Social Media and Stock Market Participation*

Karsten Müller† Yuanyuan Pan‡ Carlo Schwarz§

April 18, 2024

Abstract

We investigate the effect of social media adoption on stock market participation and retail investor beliefs in the United States. Using plausibly exogenous variation in the early adoption of Twitter across counties, we show that a 10% increase in social media usage is associated with a 2.5% higher rate of stock ownership and an overall increase in stock market wealth. Consistent with the idea that social media can lower the cost of accessing information, we find that Twitter adoption is associated with a decline in the number of financial advisors and has larger effects on stock ownership in counties with lower levels of pre-existing stock market knowledge. Twitter adoption also fuels interest in "meme stocks", which tend to be more volatile and owned by retail investors. Overall, our results suggest a distinct impact of social media platforms on household portfolio choices that differs from that of other modern information technologies.

Keywords: Social Media, Stock Market Participation, Household Finance, Participation Puzzle

*We would like to thank Tony Cookson, Stephen Dimmock, Michael Haliassos (discussant), David Hirshleifer, Nicola Limodio, Kasper Roszbach (discussant), Christoph Schiller, Moto Yogo, and participants at the CEPR Eighth European Workshop on Household Finance, 2023 Boulder Summer Conference on Consumer Financial Decision Making, Swiss Winter Conference on Financial Intermediation, Society of Economics of the Household (SEHO) Meetings, National University of Singapore, University of Colorado Boulder, and Università Bocconi for their valuable comments. Müller acknowledges financial support from a Singapore MOE start-up grant and Presidential Young Professorship (WBS numbers A-0003319-01-00 and A-0003319-02-00).

†National University of Singapore, Department of Finance and Risk Management Institute, www.karstenmueller.com, kmueller@nus.edu.sg

‡National University of Singapore, Department of Finance, yuanyuan@u.nus.edu

§Bocconi University, Department of Economics, www.carloschwarz.eu, carlo.schwarz@unibocconi.it

1 Introduction

Retail trading in the stock market is on the rise. Between 2019 and 2022, the share of direct stock ownership in the United States grew from 15 to 21 percent (Aladangady et al., 2023). During the meme stock frenzy in 2021, more than 7.8 million people opened new retail trading accounts in only two months CNBC, 2021a. Many have argued that social media is an important contributor to this trend. In the United States, more than 300 million people use social media (Statista, 2022). Among 18-to-34-year-olds, social media is now the most popular source of investment advice (CNBC, 2021b). Further, almost 80% of institutional investors use social media as part of their regular workflow (Coalition Greenwich, 2015, 2019).[1] However, despite its potential relevance in explaining the increasing importance of retail investors, there is relatively limited evidence on whether and how social media adoption affects households' investment decisions and portfolio choices.

In this paper, we quantify the effect of Twitter on stock market participation and retail investor beliefs in the United States. Existing research has highlighted that acquiring information about the stock market is costly (e.g, Haliassos & Bertaut, 1995; Vissing-Jørgensen, 2002), and that these costs play an important role in explaining the "stock market participation puzzle" widely documented in the household finance literature (e.g. Campbell, 2006; Grinblatt, Keloharju, & Linnainmaa, 2011; Guiso & Sodini, 2012). The tailored nature of social media algorithms may be able to lower the costs of accessing information, whether accurate or not, which could in turn increase stock market participation. On the other hand, social media could also increase the cost of accessing *accurate* information by generating an information overload (e.g., Bernales, Valenzuela, & Zer, 2023) or propagating misinformation (e.g., DeMarzo, Vayanos, & Zwiebel, 2003). As such, the effect of social media adoption on stock ownership is a priori ambiguous.

To study the effects of social media use, we combine data on the geographical location of Twitter users with information on stock ownership from the Internal Revenue Service (IRS), survey data from Gallup, and tweets containing the hashtag "#FinTwit." "FinTwit", short for "financial Twitter", is a popular online community where investors exchange information and advice related to investing.[2] Compared to other social media platforms such as "StockTwits",

[1] Coalition Greenwich (2015) surveyed 265 corporate and public pension funds, insurance companies, endowments, and foundations across U.S., Europe, and Asia. They find that over 80% of institutional investors use social media as part of their work, and around 30% of them report that information obtained from social media directly influences their investment decisions. Coalition Greenwich (2019) report that social media is now a more important source of information for institutional investors than finance-specific trade publications.

[2] Twitter has become the go-to platform for many finance professionals to share their perspectives, including hedge fund managers, top bankers, and entrepreneurs such as Bridgewater Associates founder Ray Dalio, former

which mainly focuses on stock-picking, "FinTwit" covers a broad set of financial topics ranging from traditional financial advice to recent topics like cryptocurrencies.[3]

Estimating the casual effect of Twitter on stock market participation is challenging due to the endogenous nature of Twitter adoption. Hence, a naive regression of Twitter usage on stock market participation would likely be biased. Building on Müller and Schwarz (2023), we propose an identification strategy that creates plausibly exogenous variation in Twitter across counties. Specifically, we leverage variation generated by Twitter's surprise popularity at the South by Southwest (SXSW) Interactive Festival in March 2007, an annual event held in Austin, Texas. Although Twitter was founded in 2006, its popularity only reached a "tipping point" when it was picked to be showcased at SXSW Interactive, which put it on an explosive growth trajectory. Following a spike in new registrations during SXSW, the festival's attendees disseminated information about Twitter through their social networks, resulting in a sudden wave in sign-ups in their home counties. Today, SXSW 2007 is widely recognized as the key catalyst event that allowed Twitter to reach a critical mass of users (TechCrunch, 2011).[4]

Using SXSW as a shock to initial Twitter adoption, we confirm the result from Müller and Schwarz (2023) that a 10% higher number of SXSW attendees is associated with a 2% increase in the number of tweets sent in a county following the festival. Importantly, these counties exhibited the same trajectory of Twitter adoption before the event, supporting the parallel trends assumption. These results suggest that SXSW led to network effects, making it more attractive for people in the home counties of SXSW attendees to sign up, and these county-level network effects persisted in the long run. As we show, a higher number of SXSW 2007 attendees in a county is associated with a higher level of Twitter usage in 2015, even after controlling for a county's pre-existing interest in SXSW and a host of geographical and socioeconomic variables.[5]

PIMCO CEO Mohamed El-Erian, or Tesla and Twitter CEO Elon Musk (CNBC, 2019).

[3]In previous studies, Twitter data has been widely used to forecast stock price movements (e.g. Bollen, Mao, & Zeng, 2011), measure firm performance (e.g. Bartov, Faurel, & Mohanram, 2018), and assess trust in CEOs (Elliott, Grant, & Hodge, 2018).

[4]In response to a question on the online messaging board Quora, Twitter co-founder Evan Williams wrote about the role of SXSW in the platform's success: "I don't know what was the most important factor, but networks are all about critical mass ... And something clicked."

[5]Existing work has used SXSW as an instrument for Twitter usage to study the effect of social media on changes in hate crime around Donald Trump's presidential campaign (Müller & Schwarz, 2023) and election outcomes (Fujiwara, Müller, & Schwarz, 2023). This identification strategy is similar to Enikolopov, Makarin, and Petrova (2020), who use the number of students from each town in the cohort of the founder of the social network *VKontakte* (VK) as an instrument for social media adoption in Russia. Following their paper, we use the term SXSW "attendees" here and throughout the paper for brevity, even though we do not perfectly observe SXSW attendance. We discuss this issue in detail in the data section.

Our identification strategy exploits the plausibly exogenous residual geographic variation in the attendance of SXSW 2007 as an instrument for social media adoption across counties in an event study design. Using the year before Twitter's launch (2005) as the pre-period, our main analysis relates *changes* in stock market participation to a county-level measure of Twitter usage (measured as of 2015), instrumented with the number of SXSW 2007 attendees. This methodology exploits within-county variation over time, similar to a difference-in-differences estimator with an (instrumented) continuous treatment variable. By doing so, we effectively hold constant any time-invariant county characteristics, such as the general tendency of people towards adopting new technologies. Our analysis indicates that counties more exposed to social media adoption did *not* exhibit differential stock ownership rates dating back to 1990. These parallel trends in our main outcome alleviate concerns that residual variation in SXSW attendance may capture unobserved differences between counties correlated with other aggregate shocks.

Using this identification strategy, we find that exposure to social media leads to an economically meaningful increase in stock market participation. Specifically, a 10% higher adoption of Twitter increases stock ownership per capita by approximately 2.5% and the share of income from stock market investments by 2.2%. This effect size is somewhat smaller than that of broadband internet studied by Hvide et al. (2022). Importantly, counties with higher SXSW attendance were on parallel trends with other counties in their stock market participation rates before Twitter's launch. These parallel trends also hold for the period of the internet bubble and crash of the late 1990s and early 2000s, further indicating that the observed effects in "treated" counties are not merely reflective of higher interest in new technologies in general. Our analysis reveals a swift and permanent positive effect on stock ownership only once Twitter reached a critical mass of approximately 15 million users following a wave of new registrations during 2009-10. A range of additional robustness exercises support the validity of our findings and rule out the potential influence of the housing boom of the 2000s and the Great Recession of 2007-08.

Our identification strategy also allows us to directly address concerns about omitted variables that could drive both changes in Twitter usage and stock market participation in counties with more SXSW 2007 attendees. In particular, we conduct a placebo test based on the number of SXSW followers who had already joined Twitter *before* SXSW. The home counties of these users serve as a natural "placebo" group because, while they shared a similar interest in the SXSW festival, they did not experience an additional inflow of Twitter users in March 2007, in contrast to counties with attendees of SXSW 2007 (our instrument). As we show, counties with SXSW 2007 participants are highly similar to those with "placebo" users across a large number of observed characteristics. However, we find no changes in stock market participation in these "placebo" counties. This finding suggests that omitted variables are unlikely to explain our findings, given that any such omitted variable (e.g., a concurrent adoption of mobile banking) should also affect the "placebo" group.

We provide evidence supporting two mechanisms through which social media use could affect stock market investments. First, social media could be interpreted as a shock to the costs of acquiring information. Similar to word-of-mouth or observational learning (Bikhchandani, Hirshleifer, & Welch, 1992; Hong, Kubik, & Stein, 2004), social media could lead to an increase in social connections, facilitating information dissemination beyond geographical boundaries. Consistent with this channel, we observe more pronounced effects of social media adoption in counties with fewer sources of information about the stock market *before* the launch of Twitter. This finding holds for several proxies for access to information, including the density of financial advisors or bank branches, knowledge about stock market investments based on survey data, and the share of the population with a bachelor's degree.

Additionally, we find evidence that Twitter adoption reduced the number of financial advisors in a county. Our estimates show that a 10% increase in Twitter usage is associated with a 1% drop in financial advisors in the long run. This result has two implications. First, it suggests that digital technologies have the potential to displace white-collar jobs that have traditionally been thought of as more resilient to the adoption of new technologies. Second, it supports the notion that social media can act as a substitute for traditional financial advice by reducing the costs associated with acquiring financial information.

Consistent with the idea that social media can provide useful information to investors, we find that Twitter adoption is associated with a change in retail investor beliefs, indicating an improved understanding of financial markets. Using Gallup survey data, we show that Twitter usage is associated with an increased likelihood of perceiving stocks as the best long-term investment relative to gold and real estate, which are believed to have higher returns than they actually do by a large fraction of Americans. Given that limited stock market participation is a "puzzle" (Campbell, 2006; Grinblatt, Keloharju, & Linnainmaa, 2011; Guiso & Sodini, 2012), and that the historical risk-adjusted returns of stocks are much higher relative to gold, an increase in favoring stocks suggests that social media can change retail investors' belief. Importantly, individual-level survey data also allow us to confirm that an exogenous increase in Twitter use in a county increases the probability of owning stocks, even after controlling for a host of individual characteristics that are known to affect stock market participation.

The second mechanism we explore involves the role of social media platforms in reinforcing the transmission of ideas. Social media adoption could lead to an easier transmission of information biased towards positive past experiences (as in Han, Hirshleifer, & Walden, 2022) or a reinforcement of similar views in "echo chambers" (Cookson, Engelberg, & Mullins, 2022).[6] These factors, in turn could amplify behavioral biases (Heimer, 2016; Hirshleifer, 2020) and contribute to trading frenzies (Tengulov et al., 2021). In line with the idea that social media can facilitate risky investments in lottery-type assets (Han, Hirshleifer, & Walden, 2022; Pedersen, 2022), our findings indicate that social media increases the likelihood of investing in "meme stocks." In particular, we find that counties with higher Twitter adoption due to SXSW experienced a more pronounced spike in Google searches for GameStop, AMC, and other meme stocks following the short squeeze episode in early 2021. Importantly, these counties were on parallel trends to others before the event, which suggests a causal effect of exposure to social media on the likelihood of participating in the trading craze.

To understand the differences between social media and traditional media, we also investigate the frequency of stock mentions on Twitter relative to financial news outlets. The most overrepresented stocks on Twitter are more volatile and more likely to be owned by retail investors, which is also consistent with a role of social media in contributing to risky investments by relatively inexperienced investors. As such, our findings suggest that social media may be a "double-edged sword" for financial decision-making: while it can spur stock market participation, it may also encourage investments in volatile and risky assets.

Our paper contributes to a growing literature exploring the effects of information technology on financial markets. Previous studies have investigated the effects of internet and smartphone usage (e.g. Barber & Odean, 2002; Bogan, 2008; Choi, Laibson, & Metrick, 2002; Kalda et al., 2021). Other research has examined the relationship between various measures of social media sentiment and stock returns (e.g. Bartov, Faurel, & Mohanram, 2018; Bollen, Mao, & Zeng, 2011; Chen et al., 2014; Cookson et al., 2022; Luo, Zhang, & Duan, 2013), as well as the association between exposure to social media and investor behavior *conditional* on investors participating in the stock market. Cookson and Niessner (2020) and Cookson, Engelberg, and Mullins (2022) study disagreement and selective exposure on investor social media. Heimer (2016) studies the effect of social networks on the disposition effect. Hirshleifer, Peng, and Wang (2021) show a link between the centrality of a firm's location on social media and stock reactions to earnings surprises. Tengulov et al. (2021) analyze social media activity around the short squeeze events in the U.S. stock market in early 2021. Dessaint, Foucault,

[6] Han, Hirshleifer, and Walden (2022) define as *self-enhancing transmission bias* a situation where investors are more likely to communicate their positive investment experiences relative to their negative ones. Han, Hirshleifer, and Walden (2022) model how such self-enhancing bias in social networks can influence financial decision-making. They propose that some investors may prefer active investment strategies with higher risk, for example, because of this self-enhancing bias.

and Frésard (2021) show that exposure to social media content changes the informativeness of analysts' long-term compared to short-term forecasts. Cookson, Niessner, and Schiller (2022) show that negative social media sentiment is associated with a higher probability of merger withdrawals. Farrell et al. (2022) suggest that posts on the crowd-sourced investment platform SeekingAlpha make trades by retail investors better informed. Kuchler et al. (2021) use Facebook data to show that institutional investors are more likely to invest where they have more social ties.

To the best of our knowledge, our paper is the first to estimate the causal effect of social media adoption on participating in the stock market to begin with. A related paper by Hvide et al. (2022) studies the causal effect of broadband internet rollout on the investment decisions of individual investors in Norway. They find that broadband spurs a "democratization of finance" by increasing stock market participation for poorer and less-educated individuals. We find some complementary evidence that counties with less information about financial markets and a higher share of minorities benefit from social media adoption. However, we also show that, within a given county, stock ownership particularly increases among people with the highest incomes, who are more likely to own stocks in the first place. One potential reason for this difference relative to the results in Hvide et al. (2022) could be that, while social media provides information, it does not fundamentally alter the infrastructure of accessing investment accounts.

We also contribute to the broader literature on the stock market "participation puzzle" in household finance (e.g. Campbell, 2006; Guiso & Sodini, 2012; Haliassos & Bertaut, 1995). This literature has identified many correlates of stock market participation. Existing work would suggest that factors reducing the cost of acquiring information about financial topics could increase stock market participation (Vissing-Jørgensen, 2002). For example, there is evidence that higher levels of education (Bernheim & Garrett, 2003; Cole, Paulson, & Shastry, 2014), cognitive ability (Grinblatt, Keloharju, & Linnainmaa, 2011), and financial literacy (Van Rooij, Lusardi, & Alessie, 2011) are positively associated with stock market participation rates. These findings suggest that individuals who possess greater financial knowledge and better information processing skills are more likely to invest in the stock market. Another line of research has focused on how investor beliefs and psychological biases affect decision-making about households' financial decision-making. (e.g., Briggs et al., 2021; Guiso, Sapienza, & Zingales, 2008). Social interactions and the influence of peers have also been found to matter (e.g., Brown et al., 2008; Hong, Kubik, & Stein, 2004; Kaustia & Knüpfer, 2012; G. Li, 2014). Against this backdrop, our study presents evidence for social influences on households' financial decisions in the context of social media, where we observe how individuals communicate and exchange information in a dynamic and interactive way, which in turn may shape household beliefs and behaviors.

The remainder of this paper is organized as follows. Section 2 describes the data sources used in the paper. Section 3 introduces our empirical strategy. Section 4 presents the main results. Section 5 analyzes potential mechanisms. Section 6 concludes.

2 Data

We combine the following seven data sources for our analysis: (1) county-level data on stock market participation based on IRS individual income tax returns, (2) county-level data on Twitter usage, (3) county-level data on the number of people attending the SXSW festival in 2007, (4) individual-level survey data from Gallup, (5) DMA-level data on Google search interest in meme stocks, (6) additional county-level variables mostly used as controls, and (7) the frequency stocks are mentioned on Twitter and in traditional media. We describe these datasets in turn.

2.1 County-level stock market participation

The IRS Statistics of Income (SOI) provide aggregated individual income tax data at the county level. These data are publicly available from 1989 to 2019 and provide, among others, information about adjusted gross income (AGI) and dividend income. Dividend income identifies stock ownership because, if someone receives any income from dividends, they must own stocks. Therefore, we follow the existing literature (e.g., Brown et al., 2008; Chodorow-Reich, Nenov, & Simsek, 2021) and construct measures of stock ownership using this information on dividends. Our main variables are the ratio of total dividend income to population or the ratio of total dividend income over total AGI. Because these ratios are highly skewed, we take their log transformation in our baseline analysis. We consider other measures and transformations for robustness.

While dividend income from tax returns has been widely used as a proxy for stock ownership, it likely contains measurement error for two reasons. First, we do not observe ownership of stocks that do not pay dividends. Second, we do not observe indirectly held stocks in retirement accounts or other non-taxable entities. To address these limitations, we therefore validate our baseline measure using survey data from Gallup that asks individuals whether they directly or indirectly own stocks. Figure A.3 plots a county's dividend income per capita against the share of households reporting they own stocks in the Gallup survey data, which we describe in 2.4 below. We find a strong positive correlation between a county's dividend income per capita from IRS data and the share of households reporting stock ownership in the individual-level Gallup dataset. This correlation shows that measures based on dividend income are good proxies for actual stock market participation.

We also use two proxies for the total wealth earned by stock market investors. First, we take the measure of total stock market wealth from Chodorow-Reich, Nenov, and Simsek (2021). Their measure capitalizes the IRS SOI taxable dividend income data using a county-level age-specific CRSP price-dividend ratio and then adjust for stock wealth held in nontaxable accounts.[7] Second, we construct a proxy for the returns that stock market investors earn on their wealth. We start with data on the total capital gains from the sale of assets that are available from the IRS SOI from 2010 to 2019. To construct our proxy for county-level portfolio returns, we divide capital gains by the total stock market wealth in the previous year taken from Chodorow-Reich, Nenov, and Simsek (2021).

2.2 County-level Twitter adoption

We measure social media adoption across counties using Twitter data. Twitter was launched in 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams. Today, Twitter has around 230 million "monetizable" daily active users. To measure how Twitter usage differs across counties, we start with a large dataset of almost 500 million geo-located tweets collected between June and November in 2014 and 2015 by Kinder-Kurlanda et al. (2017). We collected profile information for the users in this dataset and assign each user to the county they tweet from most frequently. In total, we can geo-locate around 3.7 million profiles, covering approximately 7% of all US Twitter users as of 2015. These data also allow us to create measures of Twitter usage over time. Our baseline measure of Twitter usage is the natural logarithm of Twitter users in a county in 2015. Figure A.1b shows that there is substantial geographical variation in Twitter usage.

[7] Their capitalization process begins by deriving the price-dividend ratio for the value-weighted CRSP portfolio, with further adjustments that allow for variations in dividend yields across age groups. In order to calculate age-specific dividend yields, Chodorow-Reich, Nenov, and Simsek, 2021 use a sample of account-level portfolio holdings and combine it with CRSP stock and mutual fund data, and then compute average dividend yields for five age groups within each county. The resulting capitalization factor is obtained by taking age-wealth-weighted averages of age-specific dividend yields within each county based on data from the Survey of Consumer Finances (SCF). The adjustment for nontaxable stock market wealth considers the relationship between taxable dividend income and total stock wealth with demographics, also from the SCF.

2.3 County-level number of SXSW 2007 attendees

We build on Müller and Schwarz (2023) by exploiting the SXSW Interactive Festival in March 2007 as an exogenous shock to Twitter adoption. Ideally, we would like to measure the number of attendees coming from each county directly from the billing addresses used for SXSW ticket sales. Such data are unfortunately unavailable. To proxy for SXSW attendance, we follow Enikolopov, Makarin, and Petrova (2020) and use social media data for measurement. In particular, we proxy for the number of attendees coming from a particular county based on the number of Twitter followers of the SXSW account who signed up in March 2007. Given the widespread adoption of Twitter at SXSW 2007, our dataset likely covers a substantial fraction of actual attendees. To streamline our discussion, we follow Enikolopov, Makarin, and Petrova (2020) and refer to our proxy for attendance as "SXSW attendees" throughout the remainder of the paper. Similarly, we construct a placebo treatment based on the number of SXSW followers who had already signed up *before* the conference. These users will be referred to as pre-period followers or "SXSW followers, pre" for short.

The data on the followers of the SXSW account are sourced from Müller and Schwarz, 2023, who collected them using the Twitter API. User profiles are assigned to counties based on location strings mentioned in their profiles. Additionally, the profiles allow us to measure when a particular user joined Twitter. While Twitter does not provide information on when a user started following a specific account, it is reasonable to assume that the SXSW followers who signed up at the exact time of the festival in 2007 were induced to do so by the festival.[8]

Using social media data to measure SXSW attendance could introduce measurement error. However, such potential measurement error is unlikely to affect our findings for three reasons. First, classical measurement error would simply bias our first stage coefficient towards zero.[9] More importantly, it would leave the IV estimates unaffected as long as the measurement error in our instrument is uncorrelated with the error term in the second stage (see Pancost & Schaller, 2022, for a more extensive discussion). In other words, we only require that the measurement in SXSW attendance is uncorrelated with changes in stock market participation, which is plausible.

Second, while it is impossible to empirically investigate the correlates of unobserved measurement error, we can analyze the dynamics of the our estimated IV coefficients in an

[8]Note that the interpretation of our identification strategy is unchanged if people learned about and joined Twitter due to news reports covering the SXSW festival, even if they did not attend directly.

[9]For example, in the case of a regression model of the form $x = \beta z + \epsilon$, in which we only observe a mismeasured $\tilde{z} = z + u$, we would estimate $\hat{\beta} = \lambda\beta$ with $\lambda = \frac{\sigma_z^2}{\sigma_z^2+\sigma_u^2} < 1$. In other words, there is attenuation bias in the estimate $\hat{\beta}$.

event study design. As we will show in Section 4, we find no significant association between our instrument and changes in stock market participation all the way back to 1990. In our view, there is no credible reason why any measurement error should only begin to matter right at the time when Twitter rose to prominence.

Third, all of our main regressions include a placebo check based on the pre-period number of SXSW Twitter followers. This variable should be subject to exactly the same type of measurement error as our instrument. However, as we show, the coefficient estimates on this placebo variable are consistently close to 0 and statistically insignificant, which makes it unlikely that any non-classical measurement error can explain our findings.

2.4 Individual-level survey data

To analyze the mechanisms through which social media could spur stock market participation, we use individual-level survey data from the Gallup Social Survey covering the period 2006–2021. Every April, this survey asks around 1,000 respondents aged 18 and older living in all 50 states and the District of Columbia about their personal financial situation and a host of additional information. The dataset has a broad geographical coverage of around 600 counties every year, representing two thirds of the US population.

To construct a measure of stock market participation, we focus on the question "*Do you personally or jointly with a spouse, have any money invested in the stock market right now – either in an individual stock, a stock mutual fund, or in a self-directed 401-K or IRA?*." We code responses into an indicator variable equal to 1 if an individual answers "Yes", and 0 if they answer "No." As discussed in Section 2.1, the share of respondents owning stocks in the Gallup data is highly correlated with our baseline measure of stock market participation based on IRS data.

The Gallup Social Survey also allows us to infer people's opinion about different types of investments. In particular, we use the question "*Which of the following do you think is the best long-term investment – Bonds, real estate, savings accounts or CDs, stocks or mutual funds, or gold?*." Similar to the stock ownership question, we code the responses into indicator variables equal to 1 if an individual mentioned a particular investment to be the best, and 0 otherwise. Appendix Table A.1 provides summary statistics for the individual-level survey data from Gallup.

Stock ownership in the Gallup data is similar to that reported in the Survey of Consumer

Finance (SCF).[10] Figure A.4 shows that around 53% of families owned stocks directly or indirectly as of 2019 in the SCF data, a slight upwards trend from 2013. Similarly, the Gallup data in Figure A.4 reports 55% of households in the U.S. own stocks as of 2019.

2.5 Google search interest data

We test the link between social media adoption and interest in meme stocks using data from Google Trends. Da, Engelberg, and Gao (2011), for example, suggest that Google Search Volume Index (SVI) data can be used as a direct measure of retail investor attention. Data on the frequency of different search terms is available back to January 2004 on the level of 210 designated market areas (DMAs), which are largely defined based on metropolitan areas. We thus obtain measures of search interest for the tickers of several "meme stocks" (e.g., "GME" for GameStop Corporation).[11] We use tickers because, as pointed out by Da, Engelberg, and Gao (2011), using searches for company names could be problematic if investors search for companies for reasons unrelated to investing or if they use different versions of a company's name. Searches for ticker symbols, instead, are unambiguously motivated by an investment motive.

The Google SVI measures how frequently a given search term is entered into Google's search engine relative to the site's total search volume over a given period of time. In order to build a panel of SVI data that compares each DMA region and point in time, we follow the existing literature and normalize the data using a two-step process. First, each data point is divided by the total searches in a DMA and time period to measure relative popularity. Second, the resulting numbers are then scaled to range between 0 and 100. See Appendix A.1 in the online appendix for details.

[10]The SCF data is widely used to study stock market participation (Campbell, 2006; Hong, Kubik, & Stein, 2004), but it does not provide identifiers about the location of the survey participants.

[11]We construct an average SVI based on a list of keywords for "meme" stocks, including "GME" ("GameStop Corp."), "AMC" ("AMC Entertainment Holdings Inc"), "UPST" ("Upstart Holdings Inc"), "BBBY" ("Bed Bath & Beyond Inc."), "BBIG" ("Vinco Ventures Inc"), "BB" ("BlackBerry Ltd."), "NOK" ("Nokia Corp."), "SAVA" ("Cassava Sciences, Inc"), "CLOV" ("Clover Health Investments Corp"), "PLTR" ("Palantir Technologies."), "SPCE" ("Virgin Galactic Holdings"), and "LCID" ("Lucid Group, Inc. "). We construct this list of meme stocks based on articles published on the articles Investopedia and Yahoo Finance. We also collect data on Google searches for "crypto," given the importance of social media for "pump and dump" schemes (T. Li, Shin, & Wang, 2021).

2.6 Additional county-level variables

We obtain control variables to measure county-level differences in socioeconomic factors, demographics, geography, and media usage. Demographic and socioeconomic conditions are taken from the United States Census and the American Community Survey. County-level unemployment rates and industry-level employment shares are from the Bureau of Labor Statistics. Economic metrics such as county-level GDP, personal income and employment data are sourced from the Bureau of Economic Analysis. The number of bank branches is taken from the FDIC's Summary of Deposits (SOD). We take county-level data on financial advisors from Charoenwong, Kwan, and Umar (2019). We also construct a measure of knowledge of financial markets based on the National Financial Capability Study (NFCS), a large survey on various aspects of personal finance and financial literacy. In particular, we calculate the share of respondents in a state that say they have very high overall financial knowledge (7 on a scale of 1 to 7) on the state-level. Information on TV viewership patterns was collected from Simply Analytics and the Facebook Social-connectedness Index (SCI) is from the Meta Humanitarian Data Exchange.

We also use several variables to measure county exposure to the housing boom of the 2000s. Data on changes in house prices are based on the House Price Index (HPI) published by the FHFA.[12] Growth in mortgage credit is measured by aggregating loan-level data published as part of the Home Mortgage Disclosure Act (HMDA). County-level debt to income ratios are calculated using data from the Federal Reserve Bank of New York data and IRS, taken from Mian, Rao, and Sufi (2013). Table 1 presents summary statistics for the county-level dataset.

2.7 #FinTwit and Factiva data

To shed light on the extent of stock market-related topics on Twitter, we collect data using the Twitter Academic API. More specifically, we scraped all tweets that contain the hashtag "#FinTwit." This allows us to analyse the volume as well as the content of these tweets over time. Additionally, we used the Twitter count API to obtain the number of tweets containing the ticker of any stock listed on the NASDAQ or the New York Stock Exchange (NYSE) for each day in our observation period.

We explore the content and topics discussed on "financial Twitter" by visualizing a "FinTwit" word cloud. Figure A.2, based on textual tweets with the hashtag "#FinTwit" posted

[12] We assign changes in the state-level house price index to counties where the FHFA does not publish county-level data, but this does not make any difference for our results.

between December 2010 and December 2020, shows that “stock”,“market”,“trading”, “buy” and “sell” appear as frequent words. The prevalence of words such as “learn” and “community” suggest a potential role of social media in reducing the cost of acquiring information about stock markets.

To study differences between social media and traditional media sources, we calculate the number of times a stock is mentioned in the Wall Street Journal or Dow Jones Institutional News relative to Twitter between 2006 and 2020. The traditional news counts are taken from the Factiva database, where we focus on the 120 most frequently mentioned US stocks on Twitter. To study the correlates of over-representation on social media, we use data on stock returns from the Yahoo Finance API. The share of a stock owned by retail investors is inferred from the holdings of institutional investors in 13F filings.

3 Empirical Strategy

To estimate the causal effect of social media usage on stock market participation, we use the SXSW Festival in 2007 as an exogenous shock to the adoption of Twitter across US counties in an IV strategy. Section 3.1 provides some historical background on SXSW and Twitter adoption. Section 3.2 discusses the first stage results. Section 3.3 describes our IV model and the identifying assumptions.

3.1 South by Southwest and Twitter Adoption

Our identification strategy exploits plausibly exogenous variation in Twitter adoption across counties based on the platform’s popularity at the South by Southwest Festival (SXSW) in March 2007. First held in 1987, SXSW is an annual event in Austin, Texas, combining media, music, and film festivals and conferences. One part of the event, SXSW Interactive, focuses on new technologies.

Twitter was originally launched in March 2006 and had few users before SXSW 2007. At the event, Twitter put up large displays in the hallways of SXSW Interactive that showed tweets sent by the participants. The company also provided a festival-specific feature enabling attendees to create a Twitter account by sending a simple text message.

These incentives motivated many of the around 8,000 SXSW attendees to sign up, which massively increased Twitter activity. The daily volume of tweets tripled from 20,000 to 60,000 during SXSW during the festival (Gawker, 2007), as bloggers and other attendees broadcasted the platform to the outside world, urging their friends, families, and colleagues to sign up.

Today, SXSW is widely considered as the "tipping point" in Twitter's popularity that put the platform on a rapid growth trajectory.

3.2 First stage results

Müller and Schwarz (2023) and Fujiwara, Müller, and Schwarz (2023) study the impact of the SXSW festival on Twitter adoption in the United States. Immediately after the festival, the home counties of SXSW attendees exhibited higher rates of new accounting creations on Twitter relative to other counties. As a result, around 60% of Twitter's early adopters either joined during SXSW 2007 or were directly connected to someone who did.

We summarize this evidence here by presenting event study estimates that relate cumulative Twitter users per capita x_{it} in county i in half-year t to SXSW 2007 attendance z_i interacted with year dummies:

$$x_{it} = \sum_{w \neq 2006h2} \gamma_t z_i \cdot \mathbb{1}\{t = w\} + \theta_i + \theta_t + \nu_{it}, \tag{1}$$

where z_i is defined as counts in natural logarithm (with one added inside). The estimates for γ_t measure the change in Twitter users per capita relative to the period before SXSW 2007, depending on the number of festival attendees in a county. Standard errors are clustered by state, and we plot 95% confidence intervals.

Figure 1 plots the point estimates of γ_t. These estimates suggest no pre-existing trends before SXSW 2007. After the festival, the adoption effect of SXSW on Twitter usage shows an S-shaped pattern, commonly observed in the adoption of new technologies. This S-curved adoption effect moves hand in hand with the overall adoption of Twitter, proxied here by search interest for the word "Twitter" on Google.

This evidence for the effect of SXSW 2007 on Twitter adoption motivates the following first stage regression model:

$$x_i = \theta + \lambda_1 z_i + \lambda_2 w_i + \mathbf{X_i'}\phi + \upsilon_i \tag{2}$$

where x_i is the number of Twitter users in county i (in natural logarithm). z_i is our instrument: the number of SXSW 2007 attendees in natural logarithm (with one added inside). w_i is our placebo treatment based on the number of SXSW followers who joined Twitter before the festival in natural logarithm (with one added inside). As we show in Appendix Figure A.5, counties with SXSW Twitter followers in the pre-period are observationally equivalent to those with SXSW attendees based on a large number of county characteristics. $\mathbf{X_i'}$ is a vector of control variables, which we describe in more detail in Section 3.3.

The results are reported in Appendix Table A.2. Consistent with previous evidence in Müller and Schwarz (2023) and Fujiwara, Müller, and Schwarz (2023), we find that SXSW 2007 had a permanent effect on the intensity of Twitter adoption across counties. The estimate of 0.256 in column (5) suggests that a 10% increase in SXSW attendance is associated with a 2.6% increase in Twitter usage. In line with the idea that we are indeed picking up SXSW-induced increases in Twitter usage instead of a general Twitter affinity in counties with SXSW interest, the estimates of placebo treatment are always close to 0 and statistically insignificant. If counties with interest in SXSW would have adopted Twitter even in the absence of the festival, we should also observe an effect for the counties with SXSW interest in the pre-period. The next section outlines how we use this variation to identify the causal effect of social media on stock market participation.

3.3 Empirical model

To estimate the causal effect of Twitter adoption, we use the following first-difference IV model:

$$\Delta y_i = \alpha + \beta_1 \hat{x}_i + \beta_2 w_i + \mathbf{X_i'}\omega + \varepsilon_i, \tag{3}$$

where Δy_i is the difference in stock market participation between the post-event period (in our baseline, 2015) and the pre-event period (the year before Twitter's launch, 2005) for each county i. To study dynamic effects over time, we vary the event period to refer to the years between 1990 and 2019, which yields year-specific estimates for β_1 and β_2 akin to an event-study design. $\hat{x}_i$ is the predicted number of Twitter users in county i (measured as of 2015, in natural logarithm) based on our instrument, the number of SXSW 2007 attendees in natural logarithm (with one added inside). w_i again is our placebo treatment based on the number of SXSW followers who joined Twitter before the festival in natural logarithm (with one added inside). $\mathbf{X_i'}$ is a vector of control variables, which always includes population decile fixed effects to flexibly control for population size and Census division fixed effects. We also report regressions with more extensive control sets, incorporating geographic controls (population density), socioeconomic controls (ethnic composition, educational attainment, age shares, inequality, median household income), controls for other media usage (proxied using the prime-time TV viewership to population ratio and the Facebook Social Connectedness

Index), and controls for the 2000s housing boom and bust (changes in house prices, debt-to-income ratios, and growth in mortgage debt). We weight observations by county population and cluster standard errors by state.

In this specification, β_1 measures changes in the outcome y_i relative to 2005, depending on the number of SXSW 2007 attendees in a county. To interpret the IV estimate β_1 in Equation (3) as the causal effect of Twitter usage requires that, conditional on $\mathbf{X_i'}$, the instrument z_i in Equation (2) is excludable. In other words, the exclusion restriction requires that the number of SXSW 2007 attendees should only affect changes in stock market participation through its effect on Twitter adoption.[13] Because SXSW is mainly a culture, media, and tech event–and financial firms played no role in SXSW 2007–we believe the assumption of excludability is plausible.

By differencing the dependent variable over time, we implicitly control for any time-invariant county-level characteristics that drive stock market participation. As such, the identifying assumption underlying this approach is *not* that SXSW attendance in 2007 was random. Instead, it assumes that county-level stock market participation would have followed similar trends in the absence of the SXSW adoption shock. In the subsequent section, we show that there were no pre-existing trends in stock market participation before 2007 depending on the number of SXSW attendees in a county.

One potential concern regarding the exclusion restriction is that omitted variables could be correlated with selection into SXSW attendance. For example, SXSW attendees might be from counties that adopted online brokerage apps (such as Robinhood) for reasons other than social media, which may in turn have increased stock market participation. To mitigate this concern, we incorporate a large set of county characteristics as control variables in the vector $\mathbf{X_i'}$ in Equation (3). Importantly, we control for several variables capturing pre-existing interest in Twitter or SXSW. These include: (1) our placebo treatment based on the number of followers of the SXSW account before SXSW 2007, (2) search interest for SXSW on Google before SXSW 2007, and (3) the linear distance from Austin, Texas. Conditional on these controls, the identifying variation comes exclusively from the *difference* in SXSW 2007 attendance relative to what one would expect based on pre-existing interest in Twitter and SXSW (as well as other covariates).

[13]The other two assumptions of the IV model are monotonicity and instrument relevance. Figure 2 shows that Twitter usage increases monotonically with the number of SXSW attendees in a county. Table A.2 shows that a regression of Twitter usage on the number of SXSW 2007 attendees is highly statistically significant. In Section 4.2, we show that the corresponding F-statistic is usually in the range of 50; weak instrument bias is thus not a concern in our setting.

To illustrate, consider the scenario where SXSW 2007 attendance is correlated with an omitted variable, such as a county's overall affinity to new technologies. Any such omitted variable should also be correlated with interest in Twitter and SXSW *before* 2007, and this pre-existing interest should thus also predict changes in stock market participation. However, as we will show in Section 4.3, the estimates for our placebo treatment are consistently indistinguishable from 0 and also by an order of magnitude smaller than our main estimates. Since the counties of our placebo test are observationally equivalent to our main instrument, our findings are unlikely to be driven by omitted variables. In other words, our identification strategy compares two sets of counties that are observationally equivalent (see Figure A.5) and both of which are interested in SXSW. However, one set of counties receives an inflow of Twitter users due to the festival and therefore exhibits higher levels of Twitter usage (our instrument), while the other set of counties remains at the level of Twitter usage it would have had in the absence of the Twitter launch event at SXSW 2007 (our placebo). Our empirical strategy then analyzes if this shock to Twitter usage impacted stock market participation.

Furthermore, we can show that stock ownership was on parallel trends for counties with high and low numbers of SXSW participants all the way back to 1990, including during the internet bubble and subsequent bust in the late 1990s and early 2000s. In Section 4.5, we also provide extensive evidence that our results cannot be explained by differences in county exposure to the Great Recession and the boom preceding it, either. Taken together, we interpret these findings as evidence that omitted factors are unlikely to explain the link between SXSW, Twitter adoption, and stock ownership.

Another factor that could potentially affect the interpretation of our results may be that social media affects stock ownership only indirectly by changing another factor such as wealth. For example, if Twitter increases the overall income in a country, we would expect this higher income to translate into higher stock market participation. This is not a concern for our identification strategy per se, but it could change the interpretation of the coefficients of interest. However, as we show below, the number of SXSW 2007 attendees is uncorrelated with changes in GDP, personal income, employment, and the share of employment in the financial or tech sectors. As such, we interpret our estimates as likely capturing the direct effect of social media rather than an indirect effect working through an established correlate of stock ownership such as income.

4 Results

This section presents the main results on the effect of social media usage on stock market participation. Section 4.1 presents some introductory correlations. Section 4.2 describes the baseline regression results. Section 4.3 shows dynamic event study estimates. Section 4.4 differentiates between the intensive and extensive margin of stock ownership. Section 4.5 provides robustness checks and Section 4.6 looks at heterogeneous effects.

4.1 Introductory correlations

Figure 2 shows a binned scatter plot of dividends per capita (our main measure of stock market participation) for the year 2015 as a function of the number of Twitter users in a county. We create 30 bins with approximately 100 counties in each and control for a county's population. The binscatter shows that social media usage and stock ownership are highly correlated. The magnitude of this correlation is substantial. The linear regression underlying the linear fit line in the binscatter suggests that a 10% higher adoption of Twitter is associated with a 2.7% higher stock market participation.

Of course, it is not possible to draw causal conclusions from Figure 2. Counties differ along many observable and unobservable dimensions, and it is likely that some of the factors that cause people to adopt social media also make them more likely to own stocks. An obvious example could be differences in wealth. If Twitter is more popular with people living in wealthier areas, and wealthier individuals are more likely to invest in the stock market, social media may have no causal effect. In the next section, we present the results based on our IV strategy which allows us to overcome this identification challenge.

4.2 IV estimation

Table 2 presents the IV estimates from jointly estimating equations 2 and 3. The dependent variable is the change in log dividends per capita between 2005 and 2015. The reduced-form and IV coefficients are statistically significant at the 1%-level. The IV coefficient of 0.177 in the most saturated model in column (5) implies that a 10% increase in Twitter usage is associated with a 1.8% higher growth in dividends per capita. This magnitude is somewhat smaller than the estimated effect of broadband internet on stock market participation studied by Hvide et al. (2022). The F-statistics for the first-stage relation between SXSW attendance and Twitter usage range between 30 and 80. Bias as a result of a weak instrument is thus an unlikely concern for our results.

In line with the exclusion restriction, we do not find any effect on stock market participation for "placebo counties" with SXSW followers who joined Twitter before the festival. These estimates are small, statistically insignificant, and, if anything, negative. This suggests that counties with more SXSW 2007 attendees would not have seen higher stock market participation in the absence of the SXSW-induced increases in Twitter adoption. The existence of omitted variables, therefore, seems unlikely to explain our findings, as any such omitted variable should also affect the counties with interest in SXSW in the pre-period.

Table 3 shows that the results are similar when using alternative measures of stock market participation. In columns (1) and (2), we find that social media adoption also matters for changes in dividends per tax filing and the share of dividends in total income. Columns (3)-(5) look at rates of participation as of 2015. These estimates imply that a 10% higher Twitter adoption increases the level of dividends per capita or per tax filing by around 2.5%, and the share of income coming from dividends by around 2.2%.

An important question is whether social media only increases ownership of stocks or also enhances the quality of financial decision-making. Hvide et al. (2022), for example, show that the rollout of broadband internet in Norway improved portfolio performance of retail investors. Columns (6) and (7) provide some suggestive evidence that social media adoption is associated with higher wealth gains, not just higher participation rates. In column (6), we examine the change in total stock market wealth per capita between 2005 and 2015, taken from Chodorow-Reich, Nenov, and Simsek (2021). The point estimate of 0.072 is statistically significant at the 1% level and suggests a 0.8% increase in stock market wealth for every 10% increase in Twitter adoption. Column (7) proxies for the return equity investors achieve by looking at the ratio of capital gains (a measure of changes in wealth) relative to the previous year's level of stock market wealth.[14] The coefficient of 0.005 suggests that 10% higher exogenous Twitter adoption is associated with a 1.7% ($0.005 * 10\%$ increase relative to the mean of capital gain, 0.03) increase in returns on initial wealth. While this set of tests can only be interpreted as suggestive, they provide some indication that social media may improve the quality of financial decision-making, in addition to its effect on stock ownership rates. We will investigate this possibility further in Section 5.

[14]The IRS data on capital gains has several limitations as a measure of returns. First, it includes gains from selling any asset, not just stocks. Second, these data only become available in 2010, so we cannot look at their change relative to the pre-period. Third, we lack administrative data on stock market wealth we could use in the denominator, so we rely on the capitalized dividend method from Chodorow-Reich, Nenov, and Simsek (2021).

4.3 Dynamic estimates

Next, we investigate the dynamic effect of social media adoption on stock ownership. Figure 3 plots the results of estimating variants of Equation (3), where we vary the horizon of the dependent variable. For each year since the start of the IRS data, we calculate the change in stock market participation relative to 2005, akin to an event study specification. For example, 2015 marks the specification reported in column (6) of Table 2.

Our findings indicate that counties with many SXSW 2007 attendees followed similar trends in stock market participation to other counties between 1990 and 2010. These results support the parallel trends assumption, suggesting that our instrument does not capture unobserved factors correlated with SXSW attendance. If selection into attending SXSW in 2007 was driving our results, these counties should also show differences in stock market participation *before* the 2007 event. However, we find no evidence that this is the case, even during the period of the internet bubble and ensuing bust. As such, our instrument does not seem to be picking up a broader interest in new technologies, but instead the specific exposure to social media adoption.

After 2010, we observe a rapid increase in stock market participation in counties with many SXSW attendees. This timing aligns closely with the spike in Twitter accounts in 2009-10 and the effect of SXSW on Twitter adoption we documented in Figure 1. We interpret these patterns as evidence that Twitter adoption spurred by SXSW 2007 caused an increase in stock market participation in the United States.

In Appendix Figure A.6, we also show the dynamic estimates for our placebo treatment, which shows no significant changes in stock market participation.

4.4 Intensive vs Extensive Margin

We next investigate if Twitter increases investments by people who already own stocks (intensive margin) or convinces more people to purchase stocks in the first place (extensive margin). To see how our measure of stock ownership can be decomposed into the extensive and intensive margins, consider the following identity:

$$\frac{\text{Dividends paid}}{\text{\# of tax returns}} = \underbrace{\frac{\text{\# of returns with dividends}}{\text{\# of tax returns}}}_{\text{extensive margin}} \times \underbrace{\frac{\text{Dividends paid}}{\text{\# returns with dividends}}}_{\text{intensive margin}}$$

Table 4 reports the results from three regressions. Column 1 uses dividends per tax return as the dependent variable, which combines the extensive and intensive margin of stock market participation. Columns 2 and 3 report the results on the extensive and intensive margin, respectively.[15]

We find that social media increases stock market participation at both the extensive and intensive margin, but the magnitudes differ substantially. The coefficient of 0.09 in column 2 suggests that a 10% higher Twitter usage increases the number of tax filers reporting dividend income by around 0.9%. Column 3 shows a larger effect on the intensive margin of 1.5% for the same increase in Twitter usage. Put differently, we estimate that two thirds of the effect of social media on stock market participation in column 1 is driven by people who already own stocks, and only one third by people buying stocks that did not before.

4.5 Robustness and alternative explanations

In this section, we conduct a series of robustness tests and address potential alternative explanations.

As a first robustness check, we consider alternative specifications in Table A.3. In column (1), we find similar magnitudes to our baseline estimates when we do not weigh by population. In column (2), we restrict the sample to counties where we have either SXSW attendees in March 2007 or SXSW followers before 2007 Column (3) addresses concerns about outliers in our data by winsorizing the dividends per capita measure. In column (4), we apply the inverse hyperbolic sine instead of the logarithmic transformation to all dependent and independent variables.[16] None of these changes make a material difference to our estimates.

Our empirical strategy relies on the exclusion restriction, requiring that SXSW 2007 attendance affects stock market participation only through its impact on Twitter usage. To hold constant other channels, our main results are already based on a specification with a comprehensive set of county characteristics as control variables. However, there are two potential confounding factors that one might consider concerning. First, counties with SXSW 2007 attendees might be more likely to adopt new technologies, and perhaps those related to financial technology in particular. Second, the housing boom of the 2000s and the 2007-08 financial crisis might have affected counties with many SXSW attendees in a differential way. We address these concerns in turn.

[15] Note that the variable on the number of reports containing dividends is only available since 2010. As such, we restrict our analysis to the "level" of stock market participation in 2015 rather than the change relative to 2005.

[16] In contrast to the log transformation, the inverse hyperbolic sine is defined for zero values.

Technology Adoption. First, one possible scenario could be that counties with higher SXSW 2007 attendance are more likely to adopt any kind of technology in a way that increases productivity and ultimately higher labor incomes. If this is the case, we may not be able to identify the effect of social media, but would rather pick up the well-documented relation between income and stock ownership. To investigate this possibility, we look at changes in GDP, personal income per capita, and overall employment between 2005 and 2015 as dependent variables following Equation 3. The results in columns 1-3 in Table 5 suggest that the number of SXSW 2007 attendees is not correlated with changes in economic conditions over the sample period. As such, it is unlikely that the effect of social media adoption on stock market participation we document is explained by income effects.

Second, it could be that counties with many SXSW 2007 attendants tend to be more likely to adopt financial technologies in particular, such as digital brokerage apps (e.g., Robinhood). If these new technologies allow individuals to access information at a lower cost, they could increase stock market participation. If individuals adopt Robinhood *because* of Twitter, this would note be a concern per se. However, when we relate changes in employment in the finance and tech industry to our instrument using Equation 3, the results reported in Table 5 suggest that SXSW 2007 attendance is not correlated with changes in these variables over the time period we study. We also use Google Search Interest in "Robinhood" as a proxy for interest in digital brokerage apps, which we can do on the level of Designated Media Areas (instead of counties). There is essentially no correlation between SXSW 2007 attendees and searches for "Robinhood."

Third, it is worth noting that a higher adoption of (financial) technologies in the absence of the SXSW festival is inconsistent with the placebo results based on the pre-period interest in SXSW. If counties with interest in SXSW would be more likely to adopt technologies over this time period even in the absence of the SXSW festival, we should also observe a positive correlation between this placebo variable and stock ownership, but we do not.

Housing Boom and Financial Crisis. Two sets of results make it unlikely that our findings are driven by this alternative explanation. First, the event study patterns in Figure 3 are entirely inconsistent with the idea that we are picking up county differences in business cycle exposure. SXSW-induced differences in Twitter adoption across counties were uncorrelated with with trends in stock market participation before around 2010, going back all the way to 1990. This time period spans the 1990-91 recession, the dot-com bubble of the 1990s and 2001 recession, the 2000s housing boom, and the 2007-08 financial crisis. Put differently, the major macroeconomic events during the 20-year period before Twitter reached mass adoption were not associated with differences in stock market adoption in counties with higher SXSW attendance.

Second, controlling for exposure to the housing boom and bust has no effect on our estimates. Table 6 shows that our coefficient of interest remains virtually unchanged when we control for proxies of a county's exposure to the housing boom and subsequent recession. In particular, we control for the following variables: (1) the change in house prices index during the boom, 2002-2006 (2) the change in house prices index during the bust, 2006-2009 (3) the expansion in mortgage credit, (4) the household debt-to-income ratio as of 2006, and (5) the change in household debt-to-income ratios during the boom 2002-2006. We first add these metrics to our baseline specification one by one, and then control for them jointly. These variables have limited predictive ability for changes in stock market participation between 2005 and 2015.

4.6 Heterogeneous effects

The results in the previous sections suggest that social media can affect stock market participation. In this section, we explore heterogeneous effects within and across counties. This exercise provides evidence on which groups are more likely to react in their investment behavior to a shock to social media adoption, which matters from a policy perspective.

As a starting point, we investigate heterogeneity by income, a key predictor of stock ownership. We present the results of estimating the effect of social media on changes in dividends per capita across the six income buckets that are available in the IRS data. Table 7 shows that the effect of social media adoption is concentrated among people with incomes of $100,000 and above, who are already disproportionately more likely to hold stocks. The coefficient 0.134 in column (6) implies that a 10% increase in social media usage is associated with an 1.3% higher stock market participation among high-income earners. These results are consistent with the finding in section 4.4 that the majority of the social media effect is driven by the intensive margin of existing investors investing more, rather than new investors entering the market.

Next, we investigate heterogeneity across counties. We implement these tests by interacting the instrument, SXSW attendees, with county-level variables in a reduced form specification. Table 8 plots the estimates. Column (1) shows that the effect of social media usage on stock market participation is larger in areas with higher inequality. Consistent with our results in Table 7, this suggests that social media interacts with pre-existing differences in income inequality when affecting stock ownership. Similarly, column (2) shows an interaction with a higher initial poverty rate. Column (3) shows a positive interaction with city size, as measured by population, and column (4) shows a negative interaction with the share of the population identifying as white. This could indicate that Twitter might matter more for groups who are traditionally underrepresented when it comes to stock market investments.

5 Mechanisms

The existing literature is divided on whether social media makes market participants more or less informed. On one hand, there is evidence that social media can provide useful information about, among others, firm earnings (Bartov, Faurel, & Mohanram, 2018) and the likelihood of merger withdrawals (Cookson, Niessner, & Schiller, 2022), and institutional investors routinely use social media information when making investment decisions (Coalition Greenwich, 2019). On the other hand, social media has been tied to behavioral biases (Heimer, 2016), information silos (Cookson, Engelberg, & Mullins, 2022), and trading frenzies (Pedersen, 2022; Tengulov et al., 2021). In this section, we study both of these mechanisms in the context of the rollout of Twitter across U.S. counties, which allows us to speak to the real effects on investor portfolios.

5.1 Information provision, financial advice, and investor beliefs

A first potential channel through which social media could affect stock market investments is by providing information. The idea is that, by providing curated and targeted content for free, social media platforms could considerably reduce the fixed costs of getting access to financial advice.

As an initial set of tests, we run reduced-form regressions similar to those in Table 8 where we interact our instrument with several proxies for the availability of information about stock markets before Twitter's launch. If Twitter provides useful information to investors, we would expect that it has a larger effect in counties where such information was previously harder to come by. To aid interpretability, we standardize the instrument and all proxies for access to information to have a mean of zero and standard deviation of one.

Table 9 plots the results. Column (1) shows that the interaction of SXSW attendance with the number of financial advisors per capita of -0.036 is negative and statistically significant at the 5% level. The estimates imply that the effect of social media on stock ownership is 45% larger in a county with a one standard deviation lower number of financial advisors before. This finding is consistent with Twitter as a partial substitute for offline information via financial advisors.

Columns (2)-(4) use alternative proxies for the pre-existing information environment. Column (2) shows a negative interaction with the number of bank branches per capita. Because banks also serve as a common source of information about stock market investments, this corroborates our finding on financial advisors. Column (3) uses survey data from the National Financial Capability Study on the share of individuals who believe they have considerable knowledge about financial markets as a proxy. Column (4) uses the share of people with a bachelor's degree or higher. Both proxies also point in the direction of social media as providing information to previously underserved population.

To more directly test the idea that social media could partially substitute for other sources of information about the stock market, we look at the effect of Twitter adoption on financial advisors. In particular, we use data from Charoenwong, Kwan, and Umar (2019) and rerun our baseline regression model in Equation (3), where the dependent variable is now the log-difference in financial advisors in a county. Figure 4 shows event study estimates where the excluded period is 2007. Twitter adoption was not associated with differences in the number of financial advisors before the pivotal SXSW event. From 2010 onwards, however, we find an increasing and statistically significant negative effect, consistent with Twitter adoption causing a displacement of offline financial advice. Table A.4 plot regression estimates, suggesting a decline of around 1% in the number of financial advisors for every 10% increase in Twitter usage. These findings suggest that social media can provide access to information, which leads to a decrease in the demand for offline financial advice or the profitability for firms to provide it.

Do higher rates of social media adoption indeed make potential investors better informed? To shed light on this question, we turn to individual-level survey data from Gallup. These data contain two key questions that are relevant for our analysis: (i) a question on whether an individual owns stocks directly or through a fund or retirement plan, and (ii) a question on which types of assets individuals think are the "best long-term investment."

Equipped with these data, we estimate modified versions of equations 2-3, which replace the dependent variable with dummy variables for whether an individual owns stock or whether they believe an asset is the best investment choice. We estimate these models using a linear probability model. In addition to our county-level control variables, we also control for individual-level demographic information such as age, gender, income group, education, race, and marital status. Table 10 presents the resulting estimates.

Column (1) shows the results for stock ownership. The coefficient for Twitter adoption is 0.042 and significant at the 5% level. A 10% increase in Twitter usage is associated with a 0.5% increase in the probability of an individual owning a stock. This finding aligns closely with our main results based on the county-level measure of stock market participation. In columns (2)-(6), we investigate people's evaluation of different asset classes. Exposure to social media increases the share of people who regard stocks as the best long-term investment. The coefficient in column (2) suggests that a 10% increase in social media usage is associated with a 0.5% increase in the share of people who regard stocks as the best investment. For gold, real estate, and other investments, we find negative coefficients, which are statistically significant for gold and real estate.

To put these results in perspective, it is important to note that stocks are trusted considerably less than would be warranted by their expected risk-adjusted returns vis-à-vis other asset classes. Although stocks have considerably higher expected returns and Sharpe ratios than gold, they do not enjoy a much better reputation: around 24% of people believe stocks are the best investment, while 18% believe gold is. As such, our interpretation of the evidence in Table 10 is that social media can potentially facilitate information acquisition about the return profile of different asset classes.

5.2 Echo chambers and trading frenzies

Social media differs from the traditional news media with respect to the production and dissemination of information. While traditional media rely on strongly moderated editorial oversight, social media facilitate the peer-to-peer transmission of information with limited fact-checking or accountability for the accuracy of information (Chawla et al., 2021; Jiao, Veiga, & Walther, 2020). Taking Twitter as an example, users acquire information from "tweets" posted by accounts they follow, and can subsequently propagate this information via retweeting it. Such information diffusion patterns could lead to what Han, Hirshleifer, and Walden (2022) call "self-enhancing transmission bias", where an information sender is more likely to share investment strategies when their performance is good than bad, but an information receiver may neglect such selection bias. Social media could also enable the formation of "echo chambers" (Cookson, Engelberg, & Mullins, 2022), where information receivers self-select into information that resonates with their own view and interpret repetitive information as genuinely new information. This dynamic could boost the confidence of behavioral traders and give rise to persistent disagreement (Cookson, Engelberg, & Mullins, 2022; DeMarzo, Vayanos, & Zwiebel, 2003). As such, investors may overreact to information found on social media, leading to elevated trading volumes and increased return volatility (Jiao, Veiga, & Walther, 2020).This phenomenon could be particularly pronounced among retail investors, who tend to be relatively inexperienced and more subject to behavioral biases (e.g., Barber & Odean, 2000, 2001; Coval & Shumway, 2005; Odean, 1998).

We test the hypothesis that an exogenous adoption of social media usage is associated with higher investment in volatile, lottery-type assets by focusing on "meme stocks." These stocks are arguably an example of investments where trading volume and price movements are often driven by sentiment and heated discussions on social media rather than by fundamental information. Meme stocks such as GameStop and AMC gained popularity in early 2021 because of an investment craze initially started by trading communities on social media platforms like Reddit, Twitter, and Facebook. At the time, these stocks were traded mostly based on their popularity among retail traders rather than their underlying fundamentals, which may exacerbate deviations from the stocks' true value and bring about periods of high volatility (Tengulov et al., 2021). In January 2021, a short squeeze of GameStop investors targeting the short selling of certain hedge funds was triggered by users posting on the subreddit r/WallStreetBets, causing the price of GameStop to spike within a short time period (The Guardian, 2021). Another widely-discussed topic popular among these communities is cryptocurrencies, with considerable investor attention fueled by discussion and interaction on social media (T. Li, Shin, & Wang, 2021).

Figure 5 plots event study estimates where we relate SXSW-induced Twitter adoption to Google searches for "GME" and "AMC," the tickers of the two most prominent meme stocks around the onset of the short squeeze episode in early 2021. These figures show a significant increase in search interest in counties with higher Twitter usage starting from January 2021 onward, when the GME short squeeze gained widespread attention on social media and the news media. Before this shock to the salience of meme stocks, we find no statistically significant difference depending on a county's Twitter adoption. Table 11 shows the results for regressions of the type in Equations 2-3 on the DMA-level, where the dependent variable is the 2021 average re-normalized search volume index (SVI) for a list of meme stocks, including GME and AMC, as well as the term "crypto." The IV estimate of 0.335 in column (1) suggests that a 10% increase in Twitter usage is associated with a 3.3% higher interest in meme stocks.

To shed more light on the idea that higher exposure to social media may be associated with higher retail ownership and increased market volatility through an "echo chamber" effect, we analyze the characteristics of stocks that are especially frequently discussed on Twitter. In particular, we collect data on how often stocks are mentioned on Twitter relative to financial news papers. The difference in the mentions of stocks on Twitter relative to financial newspapers gives us an idea of how "overrepresented" stocks are on social media, and we relate this simple measure of excess mentions to stock characteristics.

Figure 6 plots a stock's average return volatility and retail ownership share against excess Twitter mentions relative to newspapers. Stocks that are more likely to be mentioned on social media are considerably more volatile and have a higher share of retail investors. Together with our findings above, this suggests that social media may particularly attract retail investors who lack experience and tend to be more subject to behavioral biases.

6 Conclusion

Social media has become an increasingly common source of information for retail and institutional investors. However, our understanding of whether and how it affects households' personal finance decisions, including portfolio choice, is limited. This paper uses exogenous variation in Twitter adoption across counties in the United States to show that social media increases stock market participation, particularly for people with higher incomes who are likely to own stocks in the first place.

We document that Twitter-fuelled stock market investments are a double-edged sword. On one hand, it increases stock ownership in areas that traditionally had less access to financial information. It may also shift people's beliefs about the stock market in a way that is more aligned with financial theory: after a shock to social media adoption, retail investors evaluate stocks more positively relative to gold or real estate. On the other hand, we also find that social media increases the appetite to invest in risky, volatile assets such as meme stocks. While we find suggestive evidence of an increase in returns on initial wealth, social media may fuel investments in particularly volatile asset classes, which could also have implications for financial markets as a whole.

It is important to highlight that the welfare implications of our results on social media are not clear-cut. The literature on prize-linked savings accounts, for example, finds that gambling incentives can be a powerful instrument to get people to save more (see e.g. Cole, Iverson, & Tufano, 2022; Cookson, 2018). Even if they lose some money initially because of inexperienced trading in volatile stocks, social media could induce people that otherwise would not have had sufficient information to earn the equity premium, which may outweigh any potential "gambling losses" in the long-run.

Another important question is how our results should be interpreted given that, over the time period we study, stock market participation in the United States was flat or even slightly decreased.[17] One possible interpretation is that stock ownership would have decreased even more in the absence of social media.

Social media is likely here to stay. As such, a better understanding of the channels through which it can affect financial decision-making is important. While our paper makes a first step in this direction by providing evidence of its effect on participation in the stock market, we hope that future work will provide additional insights from other settings.

[17] We would like to thank Andreas Fuster for raising this point.

References

Aladangady, A., Bricker, J., Chang, A. C., Goodman, S., Krimmel, J., Moore, K. B., Reber, S., Volz, A. H., & Windle, R. A. (2023). Changes in us family finances from 2019 to 2022: Evidence from the survey of consumer finances. *Federal Reserve Board Publication*.

Barber, B. M., & Odean, T. (2000). Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors. *The Journal of Finance*, *55*(2), 773–806.

Barber, B. M., & Odean, T. (2001). Boys Will Be Boys: Gender, Overconfidence, And Common Stock Investment. *The Quarterly Journal of Economics*, *116*, 261–292.

Barber, B. M., & Odean, T. (2002). Online Investors: Do the Slow Die First? *The Review of Financial Studies*, *15*(2), 455–488.

Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can Twitter Help Predict Firm-Level Earnings and Stock Returns? *The Accounting Review*, *93*(3), 25–57.

Bernales, A., Valenzuela, M., & Zer, I. (2023). *Effects of Information Overload on Financial Markets: How Much Is Too Much?* (International Finance Discussion Papers No. 1372). Board of Governors of the Federal Reserve System (U.S.)

Bernheim, B. D., & Garrett, D. M. (2003). The Effects of Financial Education in the Workplace: Evidence From a Survey of Households. *Journal of Public Economics*, *87*(7-8), 1487–1519.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Bogan, V. (2008). Stock Market Participation and the Internet. *Journal of Financial and Quantitative Analysis*, *43*(1), 191–211.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, *2*(1), 1–8.

Briggs, J., Cesarini, D., Lindqvist, E., & Östling, R. (2021). Windfall Gains and Stock Market Participation. *Journal of Financial Economics*, *139*(1), 57–83.

Brown, J. R., Ivković, Z., Smith, P. A., & Weisbenner, S. (2008). Neighbors Matter: Causal Community Effects and Stock Market Participation. *The Journal of Finance*, *63*(3), 1509–1531.

Campbell, J. Y. (2006). Household Finance. *The Journal of Finance*, *61*(4), 1553–1604.

Charoenwong, B., Kwan, A., & Umar, T. (2019). Does Regulatory Jurisdiction Affect the Quality of Investment-Adviser Regulation? *American Economic Review*, *109*(10), 3681–3712.

Chawla, N., Da, Z., Xu, J., & Ye, M. (2021). Information Diffusion on Social Media: Does It Affect Trading, Return, and Liquidity?

Chen, H., De, P., Hu, Y. J., & Hwang, B.-H. (2014). Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *The Review of Financial Studies*, *27*(5), 1367–1403.

Chodorow-Reich, G., Nenov, P. T., & Simsek, A. (2021). Stock Market Wealth and the Real Economy: A Local Labor Market Approach. *American Economic Review*, *111*(5), 1613–1657.

Choi, J. J., Laibson, D., & Metrick, A. (2002). How Does the Internet Affect Trading? Evidence From Investor Behavior in 401 (K) Plans. *Journal of Financial Economics*, *64*(3), 397–421.

CNBC. (2019). How to Use Twitter to Get Free Investing Tips From Wall Street's Wisest. https://www.cnbc.com/2019/08/30/how-to-use-twitter-to-get-investing-tips-from-wall-streets-wisest.html

CNBC. (2021a). A large chunk of the retail investing crowd started during the pandemic, Schwab survey shows. https://www.cnbc.com/2021/04/08/a-large-chunk-of-the-retail-investing-crowd-got-their-start-during-the-pandemic-schwab-survey-shows.html

CNBC. (2021b). Social Media Is the Most Popular Source of Investment Ideas for Young Investors, CNBC Survey Finds. https://www.cnbc.com/2021/08/26/social-media-top-pick-of-young-investors-for-ideas-cnbc-survey-finds.html

Coalition Greenwich. (2015). Social Media Influencing Investment Decisions at Global Institutions. https://www.greenwich.com/press-release/social-media-influencing-investment-decisions-global-institutions

Coalition Greenwich. (2019). Investing in the Digital Age: Media's Role in the Institutional Investor Engagement Journey. https://www.greenwich.com/market-structure-technology/investing-digital-age

Cole, S., Iverson, B., & Tufano, P. (2022). Can Gambling Increase Savings? Empirical Evidence on Prize-Linked Savings Accounts. *Management Science*, *68*(5), 3282–3308.

Cole, S., Paulson, A., & Shastry, G. K. (2014). Smart money? the effect of education on financial outcomes. *The Review of Financial Studies*, *27*(7), 2022–2051.

Cookson, J. A. (2018). When Saving is Gambling. *Journal of Financial Economics*, *129*(1), 24–45.

Cookson, J. A., Engelberg, J. E., & Mullins, W. (2022). Echo Chambers. *The Review of Financial Studies*, *36*(2), 450–500.

Cookson, J. A., Lu, R., Mullins, W., & Niessner, M. (2022). The Social Signal. *Available at SSRN: https://ssrn.com/abstract=4241505 or http://dx.doi.org/10.2139/ssrn.4241505*.

Cookson, J. A., & Niessner, M. (2020). Why Don't We Agree? Evidence from a Social Network of Investors. *The Journal of Finance*, *75*(1), 173–228.

Cookson, J. A., Niessner, M., & Schiller, C. M. (2022). *Can Social Media Inform Corporate Decisions? Evidence from Merger Withdrawals* (tech. rep. 56yrj). Center for Open Science.

Coval, J. D., & Shumway, T. (2005). Do Behavioral Biases Affect Prices? *The Journal of Finance*, *60*(1), 1–34.

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, *66*(5), 1461–1499.

DeMarzo, P. M., Vayanos, D., & Zwiebel, J. (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. *The Quarterly Journal of Economics*, *118*(3), 909–968.

Dessaint, O., Foucault, T., & Frésard, L. (2021). *Does Alternative Data Improve Financial Forecasting? The Horizon Effect* (CEPR Discussion Papers No. 15786). C.E.P.R. Discussion Papers. https://ideas.repec.org/p/cpr/ceprdp/15786.html

Elliott, W. B., Grant, S. M., & Hodge, F. D. (2018). Negative News and Investor Trust: The Role of $ Firm And# CEO Twitter Use. *Journal of Accounting Research*, *56*(5), 1483–1519.

Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social Media and Protest Participation: Evidence From Russia. *Econometrica*, *88*(4), 1479–1514.

Farrell, M., Green, T. C., Jame, R., & Markov, S. (2022). The Democratization of Investment Research and the Informativeness of Retail Investor Trading. *Journal of Financial Economics*, *145*(2, Part B), 616–641.

Fujiwara, T., Müller, K., & Schwarz, C. (2023). *The effect of social media on elections: Evidence from the united states* (tech. rep.). Oxford University Press.

Gawker. (2007). Twitter Blows Up at SXSW Conference. https://gawker.com/243634/twitter-blows-up-at-sxsw-conference%20(accessed%20March%203rd,%202018)

Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2011). IQ and Stock Market Participation. *The Journal of Finance*, *66*, 2121–2164.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the Stock Market. *The Journal of Finance*, *63*(6), 2557–2600.

Guiso, L., & Sodini, P. (2012). Household Finance: An Emerging Field. *Handbook of the Economics of Finance*, *2*.

Haliassos, M., & Bertaut, C. C. (1995). Why Do So Few Hold Stocks? *The Economic Journal*, *105*(432), 1110–1129.

Han, B., Hirshleifer, D., & Walden, J. (2022). Social Transmission Bias and Investor Behavior. *Journal of Financial and Quantitative Analysis*, *57*(1), 390–412.

Heimer, R. Z. (2016). Peer Pressure: Social Interaction and the Disposition Effect. *The Review of Financial Studies*, *29*(11), 3177–3209.

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*, *75*(4), 1779–1831.

Hirshleifer, D., Peng, L., & Wang, Q. (2021). Social Networks and Market Reactions to Earnings News. *Available at SSRN 3824022*.

Hong, H., Kubik, J. D., & Stein, J. C. (2004). Social Interaction and Stock-Market Participation. *The Journal of Finance*, *59*, 137–163.

Hvide, H. K., Meling, T. G., Mogstad, M., & Vestad, O. L. (2022). *Broadband Internet and the Stock Market Investments of Individual Investors* (Working Paper No. 30383). National Bureau of Economic Research.

Jiao, P., Veiga, A., & Walther, A. (2020). Social Media, News Media and the Stock Market. *Journal of Economic Behavior & Organization*, *176*, 63–90.

Kalda, A., Loos, B., Previtero, A., & Hackethal, A. (2021). *Smart(Phone) Investing? A within Investor-time Analysis of New Technologies and Trading Behavior* (Working Paper No. 28363). National Bureau of Economic Research.

Kaustia, M., & Knüpfer, S. (2012). Peer Performance and Stock Market Entry. *Journal of Financial Economics*, *104*(2), 321–338.

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, *4*(2), 2053951717736336.

Kuchler, T., Li, Y., Peng, L., Stroebel, J., & Zhou, D. (2021). Social Proximity to Capital: Implications for Investors and Firms. *The Review of Financial Studies*, *35*(6), 2743–2789.

Li, G. (2014). Information Sharing and Stock Market Participation: Evidence from Extended Families. *Review of Economics and Statistics*, *96*(1), 151–160.

Li, T., Shin, D., & Wang, B. (2021). Cryptocurrency Pump-and-Dump Schemes. *Available at SSRN: https://ssrn.com/abstract=3267041 or http://dx.doi.org/10.2139/ssrn.3267041*.

Luo, X., Zhang, J., & Duan, W. (2013). Social Media and Firm Equity Value. *Information Systems Research*, *24*(1), 146–163. Retrieved April 12, 2023, from http://www.jstor.org/stable/42004274

Mian, A., Rao, K., & Sufi, A. (2013). Household Balance Sheets, Consumption, and the Economic Slump. *The Quarterly Journal of Economics*, *128*(4), 1687–1726.

Müller, K., & Schwarz, C. (2023). From Hashtag to Hate Crime: Twitter and Anti-minority Sentiment. *American Economic Journal: Applied Economics*, *15*(3), 270–312.

Odean, T. (1998). Are Investors Reluctant to Realize Their Losses? *The Journal of Finance*, *53*(5), 1775–1798.

Pancost, N. A., & Schaller, G. (2022). Measuring measurement error. *Available at SSRN 4045772*.

Pedersen, L. H. (2022). Game On: Social Networks and Markets. *Journal of Financial Economics*, *146*(3), 1097–1119.

Statista. (2022). Social Media - Statistics & Facts. https://www.statista.com/topics/1164/social-networks/#dossierKeyfigures

TechCrunch. (2011). Twitter And Foursquare Explain Their SXSW Explosions: Hustle, Buzz, And Maybe $11K. https://techcrunch.com/2011/01/04/twitter-foursquare-sxsw/

Tengulov, A., Allen, F., Nowak, E., & Pirovano, M. (2021). *Squeezing Shorts Through Social News Platforms* (Swiss Finance Institute Research Paper Series No. 21-31). Swiss Finance Institute.

The Guardian. (2021). GameStop: How Reddit Amateurs Took Aim at Wall Street's Short-Sellers, by Rob Davies.

Van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial Literacy and Stock Market Participation. *Journal of Financial Economics*, *101*(2), 449–472.

Vissing-Jørgensen, A. (2002). Limited Asset Market Participation and the Elasticity of Intertemporal Substitution. *Journal of Political Economy*, *110*(4), 825–853.

Table 1: Descriptive Statistics

	Mean	Std.	Min.	25%	50%	75%	Max.	N
Dividend, Income and Twitter variables								
Δ Log(Dividend)	0.34	0.38	-2.15	0.16	0.35	0.54	3.85	3104
Δ Log (Dividend(%))	-0.02	0.33	-2.17	-0.14	0.02	0.16	2.89	3104
Δ Log (Dividend p.c.)	0.30	0.35	-2.06	0.15	0.32	0.48	3.58	3104
Δ Log (Dividend per tax filers)	0.16	0.33	-2.26	0.03	0.18	0.33	3.42	3104
Δ Log (Stock wealth p.c.)	0.26	0.30	-3.06	0.17	0.26	0.36	10.55	3083
Capital wealth	0.03	0.03	0.00	0.02	0.02	0.03	1.35	3106
Log(Twitter users)	5.29	1.76	0.00	4.06	5.13	6.33	12.35	3108
Log(SXSW followers, pre)	0.02	0.18	0.00	0.00	0.00	0.00	3.61	3108
Log(SXSW 2007 attendees)	0.06	0.32	0.00	0.00	0.00	0.00	4.98	3108
Log(Google search for SXSW, pre)	0.93	0.88	0.00	0.00	1.10	1.39	4.62	3108
Demographic controls								
% aged 20-24	0.06	0.02	0.01	0.05	0.06	0.07	0.27	3108
% aged 25-29	0.06	0.01	0.03	0.05	0.06	0.07	0.15	3108
% aged 30-34	0.06	0.01	0.03	0.05	0.06	0.06	0.12	3108
% aged 35-39	0.06	0.01	0.03	0.05	0.06	0.06	0.11	3108
% aged 40-44	0.06	0.01	0.02	0.05	0.06	0.06	0.10	3108
% aged 45-49	0.06	0.01	0.02	0.06	0.06	0.07	0.09	3108
% aged 50+	0.39	0.07	0.11	0.35	0.39	0.43	0.75	3108
Geographical controls								
Population density	261.27	1733.47	0.10	17.60	45.60	114.85	69468.40	3108
Distance from Austin, TX (in miles)	1450.64	612.61	5.04	1055.25	1464.66	1863.85	3098.88	3108
Race and religion controls								
% white	0.77	0.20	0.03	0.65	0.84	0.93	0.98	3108
% black	0.09	0.14	0.00	0.01	0.02	0.10	0.85	3108
% native American	0.02	0.06	0.00	0.00	0.00	0.01	0.90	3108
% Asian	0.01	0.02	0.00	0.00	0.01	0.01	0.37	3108
% Hispanic	0.09	0.14	0.01	0.02	0.04	0.09	0.96	3108
Socioeconomic controls								
Gini index	0.44	0.03	0.33	0.42	0.44	0.46	0.65	3108
Log(Median household income)	10.72	0.24	9.87	10.56	10.71	10.86	11.72	3107
% below poverty level	16.74	6.58	1.40	12.10	16.00	20.30	53.30	3108
% adults with high school degree	34.77	7.07	7.50	30.30	35.20	39.60	54.80	3108
% adults with graduate degree	7.05	4.12	0.00	4.40	5.80	8.30	44.40	3108
Media Controls								
% watching prime time TV	0.43	0.01	0.40	0.43	0.43	0.44	0.47	3107
Log(Facebook SCI)	9.81	1.01	6.95	9.16	9.82	10.49	13.24	3107
Recession Controls								
HPI Change 2002-2006	29.57	22.10	-8.23	16.17	22.09	34.02	134.94	3108
HPI Change 2006-2009	-1.22	18.54	-154.83	-3.17	2.94	7.77	50.73	3108
Mortgage Amount Change 2002-2006	0.73	0.77	-0.33	0.21	0.56	1.03	4.17	3098
Number of Mortgage Change 2002-2006	0.42	0.50	-0.36	0.08	0.33	0.65	2.46	3098
Debt to income from NY Fed 2006	1.57	0.58	0.58	1.16	1.44	1.84	4.93	2209
Debt to income Change 2002-2006	0.30	0.19	-0.36	0.18	0.28	0.40	1.61	2209
Pre-event information environment								
Financial advisors p.c.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1946
Bank branches p.c.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3088
High financial knowledge	0.10	0.02	0.06	0.08	0.10	0.11	0.15	2919

Table 2: Social Media and Changes in Stock Market Participation

Dependent variable:	$\Delta_{'05\rightarrow'15}$Log(Dividends p.c.)				
	(1)	(2)	(3)	(4)	(5)
Log(Twitter users)	0.201***	0.208***	0.198***	0.160***	0.177***
	(0.046)	(0.047)	(0.045)	(0.042)	(0.051)
Log(SXSW followers, pre)	-0.027	-0.027	-0.033	-0.028	-0.036
	(0.044)	(0.044)	(0.042)	(0.039)	(0.044)
Population deciles	Yes	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes	Yes
Selection controls		Yes	Yes	Yes	Yes
Geographical controls			Yes	Yes	Yes
Socioeconomic controls				Yes	Yes
Media Controls					Yes
Observations	3,104	3,104	3,104	3,104	3,104
Mean of DV	0.548	0.548	0.548	0.548	0.548
Robust F-stat	50.28	47.05	42.48	91.66	52.13

Notes: This table presents first difference regressions as in equation 3, where the dependent variable is the difference in *Log(Dividends per capita)* between 2005 and 2015. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles and Census division fixed effects. Selection controls include the linear distance from the SXSW festival location (Austin, Texas) and Google search intensity for the SXSW festival before 2007. Geographical controls is population density. Socioeconomic controls include age buckets (the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50) and ethnic composition (the share of people identifying as white, African American, Native American or Pacific Islander, Asian or Hispanic), the Gini coefficient, log median household income, the share of high school graduates, and the share of people with a graduate degree. Media controls include the prime time TV viewership to population ratio and the Facebook Social Connectedness Index (SCI). Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 3: Alternative Measures of Stock Market Participation and Returns on Wealth

Dependent variable:	$\Delta_{'05\to'15}$Log(Participation)		Log(Participation in 2015)			Returns on wealth	
	$\frac{Dividends}{Tax\ filers}$	$\frac{Dividends}{Income}$	$\frac{Dividends}{Population}$	$\frac{Dividends}{Tax\ filers}$	$\frac{Dividends}{Income}$	$\Delta\text{Log}(\frac{St.\ wealth}{Pop.})$	$\frac{\text{Capital gains}}{\text{Stock wealth}}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(Twitter users)	0.169***	0.136***	0.251**	0.241**	0.216***	0.072**	0.005***
	(0.050)	(0.044)	(0.097)	(0.092)	(0.079)	(0.027)	(0.001)
Log(SXSW followers, pre)	-0.030	-0.029	-0.037	-0.043	-0.043	-0.024	-0.000
	(0.043)	(0.035)	(0.070)	(0.068)	(0.053)	(0.024)	(0.001)
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Media Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,104	3,104	3,104	3,104	3,104	3,083	3,105
Mean of DV	0.377	0.192	-0.593	0.178	-3.971	0.274	0.030
Robust F-stat	52.13	52.13	52.13	52.13	52.13	52.41	52.37

Notes: This table presents regressions as in equation 3. The dependent variables are different measures of either changes in stock market participation between 2005 and 2015 in columns 1-2 or the level of stock market participation as of 2015 in columns 3-5. In column 6, we examine changes in stock market wealth per capita between 2005 and 2015 using data from Chodorow-Reich, Nenov, and Simsek (2021). In column 7, the dependent variable is the ratio of capital gains in 2015 relative to stock market wealth in 2014, a proxy for portfolio returns. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the full set of county-level controls in Table 2. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4: Intensive vs. Extensive Margin of Stock Market Participation

Dependent variable:	Log(Participation in 2015)		
	$\frac{Dividend}{Tax\ filer}$	$\frac{No.\ Divd}{Tax\ filer}$	$\frac{Dividend}{No.\ Divd}$
	(1)	(2)	(3)
Log(Twitter users)	0.241**	0.090***	0.151**
	(0.092)	(0.025)	(0.074)
Log(SXSW followers, pre)	-0.043	-0.034*	-0.009
	(0.068)	(0.018)	(0.059)
Population deciles	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes
Media Controls	Yes	Yes	Yes
Observations	3,104	3,104	3,104
Mean of DV	0.178	-1.781	1.959
Robust F-stat	52.13	52.13	52.13

Notes: This table presents county-level regressions as in equation 3, where we differentiate between the effect of social media on the extensive and intensive margin of stock market participation. The dependent variable is a measure of stock market participation as of 2015. Column (1) shows estimates where the dependent variable is dividends per tax filer (in logarithm), which combines the extensive and intensive margins of stock ownership. Column (2) plots estimates for the extensive margin, the number of tax filings with reporting any dividend income relative to all tax filings. Column (3) focuses on the intensive margin, total dividend income per tax filing declaring any dividend income. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the full set of county-level controls in Table 2. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 5: Social Media and Other Predictors of Stock Ownership

Dependent variable:	GDP	Income p.c.	Employment	Emp. in fin. sector	Emp. in tech. sector	"Robinhood" searches
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Twitter users)	0.011	-0.007	-0.004	-0.013	0.040	-0.352
	(0.023)	(0.014)	(0.012)	(0.029)	(0.027)	(0.441)
Log(SXSW followers, pre)	0.021	0.025**	0.004	0.019	-0.017	0.402*
	(0.019)	(0.012)	(0.013)	(0.026)	(0.023)	(0.217)
Selection controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
Media Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,055	3,055	3,055	3,055	3,055	189
Mean of DV	0.121	0.284	0.097	0.193	0.143	-0.086
Robust F-stat	40.48	40.48	40.48	40.48	40.48	17.28

Notes: This table presents first difference regressions as in equation 3. In columns 1-5,the dependent variable is the difference in GDP, personal income per capita, total employment, employment in finance and insurance, employment in professional, scientific, information and technical services between 2005 and 2015 (all in logs). In Column 6, the dependent variable is the difference in the rescaled Google Search Index for "Robinhood" between 2005 and 2015 in natural logarithm with one added inside. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles and Census division fixed effects. Selection controls include the linear distance from the SXSW festival location (Austin, Texas) and Google search intensity for the SXSW festival before 2007. Geographical controls is population density. Socioeconomic controls include age buckets (the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50) and ethnic composition (the share of people identifying as white, African American, Native American or Pacific Islander, Asian or Hispanic), the Gini coefficient, log median household income, the share of high school graduates, and the share of people with a graduate degree. Media controls include the prime time TV viewership to population ratio and the Facebook Social Connectedness Index (SCI). Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 6: Robustness Tests – 2000s Housing Boom and Bust

Dependent variable:	$\Delta_{'05\rightarrow'15}$Log(Dividends p.c.)				
	(1)	(2)	(3)	(4)	(5)
Log(Twitter users)	0.177***	0.177***	0.181***	0.174***	0.193***
	(0.051)	(0.049)	(0.052)	(0.052)	(0.057)
HPI Change 2006-2009		0.000			0.000
		(0.001)			(0.001)
HPI Change 2002-2006		-0.000			-0.001
		(0.001)			(0.001)
Mortgage Amount Change 2002-2006			-0.014		0.083
			(0.046)		(0.071)
Number of Mortgage Change 2002-2006			0.057		-0.055
			(0.086)		(0.106)
Debt to income from NY Fed 2006				-0.054	-0.051
				(0.039)	(0.052)
Debt to income Change 2002-2006				0.017	0.036
				(0.073)	(0.084)
Population deciles	Yes	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes
Observations	3,104	3,104	3,097	2,209	2,208
Mean of DV	0.548	0.548	0.548	0.554	0.554
Robust F-stat	52.13	69.43	52.36	53.82	63.41

Notes: This table presents first difference regressions as in equation 3 with additional control variables for the 2000s housing boom and bust, including the change in a county's house price index, growth in mortgage debt, and changes in household debt-to-income ratios. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the full set of county-level controls in Table 2. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 7: Social Media and Stock Market Participation Across Income Groups

Dependent variable:	$\Delta_{'05\rightarrow'15}$Log(Dividends p.c.)				
	<$25k	$25-50k	$50-75k	$75-100k	>$100k
	(1)	(2)	(3)	(4)	(5)
Log(Twitter users)	0.002	0.001	-0.004	0.010	0.134**
	(0.008)	(0.006)	(0.011)	(0.016)	(0.057)
Log(SXSW followers, pre)	0.012**	0.009**	0.011	-0.004	-0.063*
	(0.005)	(0.004)	(0.008)	(0.010)	(0.036)
Population deciles	Yes	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes
Media Controls	Yes	Yes	Yes	Yes	Yes
Observations	3,106	3,106	3,106	3,106	3,105
Mean of DV	-0.164	-0.037	-0.019	0.007	0.030
Robust F-stat	52.13	52.13	52.13	52.13	52.13

Notes: This table presents first difference regressions as in equation 3, where the dependent variable is the difference in *Log(Dividends per capita)* between 2005 and 2015. We estimate seperate regressions for five income groups. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the full set of county-level controls in Table 2. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 8: Heterogeneity Tests

Dependent variable:	$\Delta_{'05\rightarrow'15}$Log(Dividends p.c.)			
	(1)	(2)	(3)	(4)
Log(SXSW 2007 attendees)	-0.675	-0.035	-0.025	-0.083
	(0.458)	(0.086)	(0.040)	(0.108)
Log(SXSW 2007 attendees) × Gini coefficient	1.739*			
	(1.024)			
Log(SXSW 2007 attendees) × Poverty rate		0.012*		
		(0.006)		
Log(SXSW 2007 attendees) × White Population (%)			-0.119***	
			(0.034)	
Log(SXSW 2007 attendees) × Log(Population)				0.082*
				(0.046)
Population deciles	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes
Observations	3,104	3,104	3,104	3,104
Mean of DV	0.548	0.548	0.548	0.548

Notes: This table presents reduced-form heterogeneity tests where we regress the difference in *Log(Dividends per capita)* between 2005 and 2015 directly on our instrument and interactions with county characteristics. *Log(SXSW 2007 attendees)* is the number of SXSW attendees in 2007 in natural logarithm (with one added inside). *Log(SXSW followers, Pre)* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the county-level controls in column 3 of Table 2 interacted with the interaction variable of interest. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 9: Social Media and Information Provision

Dependent variable:	$\Delta_{'05\rightarrow'15}$Log(Dividends p.c.)			
	(1)	(2)	(3)	(4)
Log(SXSW 2007 attendees)	0.090***	-0.025	0.081***	0.144***
	(0.019)	(0.054)	(0.018)	(0.037)
Log(SXSW 2007 attendees) × Financial advisors p.c.	-0.036*			
	(0.018)			
Log(SXSW 2007 attendees) × Bank branches p.c.		-0.139*		
		(0.070)		
Log(SXSW 2007 attendees) × High financial knowledge			-0.038**	
			(0.016)	
Log(SXSW 2007 attendees) × Share bachelor's degree				-0.041*
				(0.022)
Population deciles	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes
Media Controls	Yes	Yes	Yes	Yes
Observations	1,946	3,088	2,915	3,104
Mean of DV	0.565	0.548	0.550	0.548

Notes: This table presents reduced-form heterogeneity tests where we regress the difference in *Log(Dividends per capita)* between 2005 and 2015 directly on our instrument and interactions with variables proxy for the availability of information about stock markets before Twitter's launch. *Log(SXSW 2007 attendees)* is the number of SXSW attendees in 2007 in natural logarithm (with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the county-level controls in column 5 of Table 2 interacted with the interaction variable of interest. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 10: Survey Evidence – Stock Ownership and Investment Preferences

	Stock	What is the best long-term investment?				
	ownership	Stock	Real Estate	Fixed Income	Gold	Others
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Twitter users)	0.042**	0.053*	-0.049*	0.028	-0.042*	-0.005
	(0.020)	(0.030)	(0.027)	(0.023)	(0.025)	(0.006)
Population deciles $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Census division $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic controls	Yes	Yes	Yes	Yes	Yes	Yes
Media Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,535	12,421	12,421	12,421	7,958	12,421
Mean of DV	0.563	0.246	0.330	0.293	0.184	0.013
Robust F-stat	40.358	38.418	38.418	38.418	38.980	38.418

Note: This table presents individual-level linear probability estimates from regressing dummy variables measuring stock ownership or investment preferences on Twitter usage (instrumented with SXSW attendees). The sample period is 2007 to 2020. The dependent variable in column (1) is a dummy variable that equals to 1 if the respondent of the Gallup survey answers *Yes* to the question *Do you personally or jointly with a spouse, have any money invested in the stock market right now – either in an individual stock, a stock mutual fund, or in a self-directed 401-K or IRA?*, and 0 otherwise. The dependent variables from column (2) to column (6) are dummy variables that capture individuals' preferences for investing in different asset classes. The dependent variable *Stock* in column (2) equals to 1 if the respondent believes stock market or mutual funds are the best long-term investment compared to other asset classes when answering the question *Which of the following do you think is the best long-term investment – [bonds, real estate, savings accounts or CDs, (or) stocks or mutual funds]? Or gold?*. The dependent variables in columns (3) to (6) are defined similarly. All regressions control for population decile $\times$ year fixed effects, Census division $\times$ year fixed effects, and the full set of county-level controls in Table 2. We also include individual control variables for age groups, race, gender, income groups, education, and employment status. Standard errors in parentheses are clustered by state and we apply survey weights. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

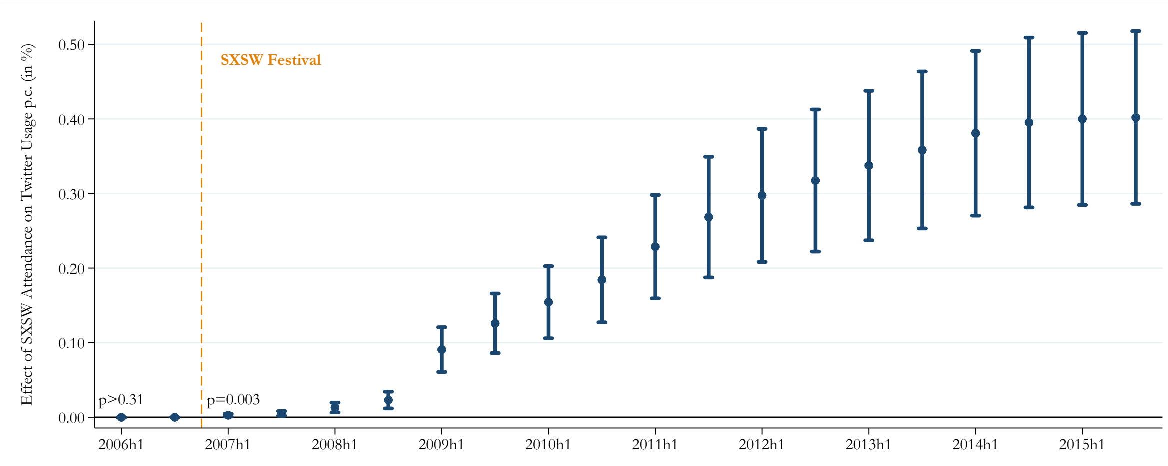
Table 11: Social Media and Search Interest in Meme Stocks

Dependent variable:	Log(Google searches for meme stocks)				
	(1)	(2)	(3)	(4)	(5)
Log(Twitter users)	0.335**	0.344**	0.366**	0.232**	0.225**
	(0.132)	(0.135)	(0.174)	(0.113)	(0.106)
Log(SXSW followers, pre)	0.038	0.026	0.032	-0.026	-0.012
	(0.056)	(0.057)	(0.051)	(0.041)	(0.041)
Population deciles	Yes	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes	Yes
Selection controls		Yes	Yes	Yes	Yes
Geographical controls			Yes	Yes	Yes
Socioeconomic controls				Yes	Yes
Media Controls					Yes
Observations	205	205	205	205	205
Mean of DV	2.500	2.500	2.500	2.500	2.500
Robust F-stat	18.66	19.70	16.79	15.53	14.13

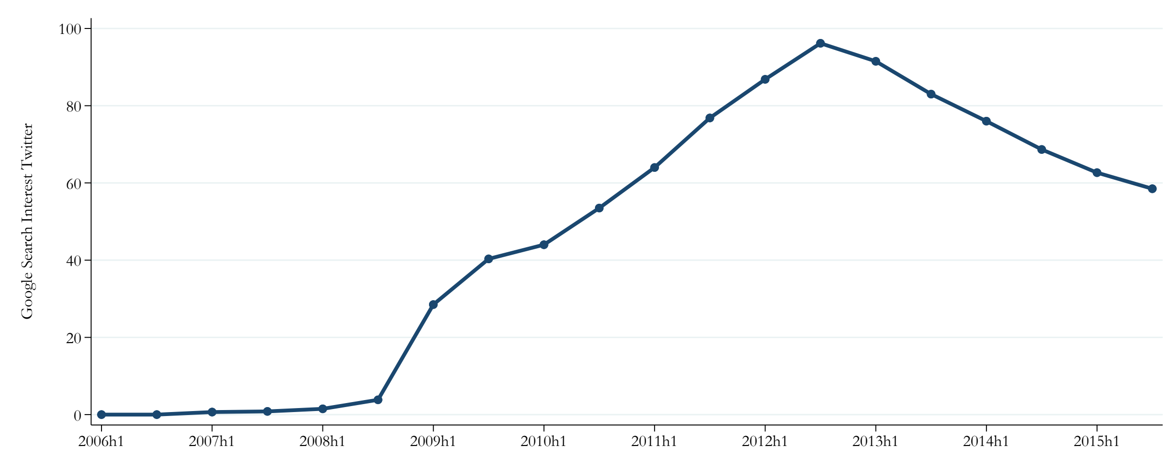
Notes: This table presents DMA-level regressions where the dependent variable is the average Google Trends SVI for a list of meme stocks in 2021, including "GME" ("GameStop Corp."), "AMC" ("AMC Entertainment Holdings Inc"), "UPST" ("Upstart Holdings Inc"), "BBBY" ("Bed Bath & Beyond Inc."), "BBIG" ("Vinco Ventures Inc"), "BB" ("BlackBerry Ltd."), "NOK" ("Nokia Corp."), "SAVA" ("Cassava Sciences, Inc"), "CLOV" ("Clover Health Investments Corp"), "PLTR" ("Palantir Technologies."), "SPCE" ("Virgin Galactic Holdings"), "LCID" ("Lucid Group, Inc. "), and "Crypto." *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the full set of county-level controls in Table 2. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Figure 1: South by Southwest 2007 and Twitter Adoption

(a) Effect of SXSW on Twitter Adoption

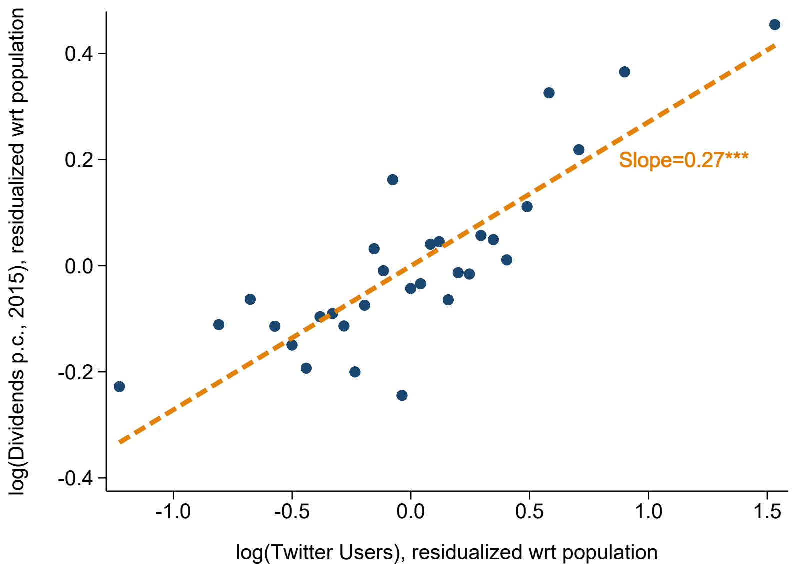


(b) Search Interest for Twitter



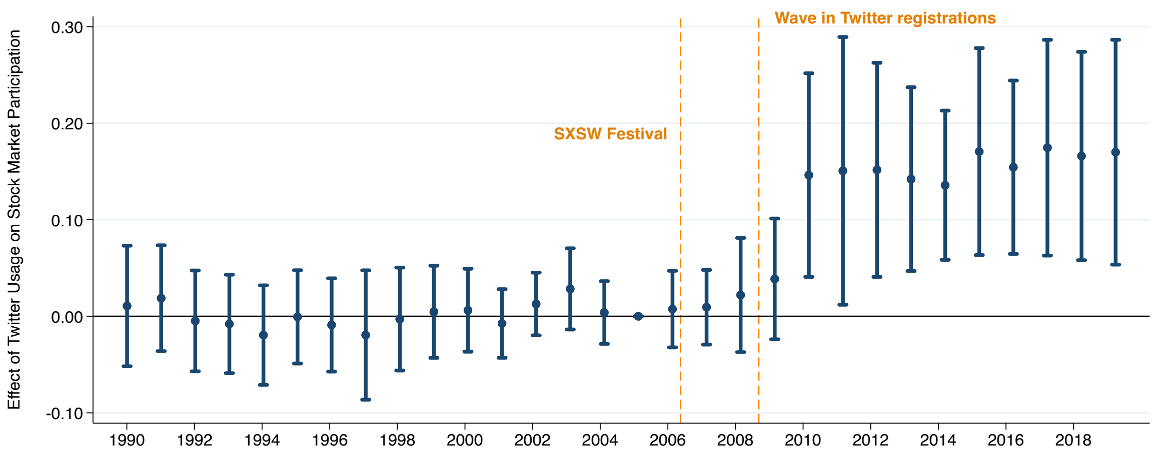
Notes: This figure provides evidence on the link between SXSW attendance in 2007 and the geography of Twitter adoption. Panel (a) shows the estimates of γ_t from a panel regression $x_{it} = \sum_{w \neq 2006h2} \gamma_t z_i \cdot \mathbb{1}\{t = w\} + \theta_i + \theta_t + \nu_{it}$, where x_{it} is Twitter users per capita in county i for the halfyear indicated on the x-axis, z_i is the natural logarithm of SXSW 2007 attendees (with one added inside). We plot 95% confidence intervals based on standard errors are clustered by county. Panel (b) plots the search interest for Twitter on Google between 2006 and 2015.

Figure 2: Correlation between Twitter Usage and Stock Market Participation



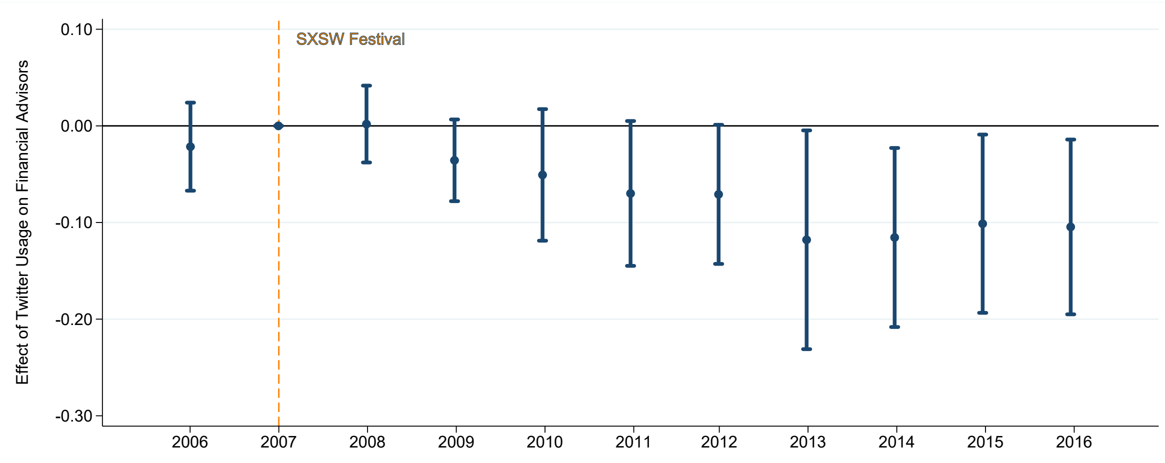
Notes: This figure shows a binned scatter plot where dividends per capita (in logs, as of 2015) are binned into 30 buckets of log(Twitter usage). Both variables are residualized with respect to population deciles.

Figure 3: Dynamic IV Estimates



Notes: This figure plots the estimates β_1 from cross-sectional IV regressions as in Equation (3), where the dependent variable is the change in dividends per capita relative to 2005 (the year before Twitter's launch). We include the full set of controls as in column 5 of Table 2. Standard errors are clustered by state, and we plot 95% confidence intervals.

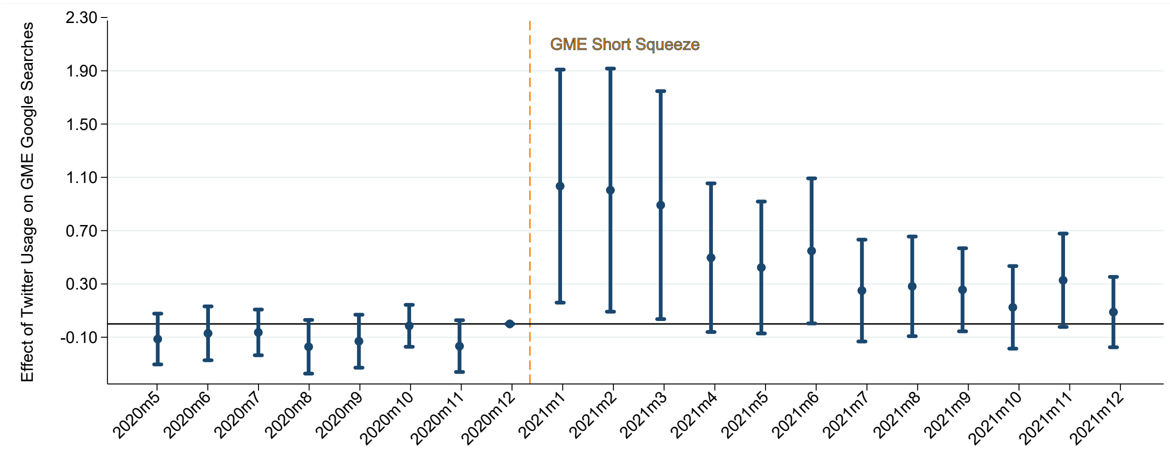
Figure 4: Social Media Adoption and Financial Advisor Displacement



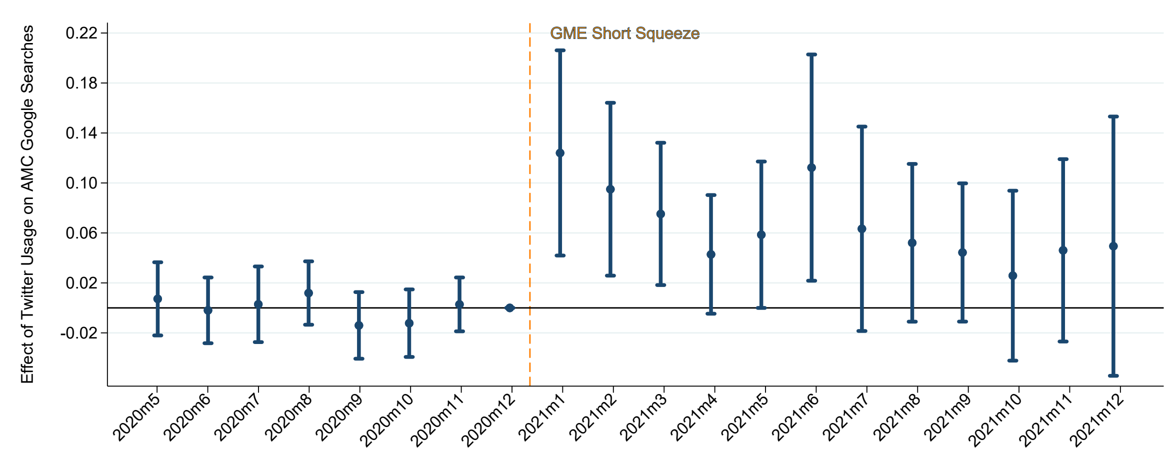
Notes: This figure plots the estimates β_1 from cross-sectional IV regressions as in Equation (3), where the dependent variable is the change in the number of financial advisors (in logs with 1 added inside) relative to 2006 (the first year for which we have data). The data on financial advisors is taken from Charoenwong, Kwan, and Umar (2019). We include the same controls as in column 4 of Table 2. Standard errors are clustered by state, and we plot 95% confidence intervals.

Figure 5: Social Media and the GameStop Short Squeeze Episode

(a) Google Searches for GME



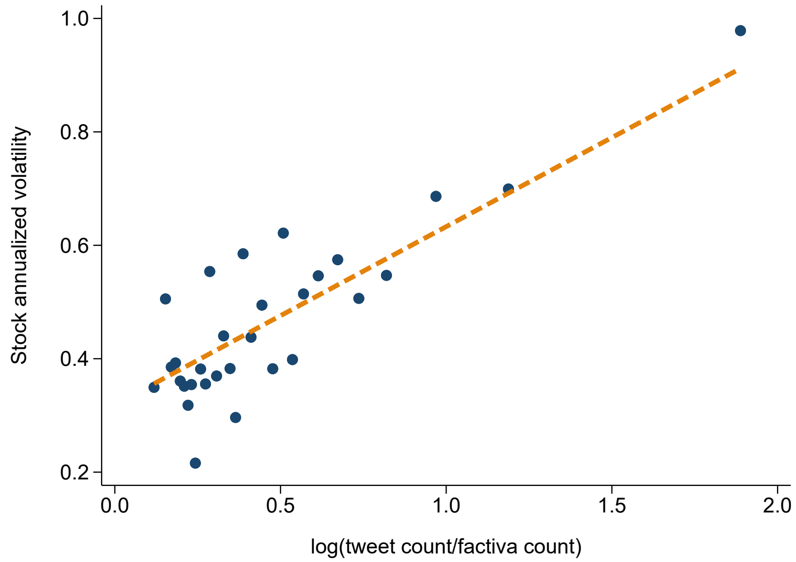
(b) Google Searches for AMC



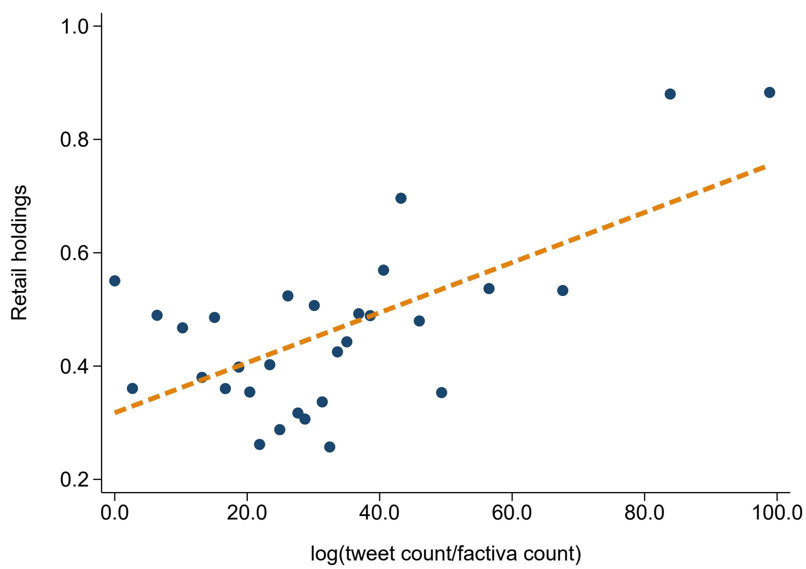
Notes: This figure plots the estimates β_1 from cross-sectional IV regressions as in Equation (3), where the dependent variable is the change in Google search index for "GME" and "AMC" relative to 2005 in panel (a) and (b), respectively. We include the full set of controls as in column 5 of Table 2. Standard errors are clustered by state, and we plot 95% confidence intervals.

Figure 6: Which Stocks Are Overrepresented On Social Media?

(a) Stock volatility



(b) Retail Ownership



Notes: These figures plots stock volatility and retail ownership against the difference between the share of a stock's mentions on Twitter and traditional media sources (the Wall Street Journal or Dow Jones Institutional News).

A Online Appendix

Table A.1: Descriptive Statistics – Gallup Data Variables

	Mean	Std.	Min.	25%	50%	75%	Max.	N
Stock ownership								
Yes	0.57	0.64	0.00	0.00	0.50	0.84	4.12	22341
No	0.42	0.74	0.00	0.00	0.00	0.63	4.12	22341
Best Long-term Investment								
Stocks or mutual funds	0.16	0.45	0.00	0.00	0.00	0.00	4.12	22341
Fixed income	0.19	0.53	0.00	0.00	0.00	0.00	4.12	22341
Gold	0.08	0.34	0.00	0.00	0.00	0.00	4.12	22341
Real Estate	0.22	0.52	0.00	0.00	0.00	0.00	4.12	22341
Others/DK	0.03	0.20	0.00	0.00	0.00	0.00	3.79	22341
Gender								
Male	0.49	0.66	0.00	0.00	0.28	0.75	4.12	22341
Female	0.51	0.73	0.00	0.00	0.00	0.79	4.12	22341
Age								
< 20	0.05	0.33	0.00	0.00	0.00	0.00	4.12	22341
20-24	0.08	0.39	0.00	0.00	0.00	0.00	4.12	22341
25-34	0.16	0.51	0.00	0.00	0.00	0.00	4.12	22341
35-44	0.17	0.51	0.00	0.00	0.00	0.00	4.12	22341
45-54	0.18	0.47	0.00	0.00	0.00	0.00	4.12	22341
55-64	0.16	0.39	0.00	0.00	0.00	0.00	4.05	22341
65+	0.19	0.40	0.00	0.00	0.00	0.30	4.05	22341
Education								
High School or Less	0.39	0.75	0.00	0.00	0.00	0.59	4.12	22341
Some college	0.30	0.59	0.00	0.00	0.00	0.44	4.12	22341
College Grad only	0.16	0.37	0.00	0.00	0.00	0.00	4.12	22341
Post-grad	0.15	0.35	0.00	0.00	0.00	0.00	4.12	22341
Martial Status								
Currently Married	0.50	0.65	0.00	0.00	0.34	0.79	4.12	22341
Divorced	0.10	0.35	0.00	0.00	0.00	0.00	4.05	22341
Living together with Partner	0.08	0.36	0.00	0.00	0.00	0.00	4.12	22341
Never Married	0.20	0.57	0.00	0.00	0.00	0.00	4.12	22341
Separated	0.03	0.22	0.00	0.00	0.00	0.00	4.05	22341
Widowed	0.07	0.28	0.00	0.00	0.00	0.00	4.12	22341

Table A.1: Summary Statistics – Gallup Data Variables (Continued)

	Mean	Std.	Min.	25%	50%	75%	Max.	N
Employment								
Employed	0.57	0.73	0.00	0.00	0.40	0.90	4.12	22341
Unemployed	0.07	0.34	0.00	0.00	0.00	0.00	4.12	22341
Homemaker	0.06	0.31	0.00	0.00	0.00	0.00	4.12	22341
Retired	0.22	0.42	0.00	0.00	0.00	0.36	4.05	22341
Student	0.06	0.35	0.00	0.00	0.00	0.00	4.12	22341
Disabled	0.03	0.21	0.00	0.00	0.00	0.00	4.05	22341
Race								
Asian	0.02	0.17	0.00	0.00	0.00	0.00	4.12	22341
Hispanic	0.11	0.46	0.00	0.00	0.00	0.00	4.12	22341
Non-Hispanic White	0.73	0.66	0.00	0.30	0.59	0.98	4.12	22341
Non-Hispanic Black	0.11	0.44	0.00	0.00	0.00	0.00	4.12	22341
Other	0.03	0.19	0.00	0.00	0.00	0.00	3.59	22341
Income								
$< 10K$	0.04	0.26	0.00	0.00	0.00	0.00	4.12	22341
10-20K	0.07	0.34	0.00	0.00	0.00	0.00	4.12	22341
20-30K	0.09	0.39	0.00	0.00	0.00	0.00	4.12	22341
30-40K	0.10	0.38	0.00	0.00	0.00	0.00	4.12	22341
40-50K	0.09	0.35	0.00	0.00	0.00	0.00	4.12	22341
50-75K	0.13	0.42	0.00	0.00	0.00	0.00	4.12	22341
75-99K	0.09	0.34	0.00	0.00	0.00	0.00	4.12	22341
100-149K	0.08	0.32	0.00	0.00	0.00	0.00	4.12	22341
150-249K	0.05	0.23	0.00	0.00	0.00	0.00	3.69	22341
250-500K	0.02	0.14	0.00	0.00	0.00	0.00	3.04	22341
500K+	0.00	0.08	0.00	0.00	0.00	0.00	4.05	22341

Notes: This table presents summary statistics of individual-level survey data from Gallup. We apply survey weights.

Table A.2: First Stage – South by Southwest 2007 and Twitter Diffusion

Dependent variable:	Log(Twitter usage)				
	(1)	(2)	(3)	(4)	(5)
Log(SXSW 2007 attendees)	0.733***	0.721***	0.710***	0.615***	0.491***
	(0.103)	(0.105)	(0.109)	(0.064)	(0.068)
Log(SXSW followers, pre)	0.096	0.089	0.079	0.041	0.133
	(0.122)	(0.123)	(0.121)	(0.102)	(0.102)
Population deciles	Yes	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes	Yes
Selection controls		Yes	Yes	Yes	Yes
Geographical controls			Yes	Yes	Yes
Socioeconomic controls				Yes	Yes
Media Controls					Yes
Observations	3,108	3,108	3,108	3,107	3,106
Mean of DV	8.368	8.368	8.368	8.368	8.369
R^2	0.911	0.912	0.913	0.936	0.948

Notes: This table presents county-level first stage regressions as in Equation (2) where the dependent variable is the number of Twitter users as of 2015 (in natural logarithm). *Log(SXSW 2007 attendees)* is the number of SXSW attendees in 2007 in natural logarithm (with one added inside), our instrument. *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles and Census division fixed effects. Selection controls include the linear distance from the SXSW festival location (Austin, Texas) and Google search intensity for the SXSW festival before 2007. Geographical controls is population density. Socioeconomic controls include age buckets (the share of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and those over 50) and ethnic composition (the share of people identifying as white, African American, Native American or Pacific Islander, Asian or Hispanic), the Gini coefficient, log median household income, the share of high school graduates, and the share of people with a graduate degree. Media controls include the prime time TV viewership to population ratio and the Facebook Social Connectedness Index (SCI). Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table A.3: IV Estimation – Robustness Tests

Dependent variable:	ΔLog(Dividends p.c.)			
	No weights (1)	SXSW counties only (2)	Winsorize outcome (3)	Inverse hyperbolic sine (4)
Log(Twitter users)	0.136***	0.223***	0.172***	0.070**
	(0.037)	(0.075)	(0.046)	(0.034)
Population deciles	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes
Selection controls	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes
Observations	3,104	172	3,104	3,106
Mean of DV	0.305	0.736	0.539	0.255
Robust F-stat	89.68	14.26	33.01	78.44

Notes: This table presents regressions as in equation 3, where the dependent variable is the change in Log(Dividends per capita) between 2005 and 2015 unless otherwise specified. Column (1) reruns our baseline estimates without weighting by population. Column (2) restricts the sample to counties where we have either SXSW attendees or other SXSW followers before 2007. Column (3) winsorizes the dependent variable at the 1st and 99th percentile. In Column (4), we use the inverse hyperbolic sine instead of the logarithmic transformation of dividends per capita, Twitter users, and SXSW attendees. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the full set of county-level controls in Table 2. Standard errors in parentheses are clustered by state, and observations are weighted by county population except in column 1. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

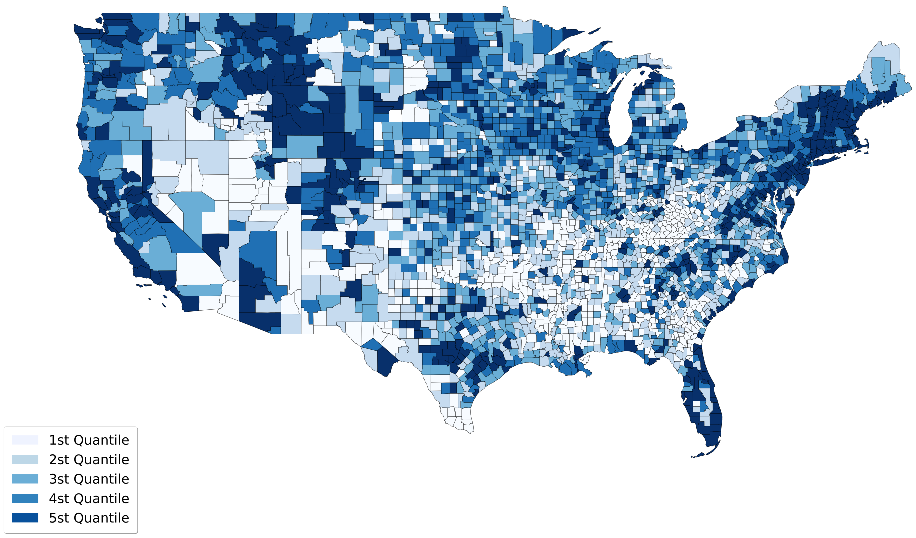
Table A.4: Social Media and Changes in Financial Advisors

Dependent variable:	$\Delta_{'07\rightarrow'15}$Log(# Financial Advisors)				
	(1)	(2)	(3)	(4)	(5)
Log(Twitter users)	-0.098**	-0.096**	-0.104**	-0.101**	-0.179**
	(0.047)	(0.046)	(0.050)	(0.046)	(0.069)
Log(SXSW followers, pre)	0.112**	0.107**	0.103*	0.079*	0.113*
	(0.054)	(0.050)	(0.052)	(0.046)	(0.057)
Population deciles	Yes	Yes	Yes	Yes	Yes
Census division FE	Yes	Yes	Yes	Yes	Yes
Selection controls		Yes	Yes	Yes	Yes
Geographical controls			Yes	Yes	Yes
Socioeconomic controls				Yes	Yes
Media Controls					Yes
Observations	1,946	1,946	1,946	1,946	1,946
Mean of DV	0.355	0.355	0.355	0.355	0.355
Robust F-stat	50.41	46.70	42.14	84.06	46.98

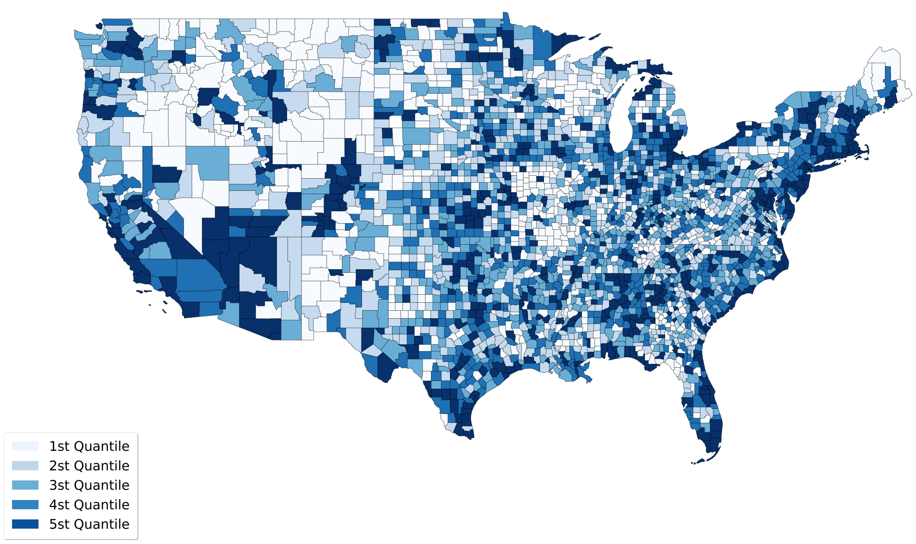
Notes: This table presents first difference regressions as in equation 3, where the dependent variable is the difference in the number of financial advisors (in natural logarithm) between 2007 and 2015. *Log(Twitter usage)* is the number of unique Twitter users in a county in natural logarithm, instrumented with *Log(SXSW 2007 attendees)* (the number of SXSW attendees in 2007 in natural logarithm with one added inside). *SXSW followers, Pre* is the number of SXSW followers who signed up to Twitter before the event in 2007. All regressions control for population deciles, Census division fixed effects, and the full set of county-level controls in Table 2. Standard errors in parentheses are clustered by state, and observations are weighted by county population. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Figure A.1: Geographical Variation of Main Variables

(a) Dividends per capita



(b) Twitter Usage per capita

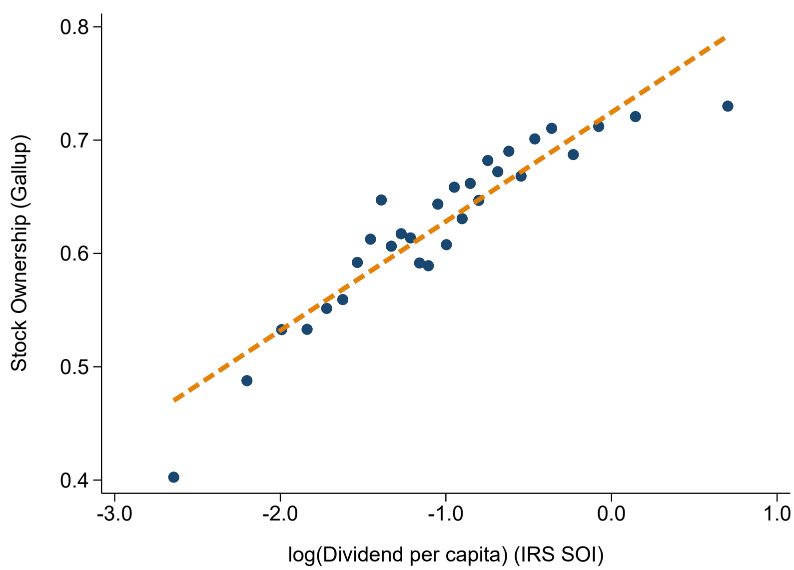


Notes: These figures visualize the geographical variation of our main measure of stock market participation (dividends per capita) and Twitter usage across the United States. Panel (a) presents quantiles of dividends per capita measured in 2015. Panel (b) plots the number of Twitter users per capita based on a large dataset of geo-located tweets collected between June and November in 2014 and 2015 by Kinder-Kurlanda et al. (2017).

Figure A.2: #FinTwit Word Cloud

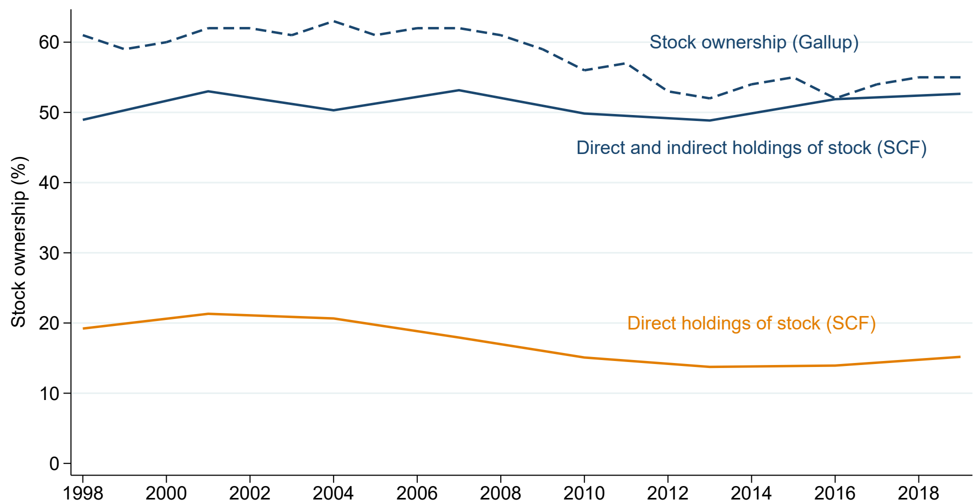
Notes: This figure visualizes the keywords discussed in ”financial Twitter” by plotting word cloud generated based on tweets with the hashtag ”FinTwit” from Dec 2010 to Dec 2020.

Figure A.3: Correlation of Dividend Income and Stock Ownership



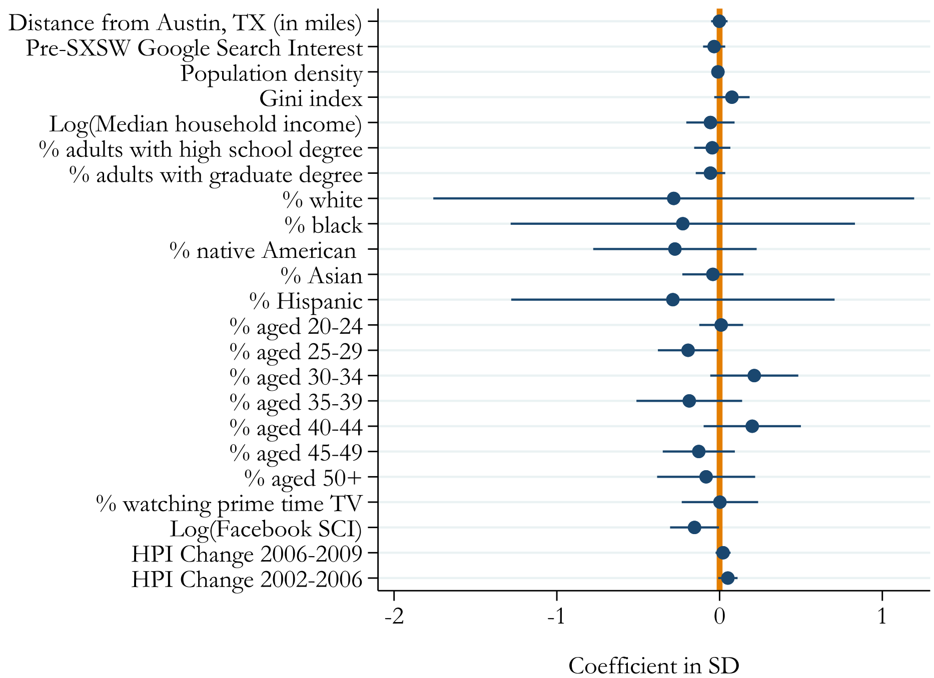
Notes: This figure shows a binned scatter plot of the ratio of dividend income per capita from the IRS SOI data against the share of individuals saying they own stocks from the Gallup Social Survey. The underlying combined dataset has 11,473 observations covering 2,354 counties for the period from 2001 to 2019.

Figure A.4: Comparing Gallup and SCF Data on Stock Ownership



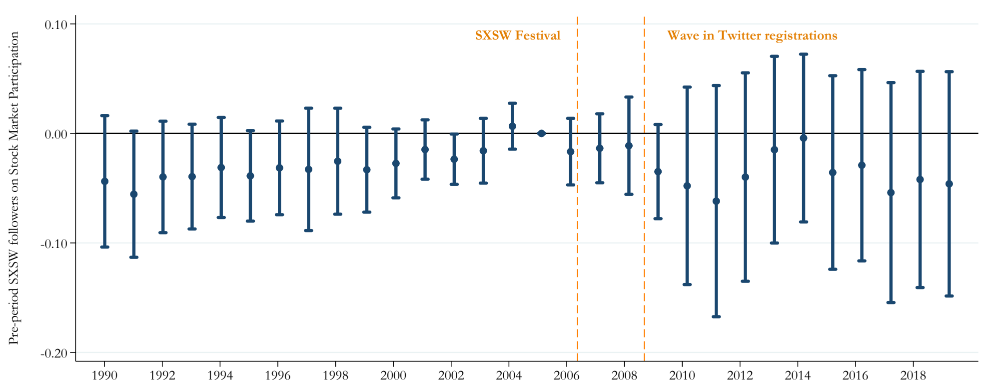
Notes: This figure compares U.S. stock ownership rates from 1998 to 2019 based on data from the Survey of Consumer Finances (SCF) and Gallup Social Survey.

Figure A.5: Observable Balancedness: Placebo Test



Notes: This figure plots the coefficients of multivariate regression where the dependent variable is a dummy variable equals 1 if a county receives an inflow of Twitter users due to the SXSW 2007 festival (our instrument), and 0 if a county has SXSW followers before the 2007 festival ("SXSW followers, pre") but did not receive any inflow of Twitter users in March 2007. We relates this dummy variable to a comprehensive sets of observable characteristics of counties. To aid interpretability, we standardize all the observable county characteristics to have a mean of zero and standard deviation of one.

Figure A.6: Dynamic IV Estimates for Placebo Treatment



Notes: This figure plots the estimates β_2 from cross-sectional IV regressions as in Equation (3), where the dependent variable is the change in dividends per capita relative to 2005 (the year before Twitter's launch). We include the full set of controls as in column 5 of Table 2. Standard errors are clustered by state, and we plot 95% confidence intervals.

A.1 Google Search Index Renormalization

We collect data on the frequency of Google searches (Google Search Volume Index, or SVI) using the Google Search API. To construct SVI time series for keywords of interest from 2004 to 2021, we first set the location to be one of the 210 Designated Market Area (DMA) in the United States. The monthly Google SVI is then computed as the normalized query share for the relevant keyword at each DMA location and at a specific point in time. For example, the SVI value for "GME" in Atlanta on 2021-12-01 is the number of queries for "GME" during December, 2021 divided by the highest number of queries for "GME" between 2004 and 2021 in Atlanta. As such, a decrease in the SVI means that a search term is becoming less popular in the specified location over time. We denote the SVI at each time point t and location s for stock i as $SVI\_T_{i,t,s}$.

To retrieve regional SVI data, we call the Google Search API in the following way. Given our time period of interest is fixed as 2004 – 2021, the DMA-level SVI values are computed as the number of queries for the relevant stock ticker in each region, e.g., Palm Springs, CA, from 2004 to 2021, normalized by the highest number of queries for that stock ticker across all DMAs. A relatively higher SVI value thus represents higher investor attention in the specific region within a given time period. We denote the SVI for each stock i at location s as $SVI\_R_{i,s}$ from 2004 – 2021.

To compare investor attention across DMAs over time, we need to renormalize the $SVI\_T_{i,t,s}$ and $SVI\_R_{i,s}$ as described above to make it comparable in each region and at each time point of interest. The process works as follows. For each stock ticker, it is trivial to find the location S where $SVI\_R_{i,s}$ equals to 100 (the maximum). Let $V_{i,S}$ be the total number of queries for stock ticker i in top state S from 2004 to 2021. Then, the total query volume for each DMA can be expressed as $\frac{SVI\_R_{i,s} * V_{i,S}}{100}$. We can also derive the total query volume for each DMAusing time series data as $\sum_t \frac{SVI\_T_{i,t,s} * W_{i,s}}{100}$, where $W_{i,s}$ is the highest total search volume for stock i in DMA s from 2004 to 2021. Intuitively, $W_{i,s}$ can be rewritten as:

$$W_{i,s} = \frac{SVI\_R_{i,s} * V_{i,S}}{\sum_t SVI\_T_{i,t,s}} \tag{4}$$

To compute comparable values of the SVI for each region and each point in time, we renormalize it as a percentage form of $W_{i,s}$:

$$S\tilde{VI}_{i,t,s} = SIV_{Renormalize} = SVI\_T_{i,t,s} * \frac{SVI\_R_{i,s}}{\sum_t SVI\_T_{i,t,s}} \tag{5}$$