is a negative word?

The standard may be different today

Two caveats:

1. **The LM word lists may be noisy**. Garcia, Hu and Rohrer (2023) make this point with more modern ML methods

ELSEVIER

Journal of Financial Economics
Volume 147, Issue 3, March 2023, Pages 525-549

The colour of finance words ☆

Diego García [a], Xiaowen Hu [b], Maximilian Rohrer [c]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.jfineco.2022.11.006 Get rights and content

Full text access

Abstract

Our paper relies on stock price reactions to colour words, in order to provide new dictionaries of positive and negative words in a finance context. We extend the machine learning algorithm of Taddy (2013), adding a cross-validation layer to avoid over-fitting. In head-to-head comparisons, our dictionaries outperform the standard bag-of-words approach (Loughran and McDonald, 2011) when predicting stock price movements out-of-sample. By comparing their composition, word-by-word, our method refines and expands the sentiment dictionaries in the literature. The breadth of our dictionaries and their ability to disambiguate words using bigrams both help to colour finance discourse better.

Positive words				Negative words			
Token	Cov.	% Pos	% Neg	Token	Cov.	% Pos	% Neg
good$^+$	35.1	99.4	0.0	question	25.6	39.2	7.0
strong$^+$	26.0	100.0	0.0	questions	10.5	21.8	6.4
better$^+$	15.1	92.4	0.0	decline$^-$	8.0	0.0	99.8
opportunities	12.9	58.4	4.6	loss$^-$	6.8	0.0	99.0
able	12.1	63.2	2.2	negative$^-$	4.4	0.2	96.6
opportunity	11.9	68.0	3.8	difficult	3.7	0.0	78.4
positive	10.2	62.6	2.6	against	3.6	7.8	27.4
improvement$^+$	10.0	100.0	0.0	declined$^-$	3.5	0.2	91.4
progress	7.9	56.4	5.0	restructuring	3.2	30.8	30.4
pleased$^+$	7.7	99.8	0.0	losses	2.8	6.0	69.0
improved$^+$	6.9	100.0	0.0	challenges$^-$	2.6	0.0	99.8
improve	6.7	11.0	34.0	challenging$^-$	2.4	0.2	87.0
best	6.5	25.6	10.4	recall	1.8	8.2	25.6
strength$^+$	4.8	100.0	0.0	declines$^-$	1.8	0.0	85.8
success$^+$	4.4	88.8	0.0	volatility	1.7	6.8	42.4
excited	4.4	49.8	4.6	slow	1.6	0.2	66.4
profitability	4.3	63.0	4.8	break	1.5	22.6	6.6
confident$^-$	3.9	0.4	80.4	weakness$^-$	1.4	0.0	99.8
improving$^+$	3.8	82.4	0.0	bad	1.3	6.0	44.4
favorable$^+$	3.6	86.4	0.0	challenge	1.3	0.2	77.4
improvements$^+$	3.5	89.4	0.2	problem	1.3	1.6	71.4
gain	3.4	64.0	1.2	weak	1.2	0.2	78.8
despite	3.3	3.8	33.6	claims	1.2	12.0	61.8
successful	3.2	41.2	2.4	slower$^-$	1.2	0.0	93.0
gains$^+$	3.2	82.4	0.0	negatively$^-$	1.2	0.0	96.8
stronger	3.2	72.0	0.2	lost$^-$	1.2	0.0	96.8
efficiency	3.1	68.6	1.6	cut	1.1	3.4	50.2
advantage	3.0	61.0	1.4	slowdown$^-$	1.1	0.0	96.8
achieve	3.0	32.0	6.0	impairment	1.1	1.2	81.0
innovation	2.8	57.2	6.4	missed	1.0	0.6	49.2

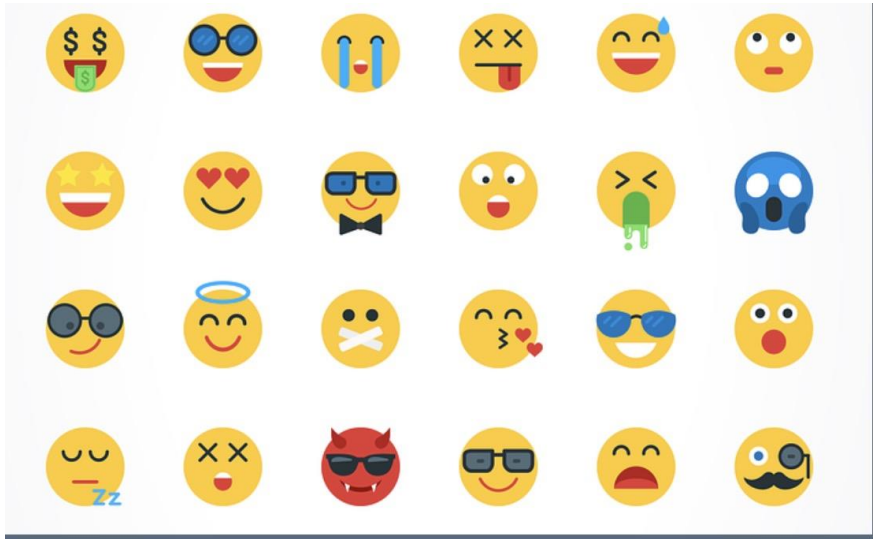
Digging into details: What is a negative word?

The standard may be different today

Two caveats:

1. **The LM word lists may be noisy**. Garcia, Hu and Rohrer (2023) make this point with more modern ML methods

2. **Social media text (outside of Seeking Alpha) is short and "has its own language."** Fox, Kelley and Paolucci (2025WP) make this point by analyzing the sentiment of emojis and extending VADER (*a social media sentiment classifier*)



How Many Words is a Picture Worth? Using Emojis From Social media to Predict Future Stock Returns

w/ Eric Kelley and Roman Paolucci

Using a new and comprehensive sample of more than 67 million Twitter posts referencing Russell 3000 firms between 2012 and 2021, we introduce a novel, unsupervised method of scoring the sentiment of emojis. Our method generates point-in-time dictionaries that map individual emojis to the contextual sentiment of recent tweets that contain them. In out-of-sample tests, we find that even controlling for the sentiment extracted from words, emoji sentiment correctly predicts future firm-level stock returns. Understanding the sentiment of emojis has become increasingly important as individuals continue to adopt these new forms of communication.

Returning to Chen et al (2014)…
How to show social media is informative

We organize our main analysis around the following regression specification:

$$ARet_{i,t+3,t+60} = \alpha + \beta_1 NegSA_{i,t} + \beta_2 NegSA\text{-}Comment_{i,t} + X\delta + \varepsilon_{i,t}. \quad (1)$$

Simple empirical strategy: Are future returns predictable by the content of Seeking Alpha, conditional on controls?

Main result

Table 4
Seeking Alpha and abnormal returns

	(1)	(2)	(3)
$NegSA_{i,t}$	−0.379	−0.332	−0.320
	(−2.24)	(−2.03)	(−1.98)
$NegSA\text{-}Comment_{i,t}$		−0.194	−0.196
		(−3.44)	(−3.55)
$I(SA\text{-}Comment_{i,t})$		0.001	0.001
		(0.25)	(0.17)

Returning to Chen et al (2014)…
How to show social media is informative

Consistent evidence for different horizons and calendar time strategies

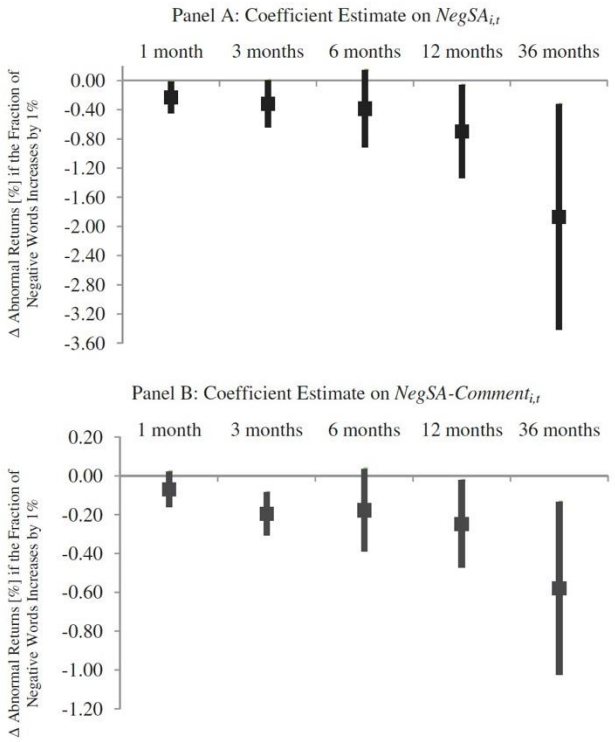
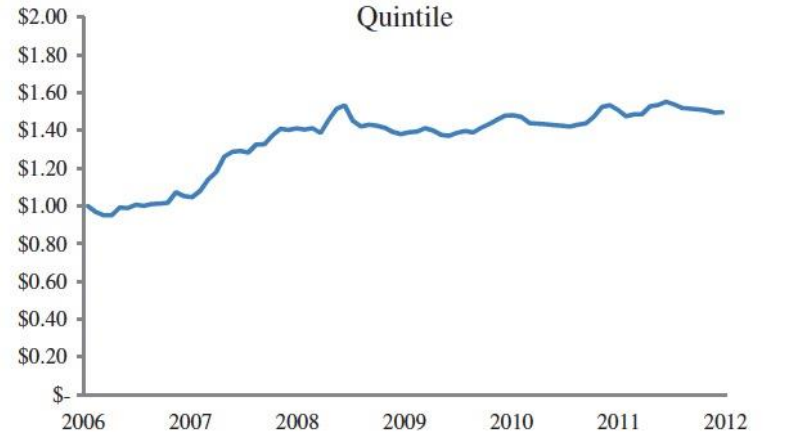
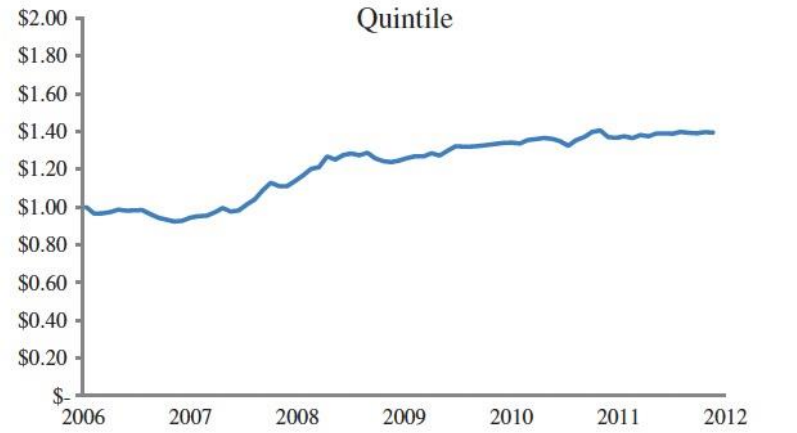


Figure 3
Seeking Alpha and abnormal returns over different holding periods



Panel B: *NegSA-Co*



Social media: Seeking Alpha

Taking stock: It is a big finding and noteworthy that SA content predicts returns.

Results are a little weak, but so were the methods.

Predictions have held up after this sample (e.g., see Russell Jame's research) and on different platforms.

JOURNAL ARTICLE

Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media

Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, Byoung-Hyoun Hwang

The Review of Financial Studies, Volume 27, Issue 5, May 2014, Pages 1367–1403,
https://doi.org/10.1093/rfs/hhu001
Published: 04 February 2014

Abstract

Social media has become a popular venue for individuals to share the results of their own analysis on financial securities. This paper investigates the extent to which investor opinions transmitted through social media predict future stock returns and earnings surprises. We conduct textual analysis of articles published on one of the most popular social media platforms for investors in the United States. We also consider the readers' perspective as inferred via commentaries written in response to these articles. We find that the views expressed in both articles and commentaries predict future stock returns and earnings surprises.

Exemplar #2.
Economic Effects of Social Networks

Key research question: Do our social networks (not necessarily online) influence real economic decisions?

The Economic Effects of Social Networks: Evidence from the Housing Market

Michael Bailey

Facebook

Ruiqing Cao

Harvard University

Theresa Kuchler

New York University

Johannes Stroebel

New York University, National Bureau of Economic Research, and Centre for Economic Policy Research

We show how data from online social networking services can help researchers better understand the effects of social interactions on economic decision making. We combine anonymized data from Facebook, the largest online social network, with housing transaction data and explore both the structure and the effects of social networks. Individuals whose geographically distant friends experienced larger recent house price increases are more likely to transition from renting to owning. They also buy larger houses and pay more for a given house. Survey data show that these relationships are driven by the effects of social interactions on individuals' housing market expectations.

Revealing Social Connections in Social Media

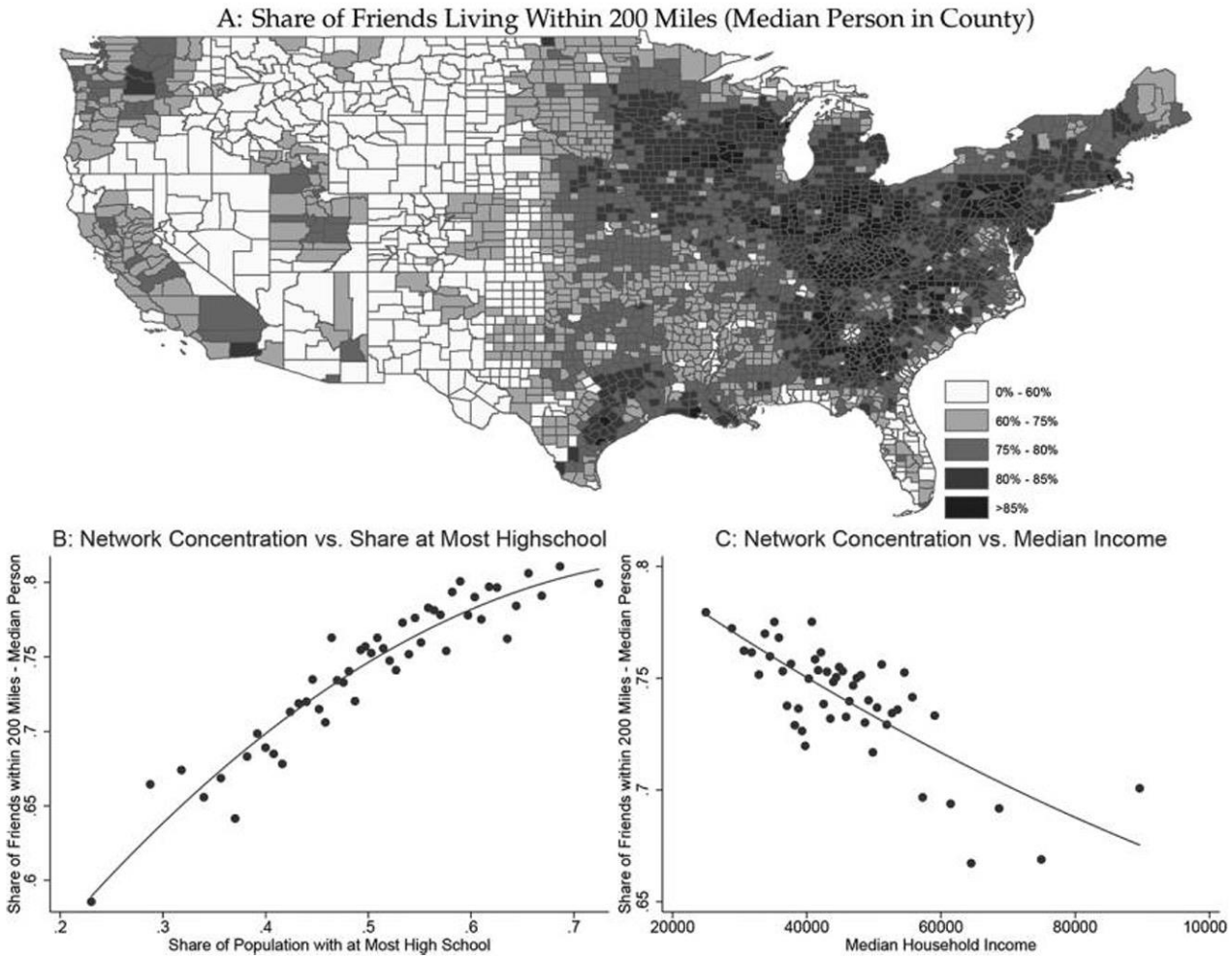
Bailey et al (2018 JPE)

- Worked with Facebook internal data, merged with all real estate transactions in Los Angeles.

- Trick is to aggregate the housing price experiences of Facebook friends, at the individual level. Does this predict house prices?
  - Yes, having friends who experienced more house price appreciation ➔ higher prices in LA market.

- Facebook snapshot in 2016, but effects in earlier years (before Facebook).

This matters because most connections are local



How to identify an effect of peer experiences on housing transactions?

$$\text{HousingInvestment}_{i,t_2} = \beta \text{FriendHPExp}^{N}_{i,t_1,t_2} + \gamma \mathbf{X}_{i,t_2} + \epsilon_{i,t_2}. \qquad (1)$$

Note: Friend exposure comes from where my friends are versus where yours are (not where I am versus you; we both live in LA in this sample; solution use out-of-state friend experiences).

Friend exposure comes from **Facebook** observed in 2016.

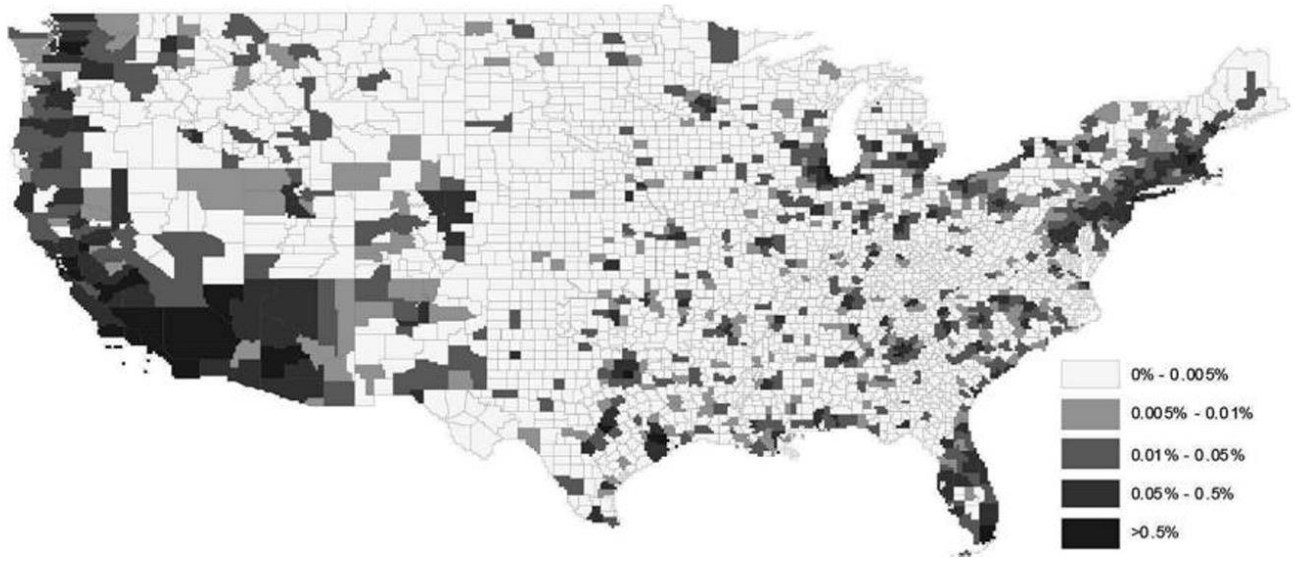


FIG. 3.—Share of friendship links of Los Angeles residents. The figure shows the abso-

More likely to buy/own a house if friends saw HP appreciation

TABLE 6
EFFECTS ON PROBABILITY OF HOME OWNERSHIP

	DEPENDENT VARIABLE: Pr(Owner in 2012)				
	(1)	(2)	(3)	(4)	(5)
	A. 2010 Renters				
Δ friend house prices, 2008–10 (%)	.608***	.511***	.501***	.544***	.672***
	(.042)	(.044)	(.169)	(.043)	(.043)
Δ friend county income, 2008–10 (%)				.332***	
				(.033)	
Δ friend house prices, 2010–12 (%)					.324***
					(.044)
Zip 2010 × zip 2012 fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Sample restriction		Stayed in same zip code	Geographically nonclustered professions		
Observations	433,836	302,686	433,836	433,836	433,836
R^2	.43	.13	.43	.43	.43
Mean dependent variable	17.8	10.3	17.8	17.8	17.8

House is likely **bigger** if friends saw HP appreciation

TABLE 8
EFFECTS ON SIZE OF PROPERTY PURCHASED

	DEPENDENT VARIABLE: 100 × Log(Property Size)					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ friend house prices past 24 months (%)	.310***	.520***	.530***	.400***	.234***	.285***
	(.053)	(.144)	(.164)	(.060)	(.079)	(.056)
Δ friend county income past 24 months (%)						.173***
						(.096)
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Buyer controls	Yes	Yes	Yes	Yes, ×year	Yes	Yes
Sample notes		Purchases since 2010	Purchases since 2010, lived in LA in 2010		Geographically nonclustered professions	
Observations	526,594	95,561	68,388	526,593	526,594	526,594
R^2	.194	.134	.126	.204	.194	.194

More likely to **pay more** (within property FE!)

TABLE 9
EFFECTS ON TRANSACTION PRICE

	DEPENDENT VARIABLE: 100 × Log(Price)				
	(1)	(2)	(3)	(4)	(5)
	A. Main Results				
Δ friend house prices, buyer—past 24 months (%)	.452***	.486***	.408***	.445***	.335***
	(.015)	(.050)	(.076)	(.015)	(.068)
Δ friend house prices, seller—past 24 months (%)				.233***	.289**
				(.059)	(.112)
Year × zip code fixed effects, controls	Yes	Yes	Yes	Yes	Yes
Sample or specification notes		Property FE	Buyer FE		Property FE
Observations	523,299	34,732	32,226	523,299	33,230
R^2	.808	.950	.948	.809	.956

Bailey et al JPE paper had individual data, but county aggregates data have enjoyed wider application

Social Connectedness: Measurement, Determinants, and Effects

Michael Bailey

Rachel Cao

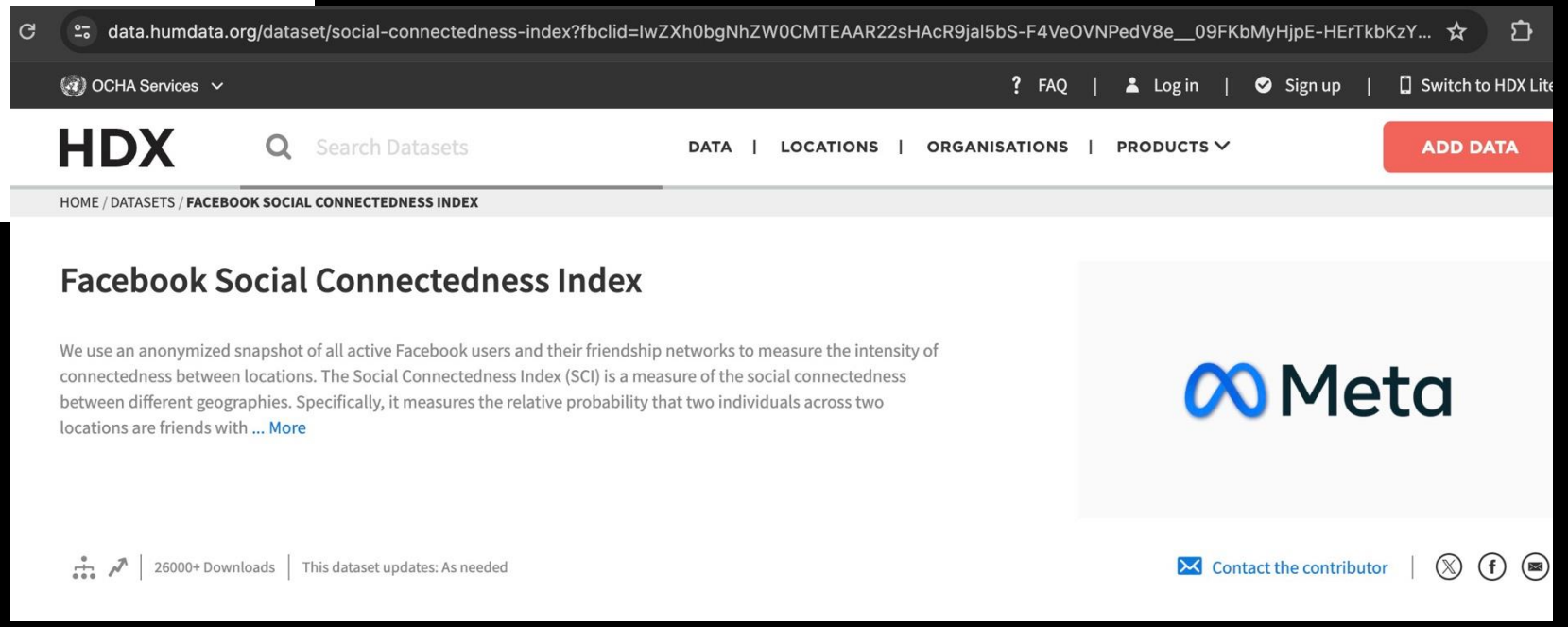
Theresa Kuchler

Johannes Stroebel

Arlene Wong

JOURNAL OF ECONOMIC PERSPECTIVES
VOL. 32, NO. 3, SUMMER 2018
(pp. 259-80)

Download Full Text PDF
(Complimentary)



More recent: Chetty et al “Social capital”

nature

Explore content ▾ About the journal ▾ Publish with us ▾

nature › articles › article

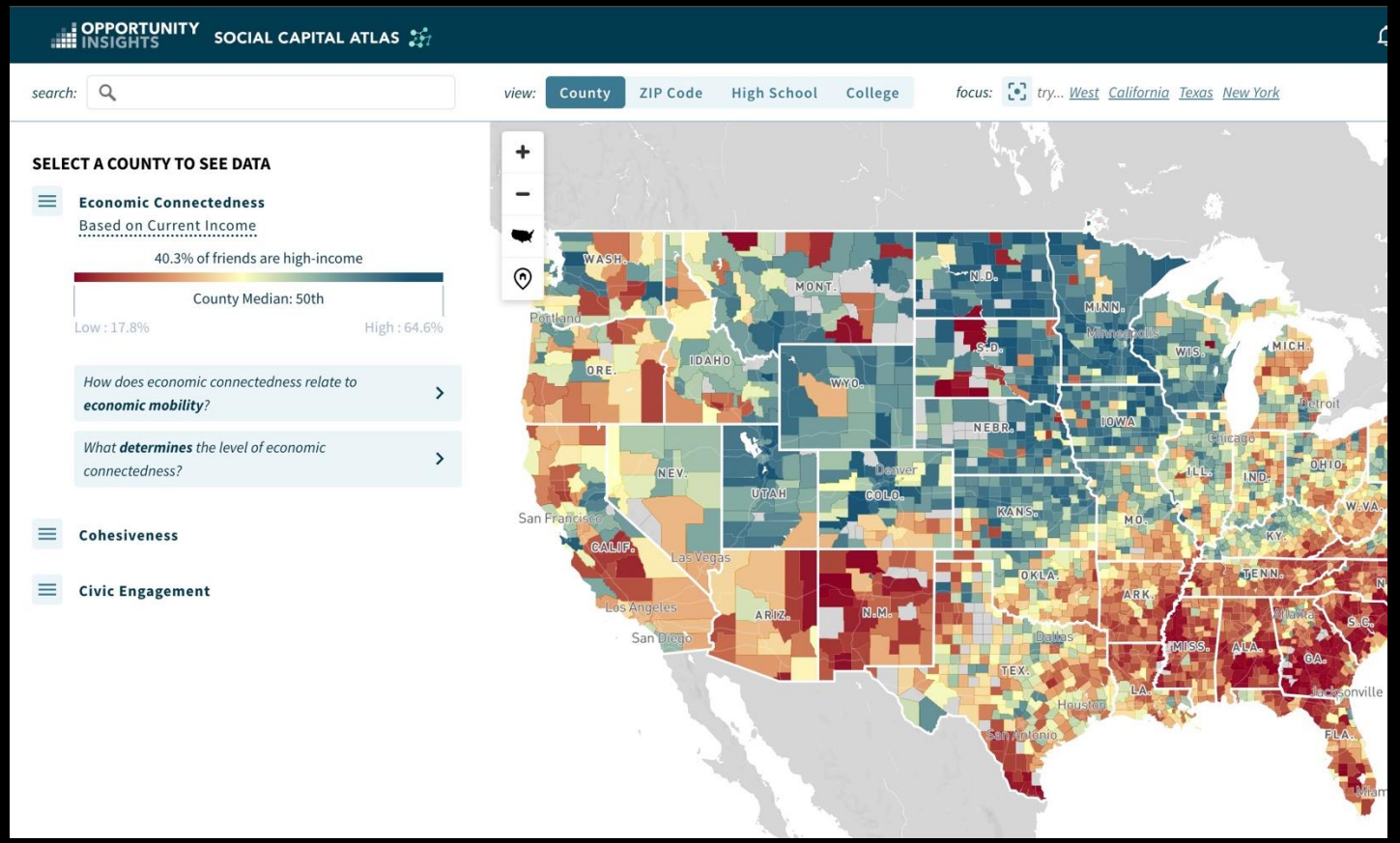
Article | Open access | Published: 01 August 2022

Social capital I: measurement and associations with economic mobility

Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole & Nils Wernerfelt

Nature **608**, 108–121 (2022) | Cite this article

189k Accesses | **177** Citations | **2318** Altmetric | Metrics



Measures in-person connections between high- and low-income people.

Easy to download and use.

The Facebook connection data are not specific to finance

Friends with Benefits: Social Capital and Household Financial Behavior

Brad Cannon, David Hirshleifer & Joshua Thornton

WORKING PAPER 32186 **DOI** 10.3386/w32186 **ISSUE DATE** March 2024

Though there are finance applications of these broader connections (e.g., see Cannon et al 2024), much of the finance literature is focused on finance-specific social networks.

Using Social Capital Data

- An example from a 2024 MFA discussion of Cannon et al (WP).
  - Note: *paper may have updated since then*.

Friends with Benefits: Social Capital and Household Financial Behavior

Brad Cannon, David Hirshleifer & Joshua Thornton

WORKING PAPER 32186 DOI 10.3386/w32186 ISSUE DATE March 2024

Friends with Benefits

Social Capital and Household Financial Behavior

Cannon, Hirshleifer and Thornton

Discussion by J. Anthony Cookson

Where does this paper fit?

- Social capital literature is by now “age old”
  - Knack and Keefer (1997), Keefer and Knack (1995)
    - Wallet drops across countries, world value surveys ==> Trust ~ growth
  - Link to financial outcomes in La Porta et al (1997) & Guiso Sapienza and Zingales (2004,2006).
    - Measures: electoral turnout or blood donation.

- More modern (tends to be narrow and focus on identification):
  - Brown, Cookson and Heimer (2019) – exposure to finance (via trust) promotes participation using “natural experiment” in exposure to finance.

  - Dupont (2022) – Clergy abuse shocks affect stock market participation via generic trust channel.

This paper's questions/measurement

Does **economic connectedness** (EC) affect **market participation** and **savings**?

- EC measured as *friendships on Facebook with SES people* (county-level).

[Social Capital Data from Opportunity Insights]

- Participation and savings measured as *% of tax returns with dividends* and *% of tax returns with interest* (county-level).

[IRS Summary of Income (SOI) data]

This paper's findings

Does economic connectedness (EC) **affect** market participation and savings?

- At county level, EC has **a strong positive correlation** with % of tax returns with dividends and (separately) interest.
- This tendency is stronger than for **cohesiveness** (are my friends likely friends with one another?) and **rates of volunteering** (traditional trust metric)
- **Decomposition**: Exposure relates more strongly than Friending Bias.
- **Identification?** Similar results based on childhood connections & effects seem to reverberate through network via non-local connections.

My take and discussion

Take

- Creative mixing of new Opportunity Insights data with public IRS data.
- Important question to think about components of social capital.
- Could use more detail, still early work.

Discussion

- Interpretation and Measurement of EC versus in Chetty et al (2023)
  - [Wow! I can download the data…]
- Some alternative specification suggestions

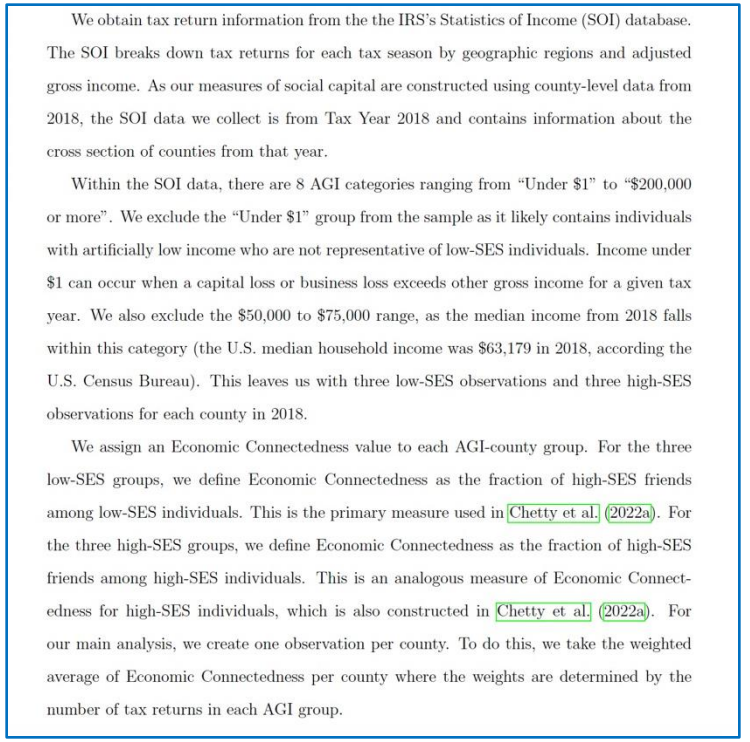
Comment #1

Measurement of Economic Connectedness

Chetty et al (2023) paper builds two EC measures

- EC – for low SES people, ~ % friends who are high SES.
- EC_high - for high SES people, ~ % friends who are high SES.

- This paper:
  - EconCon = a1*EC + a2*EC_high, where:
    - a1 = fraction of tax returns in low-income buckets.
    - a2 = fraction of tax returns in high-income buckets.



We obtain tax return information from the the IRS's Statistics of Income (SOI) database. The SOI breaks down tax returns for each tax season by geographic regions and adjusted gross income. As our measures of social capital are constructed using county-level data from 2018, the SOI data we collect is from Tax Year 2018 and contains information about the cross section of counties from that year.

Within the SOI data, there are 8 AGI categories ranging from "Under $1" to "$200,000 or more". We exclude the "Under $1" group from the sample as it likely contains individuals with artificially low income who are not representative of low-SES individuals. Income under $1 can occur when a capital loss or business loss exceeds other gross income for a given tax year. We also exclude the $50,000 to $75,000 range, as the median income from 2018 falls within this category (the U.S. median household income was $63,179 in 2018, according the U.S. Census Bureau). This leaves us with three low-SES observations and three high-SES observations for each county in 2018.

We assign an Economic Connectedness value to each AGI-county group. For the three low-SES groups, we define Economic Connectedness as the fraction of high-SES friends among low-SES individuals. This is the primary measure used in Chetty et al. (2022a). For the three high-SES groups, we define Economic Connectedness as the fraction of high-SES friends among high-SES individuals. This is an analogous measure of Economic Connectedness for high-SES individuals, which is also constructed in Chetty et al. (2022a). For our main analysis, we create one observation per county. To do this, we take the weighted average of Economic Connectedness per county where the weights are determined by the number of tax returns in each AGI group.

Main Table

	(1) P(Div)	(2) P(Div)	(3) P(Div)	(4) P(Div)	(5) P(Div)	(6) P(Div)	(7) P(Div)	(8) P(Div)
Economic Connectedness	0.751***	0.430***					0.759***	0.442***
	(60.21)	(14.56)					(45.86)	(15.46)
Cohesiveness			-0.224***	0.145***			0.056**	0.186***
			(-8.25)	(4.34)			(2.46)	(5.31)
Civic Engagement					0.296***	0.017	0.037**	0.012
					(14.93)	(1.08)	(2.57)	(0.84)
Controls		YES		YES		YES		YES
Observations	3017	3015	3088	3086	3088	3086	3017	3015
Adj. R^2	0.563	0.676	0.050	0.626	0.087	0.617	0.567	0.690

Standardized beta coefficients; t statistics in parentheses

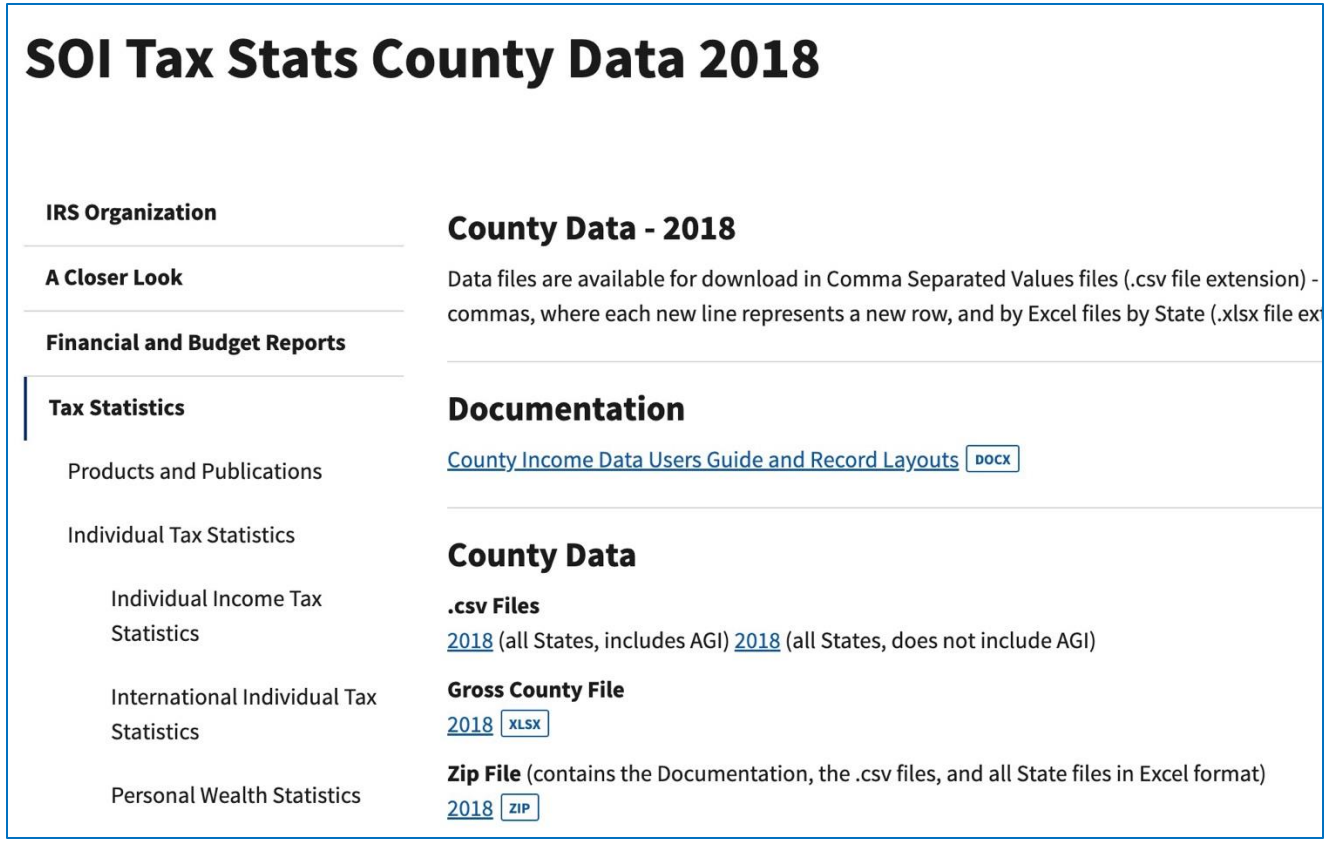
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Questions I had:

- Is the result sensitive to how EconCon is averaged across buckets?
- How much comes from low SES versus high SES buckets?
- Couldn't we just control for income using the IRS data directly?

Then, I realized… *I can download the data myself!*

The IRS SOI data are public at the county and zip code level by year!

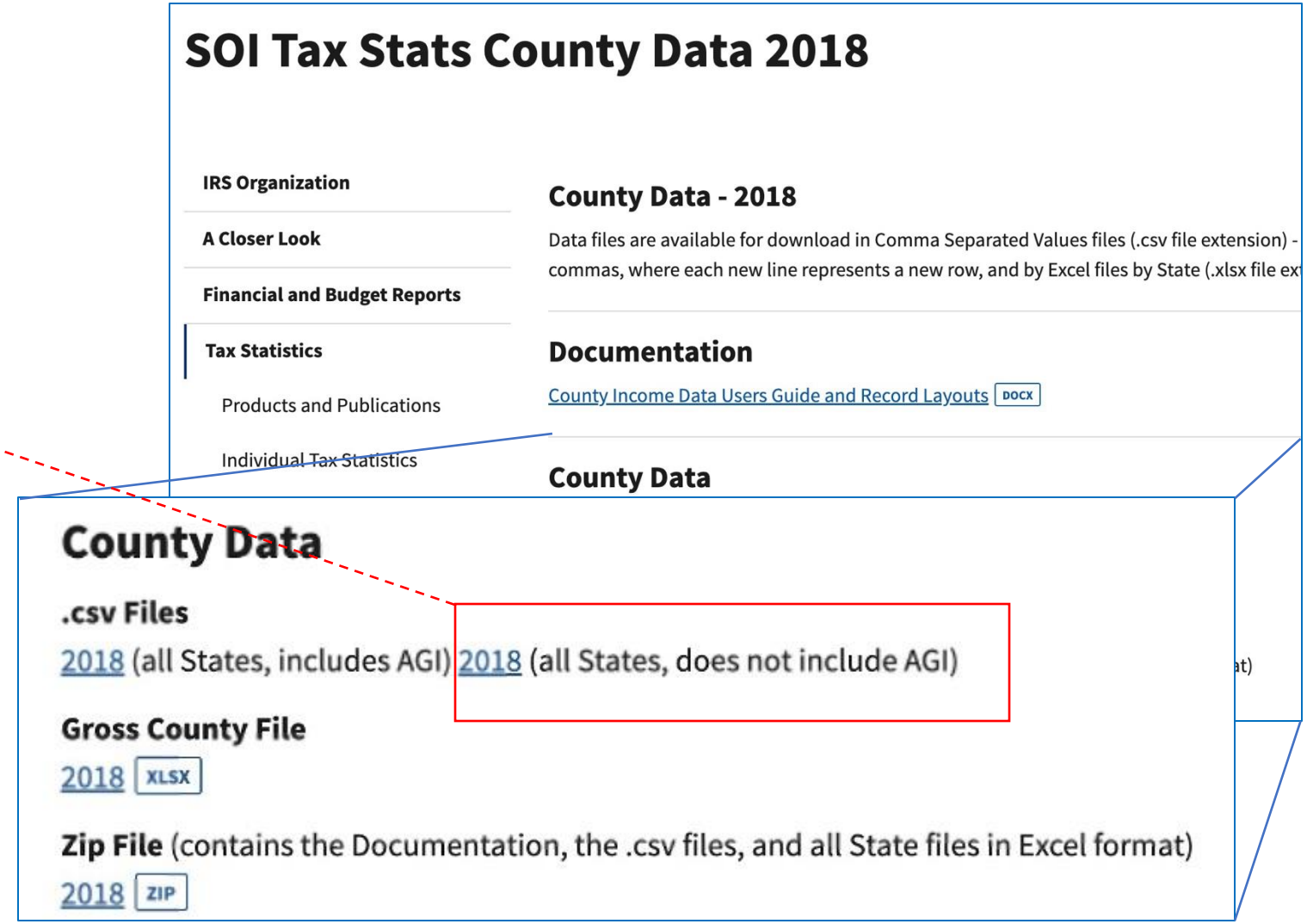


Then, I realized… *I can download the data myself!*

Observation 1:

- It is not *necessary* to work with county-income bucket data.

So, I downloaded county-level (and zip code level data) w/o AGI buckets.

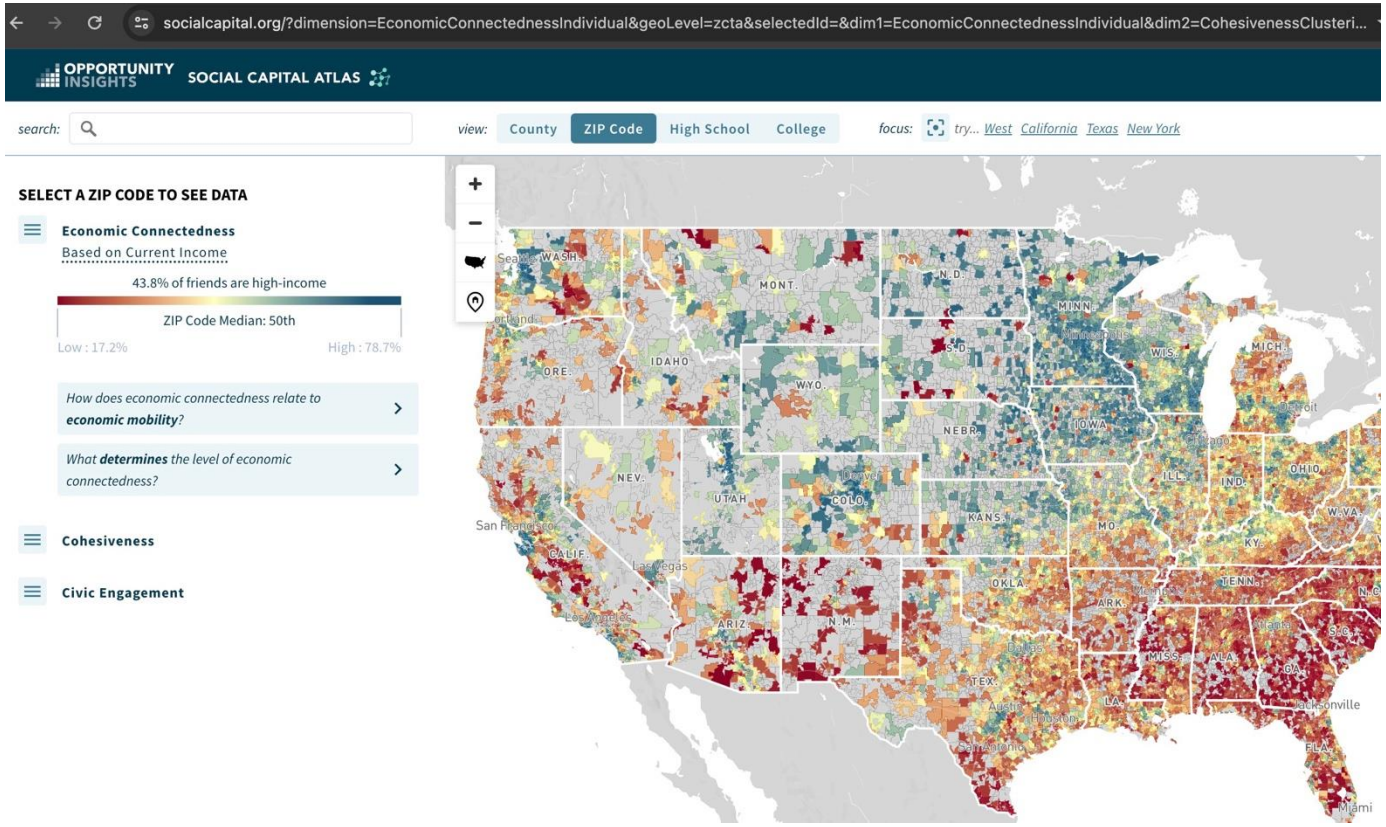


Then, I realized… *I can download the data myself!*

Observation 2:

- The Chetty data are available at county, zip, high school and college level

So, I downloaded county-level and zip code level data.



My “replication”

Differences:
EconCon = 0.5EC+0.5EC_high

I controlled for log(avg_AGI) in the IRS data and population, not the full slate of controls

	Dependent variable:					
	scale(100 * perc_div)			scale(100 * perc_int)		
	(1)	(2)	(3)	(4)	(5)	(6)
scale(ec)	0.733***	0.720***	0.469***	0.723***	0.746***	0.629***
	(0.012)	(0.015)	(0.025)	(0.012)	(0.017)	(0.021)
scale(support_ratio_county)		0.008	0.105***		0.172***	0.165***
		(0.015)	(0.018)		(0.031)	(0.019)
scale(volunteering_rate_county)		0.044***	0.081***		0.096***	0.080***
		(0.016)	(0.016)		(0.016)	(0.016)
log(pop2018)			-0.064***			-0.137***
			(0.013)			(0.013)
log(pc_income_w)			1.747***			0.838***
			(0.126)			(0.103)
Constant	-0.006	-0.005	-6.352***	-0.008	-0.008	-1.944***
	(0.012)	(0.012)	(0.443)	(0.012)	(0.012)	(0.357)
Observations	3,018	3,018	2,942	3,018	3,018	2,942
R2	0.562	0.563	0.638	0.535	0.572	0.599
Adjusted R2	0.562	0.563	0.638	0.535	0.572	0.598
Residual Std. Error	0.647 (df = 3016)	0.646 (df = 3014)	0.582 (df = 2936)	0.675 (df = 3016)	0.647 (df = 3014)	0.617 (df = 2936)
Note:						*p<0.1; **p<0.05; ***p<0.01

My “replication”

Conclusions:

1. How EC is averaged does not matter.

2. Main result seems robust to choice/depth of controls.

	Dependent				
	(1)	scale(100 * perc_div)			(8)
	P(Div)	(1)	(2)	(3)	P(Div)
scale(ec)	0.751***	0.733***	0.720***	0.469***	0.442***
	(60.21)	(0.012)	(0.015)	(0.025)	(15.46)
scale(support_ratio_c			0.008	0.105***	
			(0.015)	(0.018)	
scale(volunteering_ra			0.044***	0.081***	0.186***
			(0.016)	(0.016)	(5.31)
log(pop2018)				-0.064***	
				(0.013)	0.012
log(pc_income_w)				1.747***	(0.84)
				(0.126)	
Constant		-0.006	-0.005	-6.352***	YES
		(0.012)	(0.012)	(0.443)	
Observations	3017	3,018	3,018	2,942	3015
R2	0.563	0.562	0.563	0.638	0.690
Adjusted R2		0.562	0.563	0.638	
Residual Std. Error		[illegible]47 (df = 3016)	0.646 (df = 3014)	0.582 (df = 2936)	

Comment #2

Where is the EC result coming from?

Extension 1:
Include EC and EC_high as separate terms

EC_high >> **EC**.

The underlying mechanism is **not** "*low SES people benefit from connections to high SES people.*"

Alternative: for high SES, stock participation is greater with more high SES connections.

- Relate to Briggs et al (2021), which shows a high-income stock participation puzzle.

Suggestion: Conduct tests like this and shift the narrative from helping low SES participate and save.

	scale(100 * perc_div)		
	(1)	(2)	(3)
scale(ec_county)	0.125***	0.080***	0.029
	(0.026)	(0.026)	(0.026)
scale(ec_high_county)	0.640***	0.669***	0.484***
	(0.025)	(0.025)	(0.032)
scale(support_ratio_county)		0.036**	0.087***
		(0.017)	(0.017)
scale(volunteering_rate_county)		0.079***	0.093***
		(0.016)	(0.016)
log(pop2018)			-0.103***
			(0.013)
log(pc_income_w)			1.569***
			(0.127)
Constant	-0.006	-0.005	-5.228***
	(0.011)	(0.011)	(0.470)
Observations	3,018	3,018	2,942
R2	0.585	0.592	0.651
Adjusted R2	0.585	0.591	0.650
Residual Std. Error	0.630 (df = 3015)	0.625 (df = 3013)	0.572 (df = 2935)

Extension 2:
zip code level analysis

Advantage: can include county fixed effects (cols 2 and 4).

Disadvantage: Chetty et al add noise at finer granularity (restrict to zips with 5k people or more).

Similar conclusions on stock participation. **Might be different on saving?**

	Dependent variable:			
	scale(100 * perc_div)		scale(100 * perc_int)	
	(1)	(2)	(3)	(4)
scale(ec_zip)	0.004	0.071***	0.221***	0.273***
	(0.017)	(0.015)	(0.020)	(0.014)
scale(ec_high_zip)	0.165***	0.099***	0.065***	0.040**
	(0.019)	(0.016)	(0.019)	(0.017)
scale(support_ratio_zip)	0.063***	0.008	0.148***	0.074***
	(0.011)	(0.012)	(0.012)	(0.015)
scale(volunteering_rate_zip)	0.042***	0.025***	0.059***	0.019**
	(0.007)	(0.009)	(0.010)	(0.008)
log(pop2018)	-0.073***	0.002	-0.111***	-0.031***
	(0.008)	(0.007)	(0.011)	(0.008)
log(pc_income_w)	1.691***	1.725***	1.423***	1.468***
	(0.035)	(0.032)	(0.032)	(0.028)
Constant	-6.384***		-4.880***	
	(0.154)		(0.175)	
Observations	12,920	12,920	12,920	12,920
R2	0.855	0.924	0.800	0.925
Adjusted R2	0.855	0.906	0.800	0.908
Residual Std. Error	0.375 (df = 12913)	0.301 (df = 10523)	0.442 (df = 12913)	0.299 (df = 10523)
Note:				*p<0.1; **p<0.05; ***p<0.01

Conclusion

- Social capital embedded in our economic connections is an important and growing area.
- One reason it is growing is because of data like this paper uses.
- Lots of potential for this paper (and this whole literature).

Exemplar #3.
Social media and stock market participation

Key research question: Does social media *affect* stock market participation?

Online social interaction and discussion about stocks might overcome participation costs.

Social Media and Stock Market Participation*

Karsten Müller† Yuanyuan Pan‡ Carlo Schwarz§

June 4, 2024

Abstract

Using plausibly exogenous variation in regional Twitter adoption in the United States, we show that a 10% increase in social media usage causes a 2.5% rise in stock ownership. Consistent with lowering the costs of acquiring information, Twitter has larger effects in counties with low pre-existing stock market knowledge, improves knowledge about asset returns, and leads to a decline in the number of financial advisors. Social media also boosts interest in volatile "meme stocks" favored by retail investors. Our findings highlight the unique influence of social media on household portfolio decisions, distinct from other modern information technologies.

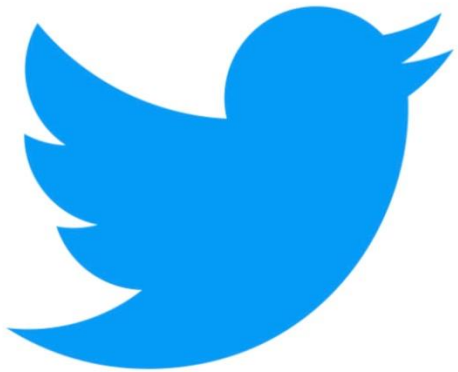
Keywords: Social Media, Stock Market Participation, Household Finance, Participation Puzzle
JEL Codes: D14, G51, D15, G11, G00, G40

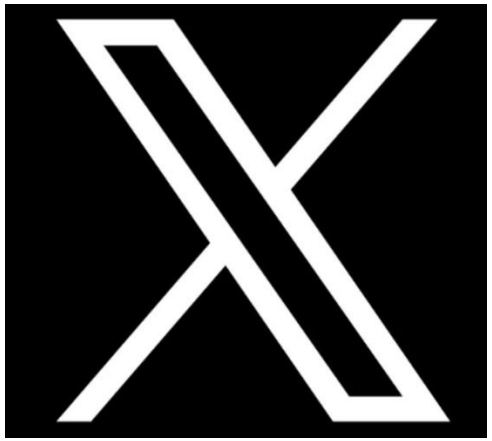
This paper shows effects using an instrument for Twitter adoption

Again, I'm going to adapt some slides (from AFA 2025) to describe and place this paper in the **social media and finance** literature.

Social Media and Stock Market Participation

By Müller, Pan and Schwarz





Discussion by J. Anthony Cookson
CU Boulder

Motivation

Dark side

- Social media and recent trading frenzies
  - GameStop (Pedersen, 2022; Bradley et al, 2023), Silicon Valley Bank (Cookson et al WP)
- A forum for biases?
  - Selective exposure (Chen and Hwang 2022, Cookson, Engelberg, Mullins 2023), biased belief updating (Chen, Peng, Zhou 2024WP)

Bright side

- Social signals are informative
  - Predict returns (Chen et al, 2014RFS, Cookson, Lu, Mullins and Niessner, 2024)
  - Feedback into corporate decisions (Cookson, Niessner and Schiller, 2024).

This paper: **Could there be a bright side for investors themselves?**

Motivation and Strategy

This paper: **Could there be a bright side for investors themselves?**

Paper finds: Social media exposure ***drives*** stock market participation.

Core strategy:

1. **Deploy South-by-Southwest (SXSW) 2007 Festival** shock to Twitter account openings to obtain exogenous geographic variation in Twitter.

2. **Measure stock market participation (in 2015)** across geographies as Dividends Tax Filings in the IRS SOI database.

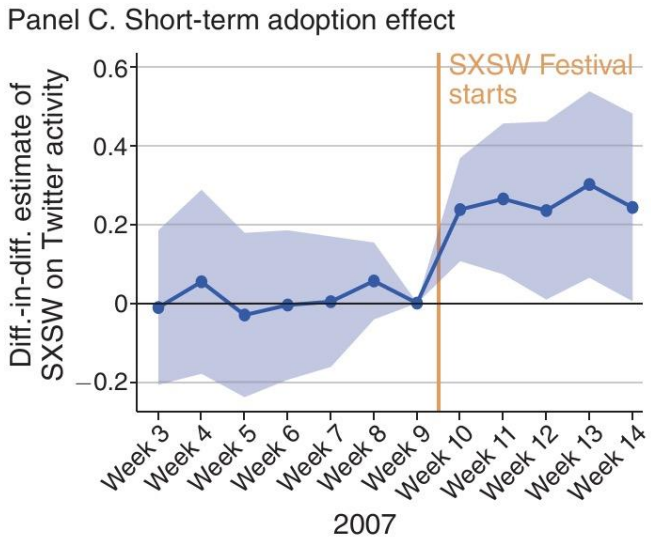
Paper deploys an instrument, developed in an earlier paper

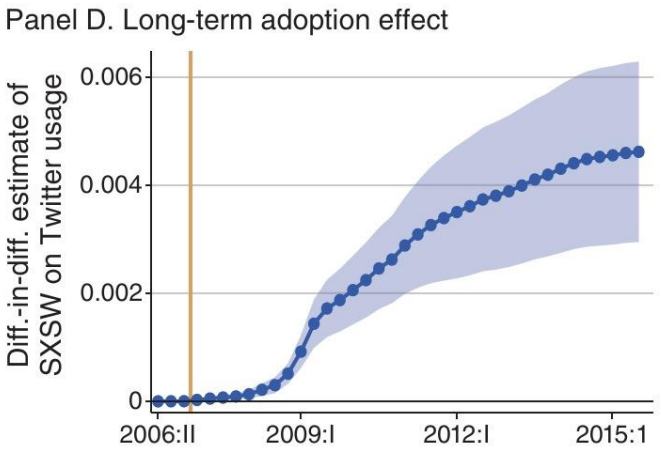
American Economic Journal: Applied Economics 2023, 15(3): 270–312
https://doi.org/10.1257/app.20210211

From Hashtag to Hate Crime: Twitter and Antiminority Sentiment†

By Karsten Müller and Carlo Schwarz*

We study whether social media can amplify antiminority sentiment with a focus on Donald Trump's political rise. Using an instrumental variable strategy based on Twitter's early adopters at the South by Southwest festival in 2007, we find that higher Twitter use in a county is associated with a sizeable increase in anti-Muslim hate crimes after the 2016 presidential primaries. Trump's tweets about Muslims predict increases in xenophobic tweets by his followers, cable news mentions of Muslims, and hate crimes on the following days. These results suggest that social media content can affect real-life outcomes. (*JEL* D72, J15, K42, L82, Z12)





Paper deploys an instrument, developed in an earlier paper

American Economic Journal: Applied Economics 2023, 15(3): 270–312
https://doi.org/10.1257/app.20210211

From Hashtag to Hate Crime: Twitter and Antiminority Sentiment†

By Karsten Müller and Carlo Schwarz*

We study whether social media can amplify antiminority sentiment with a focus on Donald Trump's political rise. Using an instrumental variable strategy based on Twitter's early adopters at the South by Southwest festival in 2007, we find that higher Twitter use in a county is associated with a sizeable increase in anti-Muslim hate crimes after the 2016 presidential primaries. Trump's tweets about Muslims predict increases in xenophobic tweets by his followers, cable news mentions of Muslims, and hate crimes on the following days. These results suggest that social media content can affect real-life outcomes. (*JEL* D72, J15, K42, L82, Z12)

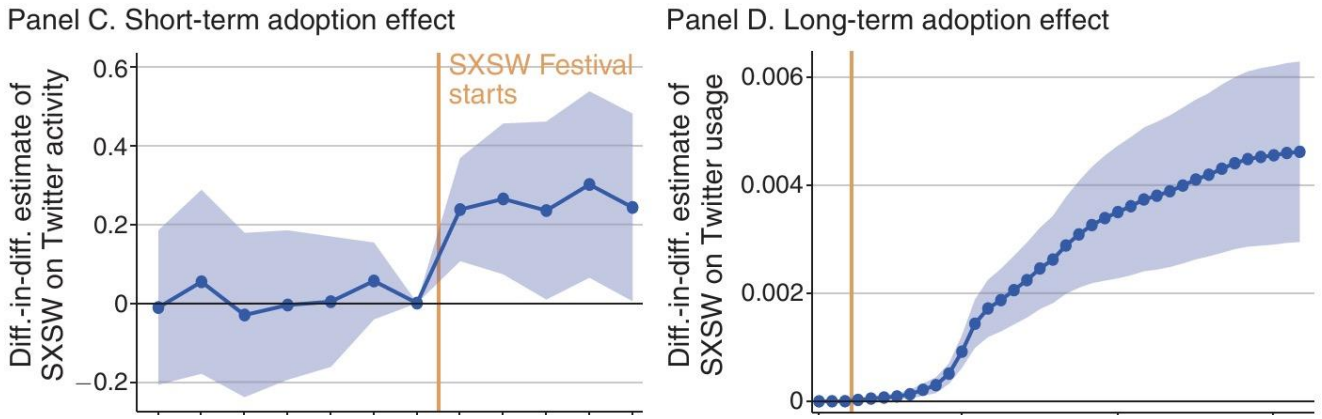


Table 1—First Stage: South by Southwest 2007 and the Diffusion of Twitter

	log(*Twitter usage*)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(*SXSW followers*, *March* 2007)	0.582	0.555	0.526	0.483	0.474	0.453	0.445
	(0.062)	(0.067)	(0.061)	(0.055)	(0.057)	(0.059)	(0.059)
log(*SXSW followers*, *Pre*)	0.226	0.172	0.118	0.115	0.109	0.099	0.091
	(0.084)	(0.078)	(0.081)	(0.077)	(0.075)	(0.074)	(0.071)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls		Yes	Yes	Yes	Yes	Yes	Yes
Race and religion controls			Yes	Yes	Yes	Yes	Yes
Socioeconomic controls				Yes	Yes	Yes	Yes
Media controls					Yes	Yes	Yes
Election control						Yes	Yes
Crime controls							Yes
Observations	3,107	3,107	3,107	3,106	3,105	3,105	3,105
R^2	0.933	0.934	0.935	0.944	0.945	0.946	0.947
Mean of DV	5.277	5.277	5.277	5.278	5.279	5.279	5.279
p-value: *March* 2007 = *Pre*	0.01	0.01	0.00	0.00	0.00	0.00	0.00

Main findings in pictures

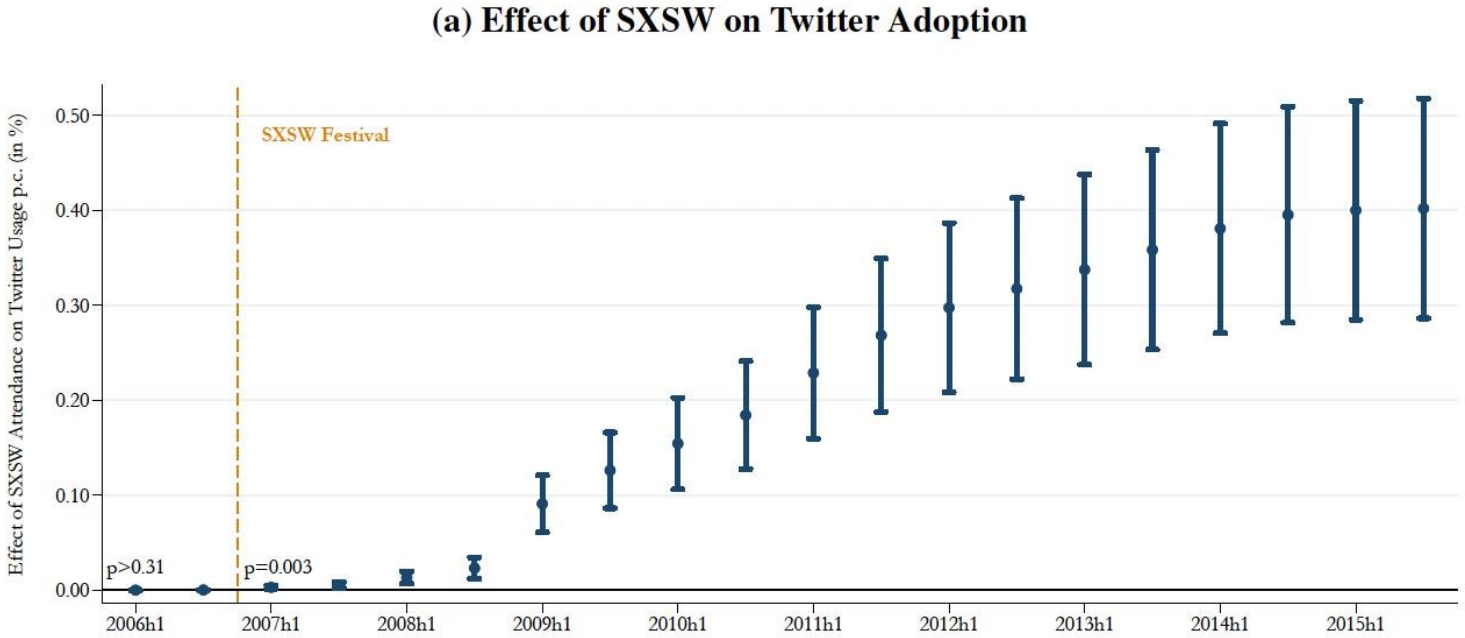
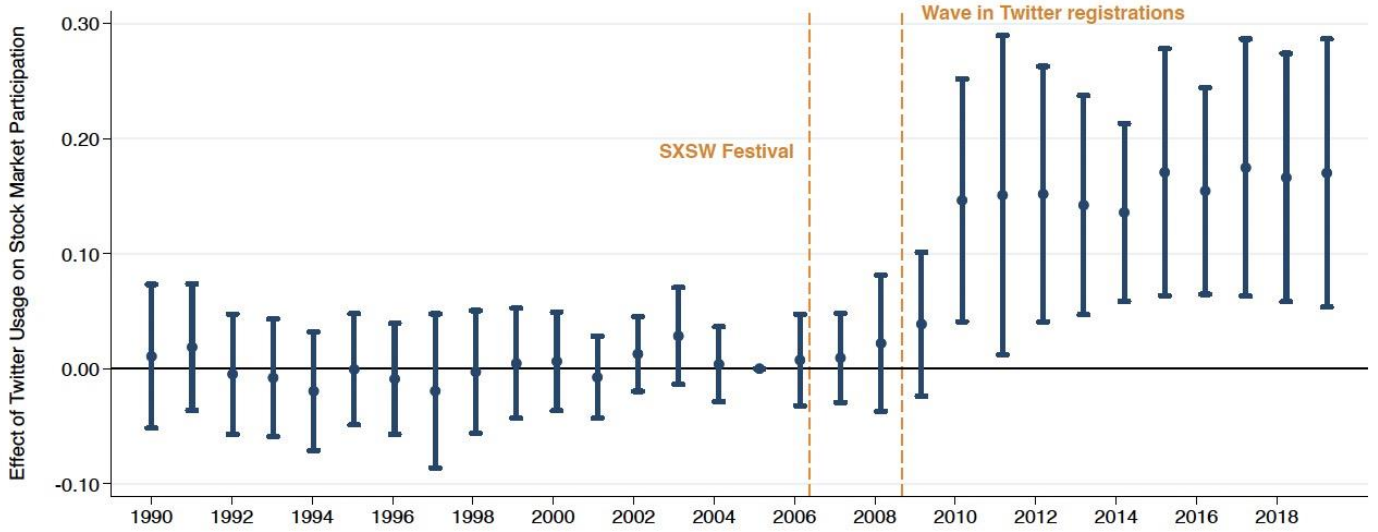


Figure 3: Dynamic IV Estimates



Not just benefits

Figure 4: Social Media Adoption and Financial Advisor Displacement

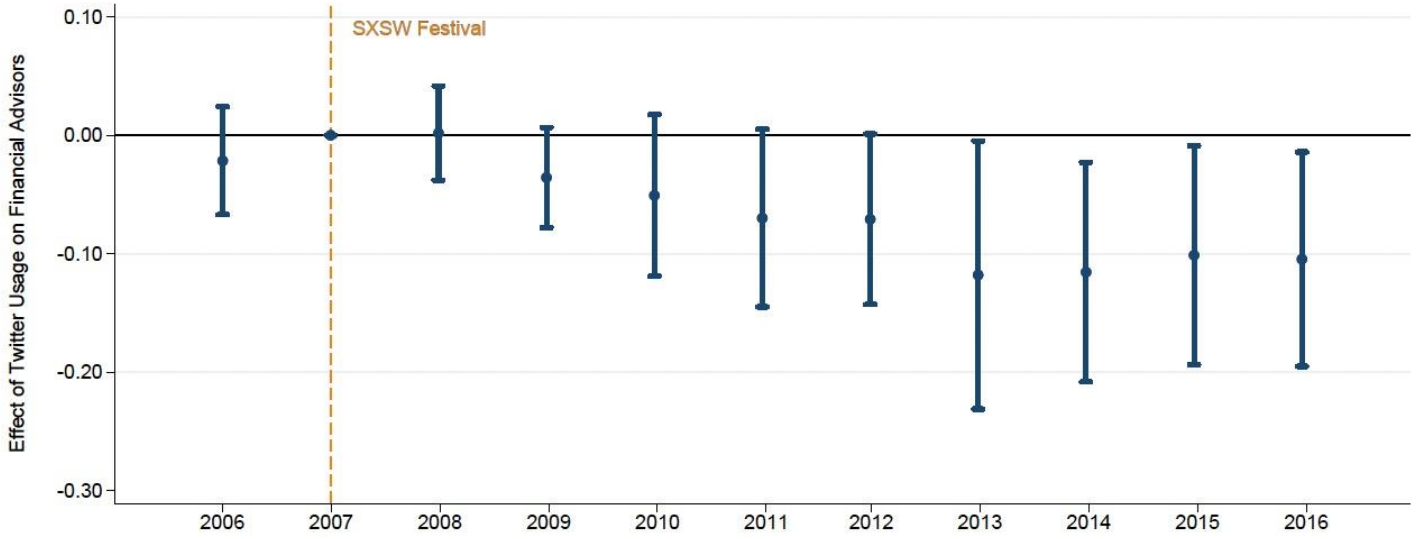
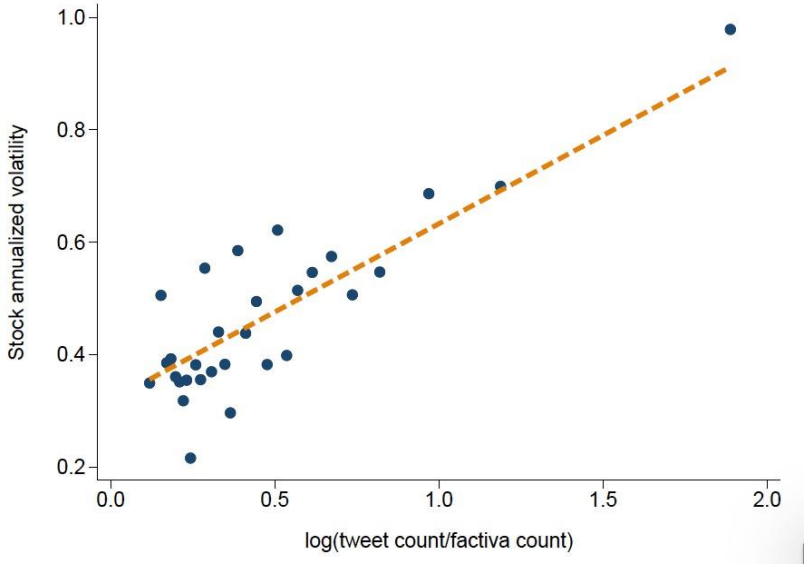
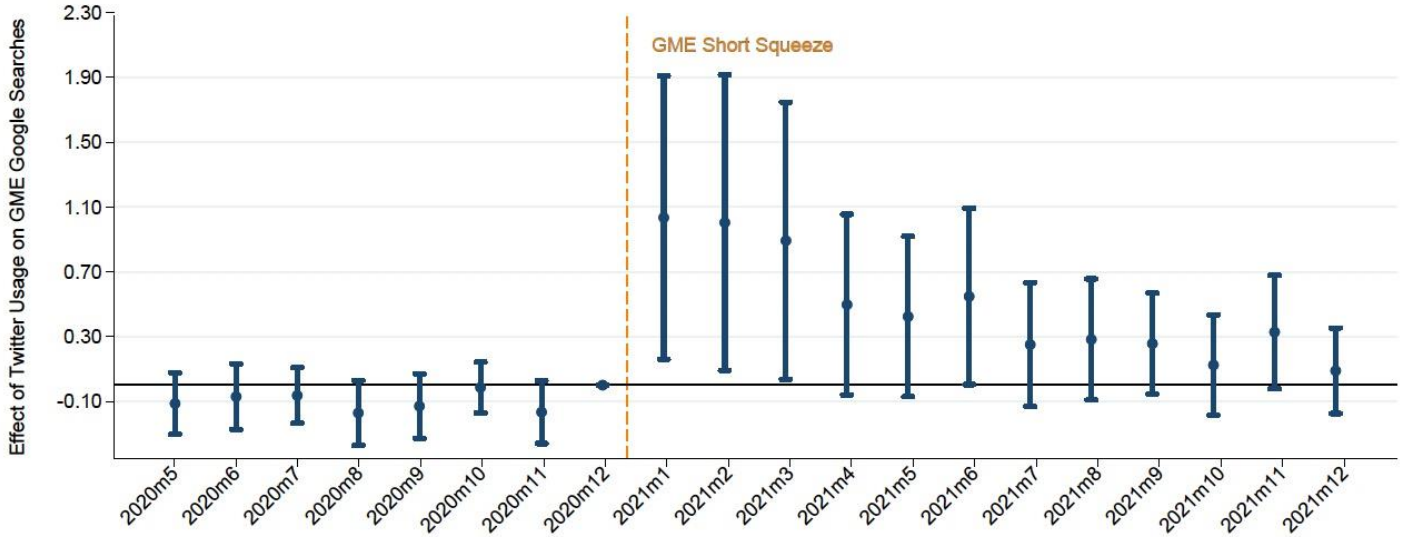


Figure 6: Which Stocks Are Overrepresented On Social Media?

(a) Stock volatility



(a) Google Searches for GME



My take and my discussion

Take

1. Welcome (causal) evidence on social media's effects on investors
2. Credible and nuanced narrative of costs **and** benefits.

Discussion

1. Measurement
2. Generalizability

Comment #1
Measurement

Key to Identification
Cross-county variation in social media exposure

South-by-Southwest Attendance Measurement

- Authors do not have a list of attendees
- Instead, they
  1. Compile a list of post-SXSW followers of @sxsw account.
  2. Identify the locations of these Twitter accounts from profile information, text strings.

First reaction: this is noisy

- We could identify < 10% of profile locations on StockTwits using something similar.

Example followers of @sxsw (from late 2024)

This is not random, but mostly to show some valid location classification (I am underrepresenting “No location”)

No location

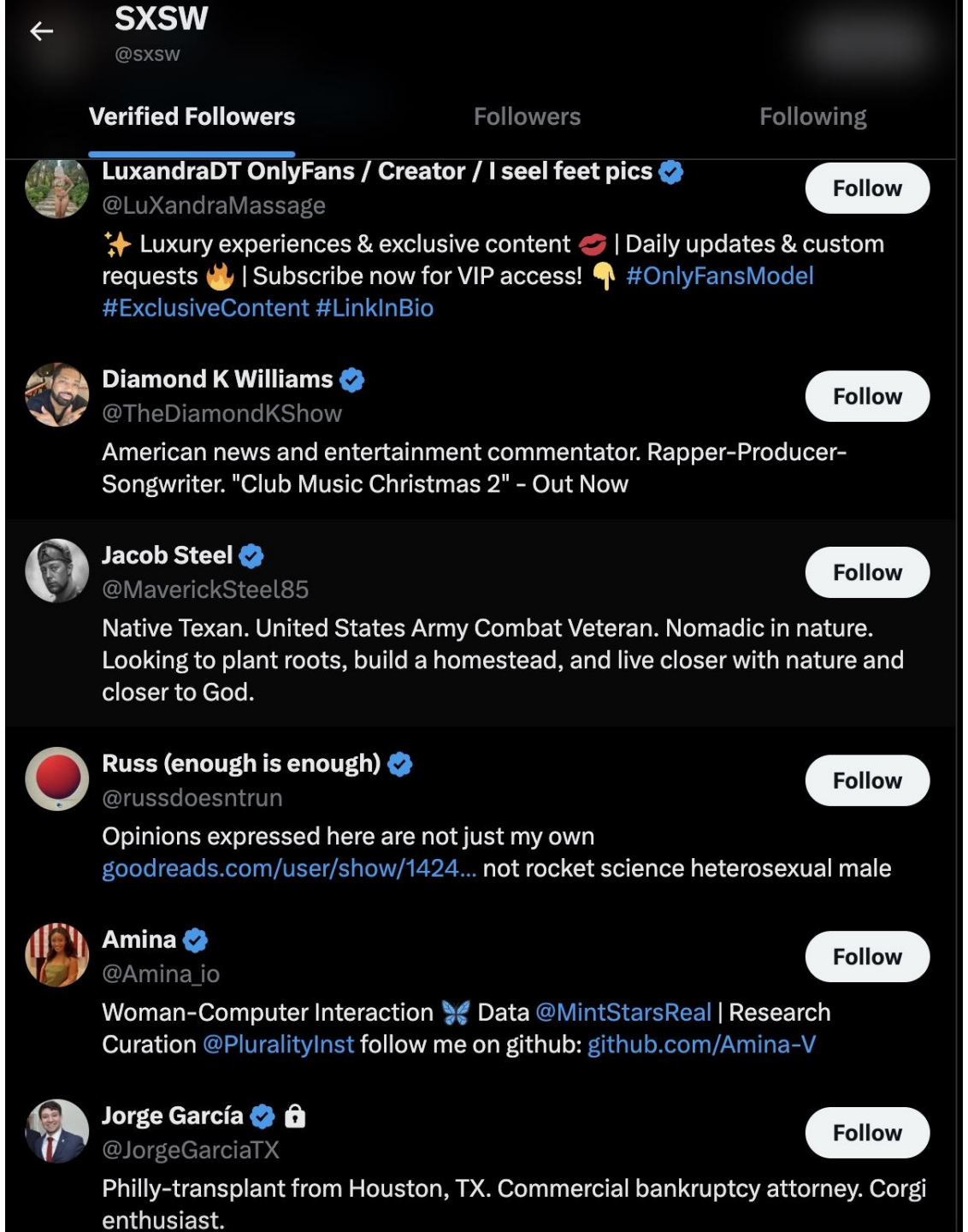
No location

Texas

No location

No location

Houston



Example followers of @sxsw (from late 2024)

Austin

Austin

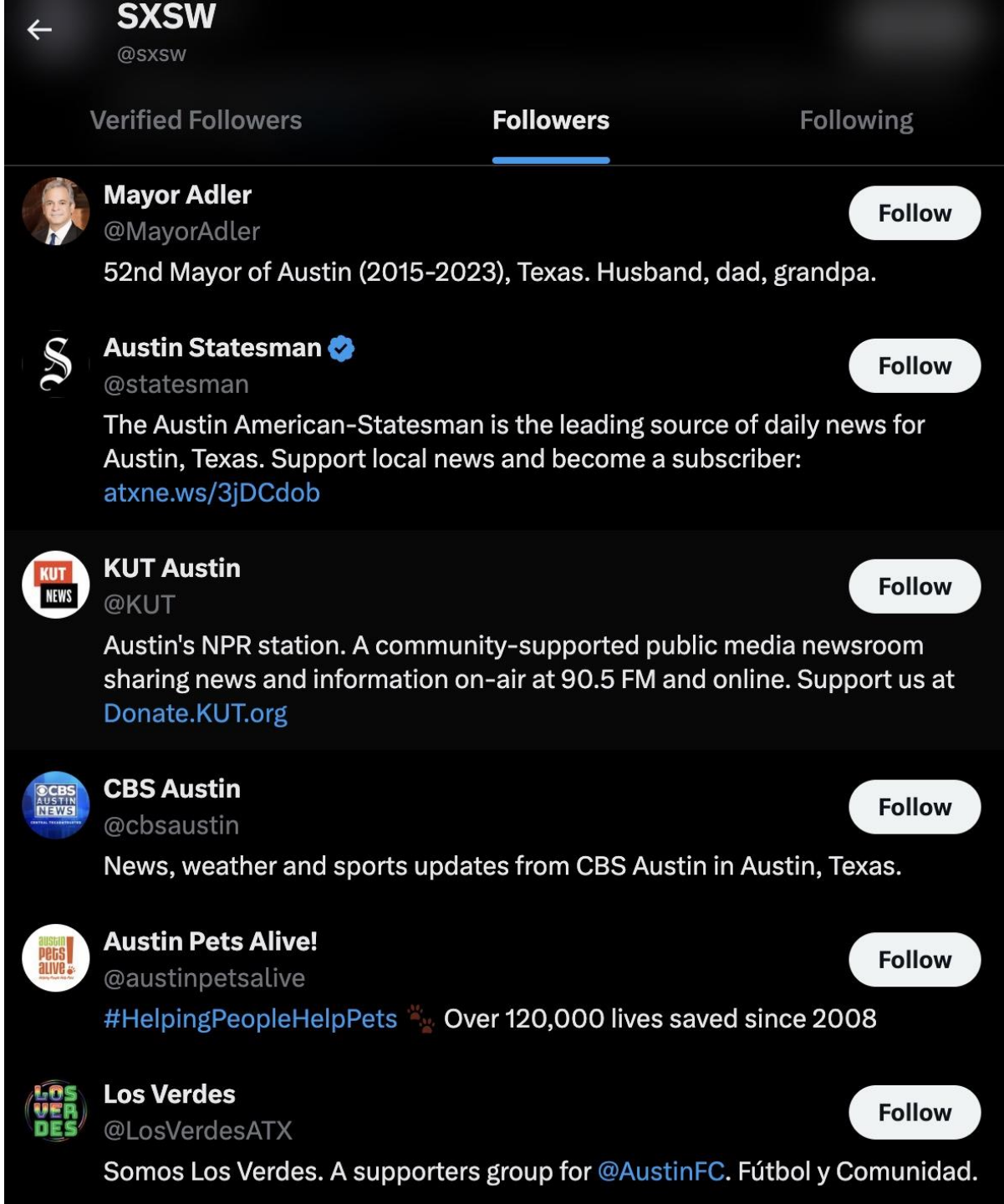
Austin

Note: paper controls for distance from Austin

Austin

Austin

Austin



How to think about this measurement issue

On one hand, *classical* measurement error biases toward zero.

But, that's true of the variable itself, not necessarily the instrument.

So, it is worth probing the nature of measurement error in this case

Some results that help:

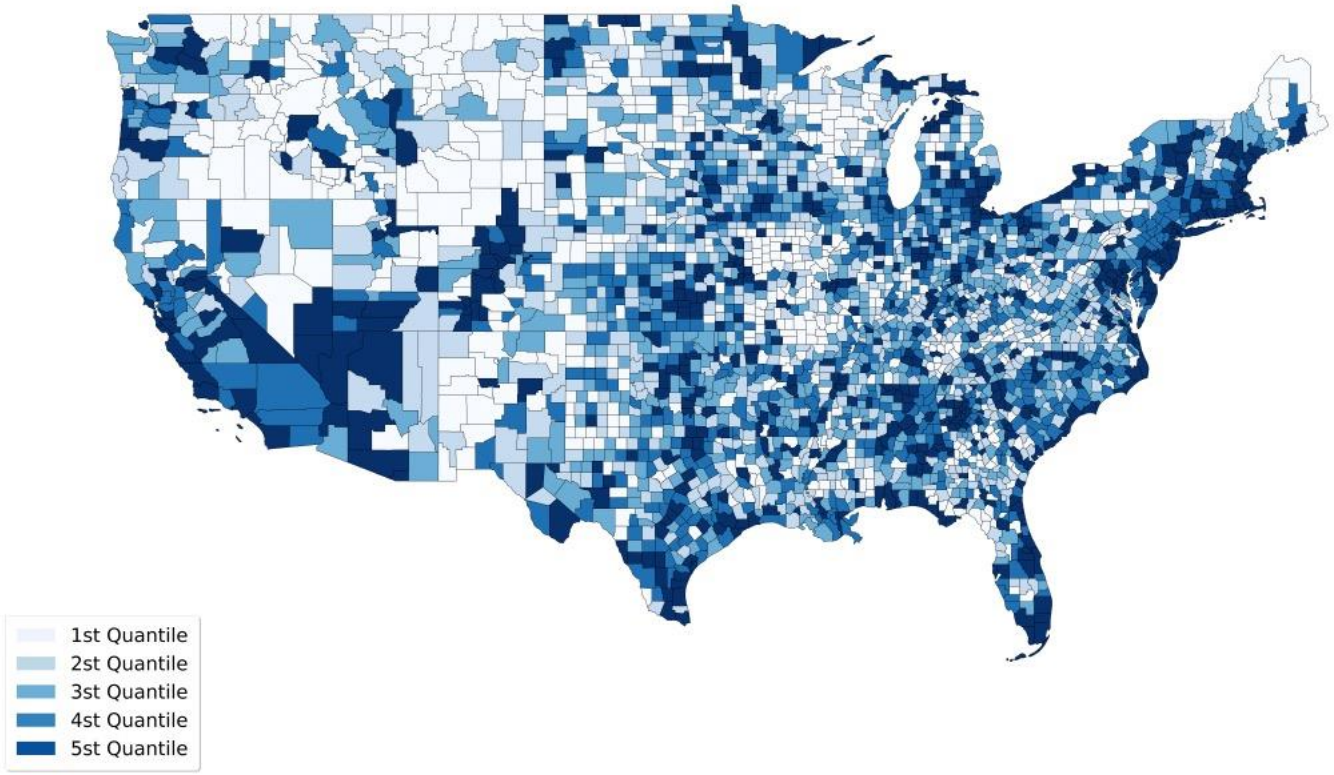
- No effect in pre-SXSW period
- No effect using pre-SXSW followers

How to think about this measurement issue

Main lingering concern

Bias toward urban areas.

Paper already controls for population, but given that cities (might mostly) identify counties, I'd like to see an *upfront* nonlinear control for population.



Comment #2
Generalizability

Comment #2: All IVs are local

- Participation in #FinTwit might have a different impact than participation in other social media.
  - In "The Social Signal," we saw big cross-platform differences in sentiment. It depends on user populations / differences across platforms.

- Could we expect the effects on stock participation to differ because…
  - **Different forums**: StockTwits? Seeking Alpha? Reddit? TikTok?
  - **Different variation**: A spike of social media interest due to meme stocks? Stimulus checks?

Comment #2: All IVs are local

What to do?

- Use heterogeneity to speak directly to this issue.
  - Variation within Twitter users' sophistication, education, etc
  - Without heterogeneity analysis, it is not obvious ex ante whether a platform like TikTok (high engagement, less sophisticated investors, low quality content) would have more benefits versus costs.

- Write specifically, upfront and humbly about claims about social media writ large, or participation on social media.
  - Internal validity for participation on FinTwit from a SXSW-like shock.
  - Needs more for broader application

Conclusion

This is a really interesting paper on an important topic.

My comments are about placing the paper and understanding its variation, but I have no doubt that the paper will publish soon (and it should be recognized for its impact).

Summing Up Day 1

Social media and finance research is widely varying. The three papers highlighted today give a broad view of this research agenda.

1. Signal in social media?

2. What does social media reflect about ourselves and markets?

3. How does social media affect trading and market participation?