SOCIAL INTERACTIONS AND LOTTERY STOCK MANIA*

Turan G. Bali David Hirshleifer Lin Peng

Yi Tang Qiguang Wang

Abstract

We find that social interactions are associated with both greater likelihood of a stock becoming more lottery-like and with greater investor overoptimism about the lottery characteristic. Heightened social media activity about a stock positively predicts the probability of an extreme daily price run-up, a *lottery event*. Lottery event stocks subject to more intense social media discussions subsequently experience greater retail buying pressure—particularly from Robinhood users—followed by lower returns. Moreover, lottery stocks of firms in more socially well-connected counties experience lower subsequent returns, and stocks with stronger lottery features trigger greater buy contagion among households in connected counties. Our findings suggest that social interactions stimulate investor excitement and asset price bubbles.

JEL classification: G10, G11, G12, G14, C13, E20, E30.

Keywords: speculation, lottery, meme stock, social interactions, social media, StockTwits, investor attention, MAX, retail investors.

---

*This draft: February 2025. This paper draws upon parts of a previous working paper that was circulated under the title "Attention, Social Interaction, and Investor Attraction to Lottery Stocks." Turan G. Bali: McDonough School of Business, Georgetown University, email: turan.bali@georgetown.edu. David Hirshleifer: Marshall School of Business, University of Southern California, email: hirshlei@marshall.usc.edu. Lin Peng: Zicklin School of Business, Baruch College/CUNY (corresponding author, lin.peng@baruch.cuny.edu). Yi Tang: Gabelli School of Business, Fordham University, email: ytang@fordham.edu. Qiguang Wang: Hong Kong Baptist University, Hong Kong, China, email: qiguangw@hkbu.edu.hk. We thank Hank Bessembinder, Huaizhi Chen (discussant), Ozgur Demirtas, Xi Dong, Joseph Engelberg (discussant), Marcel Fafchamps, Jon Garfinkel, Sanjev Goyal, Danling Jiang, Lei Jiang, Patrick Konermann, Dan Kovenock, Jens Kvarner, Weikai Li (discussant), Eric Lie, Yueran Ma, Yiming Qian, Larry Samuelson, Tyler Shumway (discussant), Johannes Stroebel, Adam Szeidl, Paul Tetlock, Alex Wolitzky, Kamil Yilmaz, Jianfeng Yu, Xiaoyan Zhang, Dexin Zhou, and Hao Zhou and seminar and conference participants at the American Finance Association Meeting, the Miami Behavioral Finance Conference, Baruch College, BI Norwegian Business School, Koc University, Rutgers University, Sabanci University, the European Networks Conference at the University of Essex, NYUAD Networks, Markets, State and Conflict conference, the State University of New York, Stony Brook, the Tsinghua Finance Workshop, the University of Iowa, and the University of International Business and Economics for helpful comments. Lin Peng acknowledges the Krell research grant for financial support and the Chinese University of Hong Kong ShenZhen for the data support she received while she was a visiting scholar.

SOCIAL INTERACTIONS AND LOTTERY STOCK MANIA

Abstract

We find that social interactions are associated with both greater likelihood of a stock becoming more lottery-like and with greater investor overoptimism about the lottery characteristic. Heightened social media activity about a stock positively predicts the probability of an extreme daily price run-up, a *lottery event*. Lottery event stocks subject to more intense social media discussions subsequently experience greater retail buying pressure—particularly from Robinhood users—followed by lower returns. Moreover, lottery stocks of firms in more socially well-connected counties experience lower subsequent returns, and stocks with stronger lottery features trigger greater buy contagion among households in connected counties. Our findings suggest that social interactions stimulate investor excitement and asset price bubbles.

JEL classification: G10, G11, G12, G14, C13, E20, E30.

Keywords: speculation, lottery, meme stock, social interactions, social media, StockTwits, investor attention, MAX, retail investors.

I. INTRODUCTION

There is extensive evidence that investors—especially retail investors—are attracted to stocks with lottery-like payoffs, resulting in high valuations of such assets and low subsequent returns.[1] Recent years have seen the rise of modern version lottery stocks—meme stocks. Many of these meme stocks, such as GameStop, AMC, Palantir, and Tupperware, repeatedly appear in the top decile when ranked by the magnitude of their largest one-day returns in a month, indicating lottery-like return distributions (Figure I).[2]

In remarks that seem to reflect these trends, Warren Buffet's annual shareholder letter released on February 24, 2024 drew parallels between today's stock market and a casino (Buffett 2024): "For whatever reasons, markets now exhibit far more casino-like behavior than they did when I was young." Buffet further noted "The casino now resides in many homes and daily tempts the occupants."

The rise of meme stocks has gone hand in hand with activity on popular investor social media platforms such as StockTwits and Reddit.[3] Stories about retail investors becoming millionaires overnight from investing meme stocks and other speculative assets have sparked intense discussions on social media platforms. While previous studies on lottery stocks have primarily focused on the role of fixed and inherent preferences for lottery characteristics at the individual investor level (Barberis and Huang 2008; Brunnermeier and Parker 2005), these recent developments raise the questions of how social interactions direct investor attention and enthusiasm toward speculative investments (Han, Hirshleifer, and Walden 2022; Pedersen 2022).

This paper provides insights into these questions using data from two types of social networks: a leading investment social media platform, StockTwits, and a general interest

[1] See, e.g., Kumar (2009), Boyer, Mitton, and Vorkink (2010), Bali, Cakici, and Whitelaw (2011), Green and Hwang (2012), Barberis, Mukherjee, and Wang (2016), and Liu et al. (2020).

[2] Meme stocks are also associated with the gamification and democratization of finance, as exemplified by Robinhood, a zero commission retail trading platform. The platform has introduced several means of gamifying stock trading, and Robinhood investors played a key role in the 2021 trading frenzy of meme stocks.

[3] Several recent studies investigate the role of social media in finance (see Antweiler and Frank 2004, Chen et al. 2014, Giannini, Irvine, and Shu 2019, Cookson and Niessner 2020, Chen and Hwang 2022), Cookson, Engelberg, and Mullins 2023.)

social media platform that is representative of the general population, Facebook. Using data from the investment social media platform, we find that more-intense social media discussions about a stock predict both the likelihood of an extreme daily price run-up (a *lottery event*) and its magnitude. Among stocks that exhibit prominent lottery features, higher social media activity predicts greater investor disagreement, higher trading volume, and stronger overvaluation, as evidenced by substantially lower subsequent returns. These stocks also attract more aggressive retail buying, followed by more negative future earnings announcement returns.

Furthermore, analysis of the general interest social media platform provides two results that reinforce the insights derived from the investment social media platform. First, lottery stocks headquartered in counties with stronger social connectivity experience greater overpricing. Second, households tend to buy more stocks with strong lotteriness following recent purchases of the same stock by households in highly-connected counties. This pattern is markedly weaker for stocks with less lotteriness. Together, these results based on both types of social media platforms suggest that social mechanisms play a crucial role in the emergence, pricing dynamics, and trading patterns of lottery stocks.

Our tests are motivated by theoretical models proposing that social interactions can propagate incorrect beliefs and naïve allocation of investor attention.[4] In Pedersen (2022), as investment ideas propagate through a social interaction, influential "fanatic" agents substantially affect equilibrium price owing to their influence on listeners. The interaction between rational agents and fanatics within the social interaction over time leads to attraction to speculative stocks, price bubbles, and subsequent reversals. This dynamic is reminiscent of that observed for meme stocks, such as the GameStop bubble in 2021.

[4]Several models have shown that social interactions can spread rumors and amplify behavior biases (DeMarzo, Vayanos, and Zwiebel 2003, Hirshleifer 2020, Han, Hirshleifer, and Walden 2022), trigger information cascades (Bikhchandani, Hirshleifer, and Welch 1992, Banerjee 1992), create free-riding incentives (Han and Yang 2013), and introduce "mutations" and transmission failures (Jackson, Malladi, and McAdams 2021). In addition, when investors exhibit Keeping-Up-with-the-Joneses preferences, social interactions can induce herding into risky securities and trigger excessive trading of local stocks (DeMarzo, Kaniel, and Kremer 2008; Gomez, Priestley, and Zapatero 2009; Roussanov 2010; Hong et al. 2014; Bali et al. 2023).

Hirshleifer (2020) provides a model of stock market bubbles caused by the social transmission of folk mental models about fundamentals or the market. This could be applied to mental models about the success of firms with moonshot projects. If there is social transmission of the mental model that certain lottery stocks are likely to have extreme successes, then social interactions will promote stock level price bubbles.

In the social transmission model of Han, Hirshleifer, and Walden (2022), high returns are disproportionately reported by signal senders and highly salient for signal receivers, leading the latter to become overoptimistic about the stocks they hear about. As a result, stocks with high skewness are overpriced and on average earn low subsequent abnormal returns.

There are also theories of attraction to stock level skewness based on non-standard preferences for lottery-like payoffs (Brunnermeier and Parker 2005; Barberis and Huang 2008). Social interactions can amplify such effects by calling investor attention to the lottery feature of stocks. In particular, word-of-mouth transmission of extreme news about a stock may attract the attention of investors who might otherwise be unaware of such stocks. So the arrival of extreme news may induce greater investor demand and overpricing of a stock.

Based on these considerations, we hypothesize that social interactions contribute to the pricing and trading of lottery stocks in two ways. First, for a stock that has lottery characteristics, social interactions can stimulate optimism about the stock's future payoffs, and can direct investor attention to the lottery characteristic, both implying retail buying and overpricing. Second, social interactions can contribute to a stock becoming more lottery-like. Once this occurs, the first channel can then promote buying. So via both channels, we predict that investor attraction to and market overpricing of lottery stocks increases with the intensity of social interactions among potential investors.

To test this hypothesis, we follow the spirit of Bali, Cakici, and Whitelaw (2011) and construct a daily measure of a stock's lottery-like characteristic based on the magnitude of its one-day returns. Specifically, a stock experiences a MAX event on a given day if its

return on that day is the highest over a trailing 21-trading-day window. This MAX-day return (MAXRET) captures the stock's lottery features, and we define a lottery event as occurring when MAXRET falls within the top 10% of the cross-sectional distribution.

Our main proxy for a stock's exposure to word-of-mouth discussions in investor social networks derives from social media activity on StockTwits, the largest finance-specific social media platform where investors share their investment opinions.[5] We obtain a dataset comprising over 10.7 million messages posted by 937,564 users between 2010 and 2022, covering 12,325 distinct stocks. We measure the degree of social media engagement among a stock's investors by the volume of StockTwits messages referencing the stock on a given day.

We first document an increase in message activity referencing a stock on the days surrounding the stock's lottery events. This suggests the possibilities that these discussions triggered such events, that such events are highly salient and heavily discussed by investors in their online social interactions, or both.

We then analyze the returns to lottery stocks. Previous studies have shown that stocks experiencing extremely large one-day returns tend to be overvalued, as evidenced by their substantially lower one-month-ahead returns. Building on this result, we examine how social interactions modulate the relationship between a measure of a stock's lotteriness and its subsequent returns.

Specifically, for stocks with MAX events, we compare the one-month-ahead-returns between stocks with elevated StockTwits message volume (above the 90th percentile of the cross-sectional distribution) in the days leading up to the MAX event and stocks without such message activity. We find that for a one-standard-deviation increase in MAXRET, such an increase in message volume is associated with much lower one-month-ahead returns, a difference of 137 basis points. A key finding is that message volume fully absorbs the effect of MAXRET itself in predicting future returns. This suggests that

[5]The dataset has been increasingly used in recent studies, including Cookson and Niessner (2020), Cookson et al. (2024), and Cookson, Engelberg, and Mullins (2023). See the review of social media and finance by Cookson, Mullins, and Niessner 2024.

social interactions may be a key source of the overvaluation of lottery stocks, as implied by the social-based models of Hirshleifer (2020), Han, Hirshleifer, and Walden (2022), and Pedersen (2022).

To assess whether social interactions promote overpricing among lottery stocks, we test for corrections at subsequent earnings announcement dates. If lottery stocks are overpriced, the subsequent arrival of fundamental news should result in corrections that are increasing with the initial level of overpricing. We find that lottery stocks experience negative abnormal returns around the next quarter's earnings announcements, with more negative returns for stocks that had higher StockTwits message volume on the MAX day. These patterns support the hypothesis that social interaction intensity contributes to investors' overly optimistic expectations for lottery stocks.

We then probe underlying mechanisms by investigating the extent to which social media discussions contribute to the likelihood of a stock experiencing a lottery event, and whether such discussions influence investors' subsequent trading of these stocks. We first show that social media discussion directly predicts the extent to which a stock becomes more lottery-like. Stocks that become the focus of intense discussions on StockTwits are 62% more likely to experience a lottery event the following day. Furthermore, conditional on a stock experiencing a lottery event, those experienced intense pre-event discussion have a event-day return that is 17% higher than the corresponding mean. This finding suggests that social media contributes to a stock's likelihood of achieving lottery status by helping direct investor attention to such stocks and sparking their optimism about the stock's future prospects.

Retail investors are more likely to be attracted by the salient features of lottery stocks and are more susceptible to the influence of social media. We therefore expect that social interactions will especially bolster demand for lottery stocks among retail investors. To examine whether the social effects are strong among retail investors, we consider data on the trading activity of users on the Robinhood platform and on aggregate retail order flows.

The Robinhood platform has been the go-to platform for retail investors central to the meme stock phenomenon. Using stock popularity data from the Robintrack website, we investigate whether the StockTwits message volume associated with a lottery event predicts Robinhood buy herding episodes, where a buy herding episode is defined as a day with a sharp increase in the number of Robinhood users holding a stock. We find that, controlling for a stock's lottery characteristic, more intense discussion on StockTwits is associated with a 38% increase in the likelihood of a Robinhood buy herding episode for the stock in the week following the lottery event, relative to the unconditional mean.

To assess whether our findings about StockTwits messages and buy herding on Robinhood are representative of the actions of retail traders in general, we examine aggregate retail order imbalances using measures proposed by Boehmer et al. (2021) and Barber et al. (2023). We find that, conditional on a lottery event, heavier discussion on StockTwits is associated with greater aggregate retail net purchases in subsequent weeks. This evidence suggests that our findings generalize to the broader retail investor population, where more intense social media discussion is associated with more aggressive retail buying of lottery stocks.

Theories of social interactions further suggest that such interactions can spread rumors, propagate incorrect beliefs, and introduce signal mutation, thereby creating scope for investor disagreement and trading.[6] We therefore investigate whether greater StockTwits message activity around lottery events is associated with greater investor disagreement and trading volume. These dimensions of investor beliefs and trading have not been analyzed in previous empirical studies of lottery stocks.

We use Google's Bidirectional Encoder Representations from Transformers (BERT) model to capture the sentiment conveyed in a message and measure disagreement as the standard deviation of message sentiment related to a stock on a given day. We find that lottery stocks that attract a large increase in StockTwits message volume exhibit a larger increase in investor disagreement in the two weeks that follow, by 6.4% and 7.7% of the

[6]See, for example, Shiller (2000), Hirshleifer (2020), Jackson, Malladi, and McAdams (2021), Pedersen (2022), and Hirshleifer, Peng, and Wang (2024).

sample mean, respectively. Similarly, these stocks also experience a larger increase in share turnover for the next two weeks relative to the sample mean, by 44% and 47%, respectively. These findings provide further evidence suggesting that social interactions affect belief updating after lottery events.

Although StockTwits is the largest finance-specific social media platform, it only constitutes a subset of investors, raising the question of how representative StockTwits is of general investor thinking, social interaction, and trading. Our results on retail trading suggest that StockTwits activities do predict real trading outcomes. To further address this generalizability concern, we use the large-scale Facebook-based social network data to construct an alternative proxy for the intensity of social interaction among investors.

The Facebook's Social Connectedness Index (SCI, Bailey et al. 2018b) captures friendship likelihood between Facebook users across counties. As the world's largest online social network, Facebook's enormous scale and coverage (over 258 million monthly active users in the US as of 2020) and the relative representativeness of its user base make SCI a unique measure of the real-world geographic structure of US social networks at the population scale (Bailey et al. 2018a,b; Chetty et al. 2022). Compared to the investing-focused StockTwits-based measure, SCI better captures long-run real-world social ties between counties. The two measures thus complement each other in testing our hypotheses.

We first investigate how the social connectedness of a county where a lottery stock's headquarters is located influences investor demand for the stock. There is extensive evidence that shows that investors are more attentive to nearby firms and are more likely to invest in and trade these firms' stocks (see, e.g., Coval and Moskowitz 1999, Huberman 2001). As investors discuss their gains from investing in local stocks, particularly those with lottery-like characteristics, news about these stocks spreads from local investors to those in other counties through word-of-mouth. Consequently, such stocks tend to be extensively discussed among a wider range of investors. Hence, we anticipate that lottery stocks based in counties with higher social connectivity will likely attract more investor interest, leading to greater overvaluation.

We define a firm's headquarters social connectedness, SCIH, as a kind of average connectedness of the firm's headquarters to all other U.S. counties. Higher headquarters connectedness implies that a lottery stock is more likely to be discussed by more investors throughout the U.S., triggering a stronger demand for the stock. Consistent with this, we find that lottery stocks headquartered in high-SCIH counties experience more-negative returns over the next three months, suggesting that greater social connectivity amplifies overvaluation.

We then apply the Facebook SCI measure as a proxy for the intensity of social interactions among individual households and test whether the trading of lottery stocks is socially transmitted. We use account-level trade data detailing the trading activity and positions of 78,000 households at a major US brokerage firm from 1991 to 1996. We test for buying contagion, wherein households increase their purchases of stocks following recent purchases by peer households from connected counties. We find that for stocks that experienced a MAX event, social connectivity is positively associated with buying contagion. This effect is especially strong for stocks that have previously experienced large MAX returns.

This evidence indicates that social interactions are associated with both overpricing and heavy trading of lottery stocks. The consistency of our findings across both the large-scale Facebook-based network data and the more granular StockTwits-based investor network lends credence to this conclusion.

To address the possibility that our results might be driven by omitted variables, we include firm and time fixed effects throughout our analysis. In the study of household trading, we employ both firm and household fixed effects. The inclusion of these fixed effects alleviates concerns that our findings are driven by omitted, time-invariant characteristics of counties, firms, or investors, as well as market-wide factors that influence investor demand for lottery stocks. Our results remain unchanged after controlling for other well-established return predictors. Additionally, we perform several robustness checks that

account for the information environment and information supply, arbitrage costs, and examine alternative definitions of lottery events.

For further identification, we examine a plausibly exogenous shock to the intensity of investor social interactions due to the COVID-19 pandemic. StockTwits experienced a sharp increase in both monthly active users and message volume in February and March of 2020, following the first confirmed COVID-19 death in the US. and the implementation of statewide lockdowns. Specifically, for stocks experiencing MAX events, the intensity of social interactions, measured by the number of likes per message received for posts made by influential users, saw a sharp increase starting in April 2020.

Motivated by this, we examine a triple interaction. Our basic interaction tests examine whether having higher StockTwits message volume than other stocks strengthens the negative relationship between a stock's lotteriness and its subsequent returns. We further test whether this modulating effect intensifies after the onset of COVID. We find that the effect is significantly strengthened in the six months following COVID's onset compared to the pre-COVID period. This evidence suggests that the sharp increase in social interaction intensity in the form of discussion on StockTwits during the pandemic increases investors' attraction to lottery stocks, leading to greater overvaluation and larger subsequent price declines.

Our paper contributes to the aforementioned literature on the overpricing of lottery characteristics and associated negative return predictability by providing key pieces of evidence consistent with the newer, social-based theories about the formation of lottery stock bubbles and investor behavior during such episodes. Our findings also provide insight into the more general literature on retail investors as a possible source of stock market anomalies (Barber and Odean 2008, Da, Engelberg, and Gao 2011, Kelley and Tetlock 2013, 2017, Andrei and Hasler 2015, Yuan 2015, and Boehmer et al. 2021), including research on Robinhood investors (Welch 2022, Barber et al. 2022, Ozik, Sadka, and Shen 2021 and Eaton et al. 2022). Our evidence suggests that retail attention is

socially transmitted, and that social networks act as a conduit that prompts the herding and purchasing of speculative stocks.

II. DATA, VARIABLE DEFINITIONS, AND PRELIMINARY ANALYSIS

Our sample consists of US common stocks (SHRCD = 10 or 11) traded on the NYSE, AMEX, and NASDAQ for the period from June 1963 through December 2022. We obtain investor social media data from StockTwits for 2010–2022, stock data from the Center for Research in Security Prices (CRSP), and other accounting and financial statement variables from the merged CRSP-Compustat database.

II.A. The Lottery Characteristic and Returns to Lottery-Based Portfolios

Bali, Cakici, and Whitelaw (2011) (BCW) use a stock's maximum daily return in a given month to measure its lottery-like characteristics and classify a stock as a lottery stock if the maximum daily return falls within the top 10% of the monthly cross-sectional distribution.

Because we focus on high-frequency variations in stocks' lottery-like features and their relation with social media and trading activities, we follow the spirit of BCW and construct a daily measure of lottery-like characteristics.[7] We define a MAX event day as when a stock's return on that day is the highest over a trailing 21-trading-day window and denote MAXRET as the MAX-event-day return. A lottery event occurs if a stock experiences a MAX event on that day and its MAXRET falls in the top decile of its daily cross-sectional distribution.

[7] An alternative way to define a stock's lotteriness uses three characteristics: low price, high idiosyncratic volatility, and high estimated idiosyncratic skewness (volatility and skewness estimated over the preceding 6 months; Kumar 2009). Since these characteristics have limited daily variation, we focus on a higher-frequency proxy to identify social interaction effects. Previous studies that study lottery-like securities include Kumar, Page, and Spalt (2011, 2016), Green and Hwang (2012), Annaerta, DeCeustera, and Verstegena (2013), Han and Kumar (2013), Boyer and Vorkink (2014), Walkshausl (2014), Barberis, Mukherjee, and Wang (2016), Zhong and Gray (2016), Carpenter, Lu, and Whitelaw (2016), Bali et al. (2017), Wang, Yan, and Yu (2017), An et al. (2020), Lin and Liu (2018), and Liu et al. (2020).

This is a natural generalization of BCW's definition to test for return predictability at a daily instead of monthly frequency. BCW identify maximum daily returns by comparing returns across all days within each month. For example, to predict returns for the month of February, the predictor is the highest daily return during the month of January. However, for tests at a daily frequency, it would be inappropriate on January 15 (for example) to use the highest return during all of January to predict returns for January 16, as this would introduce look-ahead bias. We therefore construct a predictor based upon trailing 21-day period. This approach allows us to identify lottery events without using forward looking data.

We first show that our modified maximum daily return measure, MAXRET, predicts one-month-ahead returns, as does the original BCW measure. Specifically, for each month from June 1963 to November 2022, we sort stocks into decile portfolios based on MAXRET measure and present the value-weighted long-short portfolio returns in the following month in Table A.I, Panel A. The return and alpha spreads on the long-short portfolios are negative and highly significant, ranging from 82 to 120 basis points per month, with the abnormal return (alpha) differential primarily driven by the underperformance of high-MAXRET stocks with significantly negative alpha. For comparison, untabulated results show that the long-short portfolio based on the original BCW measure generates return/alpha spreads ranging from 73 to 117 basis points, again driven by the underperformance of high-MAXRET stocks.[8]

II.B. StockTwits Data

StockTwits is a leading social media platform where investors share their investment experiences and opinions. Similar to Twitter, StockTwits allows users to post short mes-

[8]If a stock-month observation does not experience a MAX event or experience multiple MAX events, we set MAXRET equal to the highest daily return for the month. We also present the corresponding Fama-MacBeth regression results in Panel B of Table A.I in the appendix for the sample period from July 1963 to December 2022. The slope coefficients for MAXRET are significantly negative in predicting returns, with values of -0.063, -0.089, and -0.099, for one-, two-, and three-month-ahead returns, respectively. These coefficients suggest that a one-standard-deviation increase in MAXRET lowers future returns, by 59, 83, and 92 basis points, for the subsequent one, two, and three months, respectively. Our results are also robust for alternative definitions of lottery events, and we present the results in Section VII.

sages, initially limited to 140 characters until May 8, 2019, when the limit was expanded to 1,000 characters. What sets StockTwits apart from Twitter is its exclusive focus on financial markets. The platform was the first to introduce the cashtag notation, using the ticker symbol of a stock in a message (e.g., $TSLA for Tesla). The dynamic nature of this dataset allows us to incorporate firm fixed effects in our analysis, which helps control for latent, confounding factors tied to a firm that might be associated with a stock's attractiveness to investors.

Using the StockTwits API, we construct a dataset comprising over 10.7 million messages posted by 937,564 users between 2010 and 2022, covering 12,325 distinct stocks.[9] We define *Message* as the number of messages mentioning a stock and use this variable to gauge the intensity of investor social interactions regarding that stock. Higher message counts indicate greater social engagement among investors.

II.C. Other Variables

We obtain the daily and monthly return and volume data from CRSP. We adjust stock returns for delisting to avoid survivorship bias (Shumway 1997). Unless otherwise stated, all variables are measured as of the end of the previous month of a MAX event so that there is no look-ahead bias in our empirical analyses.

We use a number of well-known cross-sectional return predictors as control variables. Specifically, following Fama and French (1992), we estimate stock i's market beta (BETA), and compute the stock's size (SIZE) as the product of the price per share and the number of shares outstanding (in millions of dollars). The book-to-market equity ratio (BM) at the end of June of year t is computed as the book value of stockholders' equity, plus deferred taxes and investment tax credit (if available), minus the book value of the preferred stock for the last fiscal year ending in $t-1$, scaled by the market value of equity at the end of December of $t-1$. Following Fama and French (2015) and Hou, Xue, and Zhang (2015), a stock's investment related characteristic is measured by the annual growth rate of total assets (IA) at the end of June of year t is measured by the change in book as-

[9]The API is available at https://api.StockTwits.com/developers.

sets (Compustat item AT) for the last fiscal year ending in year $t-1$ divided by lagged AT. Similarly, quarterly operating profitability (OP) is determined by income before extraordinary items (item IBQ) for the most recent fiscal quarter, with quarterly earnings announcements made in or prior to, but no longer than six months before the portfolio formation month, divided by one-quarter-lagged book equity. Momentum (MOM) is the cumulative return of a stock over a period of 11 months ending one month prior to the portfolio formation month (Jegadeesh and Titman 1993). We define a stock's monthly co-skewness (COSKEW) following Harvey and Siddique (2000). Following Amihud (2002), a stock's monthly illiquidity (ILLIQ, scaled by 10^6) is the average daily ratio of the absolute stock return to the dollar trading volume.[10]

II.D. Preliminary Analysis and Summary Statistics

We first provide preliminary evidence that StockTwits message activity is strongly associated with a stock's lottery features. Our tests are performed at the individual stock-day level, centered around each stock's MAX event days. Each day, we sort stocks that experienced MAX events into decile portfolios based on their MAXRET, with Decile 10 corresponding to the MAX events with the top 10% MAXRET in the cross-section.

Figure II displays the average daily number of StockTwits messages for each decile portfolio around the MAX events (day 0). Notably, there is a sharp increase in message counts on day 0 for stocks in the top decile (i.e., lottery stocks), rising from below 50 messages prior to the event day to nearly 150 messages on the MAX day. These elevated message activity persists for up to five days following the event. Interestingly, the message activity for the lottery stocks in Decile 10 appears to rise even before the MAX event, particularly on the day immediately preceding the event. This suggests that the lottery events are highly salient and heavily discussed by individuals in their online social inter-

[10]We require a minimum of 24 monthly observations for market beta and co-skewness, computed from monthly data over the past five years, and a minimum of 15 daily observations for Amihud illiquidity, computed from daily data in a month. Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the book value of the preferred stock. Following Gao and Ritter (2010), we adjust for institutional features of the way that the NASDAQ and NYSE/AMEX volumes are counted. Specifically, we divide the NASDAQ volume by 2.0, 1.8, 1.6, and 1 for the periods prior to February 2001, between February 2001 and December 2001, between January 2002 and December 2003, and January 2004 and later years, respectively.

actions. In comparison, stocks in Decile 9 experience only a slight increase in message activity and there are no significant changes in message activity around the MAX events for stocks in the other deciles.

We next present the summary statistics of our key variables of interest and control variables. Panel A of Table I provides the time-series averages of the cross-sectional descriptive statistics for the aforementioned variables. It shows that the mean (median) MAXRET is 6.43 (4.42) percentage points, and there is substantial cross-sectional variation, with a standard deviation of 9.31 percentage points. The average MAXRET for lottery stocks—those with a MAXRET value at the 90th percentile or above—is 11.97 percentage points.

Regarding StockTwits message activity, we examine the variable Messages, which corresponds to the number of messages posted from day -11 up to, and including, the MAX event day (denoted as day 0). The distribution of the variable is highly right-skewed, with a median value of 12, mean of 107.99, and an average skewness of 214.0. There tend to be very low message counts for most of the stock-month observations, whereas for popular stocks in the top 10th percentile, the average message counts exceeds 123. Given this, we define an indicator variable, $\mathbf{I}_{mssg}$, which takes the value of one if message counts for the $[-11, 0]$ window rank above the 90th percentile of its cross-sectional distribution, and zero otherwise. Stocks with $\mathbf{I}_{mssg}$ equal to one are presumed to be subjected to significantly more intense social media discussions compared to those with $\mathbf{I}_{mssg}$ equal to zero. [11]

Panel B of Table I presents the time-series averages of the cross-sectional correlations among the variables used in this study. Focusing on $\mathbf{I}_{mssg}$, it shows that the variable is associated with lower one-month-ahead returns, with an average cross-sectional correlation of -1.8%. Additionally, $\mathbf{I}_{mssg}$ is positively correlated with stock characteristics such as MAXRET, market beta, size, investment growth, momentum, and co-skewness, and negatively correlated with book-to-market.

[11] In Section VII, we demonstrate that our findings remain robust when using alternative windows to define $\mathbf{I}_{mssg}$.

In our subsequent analysis, we investigate how social interaction intensity measures are associated with the returns and trading of lottery stocks, while controlling for stock characteristics that earlier studies have shown to predict future returns and trading activities.

III. SOCIAL INTERACTIONS AND RETURNS TO LOTTERY-LIKE STOCKS

As discussed in the introduction, social transmission bias can cause investors to be attracted to lottery stocks (Han, Hirshleifer, and Walden 2022) and communication can generate the "GameStop"-like price bubbles (Pedersen 2022). Furthermore, social interactions can amplify the visibility of a stock's lottery-like characteristic, thereby magnifying the effect of inherent preferences for such features on investor demand (Barberis and Huang 2008; Brunnermeier and Parker 2005). In this section, we test the pricing implications of the proposed mechanisms and examine whether social interactions amplify the overvaluation of lottery stocks by analyzing post-event returns and price reactions to subsequent earnings announcements.[12]

III.A. Lottery Stock Returns

In Section II.A, we verify the results of previous studies in our sample using univariate portfolio analysis and Fama-MacBeth regressions, which produce a strong pattern of negative abnormal returns following a large MAXRET event (lottery event). Building on this, and taking advantage of the dynamic nature of StockTwits activities and the occurrence of MAX events, we can more precisely capture the dynamics of social interactions and returns surrounding the MAX event. Accordingly, we adopt an event study framework using panel regressions and include firm and time fixed effects to account for omitted firm

[12]More broadly, a substantial literature documents peer effects on investment decisions (see Kelly and O'Grada (2000), Duflo and Saez (2002, 2003), Hong, Kubik, and Stein (2004, 2005), Brown et al. (2008), Cohen, Frazzini, and Malloy (2008), Shive (2010), Kaustia and Knüpfer (2012), Pool, Stoffman, and Yonker (2015), Heimer (2016), Ahern (2017), Crawford, Gray, and Kern (2017), Maturana and Nickerson (2018), Hong and Xu (2019), Ouimet and Tate (2020), and Huang, Hwang, and Lou (2021)). Social interactions have also been found to promote the sale of lottery tickets (Mitton, Vorkink, and Wright 2018). Kuchler and Stroebel (2021) provide a review of this literature.

characteristics that may contribute to cross-sectional variations in investors' attraction to lottery stocks.

As a baseline specification and to validate our panel regression specification, we first perform the following panel regressions of returns following a stock's MAX event for the period 2010–2022, for which the StockTwits data are available:

$$\mathrm{R}_{i,t+1} \quad = \quad \lambda_0 + \lambda_1 \mathrm{MAXRET}_{it} + \lambda_2 \mathrm{X}_{it} + \varepsilon_{i,t+1}, \tag{1}$$

where R is the cumulative returns on stock i over various post-MAX event windows starting from trading day 1 through trading days 21, 42, and 63, respectively. $\mathbf{X}$ is a vector of lagged control variables, which includes the market beta (BETA), the natural logarithm of firm size (SIZE), the natural logarithm of the book-to-market ratio (BM), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). We include firm and year-month fixed effects and use two-way robust standard errors clustered by firm and year-month.

The results, presented in Table A.II in the appendix, confirm that the findings obtained from the panel regression method are consistent with those obtained from the monthly Fama-MacBeth specifications (Panel B of the Appendix Table A.I). For example, column 1 shows that the coefficient of MAXRET is -0.105 with a t-statistic of -2.33 when the dependent variable is the cumulative return over the 21-day period following the MAX event (i.e., one-month-ahead stock returns). Economically, the coefficient indicates that a one-standard-deviation increase in MAXRET (9.31% in our sample) leads to a 98 basis point lower return in the subsequent month.

Columns 2 and 3 present the results for 42- and 63-day post-MAX event windows, respectively. The corresponding negative slope coefficients for MAXRET persists, implying 132 and 179 basis points lower returns for these longer windows for a one-standard-deviation increase in MAXRET. These economic magnitudes align with those obtained from Fama-MacBeth regressions and further validate confirm the robustness of the panel regression methodology.

Next, we examine the association between StockTwits message activity in days leading up to a MAX event and subsequent stock returns. We estimate the following panel regression by including a message activity indicator and its interaction with MAXRET:

$$\mathrm{R}_{i,t+1} = \lambda_0 + \beta \mathrm{MAXRET}_{it} \cdot \mathbf{I}_{mssg_{it}} + \lambda_1 \mathrm{MAXRET}_{it} + \lambda_2 \mathbf{I}_{mssg_{it}} + \lambda_3 \mathrm{X}_{it} + \varepsilon_{i,t+1}, \quad (2)$$

where $\mathbf{I}_{mssg}$ equals one if the number of StockTwits messages during the $[-11, 0]$ window belongs to the top 10 percent of the distribution, and zero otherwise. The variable of interest is the interaction term, MAXRET$\cdot\mathbf{I}_{mssg}$, which captures how StockTwits message activity modulates the relation between MAXRET and subsequent returns.

The results are presented in Table II. Column 1 corresponds to the case when the dependent variable is the return for the [1, 21] post-MAX window, showing that the coefficient for the interaction term MAXRET$\cdot\mathbf{I}_{mssg}$ is -0.147 and statistically significant, with a t-statistic of -3.57. Columns 2 and 3 correspond to the regression analysis with 42- and 63-trading day cumulative returns as the dependent variables, respectively. The respective coefficients for MAXRET$\cdot\mathbf{I}_{mssg}$ remain negative and statistically significant, with coefficients ranging from -0.172 to -0.254.[13]

In terms of economic significance, these coefficient estimates suggest that, for a given one standard-deviation greater value of MAXRET, stocks that attract a high level of StockTwits activity in days leading up to the MAX day are associated with an additional decrease in future returns by 137 to 236 basis points over the next one to three months.

A key finding is that including $\mathbf{I}_{mssg}$ and its interaction with MAXRET fully absorbs the effect of MAXRET itself—the slope coefficients on MAXRET are negative, ranging from -0.053 to -0.104, but they are statistically insignificant. These findings, coupled with the highly significant negative slope coefficients for MAXRET$\cdot\mathbf{I}_{mssg}$, suggest that

[13]We are grateful to Tony Cookson and Marina Niessner for sharing their dataset. As a robustness check, we repeat the panel regressions based on the dataset used in Cookson and Niessner (2023). Table A.III of the appendix shows that the slope coefficients for MAXRET $\cdot$ $\mathbf{I}_{mssg}$ are highly significant, with values of -0.124, -0.185, and -0.265 in predicting cumulative returns one, two, and three months ahead, respectively. These results are closely in line with those obtained with our own dataset.

social interactions could be a key factor contributing to the underperformance of lottery stocks in our sample.[14]

III.B. Subsequent Earnings Announcement Returns

To further test whether social interactions are associated with greater overvaluation of lottery stocks, we consider stock price reactions to earnings announcements following lottery events. We hypothesize that the arrival of fundamental news helps correct overvaluation of lottery stocks, and that the greater the overvaluation, the greater the correction for a stock on the next earnings announcement date. We therefore hypothesize that correction will be more pronounced for lottery stocks that previously experienced more intense social interactions.

To do so, We calculate the DGTW (1997) cumulative abnormal returns over the three days surrounding the earnings announcement, denoted as CAR (see, for example, Frazzini 2006; Kaniel et al. 2012). We then estimate equation (2) with the next fiscal quarter CAR as the dependent variable. The results are presented in Table III. As a benchmark for comparison, we first present the regression analysis without the StockTwits message indicator in column 1. The coefficient of MAXRET is -3.008 and highly significant, indicating that high MAXRET stocks tend to experience disappointing earnings announcements in the following quarter. These results are consistent with Engelberg, McLean, and Pontiff (2018), who find increased returns to anomaly portfolios on earnings announcement days and other corporate news days. They argue that this phenomenon indicates that information arrival plays a role in rectifying investors' biased expectations.

We then introduce the StockTwits message indicator and its interaction terms with MAXRET, with message activities measured over the $[-11, -1]$ windows. The variable of interest is the interaction term, MAXRET$\cdot\mathbf{I}_{mssg}^{pre\text{Event}}$. In Column 2, the coefficient is significantly negative, with a value of -3.186. This indicates that MAX events associated with very high pre-event message volume tend to be followed by earnings announcement

[14]The slope coefficients for size (SIZE), asset growth rate (IA), profitability (OP), and coskewness (COSKEW) are generally significant, with signs consistent with the findings documented in prior studies, such as Harvey and Siddique (2000), Fama and French (2015), and Hou, Xue, and Zhang (2015).

returns in the next quarter that are 135% (= −3.186/−2.368) more negative than the rest of the MAX events. This result supports our hypothesis that intense social media discussions in the days leading up to a MAX event contribute to more optimistic investor beliefs about the lottery stock, resulting in a greater degree of disappointment upon the future release of earnings news.

In summary, we find that MAX stocks tend to underperform in subsequent earnings announcements, and that this underperformance is more pronounced when those MAX stocks have experienced more intense social interactions. This evidence is consistent with the implications of the models of Han, Hirshleifer, and Walden (2022) and Pedersen (2022). It suggests that investors hold overly optimistic expectations for stocks with large one-day returns, an optimism caused by or further amplified by social interactions, only to be disappointed when subsequent earnings announcements are made.

We next explore the underlying mechanisms for our main findings. We investigate the extent to which social interactions contribute to the likelihood of lottery event occurrence, and to investor trading behavior of stocks around these events.

IV. MECHANISMS

In this section, we first examine whether StockTwits message activity predicts the likelihood that a stock achieves lottery status and the prominence of its lottery features. We then investigate the trading activities of retail investors around these lottery events and explore whether these social interactions are associated with excessive buying by retail investors and contribute to the herding episodes of meme stocks. Furthermore, we examine stock returns around future earnings announcements to assess the extent to which investors may have held incorrect beliefs about lottery stocks. Additionally, we present new findings on the dynamics of investor disagreement and trading volume around lottery events.

IV.A. The Making of a Lottery Stock

Previous studies of lottery stocks have mostly taken the stock's lottery feature as given and have focused on subsequent returns. There has been little exploration of the predictors of becoming a lottery stock and why lottery characteristics change over time.[15] Motivated by the models of Han, Hirshleifer, and Walden (2022) and Pedersen (2022), we propose that social interactions contribute to a stock becoming more lottery-like, and that social interactions that direct investor attention to such stocks stimulate overoptimism about its future payoffs. Our evidence in Figure II, showing a substantial rise in message activity even before a large MAXRET event, is consistent with this perspective.

In this subsection, we formally test this hypothesis by examining the extent to which social interactions, as measured by StockTwits message counts, contribute to both the likelihood of a stock becoming lottery-like (the extensive margin) and the magnitude of its lottery features (the intensive margin).

To investigate the extensive margin, we estimate the following panel regressions with all stock-day observations in our sample:

$$\mathbf{I}_{i,t+1}^{lottery} = \lambda_0 + \lambda_1 \mathbf{I}_{mssg,it}^{preEvent} + \lambda_2 \mathrm{X}_{it} + \varepsilon_{i,t+1}, \tag{3}$$

where $\mathbf{I}^{lottery}$, the lottery event indicator, takes a value of one if a stock experiences a lottery event for a given day (day 0), and zero otherwise. The primary independent variable of interest is the pre-MAX event message indicator, $\mathbf{I}_{mssg}^{preEvent}$, which equals one if the message counts during the $[-11, -1]$ window exceeds the top 10% of the cross-sectional distribution, and zero otherwise. $\mathbf{X}$ denotes $\mathrm{R}^{\mathrm{preEvent}}$, the cumulative return during the $[-11, -1]$ window and the vector of control variables specified in equation (2), lagged by one month. We include firm and year-month fixed effects and use standard errors clustered by firm and year-month.

[15] Previous studies have examined the extent to which investors' preference for lottery attributes is influenced by persistent socioeconomic factors such as education, religion, and household income (Kumar 2009; Han and Kumar 2013; Bali et al. 2023), and preference for lottery-like payoffs (Boyer, Mitton, and Vorkink 2010; Bali, Cakici, and Whitelaw 2011).

The results are presented in Table IV. The slope coefficient of $\mathbf{I}_{mssg}^{preEvent}$ is 0.228 in the univariate panel regression (column 1), and 0.262 after controlling for all other variables (column 2). Both coefficients are highly significant. The economic magnitude can be illustrated based on the coefficient estimate in column 2; an increase in message activity to the top decile increases the likelihood of a stock obtaining the lottery status on a given day by 0.262 percentage points, a substantial magnitude compared to the unconditional probability of 0.42 percentage points. The results provide strong evidence that elevated levels of social interaction intensity is a strong predictor of the rise of a lottery stock.[16]

We next investigate the intensive margin, that is, the association between StockTwits message activity and the magnitude of a stock's MAXRET on lottery event days. We estimate the following stock-month panel regressions linking the magnitude of MAXRET with the pre-lottery event message indicator ($\mathbf{I}_{mssg}^{preEvent}$):

$$\text{MAXRET}_{i,t+1} \quad = \quad \lambda_0 + \lambda_1 \mathbf{I}_{mssg,it}^{preEvent} + \lambda_2 \text{X}_{it} + \varepsilon_{i,t+1}, \tag{4}$$

where $\mathbf{X}$ denotes the vector of control variables specified in equation (3). We also include firm and year-month fixed effects.

Columns 3 and 4 of Table IV present the results, with two-way clustered standard errors by firm and year-month. The coefficient on $\mathbf{I}_{mssg}^{preEvent}$ is positive and highly significant, suggesting that more-intense social media discussions are associated with higher lottery-event day returns. Specifically, the slope coefficient on $\mathbf{I}_{mssg}^{preEvent}$ in column 4 implies that a stock that has experienced elevated StockTwits message flows before the lottery event is expected to have substantially higher levels of MAXRET, by 3.87 percentage points, which is 17% higher than the cross-sectional mean of 23.0 percentage points.

[16]For the full sample of all stock-days (columns 1−2), the coefficient of R^{preEvent} is significantly negative. This aligns with the MAX event definition, where a stock's return on that day must be highest over a 21-trading-day window. The negative relationship reflects that stocks with high cumulative returns in the $[-11,-1]$ window are less likely to experience higher returns on day 0. In contrast, columns 3−4 focus on stock-days with lottery events, where by construction, the lottery event day return exceeds daily returns in the $[-11,-1]$ window. Thus, R^{preEvent} controls for this mechanical relationship. Our findings remain robust when using an alternative, path-independent lottery event definition, as shown in Section VII.

In summary, these results show that StockTwits message activity strongly predicts both the magnitude of large one-day share price run-ups and the likelihood that a stock becomes a lottery stock. The evidence supports the theories of Han, Hirshleifer, and Walden (2022) and Pedersen (2022) that social interactions can induce overvaluation of lottery stocks. In the next two subsections, we complement the return analysis by examining investors' trading activity around lottery events.

IV.B. Retail Herding on Robinhood Platform

A notable modern type of lottery stock is the "meme stock," whose rise in early 2021 is closely tied to investors on the Robinhood trading platform. Robinhood pioneered commission-free trading, making stock market participation more accessible to a wider audience, especially younger and first-time investors. This feature was crucial in attracting a large user base who were eager to trade without the financial barriers traditionally associated with brokerage services. The platform has also introduced several means of gamifying stock trading. An example is that "new members were given a free share of stock, but only after they scratched off images that looked like a lottery ticket" (Popper 2020).

Consequently, Robinhood emerged as the preferred trading platform for retail investors who played a pivotal role in the meme stock phenomenon. Barber et al. (2022) examine stocks on Robinhood's Top Movers list, comprising of stocks with the largest absolute percentage price changes from the previous-day close. They find that Robinhood investors engage in attention-induced buying more than other retail investors, leading to increased correlation in their purchase behavior and contributing to the overvaluation of the corresponding stocks.

Motivated by the model of Han, Hirshleifer, and Walden (2022), which suggests that attention to a stock can be socially transmitted, and the pattern of a sharp increase in StockTwits message count surrounding days with large price run-ups (Figure II), we

formally investigate whether social interaction intensity contributes to Robinhood herding episodes surrounding lottery events characterized by large price run-ups.[17]

We use stock popularity data from the Robintrack website for the period between May 2, 2018 and August 13, 2020. Following Barber et al. (2022), we define a herding episode as a day when the daily increase in the number of Robinhood users for a stock ranks within the top 5% of the cross-sectional distribution of the lottery stock sample, requiring a minimum threshold of 100 new users on that day. This represents approximately 14 herding episodes per day, or 0.5% of the average number of stocks covered by the Robintrack data on a given day and comparable to the 0.5% threshold used by Barber et al. (2022) for the full Robintrack sample.

We first examine a stock's herding likelihood for days surrounding a lottery event. We classify each stock-day observations in this window into two groups: the High-message group corresponds to days when the prior-day message counts belong to the top 10% of the cross-sectional distribution, and the Low-message group corresponds to those below the top 10%. Figure III then compares the frequency of herding episodes between stocks with high and low prior-day message count groups, respectively. The results indicate that for two stocks with similar lottery status, the one that attracted more discussions on StockTwits on a given day is more likely to experience a Robinhood herding episode the following day.

Next, we formally examine the association between StockTwits message activity before and on the day that a stock obtains lottery status (day 0) and the subsequent occurrence of herding episodes. We then estimate the following panel regression for a sample of stock-day observations surrounding the lottery event day:

$$\mathrm{HR}[n]_{it} = \alpha_0 + \alpha_1 \mathrm{Message}_{it} + \alpha_2 \mathrm{X}_{it} + \varepsilon_{it}, \tag{5}$$

[17]Robintrack stopped reporting data after August 2020. Due to data limitations, we are unable to examine herding episodes specifically associated with meme stocks in the post-2020 period. Nevertheless, we believe the insights obtained during the years leading up to the meme period are useful in advancing our understanding of the role of social media in the meme stock phenomenon.

where HR measures the likelihood of a herding episode for period n, where n corresponds to the lottery event day (day 0) or the post-lottery event windows of [1, 5] and [6, 10]. HR[0] equals one if stock i experiences a herding episode on day 0 and zero otherwise. HR$[1-5]$ and HR$[6-10]$ represent the averages of daily herding likelihoods for the respective periods. Message is the StockTwits message activity indicator, $\mathbf{I}_{mssg}^{preEvent}$ or $\mathbf{I}_{mssg}^{0}$. Specifically, $\mathbf{I}_{mssg}^{preEvent}$ equals one if the message count for stock i during the pre-event window $[-11,-1]$ ranks within the top 10% of the cross-sectional distribution, and zero otherwise. Similarly, $\mathbf{I}_{mssg}^{0}$ equals one if stock i's message count on day 0 ranks within the top 10% of the cross-sectional distribution, and zero otherwise. $\mathbf{X}$ represents a vector of control variables for stock i considered by Barber et al. (2022).[18] We also control for the year-month fixed effects. Standard errors are clustered by firm and year-month. t-statistics are shown in parentheses.

Table V presents the results, with column 1 corresponding to HR[0], the herding likelihood on the lottery event day and the key explanatory variable is $\mathbf{I}_{mssg}^{pre\text{Event}}$. The lagged control variables are the pre-event averages measured over the window $[-11,-1]$. We also control for the cumulative return, R$[-11,-1]$. Columns 2 and 3 correspond to dependent variables HR$[1-5]$ and HR$[6-10]$, and the key explanatory variable is $\mathbf{I}_{mssg}^{0}$. The lagged control variables are measured as of day 0. We also control for return on day 0.

The results show that the coefficients on $\mathbf{I}_{mssg}^{preEvent}$ and $\mathbf{I}_{mssg}^{0}$, the primary variables of interest, are positive and significant across all columns except column 3. In column 1, $\mathbf{I}_{mssg}^{preEvent}$ has a coefficient of 0.070. This indicates that stocks with high StockTwits message volumes are more likely to experience a Robinhood herding episode, by 7.0 percentage points, or 27% of the average herding probability of 26.3 percentage points on the

[18]Specifically, we include the following control variables: lagged cumulative return (Return), herding indicator (HR), the natural logarithm of the number of Robinhood users for the stock(USER), and the change in the natural logarithm of the number of users (ΔUSER), abnormal trading volume (AbVol), abnormal news (AbNews), and an earnings announcement indicator (EA). AbVol is calculated as the natural logarithm of the number of share volume on a given day minus the natural logarithm of the average number of daily share volume over the past 21 trading days. AbNews is calculated as the natural logarithm of one plus the news count on a given day minus the natural logarithm of the average daily news count over the past 21 trading days. News count data are from Ravenpack. EA takes the value of one if an earnings announcement is made on the MAX day, and zero otherwise.

lottery event days. We use the coefficient estimate in column 2 to illustrate the economic magnitude of the average herding likelihood for the first five days following a lottery event. A coefficient of 0.020 indicates that a lottery event accompanied by elevated StockTwits activity is more likely to experience a Robinhood herding episode in the week that follows, by 2.0 percentage points. This effect is economically significant, corresponding to 38% of the unconditional mean of daily herding probability of 5.3 percentage points in the first five days. Column 3 shows a similar, albeit insignificant, positive association between message activity and herding probability for the [6, 10] window.[19]

Overall, our findings indicate that, given a stock's lottery features, more intense social media discussions mentioning the stock predict a higher likelihood of buy herding among Robinhood investors. This result suggests that the attention-driven herding evidence of Barber et al. (2022) is induced or amplified by social interactions.

IV.C. Aggregate Retail Trading

We next broaden our analysis of retail trading behavior around lottery events by extending the above Robinhood investor analysis to examine aggregate retail order flows.

Following Boehmer et al. (2021) and Barber et al. (2023), we define the volume-based retail order imbalance (OIB) for stock i on day d as follows:

$$\text{OIB}_{id} = \frac{\text{BVOL}_{id} - \text{SVOL}_{id}}{\text{BVOL}_{id} + \text{SVOL}_{id}}, \tag{6}$$

where BVOL and SVOL represent the number of shares bought and sold by retail investors for stock i on day d, respectively.[20]

[19]Turning to the other explanatory variables, columns 2 and 3 also show that past return, lagged herding status, and the number of Robinhood users are positively associated with herding likelihood, which are consistent with the findings documented in Barber et al. (2022). On the other hand, abnormal volume and earnings announcements are negatively associated with herding for our sample. Due to the short panel of only 27 months and limited time-series variation for this analysis, we do not use firm-fixed effects (Petersen 2009).

[20]The methodology uses off-exchange trades (with an exchange code equal to "D") from the TAQ database. A trade is classified as a retail buy if the sub-penny transaction price is above the quote midpoint and as a retail sell if the transaction sub-penny price is below the quote midpoint. We are grateful to Barber et al. (2023) for sharing the code used to calculate retail order imbalance.

We estimate equation (5) by replacing the Robinhood herding indicators with retail OIB as the dependent variable, while also incorporating lagged OIB as an additional control. Table VI presents the findings, with Panels A and B corresponding to the Robinhood and the StockTwits sample, respectively. Panel A shows that the coefficients on $\mathbf{I}_{mssg}^{preEvent}$ and $\mathbf{I}_{mssg}^{0}$, the primary variables of interest, are positive and significant for day 0 and the period of [1, 5]. Stocks subject to more intense discussions on StockTwits in the pre-lottery event period and on the lottery event day experience 2.66 and 1.47 percentage points higher net retail buys on the event day and in the subsequent five days, representing 10.2% and 9.1% of their respective sample standard deviations. The average order imbalance for the period of [6, 10], presented in column 3, is also positively correlated with $\mathbf{I}_{mssg}^{0}$, albeit statistically insignificant.

Panel B extends the analysis to the entire StockTwits universe, with eight times the number of observations used in Panel A, and omits the Robinhood-specific explanatory variables. While pre-event messages do not significantly predict event-day retail order imbalance (column 1), message activity more strongly predicts retail net buying in weeks one and two after the lottery event (columns 2 and 3). The delayed OIB response in the StockTwits sample, relative to the Robinhood sample, may reflect slower information dissemination among less visible stocks that appear in the StockTwits sample but not in the Robinhood sample.

Together, the evidence indicates that increases in StockTwits message activity predict more aggressive net buying of a lottery stock by retail investors following a lottery event for that stock. Thus, the evidence suggests that the insights obtained from the Robinhood investor analysis extend to retail investors generally, including those who do not participate in Robinhood. For both sets of investors, social interactions are strongly associated with more aggressive retail net buying of lottery stocks.

IV.D. Disagreement and Trading Volume

We have shown that social interactions among investors are associated with aggregate returns and trading activities. In this subsection, we take advantage of the rich, message-

level, information in our StockTwits data and explore unique implications of theories based upon social interaction in the context of lottery stocks.

For example, theories of social interactions suggest that word-of-mouth communication in social interactions can spread rumors, incorrect beliefs, or naïve trading strategies (Shiller 2000, Hirshleifer 2020, Han, Hirshleifer, and Walden 2022, Hirshleifer, Peng, and Wang 2024). Furthermore, Jackson, Malladi, and McAdams (2021) show that, even for rational individuals, message relaying can introduce "mutations". These effects give scope to investor disagreement regarding asset valuations. Empirically, there is evidence that social interactions trigger echo chamber effects among investors (Cookson, Engelberg, and Mullins 2023) and are associated with persistent disagreement and trading volume following earnings announcements (Hirshleifer, Peng, and Wang 2024).

Motivated by these studies, we investigate how StockTwits message activity is associated with investor disagreement and trading volume around lottery events. We begin by estimating the probability that a given message conveys positive sentiment using Google's Bidirectional Encoder Representations from Transformers (BERT) model.[21] Next, we define disagreement (DIS) as the standard deviation (in percentages) of sentiment probabilities across messages for a stock-day.

We focus on the sample of lottery events and estimate the following panel regression using daily observations around the event:

$$\text{DIS}[n]_{it} \quad = \quad \alpha_0 + \alpha_1 \text{Message}_{it} + \alpha_2 \text{X}_{it} + \varepsilon_{it}, \tag{7}$$

where DIS corresponds to message disagreement on the lottery event day ($n = 0$) and the average daily disagreement for the first and second five-day periods after the lottery

[21] We do not rely on the self-reported sentiment by StockTwits users since this variable is available for only 43.6% of the messages in our sample. We used the Bidirectional Encoder Representations from Transformers (BERT), the state-of-the-art model introduced by Google in 2018. BERT is a widely used model for sentiment analysis and has superior performance in sentiment classification (Devlin et al. 2019). Our training sample consists of StockTwits messages with self-labeled bullish/bearish indicators, with bullish messages accounting for 86% of the self-labeled messages. We calculate a continuous sentiment score for messages without self-labeled sentiment indicators. The sentiment score ranges from 0 (bearish) to 1 (bullish), and has a mean score of 0.65 and a standard deviation of 0.42.

event. Message refers to StockTwits message activity indicator, either $\mathbf{I}_{mssg}^{preEvent}$ or $\mathbf{I}_{mssg}^{0}$. To account for other potential confounding effects that may also influence post-event disagreement, we control for the following lagged variables, measured in the pre-event window $[-11,-1]$ or on the lottery event day: DIS, (cumulative) return, abnormal news (AbNews), the earnings announcement indicator, and other predictors included in equation (1). We further control for the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). We include firm and year-month fixed effects and cluster standard errors by firm and year-month.

Table VII presents the results. Columns 1 corresponds to DIS measured on the lottery event day and columns 2 and 3 correspond to the daily average disagreement over the first and second weeks after the lottery event, respectively. The coefficient on Return is positive and significant, suggesting that an extremely large return tends to trigger a significant increase in investor disagreement. More importantly, the coefficients on the message indicators, $\mathbf{I}_{mssg}^{preEvent}$ and $\mathbf{I}_{mssg}^{0}$, the key variables of interest, are consistently positive across all columns, ranging from 0.015 to 0.019, and highly significant. In terms of economic magnitude, the coefficients in columns 1–3 indicate that lottery stocks that attracted a large increase in StockTwits message activity in the pre-event window or on day 0 also experience a larger increase in investor disagreement on the lottery event day and in the two weeks that follow, by 0.015, 0.016, and 0.019, or 6.2%, 6.4% and 7.7% of the sample mean, respectively. The result suggests that social interactions can intensify disagreement on the StockTwits platform.

Next, we examine the association of StockTwits message activity on the day of a lottery event with subsequent trading volume.[22] Specifically, we estimate equation (7) by

[22]An extensive literature suggests that investor disagreement is associated with trading volume (see, e.g., Kim and Verrecchia 1991, Harris and Raviv 1993, Kandel and Pearson 1995, and Scheinkman and Xiong 2003).

replacing the dependent variable DIS with share turnover (TO), defined as the number of shares traded divided by the number of shares outstanding:

$$\mathrm{TO}[n]_{it} \quad = \quad \alpha_0 + \alpha_1 \mathrm{Message}_{it} + \alpha_2 \mathrm{X}_{it} + \varepsilon_{it}. \tag{8}$$

The results are presented in Table VIII. The coefficient of Return is positive and significant for up to ten days following the lottery event, suggesting that large lottery event-day returns tend to attract a greater trading volume over the next two weeks. The coefficients of $\mathbf{I}_{mssg}^{preEvent}$ and $\mathbf{I}_{mssg}^{0}$ are positive and highly significant across all columns, indicating that lottery stocks that experienced a large increase in message activity in the pre-event window or on the lottery event day tend to exhibit a sustained increase in trading volume, estimated by the coefficients of 0.108, 0.009, and 0.007, or 61.0%, 43.9%, and 46.7% relative to the sample mean.

The results indicate that more intense social interactions regarding a lottery stock are associated with elevated post-event investor trading of the stock, above and beyond the direct effect of the large lottery-event returns. The strong association between message activity and investor disagreement, as well as aggregate trading, presents new findings to the literature on lottery-like securities. These findings provide further evidence suggesting that social mechanisms are an important channel influencing the belief and trading dynamics associated with lottery stocks.

V. TESTS USING THE FACEBOOK SOCIAL CONNECTEDNESS INDEX

The granular, message-level StockTwits data enables us to directly map social interaction intensity with equilibrium variables and aggregate quantities such as stock return, investor herding, and aggregate trading volume. Using this dataset, we obtain results that are consistent with our hypothesis that social interactions contribute to both the rise of a lottery stock and investors' attraction to such stocks. This is informative, as StockTwits is the largest finance-specific social media platform. Nevertheless, the question arises of

how representative the activity and opinions expressed on the platform are of trading decisions in the general investor population.

In this section, we complement the StockTwits-based analysis by using an alternative measure of social interaction intensity that is representative at the population scale. We adopt the Facebook Social Connectedness Index (SCI) as an alternative measure of investors' social interaction intensity. Introduced by Bailey et al. (2018b), Facebook Social Connectedness Index (SCI) is the friendship probability of Facebook users between two counties.[23] Using this measure, we examine two issues. The first is whether the social connectedness of a lottery stock's headquarters location contributes to investor attraction to the stock. The second is how social connections influence the spread of lottery stock purchases among individual households.

V.A. Social Connectedness and Returns to Lottery Stocks

We first perform our analysis based on the geographical location of a firm's headquarters and aim to capture the social connectedness between investors of the headquarters county with those in the rest of the U.S. There is extensive evidence that shows that investors are more attentive to nearby firms and are more likely to invest in and trade these firms' stocks.[24] As investors discuss their gains from investing in local stocks, particularly those with lottery-like characteristics, news about these stocks spreads from local investors to those in other counties through word-of-mouth. Consequently, such stocks tend to be extensively discussed among a wider range of investors. As theorized in the aforementioned models (Han, Hirshleifer, and Walden 2022, Pedersen 2022), such a tendency for more intense social interactions regarding the focal stock causes greater investor attraction to stocks that have produced extreme positive returns.

[23] $\text{SCI}_{ij} = \frac{\text{Connections}_{ij}}{\text{Users}_i \cdot \text{Users}_j}$, where Connections_{ij} is the number of Facebook friendship connections between the two counties and Users_i and Users_j represent the number of Facebook users in the corresponding counties. We obtained the 2016 SCI measure from https://data.humdata.org/dataset/social-connectedness-index. This measure has been increasingly used in recent studies as a proxy for real-world social connections owing to its scale and the relative representativeness of its user base. See, e.g., Bailey et al. (2018b), Bailey et al. (2020), Bailey et al. (2021), and Chetty et al. (2022) on these points.

[24] See, for example, Coval and Moskowitz (1999), Huberman (2001), Ivković and Weisbenner (2005), Ivković and Weisbenner (2007), Hong, Kubik, and Stein (2008), Massa and Simonov (2006), Seasholes and Zhu (2010), and Hong et al. (2014). Additionally, there is a notable local bias in Google searches for firms' stock tickers (Chi and Shanthikumar 2017).

Motivated by these models, we hypothesize that lottery events experienced by stocks in highly connected counties will be associated with higher valuations and lower future returns.

To test this hypothesis, we measure the social connectedness of a stock's headquarters county (SCIH) as the sum of the SCIs between that county and all other US counties (including the headquarters county itself). SCIH therefore serves as a proxy for the strength of social ties between the stock's headquarters county and the rest of the U.S. A higher value of SCIH implies that a lottery stock is more likely to be discussed by a broader set of investors in the social network, thereby triggering a stronger demand for the stock.

We then repeat equation (2) by replacing $\mathbf{I}_{mssg}$ with SCIH for the same sample period. Consistent with the results reported in Table II, Table IX shows that the slope coefficients of MAXRET·SCHI are consistently negative in predicting one-, two-, and three-month-ahead stock returns, and are statistically significant. The coefficient of MAXRET·SCIH in column 1 suggests that a one-standard-deviation increase in MAXRET is associated with more negative one-month-ahead returns, by 81 basis points, for stocks in the highest SCIH decile compared to those in the lowest decile.[25] The corresponding values are 110 and 118 basis points for two- and three-month-ahead returns, respectively.

The result is consistent with our hypothesis that lottery stocks headquartered in socially connected areas are more likely to be overvalued and, therefore, more likely to exhibit lower subsequent returns.

V.B. Social Connectedness and Household Trading Contagion

We next examine the role of social connectedness in spreading investor trading of MAX stocks. Our hypothesis is that the social transmission of investment ideas causes contagion in the decision to buy a stock, particularly for stocks exhibiting lottery features. It follows that the amount of buying of a specific stock is greater when highly connected peers have

[25]This is calculated as the product of the standard deviation of MAXRET (9.31%), the spread in SCIH between the top and bottom SCIH deciles (3.95), and the slope coefficient of MAXRET·SCIH (−0.022) from column 1.

recently acquired that stock, particularly for stocks experiencing large MAX returns. Our test is similar to the approach of Gelman et al. (2024), who use local soccer games as a measure of social interaction intensity and provide evidence of social transmission in buy trades.

To test this hypothesis, we obtain account-level trade data detailing the trading activity and positions of 78,000 households at a major US brokerage firm. We focus on contagion in household trading of stocks that experienced MAX events. We measure contagion by the sensitivity of a given household's purchase of such stocks to the recent purchases of another randomly selected household ("peer household"). We expect a higher sensitivity for household pairs from more socially connected counties.

To keep our analysis computationally feasible, we narrow our scope to household pairs where the peer households purchased a MAX-event stock within the [0, 1] day window, and focal households have held stocks within the same SIC 4-digit category as the MAX-event stock over the past 12 months.[26]

We estimate the following panel regression model of focal household's trading of a stock following its MAX events:

$$\begin{aligned}\text{Trade}[2,5]_{kit} = \lambda_1 \text{Trade}[0,1]_{kjt} + \lambda_2 \text{Trade}[0,1]_{kjt} \cdot \text{SCI}_{ij} + \lambda_3 \text{Trade}[0,1]_{kjt} \cdot \text{SCI}_{ij} \cdot \text{MAXDCL}_{it} \\ + \lambda_4 \text{Trade}[0,1]_{kjt} \cdot \text{MAXDCL}_{it} + \lambda_5 \text{SCI}_{ij} \cdot \text{MAXDCL}_{it} + \lambda_5 \mathbf{X}_{it} + \varepsilon_{kit}, \qquad (9)\end{aligned}$$

where MAXDCL is the decile rank of MAXRET in the crossesctional distribution (where rank 10 corresponds to lottery events), $\text{Trade}[2,5]_{kit}$ is the value of focal household i's purchase of stock k during the days [2, 5], relative to its portfolio value at the start of the month. Similarly, $\text{Trade}[0,1]_{kjt}$ denotes the value of peer household j's buy orders during the days [0, 1], also relative to its portfolio value at the beginning of the month. SCI is

[26]We thank Brad Barber and Terrance Odean for sharing the data. Including all possible peer-focal household pairs for this analysis would result in a sample of 2.69 billion observations, exceeding our computational capacity. As reported by Barber and Odean (2001), the mean and median number of stocks held by the average household in this sample are four and 2.6, respectively. Hence the full peer/focal-household pairs trading activity matrix is extremely sparse. We therefore impose conditions for sample selection that reduce the number of observations while still enabling us to focus on the most relevant household pairs.

the Facebook's SCI between the counties where the two households reside and its decile rank is used in the regression.

X represents the same vector of control variables as in equation (2). We further include firm, focal household, and peer household fixed effects in all specification. The inclusion of these controls and fixed effects enables us to explore variations within firms and households, which helps address the possibility that unaccounted-for firm-level or household-level variables are responsible for our findings.

Table X presents the results, with firm, focal household, and peer household fixed effects, as well as three-way clustered standard errors by firm, focal household, and peer household across all specifications. Column 1 presents the baseline regression of focal household trading on the peer household trading. The coefficient of $\text{Trade}[2,5]_{kit}$, which measures the sensitivity of how focal household purchases respond to peer purchase amounts for arbitrary household pairs, is insignificant.

In column 2, we include the interaction term, $\text{Trade}[0,1]_{kjt} \cdot \text{SCI}_{ij}$ as well as the standalone SCI variable. The coefficient on the interaction term is 0.069 ($\times 10^{-4}$), which is statistically significant. This indicates that the focal-to-peer purchase sensitivity is significantly higher for high-SCI household pairs than pairs with low SCI. Economically, a one-decile-rank increase in SCI results in 29% (= 0.069/0.241) increase in focal-to-peer buying sensitivity compared to the baseline value.

Turning to the coefficient on SCI, since we limit our selection of peer households to those that purchased a MAX event stock in the [0, 1] window, the coefficient thus captures the association of past peer purchases with focal household purchases. The positive and significant SCI coefficient indicates that focal households increase their purchases of MAX event stocks following purchases by highly-connected peers. Hence, the evidence supports the hypothesis that social connections contribute to trading contagion among households. Column 3 includes firm-level controls, and the coefficient remains very similar.

Next, we test whether greater salience of MAX-day return amplifies the social transmission of trading. In column 4, we add MAXDCL and its interaction with $\text{Trade}[0,1]_{kjt} \cdot$

SCI. The coefficient on the triple interaction is 0.020 and is highly significant. The inclusion of this variable also renders the coefficient on $\text{Trade}[0,1]_{kjt} \cdot \text{SCI}$ insignificant. This indicates that the trading contagion mostly occurs for stocks with strong lottery features—those with higher MAX-day returns.

Overall, our findings support the hypothesis that more intense social interactions between households locations are associated with greater buying contagion following lottery events.

VI. THE COVID-19 SHOCK

We have found that social interactions are associated with high valuations of lottery-like stocks relative to non-lottery-like stocks. We next investigate whether this association intensified during the COVID-19 pandemic. The pandemic is a plausibly exogenous shock that led to increased online activity, including participation in investment-focused social media platforms.

Figure IV illustrates monthly active users (MAUs, Panel A) and message volume (Panel B) on StockTwits from August 2019 till July 2020. In the six-month period prior to February 2020, StockTwits averages 48 thousand monthly active users (MAUs) and 1.07 million messages per month. Both metrics show a large increase after February and March, following the first confirmed COVID-19 death in the U.S. and the implementation of statewide lockdowns, reaching 152 thousand MAUs and 6.22 million messages in July 2020.

We next focus on stocks experiencing MAX events and analyze the per-message intensity of social interaction regarding these stocks, especially for messages posted by influential users on StockTwits. We define influencers as users who rank in the top 10% of the cross sectional distribution based on the number of likes they receive during the past three months, excluding the most recent month. We denote the MAX event day as day 0 and examine the per-message likes for messages mentioning these stocks during the [-11, 0] window for both influencers and non-influencers.

Figure IV Panel C indicates that for stocks experiencing MAX events, the monthly average per-message likes for influencers is 2.14 for the six months pre-COVID period, which is 0.55 higher than the 1.59 likes received by non-influencers. Starting in April of 2020, there is a sharp increase in the per-message likes for influencers, by 78%, to 3.81. For comparison, the per-message likes for non-influencers remained at around 2, which shows only a slight increase relative to their pre-COVID level. This indicates that, on the days leading up to a stock's MAX event, influential users' mentions of the stock attracted substantially higher attention during the COVID period compared to earlier periods.

Motivated by this evidence, we hypothesize that the sharp increase in social media interactions during the initial months of the COVID period, as reflected in number of messages and likes on StockTwits, intensified investors' attraction to lottery stocks, resulting in greater overvaluation and larger subsequent price declines. We focus on a sample period from August 1, 2019 till July 31, 2020 and augment equation (2) to estimate the following panel regression:

$$\begin{aligned} \mathrm{R}_{i,t+1} =& \lambda_0 + \beta \mathrm{MAXRET}_{it} \cdot \mathbf{I}_{mssg_{it}} \cdot \mathbf{I}_{COVID} + \lambda_1 \mathrm{MAXRET}_{it} \cdot \mathbf{I}_{mssg_{it}} \\ &+ \lambda_2 \mathrm{X}_{it} + \varepsilon_{i,t+1}, \end{aligned} \tag{10}$$

where $\mathbf{I}_{COVID}$ equals one for days after February 1, 2020, the start of the COVID period as defined earlier, and zero otherwise. $\mathbf{I}_{mssg}$ equals one if the number of StockTwits messages during the $[-11, 0]$ window belongs to the top 10th percentile of the distribution, and zero otherwise. $\mathbf{X}$ includes lower-order interactions and firm controls. With more messages during COVID, as indicated in Figure IV Panel B, the indicator $\mathbf{I}_{mssg}$ captures stronger social interaction during this period relative to the pre-COVID period. Additionally, it is possible that the per-message strength of social interaction also increased, as reflected by the greater per-message likes shown in Figure IV Panel C.

The variable of interest is the triple interaction term, $\mathrm{MAXRET} \cdot \mathbf{I}_{mssg} \cdot \mathbf{I}_{COVID}$.The results are presented in Table XI. The coefficient on $\mathrm{MAXRET} \cdot \mathbf{I}_{mssg}$, representing the effect of messages before COVID, is negative at -0.272 for one-month-ahead returns,

consistent with the findings in Table II. The coefficient of interest is MAXRET $\cdot \mathbf{I}_{mssg} \cdot \mathbf{I}_{COVID}$, with estimates ranging from -0.635 to -1.882 and are statistically significant. This result suggests that increased social media activity during the COVID-19 pandemic intensified the overvaluation of lottery stocks. In terms of economic magnitudes, for a one-standard-deviation increase in MAXRET (7.3% during this period), high-MAXRET stocks with heightened StockTwits discussions experience subsequent one- to three-month returns that are 4.6 to 13.7 percentage points lower during the COVID period than during the pre-COVID period.[27]

Overall, the evidence indicates that increased social interactions drove increased overvaluation of lottery stocks during the COVID-19 pandemic, a period marked by a substantial increase in social media engagement.

VII. FURTHER ROBUSTNESS CHECKS

This section provides additional robustness checks. We show below that our main findings remain similar with alternative windows over which we measure message activity, after controlling for information environment, information supply, arbitrage costs, for alternative definitions of lottery events, as well as additional household trading tests.

Alternative message windows We construct versions of Table II using alternative definitions of StockTwits message indicator, $\mathbf{I}_{mssg}$, that are measured with different event windows: [−5, 0], [−3, 0], and [−1, 0]. The results are presented in Table A.IV of the appendix. The slope coefficients of MAXRET$\cdot\mathbf{I}_{mssg}$, the primary variable of interest, are negative and statistically significant, ranging from -0.106 to -0.199 for the [−5, 0] window, from -0.104 to -0.162 for the [−3, 0] window, and from -0.085 to -0.142 for the [−1, 0] window. These results are similar to those reported in Table II.

[27]For the first three months after the COVID-19 outbreak, from February 1, 2020, to May 1, 2020, stocks with high MAXRET do not experience significantly lower future returns. However, for the period from May 1, 2020 through December 30, 2022, the relationship between MAXRET and future returns is negative and significant, consistent with findings in previous literature. Our results are robust to alternative window selections, such as lengths of three months or nine months before and after February 2020. Excluding one month or three months on each side surrounding the COVID date produces qualitatively similar results.

Information environment and information supply Our findings so far suggest that social interactions help direct investor attention to lottery stocks and therefore contribute to lottery demand. We next provide evidence that this effect is not driven by a stock's information environment, as traditionally measured by CVRG, the number of analysts covering a stock (e.g., Hou and Moskowitz 2005; Hong, Torous, and Valkanov 2007).

We obtain analyst coverage data from I/B/E/S, covering the period from January 1976 to December 2022. We re-estimate equation (2), including the natural logarithm of one plus CVRG as an additional control variable. The results, presented in Panel A, Table A.V of the appendix, show that the average slope coefficients on the interaction between MAXRET and $\mathbf{I}_{mssg}$ remain significantly negative, with a magnitude similar to those reported in Table II.

To test whether our main findings are driven by the information supply channel, we obtain news coverage data from Ravenpack News for the period from January 2000 to December 2021. We measure information supply by the number of relevant news reports (NEWS) from credible sources as defined by Ravenpack. We re-estimate equation (2) including the natural logarithm of one plus NEWS as an additional control variable.

Panel B, Table A.V of the appendix shows that the average slope coefficient on the interaction between MAXRET and $\mathbf{I}_{mssg}$ is significantly negative after controlling for NEWS, and the magnitude is closely in line with those presented in Table II. Overall, the evidence provided in Table A.V, Panels A and B, indicates that our main finding on the relation between social interactions and lottery stock returns is robust to controls for information environment and information supply.

Arbitrage costs High arbitrage barriers discourage investors from trading, thereby allowing mispricing to persist (see, for example, Shleifer and Vishny 1997). We therefore consider the extent to which our result in Table II is driven by arbitrage costs.

We measure arbitrage costs in two ways: idiosyncratic volatility (IVOL) and an arbitrage cost index (COST).[28] We then estimate equation (2) with the inclusion of IVOL or COST as an additional control variable and present the results in Panels C and D of Table A.V in the appendix, respectively. We find that the average slope coefficients on MAXRET·$\mathbf{I}_{mssg}$ are significantly negative and remain similar to the corresponding results reported in Table II. Hence, the results suggest that arbitrage costs do not explain our main finding.

Alternative lottery event definition So far, our definition of a lottery event on a given day depends on a stock experiencing a MAX event on that day, where MAX event occurs when a stock's return on that day is the highest over a trailing 21-trading-day window. As explained in Subsection II.A, we use this definition of MAX events because it is the natural generalization of the specification used in previous papers that examine lottery effects at monthly frequencies (i.e., BCW) while only using backward-looking data to construct the predictor. This allows us to compare our results with the previous literature and evaluate the role of social interaction variables in modulating the relation between MAXRET and future returns (Table II).

A potential concern with this measure is that it is possible that omitted shocks to the stock over the pre-event period may influence both MAXRET and message volume and therefore drive some of our findings, such as the relationship between message activities and lottery event likelihood. To address this possible source of endogeneity, we perform robustness checks using two alternative definitions of lottery event.

The first is a path-independent lottery event definition, with a lottery event occurring if a stock's return on a day ranks in the top 10% of the cross-sectional distribution of the

[28]Idiosyncratic volatility (IVOL) has been utilized as a proxy for risk by arbitrageurs seeking to exploit mispricing (e.g., Pontiff 2006; Stambaugh, Yu, and Yuan 2015). Following Ang et al. (2006), IVOL is estimated from stock-level time-series regressions of daily excess stock returns (with a minimum of 15 daily observations) against the daily Fama-French market, size, and book-to-market factors during the portfolio formation month. We measure COST following Stambaugh, Yu, and Yuan (2012, 2015): for each month, stocks are independently sorted into decile portfolios based on institutional ownership, market capitalization, analyst coverage, illiquidity, idiosyncratic volatility, and the number of months a stock is present on the CRSP database. COST is the arithmetic average of the decile ranks of the six stock characteristics, with the requirement that a minimum of three characteristics available.

highest daily returns in the trailing 21 trading days. This definition does not depend on the stock's own past returns, thus alleviating the endogeneity concern mentioned in the paragraph above.

The second alternative definition introduces a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event is defined as occurring when a stock's return on day t is equal to or greater than its returns over the $[t-31, t-11]$ window. A stock experiences a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. This definition ensures the past return window does not coincide with the window over which we measure pre-lottery event message activity.

We reproduce our key tests, Tables II—VI, using the alternative definitions and present the results in Online Appendix Tables S.I—S.IV. These tables show patterns that are similar to the corresponding tables presented in the main text. This indicates that our main findings are robust to the alternative lottery definitions.

VIII. CONCLUSION

We explore here the hypothesis that social interactions promote investor attention to and demand for lottery stocks, and thereby the formation and amplification of lottery stock bubbles. Based on a variety of tests using data from two types of social media and multiple outcome variables (extreme return realizations, investor disagreement proxies, messaging activity, volume of trade, and subsequent stock returns), we find evidence that is consistent with recent social finance models of bubbles.

Whereas most previous studies take a stock's lotteriness as given, we show that social interactions are associated with a greater likelihood of a stock becoming a lottery stock. Furthermore, we provide new insights into the source of return predictive power of a lottery characteristic identified in past literature, the maximum recent daily return. We find that the power of this variable to predict returns is primarily driven by stocks with high StockTwits message volume. This indicates that social interactions modulate the

relationship between a stock's lottery characteristic and its subsequent returns and are an important source of the cross-sectional predictive power of the lottery characteristic.

We find that a higher volume of social media messages associated with a lottery event is also more likely to trigger a sharp increase in the number of Robinhood users buying the stock ("Robinhood herds") and to prompt greater aggregate net buying among general retail investors. Furthermore, we use the onset of the COVID-19 pandemic as a plausibly exogenous shock to the intensity of investor social interactions. Our evidence indicates that the sharp increase in social interaction intensity on StockTwits during the pandemic strengthened investors' attraction to lottery stocks, leading to greater overvaluation and larger subsequent price declines.

We complement the StockTwits-based analysis with an alternative measure of social interactions, using Facebook's social connectivity data between users' county locations. We find that lottery stocks headquartered in counties with stronger social connectivity experience more negative returns in the subsequent month compared to those in less-connected counties. Furthermore, households tend to buy more stocks with strong lottery features following recent purchases by households in highly-connected counties. This association is weaker for stocks with less lotteriness.

Overall, our findings are consistent with models in which social interactions promote investor attraction to lottery stocks. This raises the interesting possibility that social mechanisms underlie other patterns of trading behavior and stock market pricing anomalies. Social mechanisms are likely to be particularly important in the era of increased online social interactions and the rise of zero-commission trading platforms.

References

Ahern, K. R. 2017. Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics* 125:26–47.

Amihud, Y. 2002. Illiquidity and stock returns: Cross-section and time series effects. *Journal of Financial Markets* 5:31–56.

An, L., H. Wang, J. Wang, and J. Yu. 2020. Lottery-related anomalies: The role of reference-dependent preferences. *Management Science* 66:473–501.

Andrei, D., and M. Hasler. 2015. Investor attention and stock market volatility. *Review of Financial Studies* 28:33–72.

Ang, A., R. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–99.

Annaerta, J., M. DeCeustera, and K. Verstegena. 2013. Are extreme returns priced in the stock market? european evidence. *Journal of Banking and Finance* 37:3401–11.

Antweiler, W., and M. Z. Frank. 2004. Is all that talk just noise? the information content of internet stock message boards. *Journal of Finance* 59:1259–94.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel. 2018a. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy* 126:2224–76.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong. 2018b. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspective* 32:259–80.

Bailey, M., A. Gupta, S. Hillenbrand, T. Kuchler, R. Richmond, and J. Stroebel. 2021. International trade and social connectedness. *Journal of International Economics* 129:103418.

Bailey, M., D. Johnston, T. Kuchler, D. Russel, B. State, and J. Stroebel. 2020. The determinants of social connectedness in Europe. In S. Aref, K. Bontcheva, M. Braghieri, F. Dignum, F. Giannotti, F. Grisolia, and D. Pedreschi, eds., *Social Informatics*, 1–14. Cham: Springer International Publishing.

Bali, T. G., S. Brown, S. Murray, and Y. Tang. 2017. A lottery demand-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis* 52:2369–97.

Bali, T. G., N. Cakici, and R. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–46.

Bali, T. G., A. D. Gunaydin, T. Jansson, and Y. Karabulut. 2023. Do the rich gamble in the stock market? low risk anomalies and wealthy households. *Journal of Financial Economics* 150:103715–.

Banerjee, A. V. 1992. A simple model of herd behavior. *Quarterly Journal of Economics* 107:797–817.

Barber, B., and T. Odean. 2008. All that glitters: The effect of attention on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz. 2022. Attention-induced trading and returns: Evidence from robinhood users. *Journal of Finance* 77:3141–90.

———. 2023. A (sub)penny for your thoughts: Tracking retail investor activity in taq. *Journal of Finance* Forthcoming.

Barber, B. M., and T. Odean. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics* 116:261–92.

Barberis, N., and M. Huang. 2008. Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98:2066–100.

Barberis, N., A. Mukherjee, and B. Wang. 2016. Prospect theory and stock returns: An empirical test. *Review of Financial Studies* 29:3068–107.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100:992–1026.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang. 2021. Tracking retail investor activity. *Journal of Finance* 76:2249–305.

Boyer, B., T. Mitton, and K. Vorkink. 2010. Expected idiosyncratic skewness. *Review of Financial Studies* 23:169–202.

Boyer, B., and K. Vorkink. 2014. Stock options as lotteries. *Journal of Finance* 69:1485–528.

Brown, J. R., Z. Ivković, P. A. Smith, and S. Weisbenner. 2008. Neighbors matter: Causal community effects and stock market participation. *Journal of Finance* 63:1509–31.

Brunnermeier, M. K., and J. A. Parker. 2005. Optimal expectations. *American Economic Review* 95:1092–118.

Buffett, W. 2024. 2024 letter to Berkshire Hathaway shareholders.

Carpenter, J., F. Lu, and R. Whitelaw. 2016. The real value of china's stock market. Working paper, New York University.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang. 2014. Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *Review of Financial Studies* 27:1367–403. ISSN 0893-9454. doi:10.1093/rfs/hhu001.

Chen, H. C., and B.-H. Hwang. 2022. Listening in on investors' thoughts and conversations. *Journal of Financial Economics* 145:426–44.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, D. Johnston, M. Koenen, E. Laguna-Muggenburg, F. Mudekereza, T. Rutter, N. Thor, W. Townsend, R. Zhang, M. Bailey, P. Barberá, M. Bhole, and N. Wernerfelt. 2022. Social capital I: Measurement and associations with economic mobility. *Nature* 608:108–21.

Chi, S. S., and D. M. Shanthikumar. 2017. Local bias in Google search and the market response around earnings announcements. *Accounting Review* 92:115–43.

Cohen, L., A. Frazzini, and C. J. Malloy. 2008. The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy* 116:951–79.

Cookson, J. A., J. Engelberg, and W. Mullins. 2023. Echo chambers. *Review of Financial Studies* Forthcoming.

Cookson, J. A., R. Lu, W. Mullins, and M. Niessner. 2024. The Social Signal. *Journal of Financial Economics* 158:103870–.

Cookson, J. A., W. Mullins, and M. Niessner. 2024. Social media and finance. *Available at SSRN 4806692* .

Cookson, J. A., and M. Niessner. 2020. Why don't we agree? Evidence from a social network of investors. *Journal of Finance* 75:173–228.

———. 2023. Investor disagreement: Daily measures from social media. Working Paper.

Coval, J. D., and T. J. Moskowitz. 1999. Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance* 54:2045–73.

Crawford, S. S., W. R. Gray, and A. E. Kern. 2017. Why do fund managers identify and share profitable ideas? *Journal of Financial and Quantitative Analysis* 52:1903–26.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *Journal of Finance* 66:1461–99.

Daniel, K. D., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–58.

DeMarzo, P. M., R. Kaniel, and I. Kremer. 2008. Relative Wealth Concerns and Financial Bubbles. *Review of Financial Studies* 21:19–50. ISSN 0893-9454.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel. 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118:909–68.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, and T. Solorio, eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–86. Minneapolis, Minnesota: Association for Computational Linguistics.

Duflo, E., and E. Saez. 2002. Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics* 85:121–48.

———. 2003. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics* 118:815–42.

Eaton, G. W., T. C. Green, B. S. Roseman, and Y. Wu. 2022. Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics* 146:502–28.

Engelberg, J., D. McLean, and J. Pontiff. 2018. Anomalies and news. *Journal of Finance* 73:1971–2001.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 46:427–66.

———. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.

Frazzini, A. 2006. The disposition effect and under-reaction to news. *Journal of Finance* 61:2017–46.

Gao, X., and J. R. Ritter. 2010. The marketing of seasoned equity offerings. *Journal of Financial Economics* 97:33–52.

Gelman, M., D. Hirshleifer, Y. Levi, and L. Reiter Gavish. 2024. Social interaction intensity and investor behavior.

Giannini, R., P. Irvine, and T. Shu. 2019. The convergence and divergence of investors' opinions around earnings news: Evidence from a social network. *Journal of Financial Markets* 42:94–120.

Gomez, J.-P., R. Priestley, and F. Zapatero. 2009. Implications of keeping-up-with-the-joneses behavior for the equilibrium cross section of stock returns: International evidence. *Journal of Finance* 64:2703–37.

Green, T. C., and B.-H. Hwang. 2012. IPOs as lotteries: Skewness preference and first-day returns. *Management Science* 58:432–44.

Han, B., D. Hirshleifer, and J. Walden. 2022. Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis* 57:390–412.

Han, B., and A. Kumar. 2013. Speculative retail trading and asset prices. *Journal of Financial and Quantitative Analysis* 377–404.

Han, B., and L. Yang. 2013. Social networks, information acquisition, and asset prices. *Management Science* 59:1444–57.

Harris, M., and A. Raviv. 1993. Differences of opinion make a horse race. *Review of Financial Studies* 6:475–506.

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263–95.

Heimer, R. 2016. Peer pressure: Social interaction and the disposition effect. *Review of Financial Studies* 29:3177–209.

Hirshleifer, D. 2020. Presidential address: Social transmission bias in economics and finance. *Journal of Finance* 75:1779–831.

Hirshleifer, D., L. Peng, and Q. Wang. 2024. News diffusion in social networks and stock market reactions. *Review of Financial Studies* hhae025.

Hong, H., W. Jiang, N. Wang, and B. Zhao. 2014. Trading for Status. *Review of Financial Studies* 27:3171–212.

Hong, H., J. Kubik, and J. Stein. 2004. Social interaction and stock market participation. *Journal of Finance* 59:137–63.

———. 2005. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *Journal of Finance* 60:2801–24.

———. 2008. The only game in town: Stock-price consequences of local bias. *Journal of Financial Economics* 90:20–37.

Hong, H., W. Torous, and R. Valkanov. 2007. Do industries lead the stock market? *Journal of Financial Economics* 83:367–96.

Hong, H., and J. Xu. 2019. Inferring latent social networks from stock holdings. *Journal of Financial Economics* 131:323–44.

Hou, K., and T. Moskowitz. 2005. Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies* 18:981–1020.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28:650–705.

Huang, S., B.-H. Hwang, and D. Lou. 2021. The rate of communication. *Journal of Financial Economics* 141:533–50.

Huberman, G. 2001. Familiarity breeds investment. *Review of Financial Studies* 14:659–80.

Ivković, Z., and S. Weisbenner. 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. *Journal of Finance* 60:267–306.

———. 2007. Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *Review of Financial Studies* 20:1327–57.

Jackson, M. O., S. Malladi, and D. McAdams. 2021. The fragility of social learning in the presence of noisy messages. Working Paper, Stanford University.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.

Kandel, E., and N. D. Pearson. 1995. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy* 103:831–72.

Kaniel, R., S. Liu, G. Saar, and S. Titman. 2012. Individual investor trading and return patterns around earnings announcements. *Journal of Finance* 67:639–80.

Kaustia, M., and S. Knüpfer. 2012. Peer performance and stock market entry. *Journal of Financial Economics* 104:321–38.

Kelley, E. K., and P. C. Tetlock. 2013. How wise are crowds? insights from retail orders and stock returns. *Journal of Finance* 68:1229–65.

———. 2017. Retail short selling and stock prices. *Review of Financial Studies* 30:801–34.

Kelly, M., and C. O. O'Grada. 2000. Market contagion: Evidence from the panics of 1854 and 1857. *American Economic Review* 90:1110–24.

Kim, O., and R. E. Verrecchia. 1991. Trading volume and price reactions to public announcements. *Journal of Accounting Research* 29:302–21.

Kuchler, T., and J. Stroebel. 2021. Social Finance. *Annual Review of Financial Economics* 13:37–55.

Kumar, A. 2009. Who gambles in the stock market? *Journal of Finance* 64:1889–933.

Kumar, A., J. K. Page, and O. G. Spalt. 2011. Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics* 102:671–708.

———. 2016. Gambling and comovement. *Journal of Financial and Quantitative Analysis* 51:85–111.

Lin, T.-C., and X. Liu. 2018. Skewness, individual investor preference, and the cross-section of stock returns. *Review of Finance* 22:1841–76.

Liu, B., H. Wang, J. Yu, and S. Zhao. 2020. Time-varying demand for lottery: Speculation ahead of earnings announcements. *Journal of Financial Economics* 138:789–817.

Massa, M., and A. Simonov. 2006. Hedging, familiarity and portfolio choice. *Review of Financial Studies* 19:633–85.

Maturana, G., and J. Nickerson. 2018. Teachers teaching teachers: The role of workplace peer effects in financial decisions. *Review of Financial Studies* 32:3920–57.

Mitton, T., K. Vorkink, and I. Wright. 2018. Neighborhood effects on speculative behavior. *Journal of Economic Behavior & Organization* 151:42–61.

Ouimet, P. P., and G. A. Tate. 2020. Learning from coworkers: Peer effects on individual investment decisions. *Journal of Finance* 75:133–72.

Ozik, G., R. Sadka, and S. Shen. 2021. Flattening the illiquidity curve: Retail trading during the covid-19 lockdown. *Journal of Financial and Quantitative Analysis* 56:2356–2388–.

Pedersen, L. H. 2022. Game on: Social networks and markets. *Journal of FinancialEconomics* 146:1097–119.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22:435–80.

Pontiff, J. 2006. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics* 42:35–52.

Pool, V. K., N. Stoffman, and S. E. Yonker. 2015. The people in your neighborhood: Social interactions and mutual fund portfolios. *Journal of Finance* 70:2679–732.

Popper, N. 2020. Robinhood has lured young traders, sometimes with devastating results. *The New York Times* .

Roussanov, N. 2010. Diversification and its discontents: Idiosyncratic and entrepreneurial risk in the quest for social status. *Journal of Finance* 65:1755–88.

Scheinkman, J. A., and W. Xiong. 2003. Overconfidence, short-sale constraints, and bubbles. *Journal of Political Economy* 111:1183–219.

Seasholes, M. S., and N. Zhu. 2010. Individual investors and local bias. *Journal of Finance* 65:1987–2010.

Shiller, R. J. 2000. Measuring bubble expectations and investor confidence. *Journal of Psychology and Financial Markets* 1:49–60.

Shive, S. 2010. An epidemic model of investor behavior. *Journal of Financial and Quantitative Analysis* 45:169–98.

Shleifer, A., and R. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52:35–55.

Shumway, T. 1997. The delisting bias in crsp data. *Journal of Finance* 52:327–40.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104:288–302.

———. 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance* 70:1903–48.

Walkshausl, C. 2014. The max effect: European evidence. *Journal of Banking and Finance* 42:1–10.

Wang, H., J. Yan, and J. Yu. 2017. Reference-dependent preferences and the risk-return tradeoff. *Journal of Financial Economics* 123:395–414.

Welch, I. 2022. The wisdom of the robinhood crowd. *Journal of Finance* 77:1489–527.

Yuan, Y. 2015. Market-wide attention, trading, and stock returns. *Market-Wide Attention, Trading, and Stock Returns* 116:548–64.

Zhong, A., and P. Gray. 2016. The max effect: An exploration of risk and mispricing explanations. *Journal of Banking and Finance* 65:76–90.

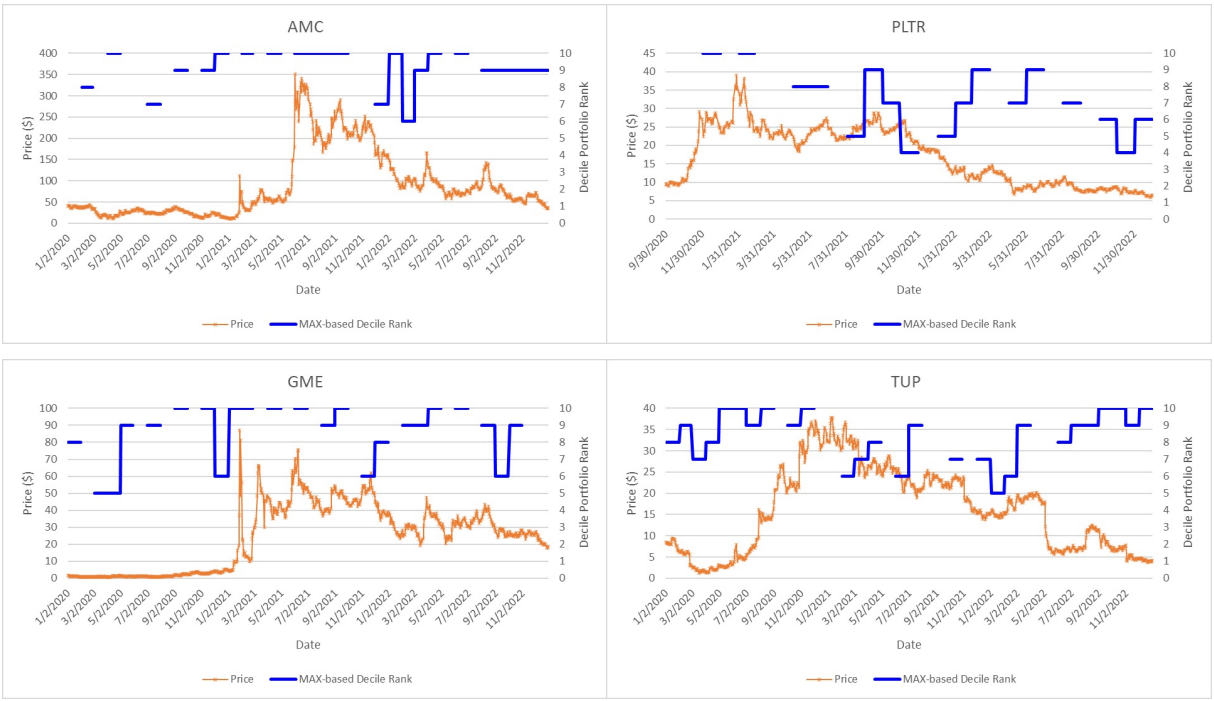


Figure I. Meme Stocks and Their Lottery-Characteristics This figure shows the prices and lottery decile rankings for AMC, GameStop (GME), Palantir (PLTR), and Tupperware (TUP) from 2020 to 2022. A stock experiences a MAX event on a given day if its return on that day is the highest over a trailing 21-trading-day window. We measure lottery characteristics using MAXRET, defined as the MAX event day return. Each month, stocks are sorted into decile portfolios based on MAXRET, with Deciles 10 and 1 representing the highest and lowest returns, respectively. If a month has multiple MAX events or none, we use the month's highest daily return as MAXRET.

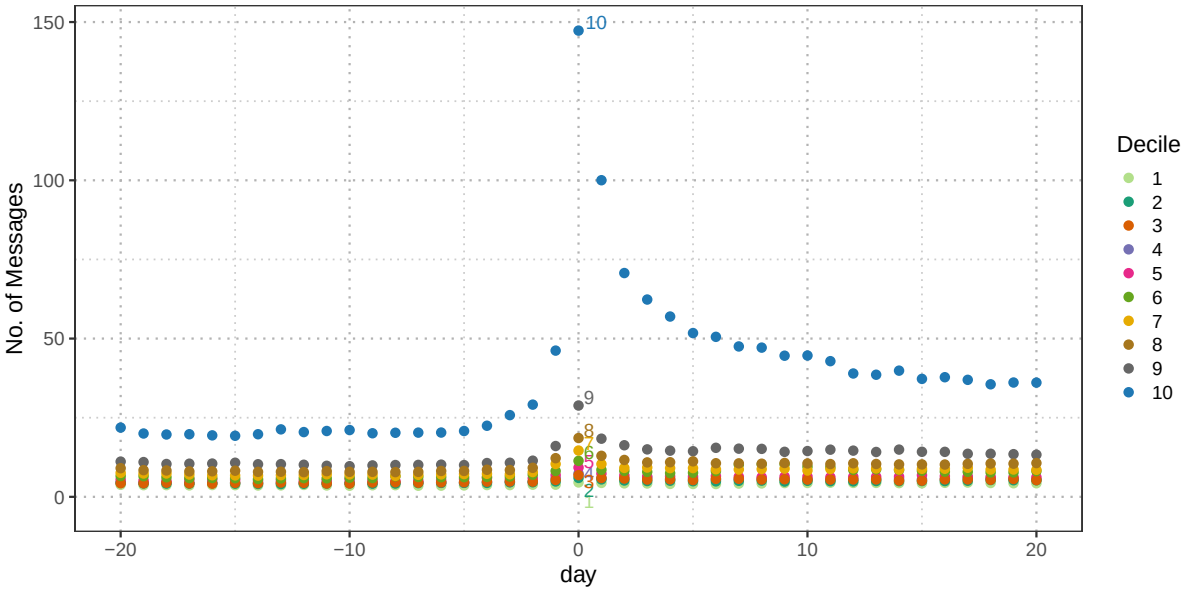


Figure II. StockTwits Activity around MAX Events This figure shows the average StockTwits message volume around MAX event days (day 0) for MAX decile portfolios. A stock experiences a MAX event on a given day if its return on that day is the highest over a trailing 21-trading day window. Decile 10 corresponds to MAX events with the top 10% MAX-day returns in the cross-section and decile 1 corresponds to events with the bottom 10% MAX returns. The sample period is from 2010 to 2022.

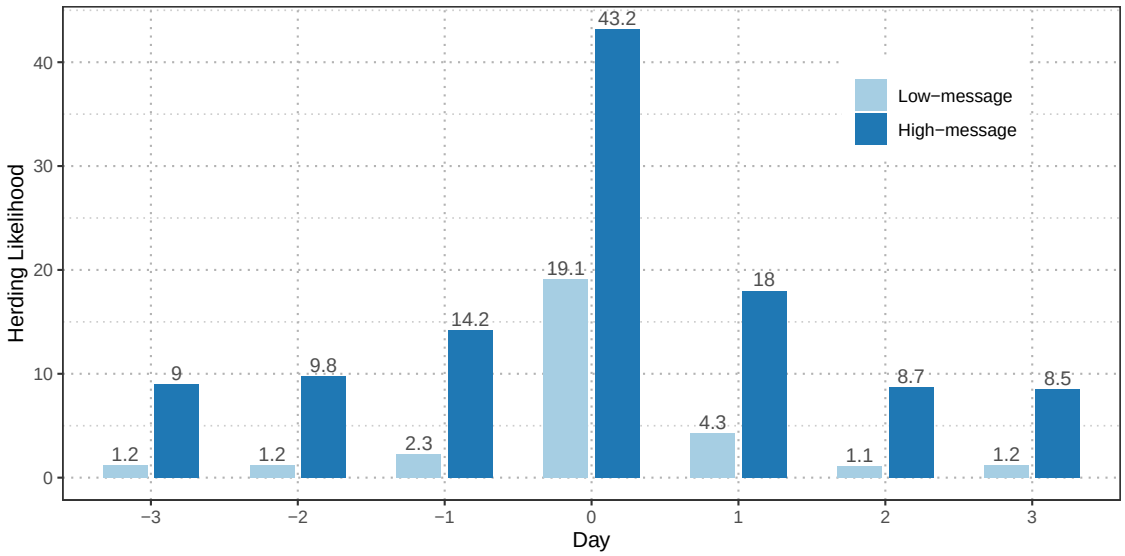
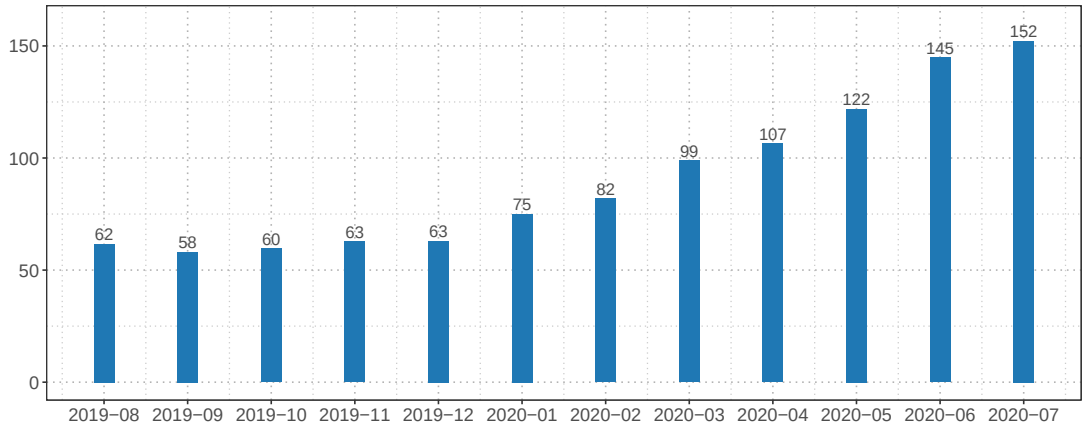
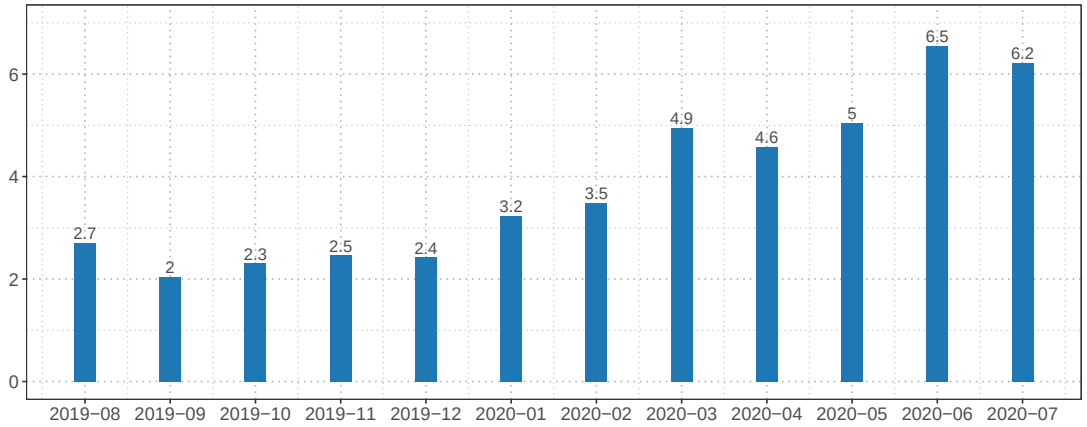


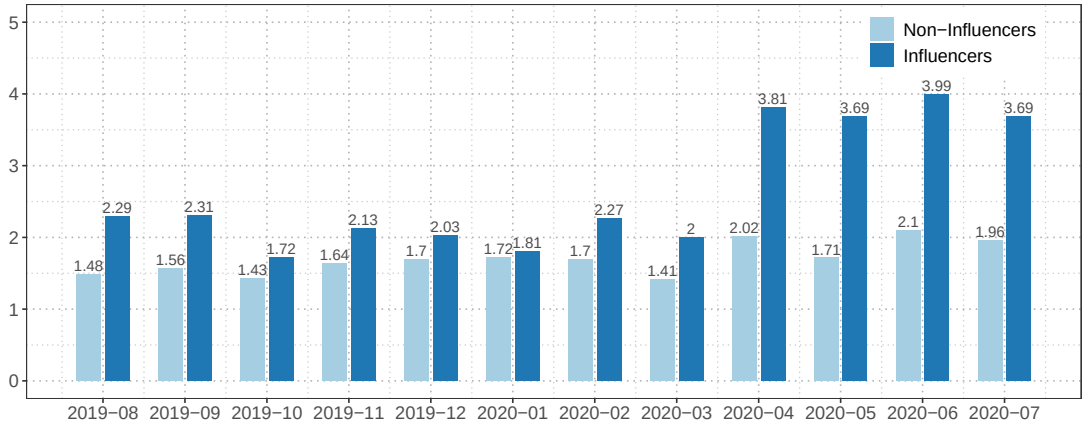
Figure III. StockTwits Activity and Robinhood Herding around Lottery Events This figure shows daily herding likelihood around lottery events (day 0). A stock experiences a MAX event on a given day if its return on that day is the highest over a trailing 21-trading-day window. A stock is then defined as experiencing a lottery event if it has a MAX event and its MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. We define a herding episode as a day when the daily increase in the number of Robinhood users for a stock ranks within the top 5% of the cross-sectional distribution of the lottery stock sample, requiring a minimum threshold of 100 new users on that day. We classify each stock-day observation in the $[-3, +3]$ window around the lottery event into two groups: High-message group (prior-day messages in top 10% of the cross-sectional distribution), and the Low-message group (all others). The figure plots herding frequency for both groups.



(A) Monthly Active Users (in thousands)



(B) Monthly Messages (in millions)



(C) Likes Per Message

Figure IV. StockTwits Activity around COVID-19 Outbreak This figure shows the monthly active users (in thousands, Panel A) and message volume (in millions, Panel B) on StockTwits from August 2019 to July 2020. Panel C focuses on stocks experiencing MAX events (day 0) and the messages mentioning these stocks during the [-11, 0] window. We present the monthly average number of likes per message for influencers (dark blue bars) and non-influencers (light blue bars), respectively.

Table I
SUMMARY STATISTICS

Panel A presents the summary statitics of each main variable utilized in the paper. Panel B displays the time series averages of the monthly cross-sectional correlations (multiplied by 100) among the variables. A stock experiences a MAX event on a given day (day 0) if its return on that day is the highest over a trailing 21-trading day window. MAXRET refers to the MAX-event day return. R represents the cumulative return (in percentages) over the window [1, 21]. Messages is the number of StockTwits messages over the [−11, 0] window. $\mathbf{I}_{mssg}$ equals one when Message ranks in the top 10th percentile of the cross-sectional distribution, and zero otherwise. The set of stock return predictors includes the market beta (BETA), the logarithm of market capitalization measured in millions of dollars (SIZE), the logarithm of book-to-market (BM), annual asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW).

A. Descriptive Statistics

				Percentile			
Variable	Mean	Median	Stdev	10^{th}	25^{th}	75^{th}	90^{th}
R	0.80	0.28	17.85	-14.59	-6.24	6.69	15.07
MAXRET	6.43	4.42	9.31	2.00	2.86	7.24	11.97
Messages	107.99	12.00	1628.41	0.00	3.00	40.00	123.00
$\mathbf{I}_{mssg}$	0.10	0.00	0.30	0.00	0.00	0.00	0.00
BETA	1.20	1.13	0.75	0.41	0.74	1.57	2.08
SIZE	6.56	6.55	2.16	3.74	4.99	8.02	9.39
BM	-0.70	-0.58	0.96	-1.91	-1.23	-0.09	0.32
IA	0.18	0.05	1.36	-0.13	-0.03	0.16	0.41
OP	-0.01	0.02	5.11	-0.13	-0.02	0.04	0.08
MOM	0.13	0.06	0.69	-0.42	-0.17	0.30	0.63
ILLIQ	3.06	0.00	65.48	0.00	0.00	0.05	0.63
COSKEW	-0.01	-0.01	0.22	-0.28	-0.15	0.13	0.26

B. Correlation Structure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) R	100											
(2) MAXRET	-4.2	100										
(3) Messages	0.1	10.2	100									
(4) $\mathbf{I}_{mssg}$	-1.8	18.9	15.7	100								
(5) BETA	1.1	9.5	2.0	8.5	100							
(6) SIZE	1.0	-25.1	3.7	18.5	0.6	100						
(7) BM	0.4	1.5	-3.7	-14.7	-1.4	-33.6	100					
(8) IA	-1.2	2.1	1.7	5.2	1.8	0.5	-5.4	100				
(9) OP	0.1	-0.5	0.2	0.0	0.0	0.6	1.4	-0.1	100			
(10) MOM	0.8	-5.3	8.2	3.3	6.2	5.2	2.9	-1.2	0.7	100		
(11) ILLIQ	0.0	5.4	-0.3	-1.4	-2.6	-7.8	3.7	-0.5	-0.2	-1.8	100	
(12) COSKEW	-1.1	0.7	1.5	5.3	16.7	3.9	-2.6	1.2	0.1	2.0	-1.0	100

Table II
SOCIAL INTERACTIONS AND LOTTERY STOCK RETURNS

This table reports the results of panel regression analysis regressing post-MAX returns on lagged explanatory variables. A stock experiences a MAX event on a given day (day 0) if its return on that day is the highest over a trailing 21-trading day window. MAXRET refers to the MAX-event day return. Cumulative returns are reported in percentages and are measured over windows [1, 21], [1, 42], and [1, 63], respectively. $\mathbf{I}_{mssg}$ equals one when the StockTwits message count for the stock over the [−11, 0] window ranks in the top 10th percentile of the cross-sectional distribution, and zero otherwise. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. *p <.1; **p <.05; ***p <.01.

	Cumulative Returns		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET $\cdot\mathbf{I}_{mssg}$	-0.147***	-0.172***	-0.254***
	(-3.57)	(-3.49)	(-3.72)
MAXRET	-0.053	-0.080	-0.104
	(-1.07)	(-1.35)	(-1.39)
$\mathbf{I}_{mssg}$	0.004	0.003	0.009
	(1.11)	(0.67)	(0.93)
BETA	0.001	0.002	0.002
	(0.59)	(0.61)	(0.39)
SIZE	-0.026***	-0.054***	-0.080***
	(-11.01)	(-11.62)	(-10.64)
BM	0.001	0.001	0.004
	(0.37)	(0.46)	(1.31)
IA	-0.002	-0.007**	-0.008*
	(-1.09)	(-2.09)	(-1.81)
OP	0.035***	0.057***	0.058***
	(6.19)	(5.63)	(3.65)
MOM	-0.008***	-0.020***	-0.030***
	(-3.09)	(-5.36)	(-6.11)
ILLIQ	0.000	0.000	0.000
	(-0.71)	(-0.47)	(-0.07)
COSKEW	-0.009	-0.017*	-0.019*
	(-1.56)	(-1.83)	(-1.83)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.20	15.54	17.15

Table III
SOCIAL INTERACTIONS AND LOTTERY STOCK RETURNS AROUND FUTURE EARNINGS ANNOUNCEMENTS

This table presents results of a panel regression examining the relationship between StockTwits message activity surrounding MAX event days and subsequent earnings announcement returns. MAX events are defined in Table II and MAXRET is the MAX-day return. The dependent variable is the DGTW-adjusted cumulative abnormal returns (in percentages), for the three days surrounding the subsequent quarter's earnings announcement. $\mathbf{I}_{mssg}^{pre\text{Event}}$ equals one if the stock ranks in the top 10% for messages posted during the $[-11,-1]$ window. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and the resultant *t*-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	CAR$[-1,1]$	
	(1)	(2)
MAXRET $\cdot$ $\mathbf{I}_{mssg}^{pre\text{Event}}$		-3.186**
		(-2.27)
MAXRET	-3.008***	-2.368***
	(-5.01)	(-3.73)
$\mathbf{I}_{mssg}^{pre\text{Event}}$		0.137
		(0.79)
BETA	0.070	0.075
	(0.80)	(0.85)
SIZE	-1.126***	-1.127***
	(-14.91)	(-14.84)
BM	-0.029	-0.031
	(-0.45)	(-0.49)
IA	0.026	0.026
	(0.94)	(0.94)
OP	-0.002	-0.002
	(-0.37)	(-0.36)
MOM	-0.277***	-0.267***
	(-4.68)	(-4.51)
ILLIQ	0.000	0.000
	(-0.58)	(-0.64)
COSKEW	-0.132	-0.140
	(-0.71)	(-0.76)
Obs.	128,088	128,002
Adj. R^2 (%)	6.4	6.4

Table IV
STOCKTWITS MESSAGE VOLUME AND LOTTERY EVENT PREDICTION

This table examines whether StockTwits message volume predict the likelihood of a stock achieving lottery status and the magnitude of its lottery characteristics using panel regressions. A stock experiences a MAX event on a given day if its return on that day is the highest over a trailing 21-trading-day window. A stock is then defined as experiencing a lottery event on that day if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. Columns 1 and 2 use the full sample of stock-day observations and correspond to the dependent variable, $\mathbf{I}^{lottery}$, which equals one if a stock experiences a lottery event on a given day (day 0), and zero otherwise. Columns 3 and 4 focus on a sample of stock-lottery event day observations, with the dependent variable being the lottery event day return. $\mathbf{I}_{mssg}^{preEvent}$ equals one if the StockTwits message counts mentioning the stock during the [−11,−1] day window rank in the top decile of its cross-sectional distribution. R^{preEvent} is the cumulative return during the [−11, −1] window. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Coefficients are multiplied by 100. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. *p <.1; **p <.05; *p <.01.

	Sample: All days $\mathbf{I}^{lottery}$		Sample: Lottery event days Lottery-event day return	
	(1)	(2)	(3)	(4)
$\mathbf{I}_{mssg}^{\text{preEvent}}$	0.228***	0.262***	3.142***	3.867***
	(10.82)	(12.87)	(5.81)	(7.16)
R^{preEvent}	-1.033***	-1.095***	11.764***	11.259***
	(-6.81)	(-7.30)	(6.73)	(6.45)
BETA		0.044***		-0.108
		(3.52)		(-0.29)
SIZE		-0.285***		-3.119***
		(-24.18)		(-13.62)
BM		-0.008		-0.648**
		(-0.90)		(-2.58)
IA		-0.022*		-0.656**
		(-1.92)		(-2.26)
OP		-0.150***		-1.199*
		(-4.66)		(-1.68)
MOM		-0.166***		-1.958***
		(-10.18)		(-7.88)
ILLIQ		0.014***		0.065**
		(9.24)		(2.17)
COSKEW		-0.014		-1.872
		(-0.50)		(-1.65)
Obs.	9,089,756	9,089,756	38,035	38,035
Adj. R^2 (%)	0.72	0.80	6.63	7.76

Table V
SOCIAL INTERACTIONS AND RETAIL HERDING IN LOTTERY STOCKS

This table presents the results of panel regressions analyzing the relationship between StockTwits message volume around lottery events and Robinhood herding episodes. Lottery events are defined in Table IV. The dependent variable in Column 1 is HR[0], an indicator for whether stock i experiences a herding episode on the lottery event day (day 0). The key explanatory variable is $\mathbf{I}^{preEvent}_{mssg,i}$, which equals one if the message volume for stock i during the pre-event window $[-11,-1]$ ranks within the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use HR$[1-5]$ and HR$[6-10]$ as dependent variables, measuring the average daily herding likelihood over days [1, 5] and [6, 10]. The key explanatory variable is $\mathbf{I}^{0}_{mssg}$, an indicator for stock i's message volume on day 0 ranking in the top decile cross-sectionally. $\mathbf{X}$ contains lagged controls for stock i: cumulative return (Return), herding likelihood (HR), log number of users (USER), log user change (ΔUSER), abnormal trading volume (AbVol), abnormal news (AbNews), and an indicator for earnings announcements coinciding with the lottery event. For column 1, controls are averaged over days $[-11,-1]$, with Return measured as the cumulative return over the same window. For columns 2 and 3, controls are measured on day 0. We include year-month fixed effects. Standard errors are clustered by firm and year-month. t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	HR[0]	HR$[1-5]$	HR$[6-10]$
	(1)	(2)	(3)
$\mathbf{I}^{preEvent}_{mssg}$ (or $\mathbf{I}^{0}_{mssg}$)	0.070***	0.020***	0.008
	(3.34)	(2.81)	(1.47)
Return	0.100***	0.030***	0.005
	(3.78)	(4.62)	(1.05)
HR	0.606***	0.076***	0.022***
	(19.43)	(7.50)	(4.32)
USER	0.033***	0.014***	0.006***
	(7.34)	(11.00)	(6.53)
ΔUSER	0.341**	0.003	-0.006
	(2.13)	(0.42)	(-1.06)
AbVol	0.009	-0.006***	0.001
	(0.77)	(-3.20)	(0.83)
AbNews	-0.013	-0.003	-0.002*
	(-0.88)	(-1.67)	(-1.77)
EA	-0.020	-0.016***	-0.014***
	(-1.13)	(-3.26)	(-5.61)
Obs.	7,981	7,947	7,946
Adj. R^2 (%)	20.66	18.41	6.30

Table VI
SOCIAL INTERACTIONS AND RETAIL TRADING OF LOTTERY STOCKS

This table presents the results of panel regressions examining the relationship between StockTwits message volume around lottery events and subsequent retail trading. Lottery events are defined in Table IV. The dependent variable in Column 1 is OIB[0], the retail order imbalance on the lottery event day (day 0). The key explanatory variable is $\mathbf{I}^{preEvent}_{mssg,i}$, which equals one if the message volume for stock i during the pre-event window $[-11,-1]$ ranks within the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use OIB$[1-5]$ and OIB$[6-10]$ as dependent variables, measuring the average daily retail order imbalance over days [1, 5] and [6, 10]. The key explanatory variable is $\mathbf{I}^{0}_{mssg}$, an indicator for stock i's message volume on day 0 ranking in the top decile cross-sectionally. $\mathbf{X}$ contains lagged controls for stock i: retail order imbalance (OIB), cumulative return (Return), herding likelihood (HR), log number of users (USER), log user change (ΔUSER), abnormal trading volume (AbVol), abnormal news (AbNews), and an indicator for earnings announcements coinciding with the lottery event. For column 1, controls are averaged over days $[-11,-1]$, with Return measured as the cumulative return over the same window. For columns 2 and 3, controls are measured on day 0. In Panel A, we use the Robinhood sample and include year-month fixed effects. In Panel B, we use the StockTwits sample and include firm and year-month fixed effects. Standard errors are clustered by firm and year-month. t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

A. Robinhood Sample: 06/2018−08/2020			
	OIB[0]	OIB$[1-5]$	OIB$[6-10]$
	(1)	(2)	(3)
$\mathbf{I}^{preEvent}_{mssg}$ (or $\mathbf{I}^{0}_{mssg}$)	2.656***	1.470***	0.705
	(3.29)	(3.46)	(1.54)
Return	-3.325***	2.754***	0.800
	(-2.91)	(6.43)	(1.57)
OIB	0.039	0.043***	0.032*
	(1.16)	(4.45)	(1.78)
HR	0.988**	0.894**	1.604***
	(2.74)	(2.32)	(3.24)
USER	-0.779***	1.093***	0.804***
	(-2.94)	(9.03)	(5.51)
ΔUSER	4.260	0.427	0.365
	(1.57)	(1.12)	(1.30)
AbVol	-1.585*	-1.420***	-1.253***
	(-1.75)	(-6.78)	(-5.77)
AbNews	-3.790***	-0.292	-0.063
	(-3.30)	(-1.03)	(-0.28)
EA	1.467*	0.071	1.278
	(1.82)	(0.12)	(1.67)
Obs.	7,981	7,980	7,978
Adj. R^2 (%)	1.12	2.99	1.90

B. StockTwits Sample: 01/2010−12/2022			
	OIB[0]	OIB[1 − 5]	OIB[6 − 10]
	(1)	(2)	(3)
$\mathbf{I}_{mssg}^{preEvent}$ (or $\mathbf{I}_{mssg}^{0}$)	-0.182	1.590***	1.302***
	(-0.51)	(7.14)	(5.28)
OIB	0.042***	0.033***	0.030***
	(3.16)	(7.09)	(5.82)
Return	-2.565***	1.046***	0.009
	(-3.94)	(2.72)	(0.03)
AbVol	-0.217	-0.719***	-0.542***
	(-0.64)	(-7.39)	(-6.30)
AbNews	-0.228	-0.695***	-0.622***
	(-0.48)	(-6.52)	(-6.45)
EA	-0.766*	1.004***	2.024***
	(-1.78)	(3.25)	(5.87)
Obs.	48,327	48,094	48,041
Adj. R^2 (%)	6.73	8.13	6.71

Table VII
SOCIAL INTERACTIONS AND INVESTOR DISAGREEMENT FOLLOWING LOTTERY EVENTS

This table presents the results of panel regressions examining the relationship between StockTwits message volume around lottery events and subsequent investor disagreement. Lottery events are defined in Table IV. The dependent variable in Column 1 is DIS[0], the standard deviation of the sentiment scores of messages on the lottery event day (day 0). The key explanatory variable is $\mathbf{I}_{mssg}^{preEvent}$, which equals one if the message volume for stock i during the pre-event window $[-11,-1]$ ranks within the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use DIS$[1-5]$ and DIS$[6-10]$ as dependent variables, measuring the average daily disagreement over days [1, 5] and [6, 10]. The key explanatory variable is $\mathbf{I}_{mssg}^{0}$, an indicator for stock i's message volume on day 0 ranking in the top decile cross-sectionally. $\mathbf{X}$ contains lagged controls for stock i: investor disagreement (DIS), cumulative return (Return), abnormal trading volume (AbVol), abnormal news (AbNews), and the indicator variable for whether an earnings announcement coincides with the lottery event day. We further control for the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). For column 1, controls are averaged over days $[-11,-1]$, with Return measured as the cumulative return over the same window. For columns 2 and 3, controls are measured on day 0. We include firm and year-month fixed effect and cluster the standard errors by firm and year-month. t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	DIS[0] (1)	DIS$[1-5]$ (2)	DIS$[6-10]$ (3)
$\mathbf{I}_{mssg}^{preEvent}$ (or $\mathbf{I}_{mssg}^{0}$)	0.015***	0.016***	0.019***
	(6.09)	(5.77)	(6.55)
DIS	0.119***	0.176***	0.124***
	(12.18)	(15.28)	(9.86)
Return	0.011***	0.027***	0.023***
	(5.01)	(6.58)	(5.39)
AbVol	0.003	0.008***	0.001
	(1.44)	(7.22)	(1.20)
AbNews	0.011***	0.001	-0.002
	(4.07)	(1.04)	(-1.47)
EA	-0.003	-0.002	-0.003
	(-1.09)	(-0.43)	(-0.72)
BETA	-0.002	0.000	-0.001
	(-1.47)	(-0.07)	(-0.32)
SIZE	0.001	0.002	0.001
	(0.95)	(1.04)	(0.53)
BM	-0.003*	-0.004***	-0.004**
	(-1.92)	(-2.68)	(-2.29)
IA	0.000	0.000	0.000
	(0.26)	(0.82)	(0.78)
OP	0.000	0.000**	0.000
	(-0.44)	(-2.36)	(0.86)
MOM	0.001**	0.003***	0.003***
	(2.49)	(4.10)	(3.15)
ILLIQ	0.000	0.000	0.000
	(-0.09)	(-0.31)	(0.49)
COSKEW	-0.006	-0.003	0.004
	(-0.84)	(-0.50)	(0.52)
Obs.	29,991	23,408	20,691
Adj. R^2 (%)	29.89	31.23	27.48

Table VIII
SOCIAL INTERACTIONS AND SHARE TURNOVER FOLLOWING LOTTERY EVENTS

This table presents the results of panel regressions examining the relationship between StockTwits message volume around lottery events and subsequent share turnover. Lottery events are defined in Table IV. The dependent variable in Column 1 is TO[0], the share turnover of stock i on the lottery event day (day 0). The key explanatory variable is $\mathbf{I}_{mssg}^{preEvent}$, which equals one if the message volume for stock i during the pre-event window $[-11,-1]$ ranks within the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use TO$[1-5]$ and TO$[6-10]$ as dependent variables, measuring the average daily turnover over days [1, 5] and [6, 10]. The key explanatory variable is $\mathbf{I}_{mssg}^{0}$, an indicator for stock i's message volume on day 0 ranking in the top decile cross-sectionally. $\mathbf{X}$ contains lagged controls for stock i: tunrover (TO), cumulative return (Return), abnormal trading volume (AbVol), abnormal news (AbNews), and the indicator variable for whether an earnings announcement coincides with the lottery event day. We further control for the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). For column 1, controls are averaged over days $[-11,-1]$, with Return measured as the cumulative return over the same window. For columns 2 and 3, controls are measured on day 0. We include firm and year-month fixed effect and cluster the standard errors by firm and year-month. t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	TO[0] (1)	TO$[1-5]$ (2)	TO$[6-10]$ (3)
$\mathbf{I}_{mssg}^{preEvent}$ (or $\mathbf{I}_{mssg}^{0}$)	0.108***	0.009***	0.007***
	(6.18)	(15.35)	(13.07)
Return	0.104***	0.012***	0.010***
	(4.34)	(10.35)	(10.39)
TO	1.782***	0.385***	0.260***
	(5.06)	(50.31)	(32.37)
AbVol	0.028***	-0.001***	-0.001***
	(2.98)	(-5.79)	(-10.61)
AbNews	0.008	-0.001***	-0.001***
	(0.90)	(-5.88)	(-8.15)
EA	-0.019*	-0.002***	-0.001***
	(-1.69)	(-7.76)	(-5.57)
BETA	0.007	0.000	0.000
	(0.83)	(1.19)	(1.24)
SIZE	-0.080***	-0.001***	-0.001***
	(-9.11)	(-5.75)	(-5.86)
BM	0.001	0.000*	0.000
	(0.16)	(-1.68)	(-1.36)
IA	-0.001	0.000*	0.000**
	(-0.44)	(1.93)	(2.59)
OP	0.000	0.000	0.000
	(-1.21)	(0.68)	(0.40)
MOM	-0.018***	0.001***	0.001***
	(-4.24)	(7.19)	(7.12)
ILLIQ	0.000**	0.000	0.000
	(-2.27)	(-1.07)	(-0.83)
COSKEW	-0.072**	0.000	0.001
	(-2.24)	(0.05)	(0.62)
Obs.	35,107 59	35,084	35,024
Adj. R^2 (%)	26.29	68.31	56.64

Table IX
HEADQUARTERS COUNTY SOCIAL CONNECTEDNESS AND LOTTERY STOCK RETURNS

This table reports the results of panel regression analysis regressing post-MAX returns on lagged explanatory variables. MAX events are defined in Table II and MAXRET is the MAX-day return. Cumulative returns are reported in percentages and are measured over windows [1,21], [1,42], and [1,63], respectively. SCIH is the social connectedness of the stock's firm headquarters county. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
	(1)	(2)	(3)
MAXRET · SCIH	-0.022**	-0.030**	-0.032*
	(-2.23)	(-2.02)	(-1.82)
MAXRET	-0.079	-0.101	-0.133*
	(-1.57)	(-1.57)	(-1.76)
SCIH	0.005***	0.007**	0.008**
	(3.70)	(2.58)	(1.99)
BETA	0.000	0.000	0.000
	(0.23)	(0.09)	(-0.00)
SIZE	-0.027***	-0.054***	-0.080***
	(-10.87)	(-12.33)	(-12.11)
BM	0.000	0.000	0.002
	(0.14)	(-0.12)	(0.67)
IA	0.000	-0.001	-0.003*
	(-0.06)	(-1.25)	(-1.71)
OP	0.003***	0.004**	0.003*
	(3.53)	(2.58)	(1.91)
MOM	-0.007***	-0.015***	-0.024***
	(-3.55)	(-4.91)	(-5.58)
ILLIQ	0.000	0.000	0.000
	(-0.28)	(-0.22)	(0.25)
COSKEW	-0.009	-0.016*	-0.017*
	(-1.57)	(-1.75)	(-1.80)
Obs.	423,531	422,514	421,351
Adj. R^2 (%)	11.47	16.87	18.67

Table X
SOCIAL CONNECTIVITY AND HOUSEHOLD TRADING FOLLOWING MAX EVENTS

This table analyzes social transmission of household trading following MAX events. MAX events (day 0) are defined in Table II. The dependent variable, Trade[2, 5]$_i$, corresponds to a focal household's relative buy size in the window of [2, 5]. MAXDCL is the decile rank of MAXRET in the crossesctional distribution (where rank 10 corresponds to lottery events), Trade[0, 1]$_j$ is the relative buy size of the peer household. SCI (in decile rank) is the social connected index between the locations of the firm and the household. MAX is the decile MAX-day return as defined in Table II. We also include lagged firm control variables.In all columns, we control for firm, focal household, and peer household fixed effects. Coefficients are multiplied by 10^4. Standard errors are three-way clustered by firm, focal household, and peer household, and the resultant *t*-statistics are shown in parentheses. *p <.1; **p <.05; ***p <.01.

	Trade[2, 5]$_i$				
	(1)	(2)	(3)	(4)	(5)
Trade[0, 1]$_j$	0.241	-0.147	-0.089	0.499	0.516
	(0.62)	(-0.40)	(-0.22)	(0.69)	(0.66)
Trade[0, 1]$_j$·SCI		0.069**	0.067**	-0.040	-0.044
		(2.34)	(2.03)	(-0.80)	(-0.83)
SCI		0.059***	0.059***	0.110***	0.107***
		(4.42)	(4.12)	(4.29)	(3.97)
Trade[0, 1]$_j$·SCI· MAXDCL				0.020***	0.022**
				(2.64)	(2.46)
Trade[0, 1]$_j$· MAXDCL				-0.117	-0.113
				(-1.34)	(-1.13)
SCI · MAXDCL				-0.009**	-0.009**
				(-2.29)	(-2.03)
MAX				0.161***	0.162***
				(3.40)	(2.99)
SIZE			0.010		0.044
			(0.04)		(0.17)
BM			-0.512*		-0.518*
			(-1.92)		(-1.96)
BETA			1.402**		1.418**
			(2.52)		(2.55)
IVOL			-20.380**		-21.002**
			(-2.45)		(-2.54)
ILLIQ			0.001		0.000
			(0.28)		(0.06)
MOM			0.681***		0.693***
			(5.00)		(5.05)
IA			0.467***		0.468***
			(3.04)		(3.02)
OP			0.008		0.006
			(0.43)		(0.33)
Obs.	53,632,948	53,632,948	47,326,147	53,632,948	47,326,147
Adj. R^2 (%)	1.76	1.76	1.91	1.76	1.91

Table XI
SOCIAL INTERACTIONS AND LOTTERY STOCK RETURNS: THE COVID-19 SHOCK

This table reports the results of panel regression analysis regressing post-MAX returns on lagged explanatory variables around COVID-19 pandemic breakout. The sample period is from August 1, 2019 till July 31, 2020. MAX events are defined in Table II and MAXRET is the MAX-day return. Cumulative returns are reported in percentages and are measured over windows [1, 21], [1, 42], and [1, 63], respectively. $\mathbf{I}_{mssg}$ equals one when the StockTwits message count for the stock over the [−11, 0] window ranks in the top 10th percentile of the cross-sectional distribution, and zero otherwise. $\mathbf{I}_{COVID}$ equals one for days after February 1, 2020 and zero otherwise. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and *t*-statistics are shown in parentheses. *p <.1; **p <.05; ***p <.01.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET · $\mathbf{I}_{mssg} \cdot \mathbf{I}_{COVID}$	-0.635**	-1.208***	-1.882**
	(-2.21)	(-3.25)	(-3.05)
MAXRET · $\mathbf{I}_{mssg}$	-0.272**	0.059	0.406
	(-2.21)	(0.24)	(0.97)
MAXRET · $\mathbf{I}_{COVID}$	0.832**	1.543***	2.189***
	(2.23)	(3.26)	(3.66)
MAXRET	-0.367***	-0.851***	-1.330***
	(-3.13)	(-4.19)	(-4.21)
$\mathbf{I}_{mssg} \cdot \mathbf{I}_{COVID}$	0.125**	0.193***	0.253***
	(2.83)	(3.22)	(4.12)
$\mathbf{I}_{mssg}$	0.007	-0.023	-0.050
	(0.33)	(-0.52)	(-0.88)
BETA	-0.009	-0.065	-0.087
	(-0.23)	(-0.99)	(-1.12)
SIZE	-0.035	-0.112***	-0.120***
	(-1.73)	(-3.65)	(-4.02)
BM	0.031***	0.018	0.040*
	(3.28)	(1.26)	(1.93)
IA	0.018**	0.021*	0.027*
	(2.39)	(2.13)	(1.80)
OP	0.042	0.049	-0.011
	(1.73)	(1.52)	(-0.25)
MOM	-0.108***	-0.213***	-0.336***
	(-3.94)	(-6.58)	(-10.17)
ILLIQ	-0.002	-0.002	0.000
	(-1.36)	(-1.41)	(0.05)
COSKEW	0.005	0.006	0.028
	(0.07)	(0.05)	(0.21)
Obs.	33,963	33,898	33,826
Adj. R^2 (%)	15.33	29.52	33.17

Appendix

- Table A.I presents monthly analyses of MAXRET and future stock returns using univariate portfolio sorts and Fama-MacBeth regressions.
- Table A.II presents panel regression analyses of MAXRET and future stock returns.
- Table A.III repeats the analyses in Table II using alternative StockTwits data shared by Cookson and Niessner (2023), covering 2010–2021.
- Table A.IV reproduces Table II using alternative message windows.
- Table A.V extends Table II by controlling investor attention, arbitrage costs, information supply, and arbitrage cost.

Table A.I
MONTHLY ANALYSIS OF MAXRET AND FUTURE STOCK RETURNS

In Panel A, we report average monthly returns of portfolios sorted on modified MAXRET. A MAX day is defined when the stock's return equals its highest daily return over a trailing 21-trading day window. MAXRET is the MAX-event day return. If a stock-month observation do not experience a MAX event or experience multiple MAX events, we set MAXRET as the highest daily returns for the month. The first row reports the value-weighted MAXRET in percentage terms for each decile and the average MAXRET difference between Decile 10 (High) and Decile 1 (Low). The second row reports the value-weighted monthly excess return (RET−RF) in percentage terms for each decile and the average return difference between Decile 10 and Decile 1; the last two rows present the corresponding alphas relative to the Fama-French-Carhart-Pastor-Stambaugh model (FFCPS) and the Fama-French five-factor model (FF5), respectively. The excess market returns and the size, book-to-market, momentum, profitability, and investment factors are from Kenneth French's data library. The liquidity factor is from Lubos Pastor's data library. In Panel B, we run a Fama-MacBeth regression of the one-month-ahead stock returns on MAXRET and the control variables; the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). Newey-West adjusted t-statistics are given in parentheses. The sample period is from July 1963 to December 2022. *p <.1; **p <.05; ***p <.01.

A. Portfolio Sorting

	Low	2	3	4	5	6	7	8	9	High	High−Low
MAXRET	1.63	2.75	3.63	4.50	5.44	6.55	7.95	9.88	13.11	24.03	22.40
RET−RF	0.52***	0.57***	0.50***	0.60***	0.70***	0.62***	0.56**	0.30	0.09	-0.46	-0.99***
	(3.69)	(3.65)	(2.89)	(3.19)	(3.16)	(2.64)	(2.02)	(0.97)	(0.28)	(-1.36)	(-3.38)
FFCPS	0.06	0.09	-0.04	0.01	0.03	0.04	-0.07	-0.36***	-0.60***	-1.14***	-1.20***
	(0.72)	(1.50)	(-0.68)	(0.18)	(0.32)	(0.43)	(-0.67)	(-3.17)	(-3.93)	(-5.37)	(-5.02)
FF5	-0.04	0.00	-0.11**	0.02	0.10	0.13*	0.03	-0.17*	-0.33***	-0.86***	-0.82***
	(-0.48)	(0.00)	(-1.96)	(0.25)	(1.27)	(1.71)	(0.32)	(-1.74)	(-2.84)	(-5.21)	(-4.52)

B. Stock-Level Fama-MacBeth Regression

	Cumulative Excess Returns		
	One-month	Two-month	Three-month
	(1)	(2)	(3)
MAXRET	-0.063***	-0.089***	-0.099***
	(-9.42)	(-7.73)	(-6.36)
BETA	0.197**	0.353**	0.459**
	(2.46)	(2.28)	(2.10)
SIZE	-0.161***	-0.256***	-0.335***
	(-4.82)	(-4.20)	(-3.94)
BM	0.154***	0.372***	0.614***
	(2.75)	(3.51)	(4.07)
IA	-0.447***	-0.942***	-1.363***
	(-6.00)	(-6.46)	(-6.50)
OP	0.011	0.108	0.226*
	(0.18)	(1.07)	(1.80)
MOM	0.005***	0.011***	0.015***
	(4.47)	(5.11)	(5.18)
ILLIQ	0.020	0.019	0.042
	(1.40)	(0.81)	(1.60)
COSKEW	-0.150	-0.255	-0.320
	(-1.56)	(-1.55)	(-1.48)
Obs.	2,170,905	2,156,352	2,141,886
Adj. R^2 (%)	5.5	6.4	6.8

Table A.II
PANEL REGRESSION ANALYSIS OF MAXRET AND FUTURE FUTURES: BASELINE MODEL

This table reports the results of panel regression analysis regressing post-MAX returns on lagged explanatory variables. MAX events are defined in Table II and MAXRET is the MAX-day return. Cumulative returns are reported in percentages and are measured over windows [1, 21], [1, 42], and [1, 63], respectively. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and *t*-statistics are shown in parentheses. $^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET	-0.105**	-0.142**	-0.192***
	(-2.33)	(-2.58)	(-2.71)
BETA	0.001	0.002	0.001
	(0.52)	(0.56)	(0.35)
SIZE	-0.027***	-0.054***	-0.081***
	(-11.08)	(-11.73)	(-10.80)
BM	0.001	0.001	0.004
	(0.50)	(0.56)	(1.40)
IA	-0.002	-0.007**	-0.008*
	(-1.13)	(-2.13)	(-1.85)
OP	0.036***	0.057***	0.059***
	(6.32)	(5.70)	(3.70)
MOM	-0.008***	-0.020***	-0.030***
	(-3.26)	(-5.50)	(-6.17)
ILLIQ	0.000	0.000	0.000
	(-0.39)	(-0.20)	(0.23)
COSKEW	-0.009	-0.017*	-0.019*
	(-1.56)	(-1.83)	(-1.84)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.15	15.50	17.11

Table A.III
SOCIAL INTERACTIONS AND LOTTERY STOCK RETURNS: ALTERNATIVE STOCKTWITS SAMPLE

This table repeats Table II using the alternative StockTwits data shared by Cookson and Niessner (2023) for the period from 2010 to 2021. See Table II for regression specification and variable descriptions.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET $\cdot \mathbf{I}_{mssg}$	-0.124**	-0.185***	-0.265***
	(-2.07)	(-3.39)	(-3.70)
MAXRET	-0.066	-0.079	-0.108
	(-1.17)	(-1.19)	(-1.29)
$\mathbf{I}_{mssg}$	0.005	0.010**	0.016
	(1.16)	(2.08)	(1.51)
BETA	0.001	0.003	0.002
	(0.61)	(0.77)	(0.53)
SIZE	-0.028***	-0.058***	-0.085***
	(-10.59)	(-12.12)	(-11.39)
BM	0.001	0.002	0.005
	(0.76)	(0.80)	(1.56)
IA	0.000	-0.004	-0.005
	(-0.02)	(-1.50)	(-1.35)
OP	0.034***	0.053***	0.057***
	(5.80)	(5.15)	(3.61)
MOM	-0.010***	-0.024***	-0.036***
	(-4.07)	(-6.59)	(-7.39)
ILLIQ	0.000	0.000	0.000
	(-1.38)	(-0.94)	(-0.47)
COSKEW	-0.010	-0.018*	-0.017
	(-1.63)	(-1.88)	(-1.62)
Obs.	384,745	383,852	382,791
Adj. R^2 (%)	10.24	15.77	17.67

Table A.IV
SOCIAL INTERACTIONS AND LOTTERY STOCK RETURNS: ALTERNATIVE MESSAGE WINDOWS

This table reproduces Table II using alternative definitions of StockTwits message indicator, $\mathbf{I}_{mssg}$, that are measured with different event windows: $[-5, 0]$, $[-3, 0]$, and $[-1, 0]$, respectively. See Table II for regression specification and variable definitions.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
A. Message Window: [-5, 0]			
MAXRET · $\mathbf{I}_{mssg}$	-0.106*	-0.158***	-0.199***
	(-1.975)	(-3.496)	(-2.993)
MAXRET	-0.058	-0.071	-0.108
	(-1.069)	(-1.143)	(-1.381)
$\mathbf{I}_{mssg}$	0.003	0.004	0.009
	(0.777)	(1.040)	(1.024)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.18	15.53	17.13
B. Message Window: [-3, 0]			
MAXRET · $\mathbf{I}_{mssg}$	-0.104*	-0.145***	-0.162**
	(-1.946)	(-2.910)	(-2.361)
MAXRET	-0.049	-0.064	-0.109
	(-0.838)	(-0.956)	(-1.265)
$\mathbf{I}_{mssg}$	0.002	0.002	0.005
	(0.537)	(0.647)	(0.692)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.18	15.53	17.13
C. Message Window: [-1, 0]			
MAXRET · $\mathbf{I}_{mssg}$	-0.085	-0.139***	-0.142**
	(-1.515)	(-2.681)	(-1.987)
MAXRET	-0.056	-0.057	-0.105
	(-0.820)	(-0.760)	(-1.055)
$\mathbf{I}_{mssg}$	0.004	0.004	0.005
	(1.284)	(1.399)	(0.772)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.16	15.52	17.12

Table A.V
SOCIAL INTERACTIONS AND LOTTERY STOCK RETURNS: ADDITIONAL CONTROLS

This table extends Table II by controlling for analyst coverage (CVRG) in Panel A, firm-level news (NEWS) in Panel B , and arbitrage costs measured by idiosyncratic volatility (IVOL) in Panel C, and arbitrage cost score (COST) in Panel D, respectively. See Table II for regression specification and variable definitions.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
A. Investor Attention: CVRG			
MAXRET $\cdot$ $\mathbf{I}_{mssg}$	-0.147***	-0.172***	-0.254***
	(-3.561)	(-3.489)	(-3.708)
CVRG	0.000	0.000	-0.002
	(-0.079)	(-0.120)	(-0.464)
Obs.	428,979	425,056	420,503
Adj. R^2 (%)	10.20	15.56	17.15
B. Information Supply: NEWS			
MAXRET $\cdot$ $\mathbf{I}_{mssg}$	-0.159***	-0.195***	-0.294***
	(-3.335)	(-3.749)	(-3.765)
NEWS	-0.001	-0.003***	-0.005***
	(-1.321)	(-2.739)	(-3.110)
Obs.	357,905	357,047	356,058
Adj. R^2 (%)	10.38	15.30	17.56
C. Arbitrage Cost: IVOL			
MAXRET $\cdot$ $\mathbf{I}_{mssg}$	-0.146***	-0.170***	-0.252***
	(-3.525)	(-3.440)	(-3.696)
IVOL	-0.001	-0.001*	-0.001
	(-1.554)	(-1.777)	(-1.651)
Obs.	428,979	425,056	420,503
Adj. R^2 (%)	10.20	15.57	17.15
D. Arbitrage Cost: COST			
MAXRET $\cdot$ $\mathbf{I}_{mssg}$	-0.142***	-0.163***	-0.240***
	(-3.431)	(-3.289)	(-3.503)
COST	0.004***	0.008***	0.012***
	(2.830)	(3.224)	(3.795)
Obs.	428,979	425,056	420,503
Adj. R^2 (%)	10.21	15.59	17.18

Internet Appendix

- Table S.I examines the robustness of Table II to alternative definitions of lottery events.
- Table S.II examines the robustness of Table IV to alternative definitions of lottery events.
- Table S.III examines the robustness of Table V to alternative definitions of lottery events.
- Table S.IV examines the robustness of Table VI to alternative definitions of lottery events.

Table S.I
SOCIAL INTERACTIONS AND LOTTERY STOCK RETURNS: ALTERNATIVE LOTTERY EVENT DEFINITIONS

This table examines the robustness of the results presented in Table II to alternative lottery event definitions. Panel A corresponds to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. Since this alternatively-defined lottery events do not have an associated MAX event, it is not directly applicable to Table II. We therefore estimate a modified version of this table by replacing MAXRET with an indicator for lottery events. Panel B corresponds to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than it return over the $[-31, -11]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. See Table II for regression specification and additional variable definitions.

A. Path-Independent Lottery Event Definition

	Cumulative Returns		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
$\mathbf{I}_{lottery} \cdot \mathbf{I}_{mssg}$	-0.032***	-0.045***	-0.064***
	(-6.51)	(-7.06)	(-5.99)
$\mathbf{I}_{lottery}$	-0.014***	-0.020***	-0.020***
	(-5.16)	(-5.23)	(-4.43)
$\mathbf{I}_{mssg}$	-0.004**	-0.009**	-0.012**
	(-2.04)	(-2.55)	(-2.36)
BETA	0.002	0.003	0.004
	(1.08)	(1.15)	(1.04)
SIZE	-0.026***	-0.051***	-0.077***
	(-12.36)	(-12.20)	(-12.13)
BM	0.001	0.002	0.003
	(0.60)	(0.90)	(1.11)
IA	-0.003**	-0.007***	-0.010***
	(-2.48)	(-2.79)	(-2.85)
OP	0.034***	0.050***	0.058***
	(8.00)	(5.91)	(4.74)
MOM	-0.010***	-0.021***	-0.032***
	(-4.39)	(-5.82)	(-7.06)
ILLIQ	0.000	0.000	0.000
	(1.31)	(1.54)	(1.39)
COSKEW	-0.005	-0.012	-0.017*
	(-1.00)	(-1.28)	(-1.80)
Obs.	9,011,326	8,926,038	8,839,161
Adj. R^2 (%)	10.63	14.54	17.10

B. Lottery Event Definition, with a 10-day Gap

	Cumulative Returns		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET $\cdot \mathbf{I}_{mssg}$	-0.103***	-0.110***	-0.180***
	(-3.28)	(-2.77)	(-3.40)
MAXRET	-0.080***	-0.112***	-0.103**
	(-2.77)	(-2.97)	(-2.08)
$\mathbf{I}_{mssg}$	0.002	0.000	0.003
	(0.46)	(-0.02)	(0.36)
BETA	0.001	0.003	0.003
	(0.52)	(0.88)	(0.61)
SIZE	-0.028***	-0.054***	-0.082***
	(-10.25)	(-11.07)	(-10.58)
BM	-0.001	0.001	0.004
	(-0.46)	(0.27)	(1.33)
IA	-0.002	-0.007*	-0.006
	(-1.08)	(-1.79)	(-1.37)
OP	0.032***	0.045***	0.052***
	(4.50)	(3.91)	(3.40)
MOM	-0.006**	-0.019***	-0.030***
	(-2.50)	(-5.06)	(-5.30)
ILLIQ	0.000	0.000	0.000
	(-0.43)	(-0.59)	(-0.02)
COSKEW	-0.014*	-0.023*	-0.024*
	(-1.71)	(-1.95)	(-1.96)
Obs.	436,243	432,103	426,104
Adj. R^2 (%)	11.62	16.91	18.12

Table S.II
PREDICTING LOTTERY EVENTS: ALTERNATIVE LOTTERY EVENT DEFINITIONS

This table examines the robustness of the results presented in Table IV to alternative lottery event definitions. Panel A corresponds to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. Panel B corresponds to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than it return over the $[-31, -11]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. See Table IV for regression specification and additional variable definitions.

A. Path-Independent Lottery Event Definition

	Sample: All days $\mathbf{I}^{lottery}$		Sample: lottery event days Lottery-event day return	
	(1)	(2)	(3)	(4)
$\mathbf{I}_{mssg}^{preEvent}$	0.804***	0.859***	0.569*	0.911***
	(16.77)	(18.73)	(1.72)	(2.80)
$R^{preEvent}$	-0.305	-0.418	2.243***	1.521**
	(-0.98)	(-1.37)	(3.13)	(2.15)
BETA		0.055**		-0.053
		(2.48)		(-0.21)
SIZE		-0.517***		-2.081***
		(-23.18)		(-13.43)
BM		-0.034**		-0.427***
		(-2.12)		(-2.68)
IA		-0.041**		-0.333**
		(-2.01)		(-2.02)
OP		-0.361***		-0.588
		(-5.87)		(-1.27)
MOM		-0.245***		-1.432***
		(-7.62)		(-9.32)
ILLIQ		0.030***		0.027
		(9.34)		(1.43)
COSKEW		0.020		-1.533**
		(0.42)		(-2.03)
Obs.	9,087,250	9,087,250	61,764	61,764
Adj. R^2 (%)	1.46	1.62	5.35	6.10

B. Lottery Event Definition, with a 10-day Gap				
	Sample: All days $\mathbf{I}^{lottery}$		Sample: lottery event days Lottery-event day return	
	(1)	(2)	(3)	(4)
$\mathbf{I}_{mssg}^{preEvent}$	0.413***	0.455***	1.138**	1.522***
	(13.14)	(14.85)	(2.26)	(2.99)
$R^{preEvent}$	0.049	-0.008	0.457	-0.734
	(0.23)	(-0.04)	(0.42)	(-0.68)
BETA		0.033**		0.104
		(2.21)		(0.29)
SIZE		-0.300***		-3.128***
		(-20.63)		(-13.75)
BM		-0.007		-0.776***
		(-0.72)		(-3.23)
IA		-0.027**		-0.608**
		(-2.04)		(-2.31)
OP		-0.181***		-0.872
		(-4.75)		(-1.15)
MOM		-0.220***		-1.491***
		(-11.62)		(-6.65)
ILLIQ		0.017***		0.046
		(9.85)		(1.62)
COSKEW		-0.008		-2.175*
		(-0.26)		(-1.94)
Obs.	9,089,756	9,089,756	39,676	39,676
Adj. R^2 (%)	0.82	0.92	5.78	6.80

Table S.III
SOCIAL INTERACTIONS AND RETAIL HERDING IN LOTTERY STOCKS

This table examines the robustness of the results presented in Table V to alternative lottery event definitions. Panel A corresponds to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. Panel B corresponds to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than it return over the $[-31, -11]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. See Table V for regression specification and additional variable definitions.

A. Path-Independent Lottery Event Definition

	HR[0]	HR$[1-5]$	HR$[6-10]$
	(1)	(2)	(3)
$\mathbf{I}_{mssg}^{pre\text{Event}}$ (or $\mathbf{I}_{mssg}^{0}$)	0.028	0.021***	0.010*
	(1.44)	(3.02)	(1.74)
Return	0.039**	0.029***	0.004
	(2.72)	(4.11)	(0.78)
HR	0.973***	0.085***	0.030***
	(14.18)	(5.65)	(4.46)
USER	0.036***	0.014***	0.007***
	(7.99)	(11.39)	(6.14)
ΔUSER	0.332**	-0.001	-0.008
	(2.66)	(-0.11)	(-1.46)
AbVol	-0.011	-0.005**	0.000
	(-1.03)	(-2.37)	(0.12)
AbNews	-0.002	-0.003*	-0.003**
	(-0.11)	(-1.85)	(-2.32)
EA	-0.017	-0.018***	-0.015***
	(-0.82)	(-3.43)	(-5.14)
Obs.	10,116	10,643	10,642
Adj. R^2 (%)	19.25	17.92	6.99

B. Lottery Event Definition, with a 10-day Gap			
	HR[0]	HR$[1-5]$	HR$[6-10]$
	(1)	(2)	(3)
$\mathbf{I}_{mssg}^{pre\text{Event}}$ (or $\mathbf{I}_{mssg}^{0}$)	0.039**	0.012*	0.005
	(2.29)	(1.81)	(1.00)
Return	0.035**	0.032***	0.005
	(2.16)	(4.84)	(1.12)
HR	1.031***	0.089***	0.029***
	(13.00)	(4.84)	(3.71)
USER	0.033***	0.014***	0.007***
	(7.65)	(12.43)	(5.51)
ΔUSER	0.323**	-0.004	-0.008
	(2.23)	(-0.50)	(-1.59)
AbVol	-0.023*	-0.005**	0.000
	(-2.00)	(-2.62)	(0.41)
AbNews	-0.016	-0.004*	-0.002*
	(-0.97)	(-1.99)	(-2.01)
EA	-0.023	-0.017***	-0.014***
	(-1.43)	(-3.69)	(-6.05)
Obs.	8,895	8,862	8,861
Adj. R^2 (%)	22.35	18.60	6.78

Table S.IV
SOCIAL INTERACTIONS AND RETAIL TRADING OF LOTTERY STOCKS: ALTERNATIVE LOTTERY EVENT DEFINITIONS

This table examines the robustness of the results presented in Table VI to alternative lottery event definitions. Panels A and B correspond to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. Panels C and D correspond to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than it return over the $[-31, -11]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. In Panel A, a stock experiences a lottery event on a given day if its return for that day equals or exceeds the 90th percentile of the cross-sectional distribution of highest daily returns over the trailing 21-trading-day window. In Panel B, a stock experiences a MAX event on a given day (day 0) if its return on that day is the highest over the $[-31, -11]$ trading-day window. A stock is then defined as experiencing a lottery event on that day if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. See Table VI for regression specification and additional variable definitions.

A. Path-Independent Lottery Event Definition

	Robinhood Sample: 06/2018–08/2020		
	OIB[0]	OIB[1 − 5]	OIB[6 − 10]
	(1)	(2)	(3)
$\mathbf{I}^{pre\text{Event}}_{mssg}$ (or $\mathbf{I}^{0}_{mssg}$)	2.141***	1.743***	0.441
	(4.01)	(3.38)	(1.27)
OIB	0.037	0.035***	0.028*
	(1.318)	(3.121)	(1.762)
Return	-0.619	2.456***	0.786*
	(-1.38)	(5.03)	(2.01)
HR	0.772**	0.824***	1.607***
	(2.07)	(2.85)	(3.25)
USER	-0.621***	1.137***	0.939***
	(-3.03)	(7.82)	(7.20)
ΔUSER	4.064	0.298	0.267
	(1.47)	(0.70)	(0.93)
AbVol	-0.429	-1.062***	-1.095***
	(-0.62)	(-7.12)	(-6.49)
AbNews	-3.612***	-0.306	0.006
	(-3.39)	(-1.22)	(0.03)
EA	1.419	0.183	1.314*
	(1.63)	(0.37)	(1.91)
Obs.	10,681	10,874	10,872
Adj. R^2 (%)	1.26	2.93	1.95

B. Path-Independent Lottery Event Definition			
	StockTwits Sample: 01/2010−12/2022		
	OIB[0]	OIB[1 − 5]	OIB[6 − 10]
	(1)	(2)	(3)
$\mathbf{I}_{mssg}^{pre\text{Event}}$ (or $\mathbf{I}_{mssg}^{0}$)	-0.432	1.448***	1.220***
	(-1.26)	(7.66)	(5.83)
Return	-1.020***	0.898**	-0.172
	(-3.46)	(2.59)	(-0.66)
OIB	0.042***	0.030***	0.027***
	(3.25)	(8.23)	(6.83)
AbVol	-0.062	-0.514***	-0.422***
	(-0.25)	(-7.09)	(-6.39)
AbNews	-0.536	-0.612***	-0.539***
	(-1.46)	(-6.81)	(-6.46)
EA	-0.385	0.853***	1.913***
	(-1.08)	(2.79)	(5.81)
Obs.	65,649	68,810	68,721
Adj. R^2 (%)	6.61	7.65	6.22

C. Lottery Event Definition, with a 10-day gap			
	Robinhood Sample: 06/2018−08/2020		
	OIB[0]	OIB[1 − 5]	OIB[6 − 10]
	(1)	(2)	(3)
$\mathbf{I}_{mssg}^{pre\text{Event}}$ (or $\mathbf{I}_{mssg}^{0}$)	2.230***	1.412**	0.583
	(3.43)	(2.63)	(1.11)
OIB	0.043	0.046***	0.034*
	(1.42)	(4.10)	(1.97)
Return	-0.738*	2.278***	0.572
	(-1.72)	(5.62)	(1.43)
HR	0.639**	1.168***	1.672***
	(2.17)	(3.16)	(3.19)
USER	-0.683**	1.130***	0.838***
	(-2.46)	(9.62)	(5.25)
ΔUSER	3.620	0.594	0.638**
	(1.33)	(1.20)	(2.14)
AbVol	0.005	-1.143***	-1.186***
	(0.01)	(-6.07)	(-4.84)
AbNews	-4.256***	-0.404	-0.055
	(-3.63)	(-1.46)	(-0.31)
EA	1.228	0.456	1.415**
	(1.02)	(0.86)	(2.24)
Obs.	8,895	9,074	9,073
Adj. R^2 (%)	0.98	2.82	1.84

D. Lottery Event Definition, with a 10-day Gap			
	StockTwits Sample: 01/2010−12/2022		
	OIB[0]	OIB[1 − 5]	OIB[6 − 10]
	(1)	(2)	(3)
$\mathbf{I}_{mssg}^{pre\text{Event}}$ (or $\mathbf{I}_{mssg}^{0}$)	-0.445	1.420***	1.338***
	(-1.22)	(5.72)	(5.92)
OIB	0.025*	0.034***	0.033***
	(1.88)	(7.04)	(7.06)
Return	-1.109***	1.120***	0.038
	(-3.05)	(3.34)	(0.14)
AbVol	-0.069	-0.700***	-0.524***
	(-0.23)	(-7.20)	(-6.62)
AbNews	-0.354	-0.635***	-0.625***
	(-0.81)	(-6.39)	(-6.33)
EA	-0.402	1.036***	1.978***
	(-0.88)	(3.42)	(6.07)
Obs.	50,297	52,648	52,585
Adj. R^2 (%)	7.02	8.30	6.79