Jin and Yu *Financial Innovation* (2022) 8:68
https://doi.org/10.1186/s40854-022-00373-2

Financial Innovation

RESEARCH

Open Access

Does communication increase investors' trading frequency? Evidence from a Chinese social trading platform

Xuejun Jin and Jiawei Yu*

*Correspondence:
jwyu@zju.edu.cn

School of Economics, Zhejiang University, 866 Yuhangtang Road, Hangzhou, China

Abstract

This study examines the impact of communication on investors' trading frequency based on a unique dataset drawn from a Chinese social trading platform. We find robust evidence that real-account portfolio owners on the platform trade more frequently under the influence of the comments posted by their leaders (the owners of portfolios they have followed). Moreover, portfolio owners are more sensitive to the quantity than to the tone of leaders' comments. Finally, both trading frequency and leaders' comments negatively impact portfolio owners' future performance. Our findings support the notion that social interaction promotes active investment strategies.

Keywords: Communication, Social interaction, Social trading platform, Trading frequency

JEL Classification: G11, G41

Introduction

Some investors trade actively and invest in individual stocks and active funds, while their strategies underperform relative to passive strategies such as holding a market index. The leading explanation for investors' active trading is individual investors' overconfidence. In the theoretical models proposed by Odean (1998) and Gervais and Odean (2001), overconfident investors, who overestimate the precision of their knowledge or private information and attribute past positive outcomes to their own abilities, tend to trade more frequently. Following such models, investor overconfidence has been verified in the psychological laboratory (Benoît et al. 2015) and helps explain the active trading behavior of individual investors in empirical studies (Barber and Odean 2000a, 2002; Zhang et al. 2019).

The theory that overconfidence leads to excessive trading explains investors' active trading behavior from the perspective of their own psychological bias. However, there is also evidence that social interaction is also related to active trading behavior, and overconfidence seems unable to explain such evidence. For example, people who interact with their neighbors, attend church, and participate in sports activities are more likely to participate in the stock market and trade actively (Hong et al. 2004; Heimer 2014).

Thus, a natural question is why people who engage in social activities are more likely to be active investors.

As Shiller (1989) pointed out, investing in speculative assets is also a social activity. Investors spend time discussing investment ideas with others, thereby influencing each other's investment decisions. According to a recent theory of social transmission bias (Hirshleifer 2020), investors transmit ideas amongst each other and suffer systematic biases in the transmission process, which ultimately promotes behavioral biases and asset pricing anomalies. Following this line of thought, Han et al. (2022) model how investors communicate with each other and propagate active investing. The authors argue that people enjoy discussing their success rather than their defeats, which is called self-enhancing transmission (SET) bias. In other words, the messages initiated by senders are overoptimistic and biased towards good outcomes in general. However, message receivers do not fully discount SET; as a result, they trade more actively under the influence of such biased conversations. The moral aspect of this story is that communication propagates active trading.

Social transmission bias theory suggests that communication about investment ideas may explain why people who engage in social activities trade more actively. However, previous empirical studies, such as those by Hong et al. (2004) and Heimer (2014), do not identify whether people generate conversations about investment ideas when interacting with others and how such communications affect their trading behavior. Therefore, in this study, we empirically investigate whether investment-related communications increase trading aggressiveness using a dataset drawn from a Chinese social trading platform. Social trading platforms allow users to share investment ideas by developing trading strategies and by posting comments, thus allowing us to observe both users' communications and trading behavior.

Specifically, we investigate the relationship between the trading frequency of real-account portfolios and the comments posted by their leaders (the owners of the portfolios they follow). By using difference-in-differences and panel regression methods, we find that portfolio owners trade more frequently under the influence of leaders' comments. In addition, the trading frequency of portfolio owners further increases with the number of comments posted by leaders, although it is not sensitive to the tone of the comments. Overall, these findings suggest that communication positively affects trading frequency.

This study contributes to several strands of research. First, it contributes to the literature on excessive trading by providing a novel explanation. Prior studies attribute excessive trading mainly to investor overconfidence (Barber and Odean 2000b, 2001). Our findings suggest that the trading frequency of real-account portfolios on a social trading platform can be increased by leaders' comments, indicating that excessive trading can also be the product of communication.

Second, our study contributes to the literature on social finance and social transmission bias. Social finance is a new paradigm that studies how social processes shape financial thinking and behavior (Akçay and Hirshleifer 2021). Social transmission bias is one of the key building blocks of social finance, an area concerned with how the transmission of investment ideas affects investor behavior and asset prices (Hirshleifer 2020). We contribute to this strand of the literature by providing individual-level empirical evidence that investment-related communications positively impact the trading frequency of real-account portfolios on a social trading platform. Our findings support the theoretical model of Han et al. (2022), which predicts that trading aggressiveness increases with communication intensity.

Third, our study contributes to the literature on social trading platforms. Prior studies on social trading platforms documented the various effects of social interactions on investor behavior, such as peer pressure (Heimer 2016) and social recognition (Pelster and Hofmann 2018; Pelster and Breitmayer 2019). Ammann and Schaub (2020) were the first to use comment data drawn from a social trading platform and document a significant relationship between posting comments and investment flows of the portfolios. We add to this strand of literature by further exploiting the relationship between leaders' comments and the trading frequency of followers.

The remainder of this paper is organized as follows. Section "Literature review" reviews the relevant literature. In Section "Data and methodology", we describe the social trading platform and the data we used in our analysis. We report the results in Section "Empirical results" and discuss further insights in Section "Discussion, future direction and practical implications". Finally, we conclude the paper in Section "Conclusion".

Literature review

Excessive trading

Excessive trading is an important topic in behavioral finance. According to Barber and Odean (2000b), investors trade too frequently, resulting in poor performance; that is, higher trading levels are detrimental to investors' wealth. To explain this phenomenon, some researchers considered overconfidence as the main reason for excessive trading. In the overconfidence models proposed by Odean (1998) and Gervais and Odean (2001), investors overweight their private information and knowledge. Consequently, they trade more aggressively than is optimal.

Empirically, Barber and Odean (2001) find that male investors trade more excessively than female investors do because men are more overconfident than women in areas such as finance. Barber and Odean (2002) show that investors trade more frequently after switching from phone-based trading to online trading. The authors argue that investors can foster an illusion of knowledge after going online and having access to vast amounts of investment data, thus becoming more overconfident. Zhang et al. (2019) also find that overconfident investors are more likely to engage in intraday arbitrage, while such trading behavior hurts their performance.

Another behavioral bias relevant to excessive trading is the disposition effect — the tendency to sell winning assets while delaying losses (Shefrin and Statman 1985). The disposition effect can be explained by prospect theory, which predicts that investors value perceived gains over perceived losses (Tversky and Kahneman 1974). Both prospect and overconfidence theories indicate that trading volume increases with investors' past performance. Statman et al. (2006) distinguish the two theories by using a vector autocorrelation (VAR) and impulse-response function methodology and finding that the trading volume of individual stocks responds more to past market returns than past stock returns, which is hard to interpret solely by the disposition effect.

Existing studies explain excessive trading behavior from the perspective of investors' psychological bias. We add to this strand of literature by finding that the trading frequency of portfolio owners on social trading platforms can also be affected by their leaders through the channel of communication after controlling for their past performance and market effects (absorbed by time fixed effects).

Communication and investor behavior

Existing studies report that interpersonal communications are of great importance in investing decisions. For example, Hong et al. (2005) and Pool et al. (2015) provide evidence that institutional investors are more likely to buy the same stocks as their neighbors, probably due to word-of-mouth communication and information sharing among neighbors.

Three theories are closely related to how communication affects investors' decisions. The first is signaling theory. According to signaling theory, in the case of online social trading, signal senders share their private information with followers (signal receivers) by setting up portfolios and posting comments. After observing signals, signal receivers can improve their returns by replicating or investing their money directly into the portfolios of signal senders, and signal senders in turn can receive monetary compensation in relation to the number of copiers and assets under management (Kromidha and Li 2019; Pelster and Breitmayer 2019).

The second is social learning theory, according to which investors update their beliefs based on observations of the actions and choices of others when communicating with each other, resulting in similar behavior (Bikhchandani et al. 2021). Therefore, social learning theory can explain the findings of Hong et al. (2005) and Pool et al. (2015) that investors make decisions similar to those of their neighbors. Moreover, social learning theory predicts that herding or information cascades may arise when individuals rationally choose identical actions, leading to bubbles and crashes in financial markets (Alevy et al. 2007).

The third is social transmission bias theory. According to the social transmission bias theory, investment ideas or signals are directionally modified when they pass from person to person, leading to behavioral bias, return anomalies, and pricing bubbles (Hirshleifer 2020). For example, Han et al. (2022) argue that people are generally subject to SET bias when sending messages to others[1] and message receivers fail to adjust to this bias. Based on these features of the message sending and receiving process, they predict that communication promotes the spread of active investment strategies.

While all three theories indicate that communication can affect investors' trading behavior, our study is more relevant to social transmission bias theory. First, users on the social trading platform that we study receive no monetary compensation from followers' actions, which does not satisfy the assumption of signaling theory that signal senders should benefit from receivers' actions (Connelly et al. 2011). Second, social learning theory focuses on the contagion of investment choice, which does not explain a tilt towards

[1] Several empirical studies provide evidence of SET bias. For example, Ammann and Schaub (2020) find that traders on social trading platforms post more comments after achieving good performance. Lane et al. (2021) report that professional investors discuss more a stock they have traded with a gain than a stock with a loss.

active behavior (which social transmission bias theory does) because either active or passive behavior can spread from person to person (Han et al. 2022).

Social trading platforms and investor behavior

With the development of the Internet in the past two decades, the behavior of participants in financial markets changed significantly, such as in terms of financing (Abdeldayem and Aldulaimi 2021) and investing. In recent years, social trading platforms became increasingly popular among investors because they allow users to trade securities and communicate with each other simultaneously (Steiger and Pelster 2020).

Existing studies demonstrate that social interactions on social trading platforms significantly affect investor behavior, including the disposition effect (Heimer 2016; Pelster and Hofmann 2018), trading frequency (Breitmayer et al. 2018; Czaja and Röder 2020), and copy trading (Ammann and Schaub 2020). Breitmayer et al. (2018) first study the trading frequency of investors on social trading platforms, finding that investors who attract attention from peers on the social trading platforms tend to trade more. Czaja and Röder (2020) study the relationship between trading frequency and self-attribution bias using data from social trading platforms and find that investors trade more aggressively under the influence of self-attribution bias. We add to this strand of the literature by showing that investment ideas shared by leaders on social trading platforms can also increase portfolio owners' trading frequency.

Data and methodology

The social trading platform

The data used in our research come from a social trading platform based in China, Xueqiu.com (referred to Snowball[2] in the remainder of this study). Social trading platforms allow users to share investment ideas by developing trading strategies and by posting comments. On such platforms, users can follow the strategies they prefer and even copy the trades of others, or invest directly in others' strategies on some platforms.

In the case of Snowball, users can post comments and develop trading strategies after registration. Each user is allowed to create at most 20 virtual trading portfolios and link at most one real money brokerage account to the platform. If the number of portfolios hits the limit, then to set up a new strategy, users must close at least one current portfolio. All these operations executed by a user can be found on the user's profile page.

When gathering information from Snowball, all information, including news, comments, and trading strategies published by the platform or other users, is freely accessible after registration. To learn from others' trading strategies, users can follow any virtual or real money brokerage portfolio. However, unlike some other platforms, Snowball does not allow investors to copy or invest directly in others' strategies (portfolios). In other words, the only way to copy others' portfolios is to imitate their trades manually. Moreover, users can view comments posted by the owners of the portfolios they have followed (called leaders in the remainder of this paper) by viewing their profile pages. Although our dataset does not include detailed information on how visitors navigate

[2] The word "Xueqi" in Chinese translates into "Snowball."

these profile pages, as Ammann and Schaub (2020) report, approximately one-third of the clicks of profile page visitors on such platforms focus on viewing traders' comments, which is very close to the clicks of current portfolio holdings, suggesting that followers care about the communications of leaders. The authors also find evidence that comment posting is associated with an increase in the net investment of followers, indicating that followers trade on the comments posted by leaders.

Social trading platforms offer researchers opportunities to study the direct effects of social interactions on users' trading behavior. However, there are several limitations in studying social trading platforms. First, users self-select into social trading platforms, which raises concerns regarding whether traders using such platforms are representative. In our case, Snowball is one of the largest stock forums and social trading platforms in China, with more than 30 million active users. The barrier of entry to Snowball is very low, and anyone with a smartphone can create an account. Thus far, we do not find evidence that portfolio owners on Snowball are different from investors using other platforms or those who do not use social trading platforms. Second, because Snowball does not allow users to invest directly in other portfolios or copy single trades, users may set up virtual portfolios for leisure. Thus, virtual portfolios may not reflect portfolio owners' real preferences and investment behavior. To address this concern, we focus on whether the behavior of real-account portfolio owners is affected by the comments of their leaders. In this setting, followers include only real-account portfolios, whereas leaders include both virtual- and real-account portfolios. Moreover, users not only self-select into social trading platforms but also self-select to follow others. Thus, the choice to follow other portfolios may introduce a selection problem. To address this concern, we follow the empirical framework of Pelster and Breitmayer (2019) and use a propensity score matching (PSM) procedure to control the difference between portfolio owners who follow at least one portfolio of another user and those who do not follow any other portfolio. We then perform a difference-in-differences analysis and panel regressions using the matched sample and treated portfolios, respectively.

Our sample covers the period from July 2016 to December 2019. The sample includes all real and virtual portfolios that have been followed at least once by the owners of the real-account portfolios. We exclude portfolios that existed for less than 180 days (six months) and execute fewer than five trades in the sample period. After the trimming process, our sample included 10,034 real-account portfolios and 19,587 virtual portfolios.

Variables

Subsequently, we describe the variables used in our empirical analysis. We obtained all data used in this study from Snowball and aggregated all variables at the weekly frequency.

Trading frequency

The key issue we explore in this study is whether communication increases investors' trading frequency. Thus, we first define two variables that reflect portfolio owners' trading frequency. First, $Trades_{i,t}$ is the (log) number of trades executed by the owner of portfolio i in week t. Second, $Turnover_{i,t}$ is the turnover ratio of portfolio i in week t.

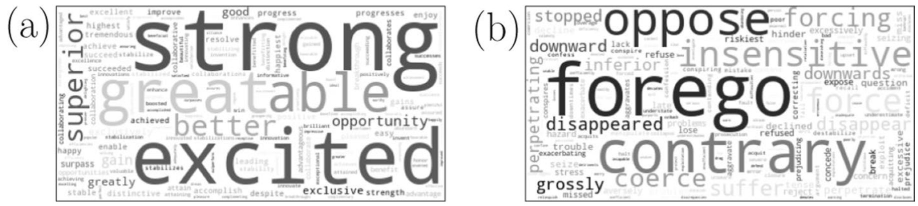

Fig. 1 This figure shows the frequency of words included in all of the studied comments, reflected by the size of each word. Panel **a** presents the word cloud for positive words. Panel **b** presents the word cloud for negative words. Definitions for positive and negative words are determined based on the Loughran and Mcdonald's dictionary. The words were translated through Google Translate

Comment characteristics

We define the variables that capture the different features of comments. We drop comments that contain fewer than five words because they were mostly uninformative.

First, we generate a dummy variable, $Leader\ Comment_{i,t}$, which takes the value of one if at least one leader of the owner of portfolio i posts at least one comment in week t, and zero otherwise. The variable $Leader\ Comment_{i,t}$ is restricted to weeks in which the portfolio owner followed at least one leader.

Second, we define $Leader\ Count_{i,t}$ as the average (log) number of comments posted by the leaders of the owner of portfolio i in week t. The variable $Leader\ Count_{i,t}$ is restricted to weeks in which at least one comment is posted by the leaders of the portfolio owner.

Third, to determine the tone of comments, we define $Leader\ Positive_{i,t}$ and $Leader\ Negative_{i,t}$ as the average fraction of positive and negative words included in all comments posted by the leaders of the owner of portfolio i in week t, respectively. We restrict both variables to the weeks in which at least one comment is posted by the leaders of the portfolio owner.

Following existing studies (see, e.g., Wang and Wu 2015; Xie and Lin 2015; Lin and Xie 2016; Zeng et al. 2018; Li et al. 2019), we first use the Jieba Tokenizer[3] to break the comments into words and then define the tone of words contained in comments based on Chinese translations of the Loughran and Mcdonald (2011) dictionary (LM dictionary). The original LM dictionary contained 353 positive and 2,337 negative words. Following Zeng et al. (2018), we translated all words into Chinese using Google Translate and retained all words in our list if one English word corresponded to several Chinese words. Hence, our Chinese word list contained 1,043 positive words and 3,422 negative words. Figure 1 shows the frequencies of positive (Panel (a)) and negative words (Panel (b)) in all comments. From Panel (a), we find that the most frequent positive words include "strong", "excited", "great", "able", and "better". In Panel (b), the most frequent negative words include "forego", "contrary", "oppose", "insensitive," and "forcing."

Although the Chinese LM dictionary is widely used in the practice of Chinese financial sentiment analysis, we must pay attention to its limitations. As Loughran and Macdonald (2016) point out, because the LM dictionary was derived in the context of 10-K filings, its accuracy cannot be guaranteed when applied to other media.

[3] Jieba is a widely used Chinese text segmentation Python package. This package is available at https://github.com/fxsjy/jieba.

Table 1 Summary statistics

	Obs	Mean	Std. dev.	P25	P50	P75
Trades	1,133,050	0.5432	0.8610	0.0000	0.0000	1.0986
Turnover	1,133,050	0.3996	1.1107	0.0000	0.0000	0.1838
Leader comment	689,206	0.6005	0.4898	0.0000	1.0000	1.0000
Leader count	413,890	1.6229	1.5679	0.4380	1.7243	2.7839
Leader positive	413,890	0.0561	0.0274	0.0488	0.0582	0.0656
Leader negative	413,890	0.0837	0.0355	0.0766	0.0885	0.0973
Return	1,133,050	− 0.0009	0.0391	− 0.0118	0.0000	0.0119
Return SD	1,133,050	0.0122	0.0405	0.0000	0.0090	0.0168
No.securities	1,133,050	1.9752	1.1127	1.0986	1.9459	2.8332
No.followers	1,133,050	0.2087	0.5431	0.0000	0.0000	0.0000
Portfolio age	1,133,050	3.8752	1.0102	3.4012	4.1431	4.6250
No.leaders	1,133,050	0.9335	0.9629	0.0000	0.6931	1.6094
Leader return	689,206	0.0010	0.0331	− 0.0127	0.0008	0.0169
Leader SD	689,206	0.0133	0.0142	0.0070	0.0114	0.0173
Leader trades	689,206	0.2858	0.4820	0.0000	0.0000	0.3973
Leader securities	689,206	1.7122	1.0548	0.8959	1.6094	2.3026
Leader followers	689,206	2.6001	1.8436	0.8100	2.3620	3.8202
Leader age	689,206	3.9980	1.1266	3.5264	4.3041	4.7769

This table presents the summary statistics of variables for real-account portfolios, including the number of observations (Obs), mean (Mean), standard deviation (Std. Dev.), the first quartile (P25), median (P50), and the third quartile (P75) of the variables

Other characteristics

We also define several variables that capture other portfolio features. We generate the following variables for each portfolio i in week t: the logarithmic return ($Return_{i,t}$), standard deviation of daily returns ($Return\ SD_{i,t}$), average (log) number of securities held ($No.securities_{i,t}$), average (log) number of followers ($No.followers_{i,t}$), and (log) number of weeks since the creation of the portfolio ($Portfolio\ Age_{i,t}$).

In addition, for the owner of portfolio i in week t, we generate the following variables to capture their leaders' characteristics, including the (log) number of leaders ($No.leaders_{i,t}$), average value of each leader's *Return* ($Leader\ Return_{i,t}$), average value of each leader's *Return SD* ($Leader\ SD_{i,t}$), average value of each leader's *Trades* ($Leader\ Trades_{i,t}$)[4], average value of each leader's *No.securities* ($Leader\ Securities_{i,t}$), average value of each leader's *No.followers* ($Leader\ Followers_{i,t}$), and average value of each leader's *Portfolio Age* ($Leader\ Age_{i,t}$).

Summary statistics

Table 1 presents the summary statistics of the variables for real-account portfolios. We observe that, on average, 60.05% of leaders post at least one comment in week t. We also find that the average return of real-account portfolios is negative, at − 0.09% per week in our sample, indicating that the owners of real-account portfolios in our sample generally lose money.

[4] We do not further define the average turnover of leaders because the number of trades and turnover are highly correlated.

Table 2 Correlation coefficients among variables

	Trades	Turnover	Leader comment	Leader count	Leader positive	Leader negative	Return	Return SD	No. securities
Trades	1								
Turnover	0.68	1							
Leader comment	0.11	0.03	1						
Leader count	0.04	−0.03	–	1					
Leader positive	−0.02	−0.03	–	0.11	1				
Leader negative	−0.01	−0.02	–	0.17	0.26	1			
Return	−0.05	−0.09	0.01	0.01	0.01	-0.02	1		
Return SD	0.08	0.09	0.02	0.01	0.00	0.00	−0.02	1	
No.securities	0.29	0.11	0.14	0.13	0.02	0.02	0.01	0.07	1
No.followers	0.03	0.00	0.09	0.03	0.01	0.03	0.02	0.02	0.14
Portfolio age	−0.10	−0.09	0.03	−0.02	0.07	0.12	0.01	0.00	−0.01
No.leaders	0.04	−0.02	0.41	0.03	0.08	0.12	0.01	0.01	0.11
Leader Return	−0.01	0.00	0.03	0.01	0.02	−0.02	0.36	−0.02	0.01
Leader SD	0.05	0.06	0.04	−0.01	0.00	0.02	−0.06	0.05	0.00
Leader trades	0.13	0.11	0.18	0.05	−0.03	−0.04	−0.02	0.02	0.04
Leader securities	0.11	0.09	0.22	0.11	0.02	0.01	−0.02	0.02	0.09
Leader followers	0.06	0.03	0.39	0.40	0.12	0.15	−0.01	0.01	0.08
Leader age	0.00	0.00	0.07	0.10	0.03	0.05	−0.02	0.01	0.03
	No. follow-ers	Portfolio age	No. leaders	Leader return	Leader SD	Leader trades	Leader securitiers	Leader followers	Leader age
No.followers	1								
Portofolio age	0.18	1							
No.leaders	0.19	0.28	1						
Leader return	0.00	−0.01	−0.01	1					
Leader SD	−0.01	−0.04	0.00	−0.04	1				
Leader trades	0.00	−0.13	−0.04	0.04	0.14	1			
Leader securities	−0.01	−0.22	−0.06	0.04	0.19	0.56	1		
Leader followers	−0.02	−0.02	0.05	0.03	0.15	0.29	0.57	1	
Leader age	−0.02	−0.01	0.02	−0.01	0.23	0.11	0.41	0.48	1

This table presents the Pearson correlation coefficients among the variables over the whole sample period

Table 2 presents the Pearson correlations among the variables[5]. As shown in the table, two trading frequency variables, *Trades* and *Turnover*, are highly correlated, with a Pearson correlation coefficient of 0.68. Therefore, we avoid including both *Trades* and *Turnover* in one regression in the remainder of this paper.

Methodology

In this study, we investigate the impact of leaders' comments on the trading frequency of real-account portfolios in two steps. First, we investigate whether leaders' comments change followers' trading behavior in a difference-in-differences setting. Second, we use panel regressions to investigate how leaders' comments affect followers' trading frequency.

[5] Note that some correlations are omitted in Table 2. For example, the correlation between *Leader Comment* and *Leader Count* is omitted, because *Leader Count* is restricted to weeks in which at least one comment is posted by the leaders of the portfolio owner (i.e., conditional on *Leader Comment* = 1), as a result, the Pearson correlation between the two variables cannot be calculated.

To begin, as we stated, investors self-select to follow others, which may introduce a systematic difference between portfolio owners who followed at least one other investment portfolio and those who do not follow any other portfolio. Thus, following Pelster and Breitmayer (2019), we apply the PSM procedure to address this concern. The treated group consists of real-account portfolios whose owners choose to follow at least one other portfolio in the sample period, whereas the control group includes portfolios whose owners do not follow any other portfolio during the entire sample period. We create a dummy variable, *Treat*, which equals one if the portfolio is in the treated group and zero otherwise. Next, following Pelster and Breitmayer (2019), we regress *Treat* on the average values of *Trades*, *Turnover*, *Return*, *Return SD*, *No.securities*, *No.followers*, and *Portfolio Age* for each portfolio using a logit model. We then paired each treated portfolio with the nearest-neighbor control portfolio based on the propensity scores obtained within a caliper of 0.01. We exclude portfolios for which we cannot find a match in the data.

Then, we conduct a difference-in-differences analysis using the matched sample. In our setting, the treatment event is set as the first time a portfolio owner receives messages (comments) from his/her leaders. It is conceivable that the treatment events for each portfolio owner will not happen at the same time. Thus, we adopt the time-varying difference-in-differences procedure proposed by Beck et al. (2010), based on the following regression:

$$Y_{i,t} = \beta \ Treatment_{i,t} + Controls_{i,t} + \gamma_i + \mu_t + \epsilon_{i,t}, \tag{1}$$

where $Y_{i,t}$ represents the trading activity variables (*Trades* and *Turnover*) and $Treatment_{i,t}$ is a dummy variable equal to one after the treatment event and zero otherwise. $Controls_{i,t}$ denotes the control variables. We include portfolio fixed effects (γ_i) to control for the portfolios' time-invariant characteristics. We also include time fixed effects (i.e., year-week dummies, μ_t) to control for the time trend and individual-invariant effects, such as market performance. Standard errors are double-clustered at the portfolio level and over time.

In step two, we study the characteristics of the leaders' comments that affect the trading frequency of the treated portfolios by estimating the following fixed-effects panel regression model:

$$Y_{i,t} = \kappa X_{i,t-1} + Controls_{i,t-1} + \gamma_i + \mu_t + \epsilon_{i,t}, \tag{2}$$

where $X_{i,t}$ represents the variables of leaders' comment characteristics. Standard errors are double-clustered at the portfolio level and over time.

One concern is that trading is often autocorrelated (Statman et al. 2006); that is, the dependent variables may depend on their previous lags. To address this dynamic endogeneity, we specify the following regression model and include the lags of the dependent variable in the model:

$$Y_{i,t} = \sum_{j=1}^{s} \rho_j Y_{i,t-j} + \kappa X_{i,t-1} + Controls_{i,t-1} + \gamma_i + \mu_t + \epsilon_{i,t}, \tag{3}$$

Table 3 Balance tests of the matched sample

	Treat	Control	Diff	*p*-value
Trades	0.5854	0.5980	− 0.0496	0.3107
Turnover	0.5095	0.5330	− 0.0235	0.2273
Return	− 0.0014	− 0.0014	0.0000	0.8316
Return SD	0.0123	0.0123	0.0000	0.7837
No.securities	1.6511	1.6496	0.0015	0.9518
No.followers	0.1016	0.1065	− 0.0049	0.5225
Portfolio age	3.4933	3.4789	0.0140	0.4154

This table reports the balance tests of the PSM procedure.

The first two columns report the average values of the covariates for treated and control groups, respectively. The last two columns report the mean differences and *p* values of *t*-tests for the mean differences of the covariates between treated and control groups, respectively

where s is the maximum number of lags. The fixed-effects estimator is biased in such a dynamic panel model. Thus, we use the system GMM approach to estimate Eq. 3 and cluster standard errors at the portfolio level.[6] Following Li et al. (2021), we assume the time dummies to be exogenous. All other regressors are assumed to be endogenous and their lags are used as instruments.

When implementing the system GMM approach, it is important to understand the number of lags in the dependent variable (i.e., s in Eq. 3) we should include in the regression. Following Wintoki et al. (2012), we run a pooled OLS regression of current trading frequency on the lags of the dependent variables and detect the number of significant lags of the dependent variables. The regression model is

$$Y_{i,t} = \sum_{j=1}^{k} \rho_j Y_{i,t-j} + Controls_{i,t-1} + \mu_t + \epsilon_{i,t}, \tag{4}$$

where k is the maximum number of lags used in the regression. Standard errors are double-clustered at the portfolio level and over time.

Empirical results

This section reports the main results of the empirical analysis. We first analyze whether leaders' comments cause a change in followers' trading frequency. We then analyze how portfolio owners' trading frequency is affected by the comments posted by their leaders.

Do leaders' comments change followers' trading behavior?

First, we investigate whether real-account portfolio owners' trading frequency is changed by the comments posted by their leaders in a difference-in-differences setting. As we discussed, there may be a systematic difference between portfolio owners who follow at least one other portfolio and those who do not follow any other portfolio. Therefore,

[6] In the system GMM approach, we do not cluster standard errors by time, because the autocorrelation test and the robust estimates of the coefficient standard errors in this approach assume no correlation across individuals (portfolios) in the idiosyncratic disturbances (Roodman 2009).

Table 4 Difference-in-differences analysis

	$Trades_{i,t}$	***$Turnover_{i,t}$***
$Treatment_{i,t}$	0.0953***	0.0938***
	(6.14)	(4.86)
$Return_{i,t}$	− 0.4857***	− 1.4245***
	(− 4.98)	(− 8.16)
Return $SD_{i,t}$	0.6393***	1.2393***
	(2.62)	(2.69)
$No.securities_{i,t}$	0.4996***	0.3862***
	(30.88)	(20.36)
$No.followers_{i,t}$	0.1041***	0.0861***
	(5.85)	(4.48)
Portfolio $age_{i,t}$	− 0.0788***	− 0.0868***
	(− 9.92)	(− 8.45)
Portfolio fixed effects	Yes	Yes
Time fixed effects	Yes	Yes
Observations	626,762	626,762
Adjusted R^2	0.3171	0.3102

This table presents the estimation of the time-varying difference-in-differences model specified in Eq. 1.

Standard errors are double-clustered at the portfolio level and over time. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

we first apply the PSM procedure outlined in Subsection "Methodology" to obtain the matched sample.

The matched sample includes 3,036 treated and 3,036 control portfolios. Table 3 reports the results of the balance tests of the covariates. From the last column of Table 3, we find that the p-values of the t-tests for the mean differences of the covariates between the treated and control portfolios are all larger than 0.1, indicating no pre-existing differences between the treated and control portfolios in our matched sample.

Next, we conduct a difference-in-differences analysis using the matched sample. In our setting, the treatment event is set as the first time a portfolio owner receives messages (comments) from his/her leaders.

Table 4 reports the estimated treatment effects. The treatment effects in both columns are positive and significant at the 1% level. On average, *Trades* and *Turnover* increase by 19.43% (=0.0953/0.4904) and 23.44% (=0.0938/0.4002), respectively, compared with the sample mean[7] after messages from leaders are received for the first time. Thus, our results indicate that leaders' communication causes a change in portfolio owners' trading behavior.

Leaders' comments and followers' trading frequency

To study this result in more detail, we investigate how the trading frequency of treated portfolio owners is affected by leaders' comments using panel regressions in the following analysis. We first run fixed-effects panel regressions and regress the trading

[7] We provide the summary statistics of the weekly observations of the matched sample and treated portfolios in Appendix Table 13.

Table 5 Leaders' comments and followers' trading frequency

	$Trades_{i,t}$		$Turnover_{i,t}$	
	(1)	(2)	(3)	(4)
$Leader\ comment_{i,t-1}$	0.0701***	0.0364***	0.0686***	0.0429***
	(7.56)	(4.37)	(5.54)	(3.66)
$Return_{i,t-1}$		0.5551***		− 0.0168
		(5.52)		(− 0.10)
$Return\ SD_{i,t-1}$		0.4938***		1.0086***
		(2.94)		(3.31)
$No.securities_{i,t-1}$		0.4209***		0.2746***
		(25.09)		(13.86)
$No.followers_{i,t-1}$		0.0735***		0.0837***
		(2.61)		(2.67)
Portfolio $age_{i,t-1}$		− 0.1006***		− 0.0856***
		(− 7.21)		(− 4.52)
$No.leaders_{i,t-1}$		0.0688***		0.0516***
		(3.90)		(2.75)
$Leader\ return_{i,t-1}$		0.1620**		0.2603**
		(2.17)		(2.41)
$Leader\ SD_{i,t-1}$		0.4760**		0.5164*
		(2.28)		(1.68)
$Leader\ trades_{i,t-1}$		0.0500***		0.0445***
		(5.13)		(2.97)
$Leader\ followers_{i,t-1}$		0.0151**		0.0112
		(2.01)		(1.16)
$Leader\ securities_{i,t-1}$		0.0137		0.0112
		(1.12)		(0.60)
$Leader\ age_{i,t-1}$		− 0.0269***		− 0.0228**
		(−4.07)		(−2.57)
Portfolio fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	262,457	262,457	262,457	262,457
Adjusted R^2	0.2861	0.3274	0.3154	0.3270

This table reports the results from the fixed-effects estimation of the panel regression model specified in Eq. 2.

The dependent variable is either the (log) number of trades of portfolios (Columns 1 and 2) or the turnover ratio of portfolios (Columns 3 and 4). Only treated real-account portfolios are included in the regressions. All explanatory variables are lagged by one week. Standard errors are double-clustered at the portfolio level and over time. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

frequency variables on a dummy variable, *Leader Comment*, which equals one if at least one comment is posted by leaders in a given week, and zero otherwise. We include only treated real-account portfolios in the regressions. All explanatory variables are lagged by one week, and standard errors are double-clustered at the portfolio level and over time. Table 5 presents the results.

In Columns (1) and (2) of Table 5, we employ *Trades* as the dependent variable. The coefficient of *Leader Comment* in Column (1) is positive and significant at the 1% level. In Column (2), we re-estimate the same regression, and include the control variables. We find that adding the control variables does not negate the relationship shown in Column (1). On average, *Trades* increases by 7.61% (=0.0364/0.4785) compared to the sample mean if leaders of the portfolio owner post at least one comment in the previous week, indicating that portfolio owners execute more trades under the influence of comments posted by leaders. Columns (3) and (4) report the results of the regressions for *Leader Comment* on *Turnover*. The coefficients of *Leader Comment* in both columns are significantly positive. On average, *Turnover* increases by 11.38% (=0.0429/0.3770) compared to the sample mean if at least one comment is posted in the previous week, indicating that portfolio owners turnover their portfolios faster under the influence of comments posted by their leaders.

The results in Table 5 show that portfolio owners trade more frequently if their leaders post comments in the previous week. To examine this finding further, we also investigate whether portfolio owners' trading behavior is affected by the characteristics of leaders' comments. We regress the trading frequency variables on the comment characteristic variables *Leader Count*, *Leader Positive*, and *Leader Negative*. We restrict the regressions to weeks in which the leaders posted at least one comment in the previous week. Table 6 reports the results.

Columns (1)-(4) of Table 6 present the results for *Trades*. In Column (1), we document a positive and significant coefficient of *Leader Count*, suggesting that portfolio owners' trading frequency increases with the number of comments posted by their leaders. On average, a one-standard deviation increase in the average number of leaders' comments increases *Trades* by 6.50% (=0.0196*1.5872/0.4785) compared to the sample mean. In Columns (2) and (3), we investigate whether the tone of leaders' comments also affects the portfolio owners' trading frequency. Interestingly, we find that the average fractions of positive and negative words are positively related to *Trades*, indicating that the fraction of sentiment words in the leaders' comments can cause portfolio owners to trade more. However, the coefficient of *Leader Positive* is not significant, while the coefficient of *Leader Negative* is only significant at the 10% level. In addition, a one-standard deviation increase in the fraction of positive (negative) words only leads to an 0.19% (0.88%) increase in *Trades* relative to the sample mean, which is not economically large. Moreover, when we include all three comment characteristic variables in the regression shown in Column (4), both the coefficients of *Leader Positive* and *Leader Negative* are insignificant. This result indicates that portfolio owners are more sensitive to the quantity than the tone of leaders' comments.

Panel B of Table 6 shows similar results for *Turnover*. The average number of leaders' comments (*Leader Count*) positively impacts the turnover ratio of followers' portfolios, whereas the tone of leaders' comments does not. On average, a one-standard deviation increase in the number of leaders' comments will increase *Turnover* by 7.75% (=0.0184*1.5872/0.3770) compared to the sample mean.

In summary, our results show that leaders' comments can increase the portfolio owners' trading frequency. In addition, when leaders post comments in the previous week, the quantity of leaders' comments further increases portfolio owners' trading frequency. However, portfolio owners seem to be insensitive to the tone of leaders' comments. This is contrary to the findings of Yang et al. (2020), who conclude that the sentiments expressed in another stock forum called Eastmoney Guba lead to the abnormal trading of individual stocks in the Chinese stock market. One potential explanation is that the sample of Yang et al. (2020) focuses only on firms listed on the Growth Enterprise

Table 6 Characteristics of leaders' comments and followers' trading frequency

	Trades$_{i,t}$				***Turnover$_{i,t}$***			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Leader count$_{i,t-1}$	0.0196***			0.0194***	0.0186***			0.0184***
	(4.31)			(4.26)	(3.25)			(3.22)
Leader positive$_{i,t-1}$		0.0319		−0.0064		0.0706		0.0373
		(0.420)		(−0.09)		(0.62)		(0.34)
Leader negative$_{i,t-1}$			0.1153*	0.0838			0.1030	0.0668
			(1.81)	(1.35)			(1.14)	(0.76)
Return$_{i,t-1}$	0.6398***	0.6413***	0.6416***	0.6401***	0.0699	0.0716	0.0717	0.0704
	(5.86)	(5.87)	(5.88)	(5.87)	(0.38)	(0.39)	(0.39)	(0.38)
Return SD$_{i,t-1}$	0.4159**	0.4187**	0.4187**	0.4159**	0.8853**	0.8879**	0.8879**	0.8852**
	(2.07)	(2.08)	(2.08)	(2.07)	(2.46)	(2.47)	(2.47)	(2.46)
No.securities$_{i,t-1}$	0.4640***	0.4647***	0.4646***	0.4640***	0.2987***	0.2993***	0.2993***	0.2986***
	(23.84)	(23.82)	(23.82)	(23.84)	(13.14)	(3.16)	(13.16)	(13.13)
No.followers$_{i,t-1}$	0.0693**	0.0701**	0.0701**	0.0693**	0.0844***	0.0852***	0.0851***	0.0844***
	(2.09)	(2.11)	(2.11)	(2.09)	(2.77)	(2.79)	(2.79)	(2.77)
Portfolio age$_{i,t-1}$	−0.1436***	−0.1438***	− 0.1439***	−0.1436***	− 0.1413***	− 0.1415***	− 0.1416***	− 0.1413***
	(−6.54)	(−6.53)	(−6.54)	(−6.54)	(−4.83)	(−4.83)	(−4.83)	(−4.83)
No.leaders$_{i,t-1}$	0.0921***	0.0972***	0.0969***	0.0918***	0.0709***	0.0757***	0.0755***	0.0707***
	(4.50)	(4.78)	(4.77)	(4.49)	(3.45)	(3.70)	(3.69)	(3.44)
Leader return$_{i,t-1}$	0.2184*	0.2339**	0.2346**	0.2188*	0.3543**	0.3681**	0.3696**	0.3537**
	(1.93)	(2.06)	(2.07)	(1.94)	(2.17)	(2.25)	(2.25)	(2.17)
Leader SD$_{i,t-1}$	0.5756	0.5826	0.5798	0.5734	0.4260	0.4322	0.4302	0.4238
	(1.55)	(1.58)	(1.57)	(1.54)	(0.85)	(0.87)	(0.86)	(0.85)
Leader trades$_{i,t-1}$	0.0506***	0.0546***	0.0549***	0.0509***	0.0365**	0.0403**	0.0405**	0.0368**
	(4.46)	(4.83)	(4.86)	(4.48)	(2.20)	(2.44)	(2.46)	(2.21)
Leader followers$_{i,t-1}$	0.0130	0.0187*	0.0184*	0.0128	0.0118	0.0171	0.0169	0.0116
	(1.23)	(1.79)	(1.76)	(1.22)	(0.89)	(1.30)	(1.28)	(0.87)
Leader securities$_{i,t-1}$	0.0172	0.0159	0.0158	0.0171	0.0364	0.0352	0.0351	0.0364
	(1.06)	(0.98)	(0.97)	(1.06)	(1.52)	(1.46)	(1.46)	(1.52)
Leader age$_{i,t-1}$	− 0.0380***	− 0.0420***	−0.0418***	−0.0379***	−0.0416***	−0.0453***	−0.0452***	−0.0415***
	(0.0110)	(− 3.83)	(− 3.82)	(− 3.46)	(− 2.95)	(− 3.19)	(− 3.18)	(− 2.94)
Portfolio fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observa-tions	150,447	150,447	150,447	150,447	150,447	150,447	150,447	150,447
Adjusted R^2	.3380	0.3377	0.3377	0.3380	0.3426	0.3425	0.3425	0.3426

This table reports the results from the fixed-effects estimation of the panel regression model specified in Eq. 2.

The dependent variable is either the (log) number of trades of portfolios (Columns 1 to 4) or the turnover ratio of portfolios (Columns 5 to 8). Only treated real-account portfolios are included in the regressions. All explanatory variables are lagged by one week. Standard errors are double-clustered at the portfolio level and over time. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

Market (GEM) on China's Shenzhen Stock Exchange, while the portfolio owners in our sample select stocks to trade based on all stocks in the Chinese stock market. As Yang et al. (2020) point out, GEM stocks are more volatile and small-cap than "mainboard" stocks listed on the Shanghai and Shenzhen Stock Exchanges. Therefore, the trading behavior of investors who invest in GEM stocks may be driven more by sentiment.

Dynamic panel analysis

Thus far, our analysis of the effect of leaders' comments on followers' trading frequency is based on static panel models. However, as Statman et al. (2006) point out, trading can be autocorrelated. In other words, the current trading frequency may depend on the past value. In this case, our results may suffer from dynamic endogeneity. Therefore, we consider the dynamic panel regression model specified in Eq. 3 to address the potential dynamic endogeneity.

First, to specify a dynamic panel regression model, we need to determine the number of lags of the dependent variable to include in the model. Following Wintoki et al. (2012), we choose the number of lags by estimating Eq. 4 and detecting the number of statistically significant lags. The results are reported in Table 7.[8]

Columns (1) and (3) of Table 7 show that the coefficients of the first five lags of *Trades* and *Turnover* are significant, at least at the 10% level, whereas the coefficients of longer lags are not. This result indicates that five lags of the dependent variable are sufficient to capture the persistence of the trading frequency. In columns (2) and (4), we drop the first five lags and include only longer lags in the regressions. In these two columns, the coefficients of the longer lags become significant. Taken together, while longer lags include relevant information, this information is subsumed by the first five lags.

Next, we used the system GMM approach to estimate the dynamic panel regression model specified in Eq. 3 with 5 lags of the dependent variable in the model. In GMM regressions, the time dummies are assumed to be exogenous. All other regressors are assumed to be endogenous and their lags are used as instruments. The results are reported in Table 8. From all four columns in Table 8, we find that the coefficients of *Leader Comment* and *Leader Count* are positive and significant at the 5% level, while the coefficients of *Leader Positive* and *Leader Negative* are not significant. On average, *Trades* (*Turnover*) increases by 6.54% (15.89%) compared with the sample mean if at least one comment is posted in the previous week. A one-standard deviation increase in *Leader Count* increases *Trades* (*Turnover*) by 7.43% (14.31%) compared to the sample mean. In addition to the coefficients on leader comment variables, the results of several important statistical tests of GMM estimation are also reported in Table 8. First, the p value of the AR(2) test in each column is greater than 0.1, indicating no serial correlation in the residuals. Second, the Hansen test of over-identification with a p value larger than 0.1 in each column suggests that the hypothesis that the instruments are exogenous cannot be rejected. Third, the difference-in-Hansen test of exogeneity with a p value larger than 0.1 in each column implies that the hypothesis that the additional subset of instruments used in the system GMM estimator is exogenous cannot be rejected.

In summary, the results of the dynamic panel models show that leaders' comments still positively impact portfolio owners' trading frequency, even after controlling for lagged dependent variables.

[8] Table 7 report the estimation results of Eq. 4 with 8 lags of the dependent variables are included in the model. In the unreported results, when we include more lags in the regressions, the results do not change.

Table 7 Lag selection procedure

	Trades$_{i,t}$		***Turnover$_{i,t}$***	
	(1)	**(2)**	**(3)**	**(4)**
$Y_{i,t-1}$	0.4220***		0.4688***	
	(49.86)		(38.29)	
$Y_{i,t-2}$	0.1360***		0.1141***	
	(25.13)		(12.63)	
$Y_{i,t-3}$	0.0571***		0.0532***	
	(12.16)		(6.36)	
$Y_{i,t-4}$	0.0336***		0.0356***	
	(7.60)		(4.39)	
$Y_{i,t-5}$	0.0078*		0.0133*	
	(1.94)		(1.87)	
$Y_{i,t-6}$	0.0047	0.1298***	− 0.0025	0.1407***
	(1.12)	(19.99)	(− 0.38)	(13.16)
$Y_{i,t-7}$	0.0023	0.0501***	0.0055	0.0534***
	(0.66)	(9.96)	(0.72)	(6.45)
$Y_{i,t-8}$	0.0031	0.0286***	− 0.0010	0.0274***
	(0.84)	(5.12)	(− 0.16)	(3.11)
Return$_{i,t-1}$	0.9580***	0.6778***	0.9424***	0.1983
	(13.92)	(7.15)	(8.67)	(1.33)
Return SD$_{i,t-1}$	0.0490	0.3879***	0.1182	0.8011***
	(0.79)	(2.74)	(1.29)	(3.11)
No.securities$_{i,t-1}$	0.1315***	0.3720***	0.0547***	0.2353***
	(21.58)	(23.62)	(8.75)	(13.45)
No.followers$_{i,t-1}$	0.0091	0.0460*	0.0085	0.0493**
	(0.85)	(1.95)	(0.90)	(2.27)
Portfolio age$_{i,t-1}$	− 0.0245***	− 0.0542***	− 0.0148	− 0.0239
	(− 2.77)	(− 2.89)	(− 1.35)	(− 0.93)
No.leaders$_{i,t-1}$	0.0062	0.0417***	0.0095	0.0388**
	(0.88)	(2.69)	(1.35)	(2.38)
Leader return$_{i,t-1}$	0.0758	0.1214*	0.0422	0.2205**
	(1.47)	(1.68)	(0.52)	(2.11)
Leader SD$_{i,t-1}$	0.0079	0.2596	− 0.1726	0.0997
	(0.06)	(1.49)	(− 1.20)	(0.40)
Leader trades$_{i,t-1}$	0.0089*	0.0459***	0.0030	0.0395***
	(1.81)	(5.12)	(0.44)	(2.87)
Leader Followers$_{i,t-1}$	0.0055*	0.0120*	0.0054	0.0076
	(1.95)	(1.89)	(1.57)	(0.92)
Leader securities$_{i,t-1}$	0.0057	0.0110	0.0071	0.0145
	(1.18)	(1.05)	(0.99)	(0.91)
Leader age$_{i,t-1}$	− 0.0073***	− 0.0206***	− 0.0070**	− 0.0184**
	(−2.96)	(−3.65)	(−2.21)	(−2.42)
Time fixed effects	Yes	Yes	Yes	Yes
Observations	252,586	252,586	252,586	252,586
Adjusted R^2	0.5445	0.3596	0.5715	0.3646

This table presents the estimation results of the lag selection procedure specified in Eq. 4.

The dependent variable is either the (log) number of trades of portfolios (Columns 1 and 2) or the turnover ratio of portfolios (Columns 3 and 4). Only real-account portfolios of treated portfolios are included in the regressions. All explanatory variables are lagged by one week. Standard errors are double-clustered at the portfolio level and over time. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

Table 8 GMM estimation

	Trades$_{i,t}$		***Turnover$_{i,t}$***	
	(1)	**(2)**	**(3)**	**(4)**
Leader comment$_{i,t-1}$	0.0313**		0.0599**	
	(2.27)		(2.23)	
Leader count$_{i,t-1}$		0.0224**		0.0340**
		(2.26)		(2.33)
Leader positive$_{i,t-1}$		− 1.1626		− 1.3631
		(− 1.09)		(− 0.95)
Leader negative$_{i,t-1}$		− 0.4159		1.1988
		(− 0.51)		(0.97)
Return$_{i,t-1}$	3.4177***	1.0144***	5.5847***	0.9183***
	(3.17)	(9.26)	(2.88)	(5.29)
Return SD$_{i,t-1}$	2.2353***	1.8947**	3.5000*	2.1011
	(2.72)	(1.98)	(1.72)	(1.20)
No.securities$_{i,t-1}$	0.0160**	0.0067	0.0085	-0.0321*
	(2.33)	(0.66)	(0.91)	(− 1.83)
No.followers$_{i,t-1}$	0.0334	0.0583**	0.0415	0.0914**
	(1.44)	(2.10)	(1.22)	(2.57)
Portfolio ae$_{i,t-1}$	− 0.0204***	−0.0172***	− 0.0228***	− 0.0169*
	(− 4.09)	(− 2.63)	(− 3.03)	(− 1.76)
No.leaders$_{i,t-1}$	0.0096	0.0087	0.0016	− 0.0276
	(0.81)	(0.50)	(0.09)	(− 1.49)
Leader return$_{i,t-1}$	0.3781	0.0660	5.0060**	0.9162
	(0.56)	(0.05)	(2.04)	(0.49)
Leader SD$_{i,t-1}$	0.4273	1.9271	1.6339	3.7328*
	(0.63)	(1.57)	(1.01)	(1.82)
Leader trades$_{i,t-1}$	0.0003	− 0.0019	− 0.0194	0.0117
	(0.02)	(− 0.08)	(− 0.95)	(0.33)
Leader followers$_{i,t-1}$	0.0069	0.0029	− 0.0071	-0.0191*
	(1.55)	(0.37)	(− 1.01)	(− 1.77)
Leader securities$_{i,t-1}$	0.0010	0.0048	− 0.0051	− 0.0004
	(0.16)	(0.40)	(− 0.49)	(− 0.02)
Leader age$_{i,t-1}$	− 0.0088	− 0.0179	− 0.0061	− 0.0034
	(− 1.54)	− 1.59)	(− 0.69)	(− 0.22)
$Y_{i,t-1}$	0.5808***	0.7511***	0.7862***	0.6949***
	(4.66)	(11.96)	(3.86)	(7.07)
$Y_{i,t-2}$	0.3931**	0.0811	0.0144	0.1073
	(2.37)	(0.57)	(0.06)	(1.42)
$Y_{i,t-3}$	− 0.1188	0.0137	0.0378	0.0557
	(− 1.39)	(0.15)	(0.46)	(0.98)
$Y_{i,t-4}$	0.0064	-0.0039	0.0036	− 0.0427
	(0.62)	(− 0.32)	(0.19)	(− 0.43)
$Y_{i,t-5}$	-0.0091*	− 0.0089	0.0072	0.0077
	(− 1.67)	(− 1.25)	(0.74)	(0.30)
Portfolio fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	243,880	140,419	243,880	140,419
AR(1) test (p value)	0.002	0.000	0.005	0.000
AR(2) test (p value)	0.170	0.620	0.581	0.447
Hansen test of over-identification (p value)	0.250	0.536	0.286	0.367
Diff-in-Hansen test of exogeneity (p value)	0.553	0.225	0.224	0.575

Table 8 (continued)

This table presents the GMM estimation results of the panel regression model specified in Eq. 3.

The dependent variable is either the (log) number of trades of portfolios (Columns 1 and 2) or the turnover ratio of portfolios (Columns 3 and 4). Only treated real-account portfolios are included in the regressions. All explanatory variables are lagged by one week. Standard errors are double-clustered at the portfolio level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

Robustness analysis

Subsequently, we conduct the following analysis to check the robustness of our results and rule out alternative explanations.

First, as we discussed, although the Chinese LM dictionary is widely used, its accuracy cannot be guaranteed when applied to social media. Therefore, we consider an alternative Chinese dictionary called the National Taiwan University Semantic Dictionary (NTUSD) to define the tone of the words. The NTUSD contains 2,810 positive words and 8,276 negative words and has been used by researchers to measure sentiment in financial articles (Chen et al. 2022; Wang et al. 2019). We reconstruct *Leader Positive* and *Leader Negative* based on the NTUSD and then re-estimate the baseline models, including static and dynamic panel models. We provide the results Table 9. We find that the coefficients of *Leader Count* are still positive and significant, while the coefficients of *Leader Positive* and *Leader Negative* are not. Therefore, using an alternative dictionary, we still find that the quantity of leaders' comments increases portfolio owners' trading frequency, whereas the tone of comments does not.

Second, previous research shows that leaders' comments can attract the fund flows of copiers (Ammann and Schaub 2020); thus, the increased level of trading frequency may result from the imitation of leaders' trading activities. As we stated, Snowball users cannot copy the trades of their leaders directly, but they can imitate leaders' trading activities manually. To address this concern, we removed imitating trades, which involve buying the same stocks as leaders within five trading days (one week), and this trimming process removed 10% of the trades in our dataset. We then re-estimate the baseline models using data without imitating trades and report the results in Table 10. The coefficients of *Leader Comment* and *Leader Count* are still positive and significant, indicating that our results are not driven by the imitation of leaders' trading activities.

Third, a fraction of our comments are firm-specific that may be affected by firm-related news and announcements. We provide examples of general and firm-specific comments automatically translated into English via Google Translate in Fig. 2. To address the possibility that our results may be driven by firm-specific news, we removed firm-specific comments, and this trimming process removed 29% of the comments in our dataset. We then re-estimated the baseline models using only general comments. The results are reported in Table 11. The coefficients of *Leader Comment* and *Leader Count* are still positive and significant, indicating that our results are not driven by firm-related news and announcements.

Taken together, we show a robust link between leaders' comments and portfolio owners' trading frequency in Tables 9, 10, 11, confirming the robustness of our findings.

Table 9 Tone of leaders' comments and followers' trading frequency - Alternative dictionary

	$Trades_{i,t}$		$Turnover_{i,t}$	
	(1)	(2)	(3)	(4)
	FE	SYS-GMM	FE	SYS-GMM
$Leader\ count_{i,t-1}$	0.0197***	0.0221**	0.0186***	0.0351**
	(4.35)	(2.13)	(3.24)	(2.48)
Leader $positive_{i,t-1}$	0.0386	− 2.0590	0.0924	− 0.7023
	(0.51)	(− 1.58)	(0.82)	(− 0.54)
$Leader\ negative_{i,t-1}$	− 0.1029	− 0.8477	− 0.0395	0.3755
	(− 1.26)	(−0.76)	(− 0.34)	(0.27)
$Return_{i,t-1}$	0.6394***	0.6918***	0.0698	0.7521***
	(5.86)	(4.24)	(0.47)	(3.86)
$Return\ SD_{i,t-1}$	0.4159**	2.0980**	0.8850***	1.9590
	(2.07)	(2.10)	(2.95)	(1.07)
$No.securities_{i,t-1}$	0.4640***	0.0092	0.2987***	− 0.0429**
	(23.84)	(0.86)	(15.03)	(− 2.55)
$No.followers_{i,t-1}$	0.0694**	0.7378**	0.0845***	0.0730*
	(2.09)	(2.44)	(2.80)	(1.91)
$Portfolio\ age_{i,t-1}$	− 0.1435***	−0.0213***	−0.1412***	−0.0200*
	(− 6.54)	(−2.92)	(−4.77)	(− 1.73)
$No.leaders_{i,t-1}$	0.0921***	0.0225	0.0708***	− 0.0096
	(4.50)	(1.20)	(3.50)	(− 0.43)
$Leader\ return_{i,t-1}$	0.2166*	0.0225**	0.3518**	1.0923
	(1.92)	(2.12)	(2.56)	(0.57)
$Leader\ SD_{i,t-1}$	0.5769	1.9864	0.4285	5.3585**
	(1.55)	(1.15)	(1.05)	(2.25)
$Leader\ trades_{i,t-1}$	0.0504***	− 0.0290	0.0365**	0.0158
	(4.43)	(− 1.02)	(2.26)	(0.38)
$Leader\ followers_{i,t-1}$	0.0132	0.0049	0.0117	− 0.0115
	(1.25)	(0.54)	(0.87)	(− 0.89)
$Leader\ securities_{i,t-1}$	0.0172	− 0.0010	0.0364	− 0.0130
	(1.06)	(− 0.08)	(1.53)	(− 0.62)
$Leader\ age_{i,t-1}$	− 0.0380***	− 0.0182	− 0.0416***	− 0.0152
	(−3.47)	(− 1.49)	(− 2.94)	(− 0.80)
$Y_{i,t-1}$		0.8229***		0.6755***
		(12.40)		(5.51)
$Y_{i,t-2}$		− 0.0780		− 0.0514
		(− 0.50)		(− 0.33)
$Y_{i,t-3}$		0.0917		0.0533
		(0.91)		(0.42)
$Y_{i,t-4}$		− 0.0078		0.1039
		(− 0.63)		(0.72)
$Y_{i,t-5}$		− 0.0058		− 0.0091
		(− 0.82)		(− 0.34)
Portfolio fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	150,447	140,419	150,447	140,419
Adjusted R^2	0.3521	–	0.3566	–
AR(1) test (p value)	–	0.000	–	0.000
AR(2) test (p value)	–	0.158	–	0.427

Table 9 (continued)

	Trades$_{i,t}$		***Turnover$_{i,t}$***	
	(1)	**(2)**	**(3)**	**(4)**
	FE	**SYS-GMM**	**FE**	**SYS-GMM**
Hansen test of over-identification (p value)	–	0.449	–	0.446
Diff-in-Hansen test of exogeneity (p value)	–	0.539	–	0.598

This table reports the results from the fixed-effects (FE) estimation of the panel regression model specified in Eq. 2 (odd columns) and the GMM (SYS-GMM) estimation of the panel regression model specified in Eq. 3 (even columns). The dependent variable is either the (log) number of trades of portfolios (Columns 1 and 2) or the turnover ratio of portfolios (Columns 3 and 4).

Leader Positive and *Leader Negative* are constructed based on NTUSD. Only treated real-account portfolios are included in the regressions. All explanatory variables are lagged by one week. In odd columns, standard errors estimated by the fixed-effects approach are double-clustered at the portfolio level and over time. In even columns, standard errors estimated by the system GMM approach are clustered at the portfolio level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

Performance implications

Finally, we investigate the performance implications of communication and excessive trading. As we demonstrated that leaders' comments can affect followers' trading behavior, it is possible that leaders' comments also impact followers' future performance. Additionally, there is abundant evidence that excessive trading lowers future returns. To investigate this issue, we perform fixed-effects panel regressions employing portfolio returns (*Return*) as the dependent variable, and report the results in Table 12.

In columns (1) and (2), the coefficients of *Leader Comment* are negative and significant, whereas in columns (3) and (4), the coefficients of the three comment characteristic variables are all insignificant. This result indicates that leaders' comments negatively affect portfolio owners' future performance, although such impacts cannot be further enhanced by the quantity and tone of leaders' comments. In general, this finding is consistent with our conjecture that leaders' comments may trigger some trading behavior by portfolio owners, which is detrimental to their future wealth.

In addition to *Leader Comment*, we find that *Trades* and *Turnover* also negatively predict the future performance of portfolio owners, which is consistent with existing studies showing that high trading frequency is associated with poor performance (Barber and Odean 2000a, b, 2002).[9]

Overall, we show that both communication and excessive trading hurt portfolio owners' future performance. However, we cannot conclude whether the increased trading frequency caused by leaders' comments further lowers portfolio owners' future returns, because various effects, such as overconfidence, can increase trading frequency. Nevertheless, the results in Table 12 show that users on social trading platforms should be aware of the potential negative impacts of communication on their performance.

[9] We do not include *Trades* and *Turnover* in one regression due to the high correlation between the two variables reported in Table 2.

Table 10 Leaders' comments and trading frequency of followers - removal of imitating trades

	Trades$_{i,t}$				***Turnover$_{i,t}$***			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	FE	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**
Leader comment$_{i,t-1}$	0.0376***	0.0312**			0.0426***	0.0615**		
	(4.54)	(2.28)			(3.67)	(2.25)		
Leader count$_{i,t-1}$			0.0194***	0.0211**			0.0186***	0.0318**
			(4.30)	(2.14)			(3.27)	(1.99)
Leader positive$_{i,t-1}$			− 0.0076	− 1.0604			0.0314	− 1.7731
			(− 0.10)	(− 1.07)			(0.28)	(− 1.19)
Leader negative$_{i,t-1}$			0.0853	−0.0543			0.0572	1.1341
			(1.37)	(− 0.07)			(0.66)	(0.82)
Return$_{i,t-1}$	0.5488***	3.0923***	0.6324***	0.9563***	− 0.0142	5.6802***	0.0765	0.7684***
	(5.48)	(2.82)	(5.83)	(9.13)	(− 0.09)	(2.92)	(0.18)	(3.97)
Return SD$_{i,t-1}$	0.4923***	2.1796**	0.4139**	2.1503**	1.0030***	3.5414*	0.8775**	2.0270
	(2.95)	(2.54)	(2.08)	(2.37)	(3.31)	(1.73)	(2.47)	(0.96)
No.securities$_{i,t-1}$	0.4182***	0.0153**	0.4605***	0.0067	0.2717***	0.0077	0.2944***	− 0.0383**
	(25.10)	(2.16)	(23.90)	(0.66)	(13.82)	(0.82)	(13.09)	(− 2.17)
No.followers$_{i,t-1}$	0.0687**	0.0327	0.0618**	0.0565**	0.0782**	0.0356	0.0749**	0.0797**
	(2.52)	(1.42)	(1.98)	(2.03)	(2.49)	(1.10)	(2.48)	(2.15)
Portfolio age$_{i,t-1}$	− 0.0973***	− 0.0199***	− 0.1394***	0.0170***	− 0.0214***	− 0.0834***	− 0.1379***	− 0.0208*
	(− 7.00)	(− 3.98)	(− 6.38)	(− 2.62)	(− 4.42)	(− 2.85)	(− 4.74)	(− 1.85)
No.leaders$_{i,t-1}$	0.0570***	0.0101	0.0759***	0.0001	0.0438**	− 0.0002	0.0595***	− 0.0183
	(3.32)	(0.86)	(3.83)	(0.01)	(2.35)	(− 0.01)	(2.91)	(− 0.85)
Leader return$_{i,t-1}$	0.1656**	0.4494	0.2279**	0.6164	0.2740**	5.1342**	0.3864**	0.7146
	(2.23)	(0.67)	(2.05)	(0.54)	(2.54)	(2.11)	(2.39)	(0.38)
Leader SD$_{i,t-1}$	0.4816**	0.3889	0.5476	1.8236	0.5091*	1.6131	0.3744	5.1628**
	(2.31)	(0.57)	(1.49)	(1.49)	(1.67)	(0.99)	(0.76)	(2.02)

Table 10 (continued)

	Trades$_{i,t}$				*Turnover$_{i,t}$*			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	FE	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**
Leader trades$_{i,t-1}$	0.0394***	− 0.0003	0.0399***	0.0011	0.0341**	− 0.0223	0.0243	0.0009
	(4.16)	(− 0.03)	(3.66)	(0.05)	(2.31)	(− 1.08)	(1.48)	(0.02)
Leader followers$_{i,t-1}$	0.0158**	0.0070	0.0123	0.0004	0.0118	− 0.0075	0.0114	− 0.0163
	(2.12)	(1.57)	(1.19)	(0.05)	(1.23)	(− 1.05)	(0.86)	(− 1.30)
Leader securities$_{i,t-1}$	0.0141	0.0009	0.0180	0.0011	0.0121	− 0.0049	0.0380	− 0.0041
	(1.16)	(0.14)	(1.11)	(0.10)	(0.64)	(− 0.47)	(1.59)	(− 0.19)
Leader age$_{i,t-1}$	−0.0271***	− 0.0084	− 0.0366***	−0.0146	−0.0228**	−0.0054	−0.0404***	− 0.0111
	(− 4.13)	(− 1.48)	(−3.40)	(− 1.32)	(− 2.59)	(− 0.60)	(− 2.89)	(− 0.60)
$Y_{i,t-1}$		0.5925***		0.7528***		0.7454***		0.6138***
		(4.65)		(12.08)		(3.87)		(5.31)
$Y_{i,t-2}$		0.3613*		0.0565		0.0553		−0.0372
		(1.85)		(0.40)		(0.26)		(−0.23)
$Y_{i,t-3}$		−0.1001		0.0308		0.0350		0.0679
		(−0.89)		(0.33)		(0.43)		(0.52)
$Y_{i,t-4}$		0.0089		− 0.0053		0.0039		0.1078
		(0.68)		(0.43)		(0.20)		(0.72)
$Y_{i,t-5}$		− 0.0116*		− 0.0088		0.0067		− 0.016
		(− 1.29)		(− 1.24)		(0.68)		(−0.06)
Portfolio fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	262,457	243,880	150,447	140,419	262,457	243,880	150,447	140,419
Adjusted R^2	0.3259	–	0.3365	–	0.3267	–	0.3421	–
AR(1) test (p value)	–	0.003	–	0.000	–	0.004	–	0.000

Table 10 (continued)

	Trades$_{i,t}$				***Turnover$_{i,t}$***			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	FE	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**
AR(2) test (p value)	–	0.332	–	0.502	–	0.687	–	0.454
Hansen test of over-identification (p value)	–	0.183	–	0.361	–	0.300	–	0.389
Diff-in-Hansen test of exogeneity (p value)	–	0.561	–	0.199	–	0.242	–	0.679

This table reports the results from the fixed-effects (FE) estimation of the panel regression model specified in Eq. 2 (odd columns) and the GMM (SYS-GMM) estimation of the panel regression model specified in Eq. 3 (even columns).

The dependent variable is either the (log) number of trades of portfolios (Columns 1 to 4) or the turnover ratio of portfolios (Columns 5 to 8), excluding imitating trades. Only real-account portfolios of treated portfolios are included in the regressions. All explanatory variables are lagged by one week. In odd columns, standard errors estimated by the fixed-effects approach are double-clustered at the portfolio level and over time. In even columns, standard errors estimated by the system GMM approach are clustered at the portfolio level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

(a) 04-30 23:21 From iPhone
Generally speaking, it is a good thing to fall before the holiday, which means it will rise after the holiday.
Forward Comments Like Collection Complaints

(b) 05-14 14:45 From Android
Ping An Bank needs to increase the amount next week, otherwise the stock price will go down. $Ping An Bank (SZ000001)$
Forward Comments Like Collection Complaints

Fig. 2 This figure provides examples of general and firm-specific comments. Panel (a) shows an example of a general comment. Panel (b) shows an example of a firm-specific comment. $Ping An Bank (SZ000001)$ stands for the stock for Ping An Bank encoded as SZ000001 and listed on the Shenzhen Stock Exchange

Discussion, future directions, and practical implications

Discussion

In traditional behavioral finance, investors' active trading behavior is attributed to inherent psychological biases such as overconfidence. However, by investigating users' investing behavior and communications on a social trading platform, we find that real-account portfolio owners trade more aggressively under the influence of comments posted by their leaders after controlling for past returns and market effects. The effect of leaders' comments on trading frequency indicates that communication also distorts trading behavior, apart from psychological bias.

In general, our findings support the prediction of social transmission bias that investment-related communications propagate active investing (Han et al. 2022). More broadly, this study follows the call of Hirshleifer (2015) to move from behavioral finance to social finance, which includes the study of how social linkages affect financial decisions and information flows and sheds new light on the role of social interactions in the stock market.

Future directions

Our study provides abundant room for future research. First, as predicted by Han et al. (2022), social interaction can increase the likelihood of transforming investors from passive to active. In addition to the higher trading frequency, other characteristics of active investors, such as higher volatility and positive skewness, also known as lottery-like features (Kumar 2009; Bali et al. 2011, 2019; Yao et al. 2019), deserve further investigation.

Second, as Fig. 2 shows, a fraction of comments in our dataset are firm-specific. It may be fruitful to identify whether message senders recommend the firms mentioned in the comments using new techniques, such as machine learning and text mining. It would be interesting to investigate whether message receivers follow such advice to buy or sell the relevant stocks.

Third, as Lane et al. (2021) point out, investors may be more easily affected by neighbors that are more tightly connected to them. Therefore, identifying the tie strength between two connected investors based on profiling and grouping analyses can also be fruitful in future studies.

Table 11 Leaders' comments and trading frequency of followers - removal of firm-specific comments

	Trades$_{i,t}$				***Turnover$_{i,t}$***			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	FE	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**
Leader comment$_{i,t-1}$	0.0266***	0.0326**			0.0246**	0.0575**		
	(3.10)	(2.04)			(2.09)	(2.10)		
Leader count$_{i,t-1}$			0.0159***	0.0260**			0.0142**	0.0317**
			(3.32)	(2.33)			(2.33)	(2.06)
Leader positive$_{i,t-1}$			− 0.0175	− 0.6226			0.0645	− 1.1496
			(− 0.64)	(− 0.94)			(0.60)	(− 1.00)
Leader negative$_{i,t-1}$			0.0654	− 1.0765			0.0044	−0.7807
			(0.99)	(− 1.13)			(0.05)	(− 0.63)
Return$_{i,t-1}$	0.5548***	7.1534***	0.6501***	1.0230***	− 0.0171	5.8654***	0.0268	0.8019***
	(5.51)	(4.78)	(5.78)	(9.58)	(− 0.10)	(3.04)	(0.14)	(4.79)
Return SD$_{i,t-1}$	0.4944***	2.2742**	0.3720**	1.7394	1.0096***	3.4614*	0.7770**	− 0.1833
	(2.94)	(2.34)	(2.01)	(1.58)	(3.31)	(1.72)	(2.40)	(− 0.10)
No.securities$_{i,t-1}$	0.4212***	0.0137*	0.4613***	−0.0011	0.2750***	0.0087	0.2853***	−0.0223
	(25.09)	(1.65)	(23.31)	(−0.09)	(13.87)	(0.93)	(12.99)	(− 1.12)
No.followers$_{i,t-1}$	0.0735***	0.0237	0.0680*	0.0811***	0.0837***	0.0428	0.0703**	0.0689*
	(2.61)	(0.98)	(1.96)	(2.60)	(2.67)	(1.26)	(2.45)	(1.81)
Portfolio age$_{i,t-1}$	−0.0159***	−0.0217***	−0.1434***	−0.0187***	−0.0221***	−0.0116***	−0.1445***	−0.0201*
	(− 7.17)	(− 3.57)	(− 6.26)	(− 2.61)	(− 2.65)	(− 2.97)	(− 4.84)	(−1.92)
No.leaders$_{i,t-1}$	0.0712***	-0.0057	0.1050***	0.0168	0.0562***	−0.0002	0.0839***	0.0364*
	(4.03)	(− 0.46)	(5.00)	(1.08)	(2.98)	(−0.01)	(4.02)	(1.74)
Leader return$_{i,t-1}$	0.1654**	− 1.3151	0.2513**	0.1281	0.2666**	4.3481*	0.3818**	1.2021
	(2.22)	(− 1.27)	(2.08)	(0.11)	(2.47)	(1.77)	(2.31)	(0.81)
Leader SD$_{i,t-1}$	0.4813**	2.2616*	0.5115	2.0306	0.5236*	1.6182	0.4045	5.3131**
	(2.30)	(1.72)	(1.31)	(1.39)	(1.69)	(1.00)	(0.75)	(2.44)

Table 11 (continued)

	$Trades_{i,t}$				$Turnover_{i,t}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	FE	SYS-GMM	FE	SYS-GMM	FE	SYS-GMM	FE	SYS-GMM
Leader trades$_{i,t-1}$	0.0508***	− 0.0102	0.0507***	−0.0169	0.0459***	−0.0172	0.0392**	−0.0570
	(5.21)	(− 0.55)	(4.19)	(−0.65)	(3.06)	(−0.85)	(2.25)	(−1.50)
Leader followers$_{i,t-1}$	0.0155**	0.0010	0.0172	0.0067	0.0121	− 0.0074	0.0148	− 0.0137
	(2.06)	(0.19)	(1.54)	(0.78)	(1.25)	(− 1.03)	(1.09)	(− 1.17)
Leader Securities$_{i,t-1}$	0.0139	0.0091	0.0224	0.0105	0.0114	− 0.0042	0.0504**	0.0340*
	(1.14)	(1.07)	(1.35)	(0.79)	(0.61)	(− 0.42)	(2.13)	(1.80)
Leader Age$_{i,t-1}$	− 0.0274***	− 0.0121	−0.0384***	− 0.0186	−0.0237***	− 0.0065	−0.0447***	− 0.0098
	(− 4.13)	(− 1.55)	(−3.35)	(−1.52)	(− 2.67)	(− 0.73)	(− 3.35)	(− 0.63)
$Y_{i,t-1}$		0.4840***		0.7346***		0.7899***		0.5833***
		(3.49)		(11.25)		(3.89)		(5.99)
$Y_{i,t-2}$		0.4169**		0.0239		0.0143		0.0158
		(2.40)		(0.17)		(0.07)		(0.11)
$Y_{i,t-3}$		− 0.0560		0.0784		0.0365		0.1006
		(− 0.55)		(0.87)		(0.45)		(0.74)
$Y_{i,t-4}$		0.0148	−0.0072			0.0036		0.0791
		(1.08)		(−0.57)		(0.19)		(0.68)
$Y_{i,t-5}$		−0.0131		-0.0083		0.0073		− 0.0070
		(− 1.45)		(− 1.15)		(0.75)		(− 0.30)
Portfolio fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	262,457	243,880	139,586	130,503	262,457	243,880	139,586	130,503
Adjusted R^2	0.3273	–	0.3389	–	0.3269	–	0.3431	–
AR(1) test (p value)	–	0.002	–	0.000	–	0.004	–	0.000

Table 11 (continued)

	Trades$_{i,t}$				***Turnover***$_{i,t}$			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	FE	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**	**FE**	**SYS-GMM**
AR(2) test (p value)	–	0.200	–	0.313	–	0.574	–	0.531
Hansen test of over-identification (p value)	–	0.893	–	0.187	–	0.236	–	0.229
Diff-in-Hansen test of exogeneity (p value)	–	0.537	–	0.215	–	0.210	–	0.102

This table reports the results from the fixed-effects (FE) estimation of the panel regression model specified in Eq. 2 (odd columns) and the GMM (SYS-GMM) estimation of the panel regression model specified in Eq. 3 (even columns). The dependent variable is either the (log) number of trades of portfolios (Columns 1 to 4) or the turnover ratio of portfolios (Columns 5 to 8). *Leader Comment*, *Leader Count*, *Leader Positive* and *Leader Negative* are constructed after removing firm-specific comments. Only real-account portfolios of treated portfolios are included in the regressions. All explanatory variables are lagged by one week. In odd columns, standard errors estimated by the fixed-effects approach are double-clustered at the portfolio level and over time. In even columns, standard errors estimated by the system GMM approach are clustered at the portfolio level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

Table 12 Performance implications

	$Return_{i,t}$			
	(1)	**(2)**	**(3)**	**(4)**
Leader comment$_{i,t-1}$	− 0.0007***	− 0.0007***		
	(− 2.80)	(− 2.73)		
Leader count$_{i,t-1}$			− 0.00004	− 0.00003
			(− 0.27)	(− 0.22)
Leader positive$_{i,t-1}$			− 0.0022	− 0.0021
			(− 0.58)	(− 0.55)
Leader negative$_{i,t-1}$			− 0.0030	− 0.0027
			(− 0.86)	(− 0.77)
Trades$_{i,t-1}$	− 0.0023***		− 0.0021***	
	(− 8.78)		(− 7.70)	
Turnover$_{i,t-1}$		−0.0025***		−0.0025***
		(− 10.57)		(− 9.35)
Return SD$_{i,t-1}$	0.0025	0.0044	− 0.0007	0.0010
	(0.34)	(0.59)	(− 0.07)	(0.10)
No.securities$_{i,t-1}$	0.0006	0.0003	0.0008	0.0006
	(0.75)	(0.45)	(0.97)	(0.72)
No.followers$_{i,t-1}$	−0.0058***	− 0.0057***	− 0.0050***	− 0.0049***
	(− 7.02)	(− 7.01)	(− 6.09)	(− 6.12)
Portfolio age$_{i,t-1}$	0.0004	0.0004	0.0018***	0.0017***
	(1.19)	(1.21)	(3.30)	(3.20)
No.leaders$_{i,t-1}$	0.0001	0.0001	− 0.0002	− 0.0002
	(0.27)	(0.18)	(− 0.43)	(− 0.51)
Leader return$_{i,t-1}$	− 0.0034	− 0.0031	− 0.0078	− 0.0074
	(− 0.58)	(− 0.53)	(− 1.03)	(− 0.97)
Leader SD$_{i,t-1}$	− 0.0070	− 0.0063	− 0.0212	− 0.0209
	(− 0.56)	(− 0.51)	(− 1.45)	(− 1.43)
Leader trades$_{i,t-1}$	0.0003	0.0003	0.0002	0.0002
	(1.07)	(1.10)	(0.63)	(0.66)
Leader followers$_{i,t-1}$	− 0.000002	− 0.00002	− 0.0001	− 0.0002
	(− 0.01)	(− 0.08)	(− 0.56)	(− 0.58)
Leader securities$_{i,t-1}$	− 0.0003	− 0.0003	− 0.0004	− 0.0003
	(− 1.10)	(− 1.11)	(− 0.93)	(− 0.82)
Leader age$_{i,t-1}$	0.0001	0.0001	0.0001	0.0001
	(0.46)	(0.51)	(0.62)	(0.56)
Portfolio fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	262,457	262,457	150,447	150,447
Adjusted R^2	0.2043	0.2062	0.2314	0.2332

This table presents the estimation results of fixed-effects panel regressions on portfolio returns. Only treated real-account portfolios are included in the regressions. All explanatory variables are lagged by one week. Standard errors are double-clustered at the portfolio level and over time. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively

Practical implications

Our study provides a novel explanation for active trading and offers useful implications for individual investors and regulatory authorities. On the one hand, individual investors, especially those engaging in social trading, need to be aware of the impact of communication on their trading behavior. Such investors should trade more cautiously because the potential behavioral changes caused by communication may be detrimental to their wealth.

On the other hand, the Chinese stock market is an immature market dominated by individual investors (Carpenter et al. 2021). Thus, irrational trading behavior stimulated by Internet posts may lead to market fluctuations. Thus, regulatory authorities should extend the scope of supervision to online posts to stabilize financial markets.

Conclusion

This study examines the relationship between investment-related communication and trading frequency based on a unique dataset drawn from a Chinese social trading platform. Using a difference-in-differences analysis, we first show that leaders' comments increase portfolio owners' trading frequency. From the panel regressions, we find that real-account portfolio owners trade more frequently if their leaders post comments in the previous week. Moreover, the trading frequency of portfolio owners further increases with the number of comments posted by their leaders, although it is not sensitive to the tone of the comments. Finally, both trading frequency and leaders' comments negatively impact portfolio owners' future performance. Overall, this study suggests that social interaction, especially communication, plays an important role in shaping investing behavior. It will be fruitful to extend traditional behavioral finance theories to social finance to study how the social process affects financial outcomes.

Appendix

See Table 13.

Table 13 Summary statistics of the matched sample and treated portfolios This table presents the mean (Mean) and standard deviation (Std. Dev.) of the weekly observations of the variables for the matched sample and treated portfolios

	Matched sample		**Treated portfolios**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Trades	0.4904	0.8308	0.4785	0.8237
Turnover	0.4002	1.1399	0.3770	1.1061
Leader comment	–	–	0.5731	0.4946
Leader count	–	–	1.5802	1.5872
Leader positive	–	–	0.0556	0.0283
Leader negative	–	–	0.0828	0.0365
Return	− 0.0013	0.0388	− 0.0013	0.0386
Return SD	0.0120	0.0440	0.0119	0.0404
No.securities	1.7720	1.0864	1.7690	1.0918
No.followers	0.1200	0.3787	0.1337	0.3939
Portfolio Age	3.8097	1.0245	3.9829	0.8854
No.leaders	–	–	1.4846	0.7607
Leader Return	–	–	0.0009	0.0337
Leader SD	–	–	0.0134	0.0127
Leader trades	–	–	0.2854	0.5005
Leader securities	–	–	1.6948	1.0646
Leader followers	–	–	2.5574	1.8616
Leader age	–	–	3.9835	1.1526

Abbreviations

SET: Self-enhancing transmission
PSM: Propensity score matching
GEM: Growth enterprise market

Acknowledgements

We thank the editors and anonymous reviewers for their valuable comments and suggestions.

Author contributions

All authors contributed equally to this work. All authors have read and approved the final manuscript.

Funding

Funding was provided by National Natural Science Foundation of China (Grant No. 7167030951).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Competing interests

The authors declare that they have no conflicts of interest.

Received: 17 September 2021 Accepted: 15 June 2022
Published online: 24 June 2022

References

Abdeldayem M, Aldulaimi S (2021) Entrepreneurial finance and crowdfunding in the middle east. Int J Organ Anal ahead-of-print(ahead-of-print)
Akçay E, Hirshleifer D (2021) Social finance as cultural evolution, transmission bias, and market dynamics. Proc Natl Acad Sci USA 118(26)
Alevy JE, Haigh MS, List JA (2007) Information cascades: evidence from a field experiment with financial market professionals. J Financ 62(1):151–180
Ammann M, Schaub N (2020) Do individual investors trade on investment-related internet postings? Manag Sci Articles in Advance:1–24
Bali TG, Cakici N, Whitelaw RF (2011) Maxing out: stocks as lotteries and the cross-section of expected returns. J Financ Econ 99(2):427–446
Bali TG, Hirshleifer DA, Peng L, Tang Y (2019) Attention, social interaction, and investor attraction to lottery stocks. Georgetown McDonough School of Business Research Paper (3343769)
Barber BM, Odean T (2000) Too many cooks spoil the profits: investment club performance. Financ Anal J 56(1):17–25
Barber BM, Odean T (2000) Trading is hazardous to your wealth: the common stock investment performance of individual investors. J Finance 55(2):773–806
Barber BM, Odean T (2001) Boys will be boys: gender, overconfidence, and common stock investment. Q J Econ 116(1):261–292
Barber BM, Odean T (2002) Online investors: do the slow die first? Rev Financ Stud 15(2):455–488
Beck T, Levine R, Levkov A (2010) Big bad banks? the winners and losers from bank deregulation in the United States. J Financ 65(5):1637–1667
Benoît JP, Dubra J, Moore DA (2015) Does the better-than-average effect show that people are overconfident?: two experiments. J Europ Econ Assoc 13(2):293–329
Bikhchandani S, Hirshleifer DA, Tamuz O, Welch I (2021) Information cascades and social learning. NBER Working Paper
Breitmayer B, Mensmann M, Pelster M (2018) Social recognition and investor overconfidence. TAF Working Paper Series
Carpenter JN, Lu F, Whitelaw RF (2021) The real value of china's stock market. J Financ Econ 139:679–696
Chen S, Liu T, Peng Q, Zhao Y (2022) Manager sentiment bias and stock returns: evidence from China. Emerg Mark Financ Trade 58(3):823–836
Connelly BL, Certo ST, Ireland RD, Reutzel CR (2011) Signaling theory: a review and assessment. J Manag 37(1):39–67
Czaja D, Röder F (2020) Self-attribution bias and overconfidence among nonprofessional traders. Q Rev of Econ Financ 78:186–198
Gervais S, Odean T (2001) Learning to be overconfident. Rev Financ Stud 14(1):1–27
Han B, Hirshleifer D, Walden J (2022) Social transmission bias and investor behavior. J Financ Quant Anal 57(1):390–412
Heimer R (2016) Peer pressure: social interaction and the disposition effect. Rev Financ Stud 29(11):3177–3209
Heimer RZ (2014) Friends do let friends buy stocks actively. J Econ Behav Organ 107:527–540
Hirshleifer D (2015) Behavioral finance. Annu Rev Financ Econ 7(1):33–59
Hirshleifer DA (2020) Presidential address: social transmission bias in economics and finance. J Financ 75(4):1779–1831
Hong H, Kubik JD, Stein JC (2004) Social interaction and stock market participation. J Financ 59(1):137–163
Hong H, Kubik JD, Stein JC (2005) Thy neighbor's portfolio: word-of-mouth effects in the holdings and trades of money managers. J Financ 60(6):2801–2824
Kromidha E, Li MC (2019) Determinants of leadership in online social trading: a signaling theory perspective. J Bus Res 97:184–197
Kumar A (2009) Who gambles in the stock market? J Financ 64(4):1889–1933

Lane JN, Lim SS, Uzzi B (2021) Biased information transmission in investor social networks: Evidence from professional traders. Working Paper
Li J, Chen Y, Shen Y, Wang J, Huang Z (2019) Measuring China's stock market sentiment. Working Paper
Li J, Ding H, Hu Y, Wan G (2021) Dealing with dynamic endogeneity in international business research. J Int Bus Stud 52:339–362
Lin L, Xie D (2016) Do investors listen for the meanings behind executives' words? An empirical analysis based on management tones (translated from mandarin). J Financ Econ 7:28–39
Loughran T, Macdonald B (2016) Textual analysis in accounting and finance: a survey. J Account Res 54(4):1187–1230
Loughran T, Mcdonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J Financ 66(1):35–65
Odean T (1998) Volume, volatility, price, and profit when all traders are above average. J Financ 53(6):1887–1934
Pelster M, Breitmayer B (2019) Attracting attention from peers: excitement in social trading. J Econ Behav Organ 161:158–179
Pelster M, Hofmann A (2018) About the fear of reputational loss: social trading and the disposition effect. J Bank Financ 94:75–88
Pool VK, Stoffman N, Yonker SE (2015) The people in your neighborhood: social interactions and mutual fund portfolios. J Financ 70(6):2679–2732
Roodman D (2009) How to do xtabond2: an introduction to difference and system gmm in stata. Stata J 9(1):86–136
Shefrin H, Statman M (1985) The disposition to sell winners too early and ride losers too long: theory and evidence. J Financ 40(3):777–790
Shiller R (1989) Market volatility. MIT Press
Statman M, Thorley S, Vorkink K (2006) Investor overconfidence and trading volume. Rev Financ Stud 19(4):1531–1565
Steiger S, Pelster M (2020) Social interactions and asset pricing bubbles. J Econ Behav Organ 179:503–522
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185(4157):1124–1131
Wang C, Wu J (2015) Media tone, investor sentiment and IPO pricing (translated from mandarin). J Financ Res 9:174–189
Wang X, Shang W, Wang S (2019) The effects of online news on the Chinese stock market (translated from mandarin). Theory Pract 39:3038–3047
Wintoki MB, Linck JS, Netter JM (2012) Endogeneity and the dynamics of internal corporate governance. J Financ Econ 105(3):581–606
Xie D, Lin L (2015) Do management tones help to forecast firms' future performance: a textual analysis based on annual earnings communication conferences of listed companies in China (translated from mandarin). Account Res 2:20–27
Yang XL, Zhu Y, Cheng TY (2020) How the individual investors took on big data: the effect of panic from the internet stock message boards on stock price crash. Pac Basin Finance J 59(101):245
Yao S, Wang C, Cui X, Fang Z (2019) Idiosyncratic skewness, gambling preference, and cross-section of stock returns: evidence from China. Pac Basin Finance J 53:464–483
Zeng Q, Zhou B, Zhang C (2018) Annual report tone and insider trading: do insiders act as what they said (translated from mandarin). Manag World 9:143–160
Zhang X, Liang J, He F (2019) Private information advantage or overconfidence? Performance of intraday arbitrage speculators in the Chinese stock market. Pac Basin Financ J 58(101):215

Publisher's Note