International Review of Financial Analysis 91 (2024) 103035

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

Investors' opinion disagreement and abnormal trading around pre-earnings announcements

Xing Li [a], Keqiang Hou [b,*]

[a] School of International Economics and Trade, Nanjing University of Finance and Economics, Nanjing, China
[b] SHU-UTS SILC Business School, Shanghai University, 99 Shangda Road, Shanghai, China

ARTICLE INFO

Keywords:
Social media
Textual analysis
Heterogeneous opinion
Pre-earnings announcement
Stock returns

ABSTRACT

We use textual analysis techniques to construct a direct measure of investor disagreement to explain the abnormal stock volume and return around earnings expectation announcement in China. We crawl scrape the most a popular Chinese stock forum for individuals' posts about their opinion on a stocks, and classify the posts with machine learning methods. We find that our measure is most correlated with return volatility and historical turnover. We then conduct an event study for abnormal stock volume around pre-earnings announcements, examining the cumulative abnormal turnover for sample groups with prior agreement or disagreement, and with a convergence or divergence of opinion. We find that given prior disagreement of opinion, both convergence and divergence of opinion causing more trading volume in the Chinese market. We also observe the sample group with a divergence of opinion has more cumulative abnormal return than the convergence group. However, this effect is found to be more profound in stocks with higher institutional ownership, which is believed to be correlated with a low short-sale constraint.

1. Introduction

With the widespread use of social media, there has been a growing amount of discussions about stocks among retail investors on these platforms to exchange financial information. These discussions not only reflect the opinions of the participants regarding the stocks but also influence the investment behaviors of other investors. Previous literature indicates that the abnormal changes in stock trading volume and returns may be explained by the dynamics of public sentiments. The participants in the Chinese stock market are predominantly small and medium-sized investors (retail investors) who demonstrate herd behavior. Hence, the impact of social media investor sentiments is likely to be larger in China than in foreign countries, and the inherent volatility is also substantial. However, currently, there is limited discussion in China regarding whether investor sentiments on various social platforms can explain abnormal trading volume and returns, and indirect indicators (such as sentiment indices provided by mainstream media) are often used instead of directly accessing the original sentiments for analysis, which may affect research outcomes.

The advancement of text analysis technology has made it possible to directly and extensively obtain the sentiment of opinions. The purpose of this study is to directly obtain investor opinions from the most popular investor online discussion forum in China and classify them into bullish, bearish, or neutral using machine learning methods. Our approach overcomes the opacity and inflexibility of indirect indicators, while also capturing the real-time dynamics of opinion changes. We focus on the typical event window of pre-earnings announcement and combine existing theories to investigate whether the dynamics of opinion changes can explain variations in trading volume and stock returns. Furthermore, our study assists in evaluating the feasibility of using open-source social platforms for financial research, inspiring quantitative trading research among institutional investors. Moreover, the discourse on interconnections between opinion movements and other metrics could assist retail investors in nimbly leveraging those metrics for investment resolutions.

In light of the aforementioned literature, this paper opts to directly glean and dissect posts regarding the 2020 annual earnings pre-announcement windows from Guba, a forum hosted by Orient Fortune, encompassing 345 firms and 124,601 posts in total, covering A-shares and Growth Enterprise Market corporations. This paper firstly harnesses the naive Bayes approach to categorize the bullish and bearish emotions within the posts as bearish (−1), neutral (0), and bullish (1). For the 5

* Corresponding author.
E-mail addresses: lixing@njue.edu.cn (X. Li), houkeqiang@shu.edu.cn (K. Hou).

https://doi.org/10.1016/j.irfa.2023.103035
Received 21 July 2023; Received in revised form 7 October 2023; Accepted 9 November 2023
Available online 14 November 2023

days preceding the earnings release, the standard deviation (SD) of post emotions is computed. Afterwards, posts surpassing the median SD are denoted as heterogeneous perspectives, while those below the median SD are marked as homogeneous perspectives, thereby bifurcating the equities into announcement homogeneous cohort (PRE_A) and heterogeneous cohort (PRE_D). Regression finds that this measurement method has a high correlation with most existing heterogeneous opinion measurement methods. In addition, the post data is daily, so this paper also uses a similar method to divide the daily heterogeneous and homogeneous opinions of stocks into four groups in the [−5, 10] window during the release period: (1) AA group, i.e. homogeneous views both before and after the release date; (2) AD group, i.e. homogeneous views before the release date and heterogeneous views after; (3) DA group, i.e. heterogeneous views before the release date and homogeneous views after; (4) DD group, i.e. heterogeneous views both before and after the release date.

This paper regresses the homogeneous and heterogeneous opinion metrics obtained against indicators in existing literature and common financial variables, and finds that opinion heterogeneity is more pronounced in listed companies with high return volatility and high historical turnover rates. In addition, our findings are in line with Bamber, Barron, and Stober (1999) theory that stocks with high trading volume during the announcement period tend to have lower returns.

Next, we explore the relationship between homogeneous and heterogeneous opinions and stock trading volume during the pre-earnings announcement period. We find that the trading volume during the announcement period is significantly higher for the PRE_D group than the PRE_A group, and the difference in cumulative abnormal turnover between the two groups reaches 0.58% on the guidance release date and the next day. The difference between the two groups is also confirmed by the coefficient of the dummy variable PRE_OP in the multivariate regression. Note that the difference remains significant in the 10 and 25 days after the release date.

Similarly, this paper compares the cumulative abnormal turnover rates of groups AD, AA, DA, and DD, and finds that group DD has the highest cumulative abnormal turnover, followed by DA and DD, and group AA has the lowest. This shows that our results are similar to those of Giannini et al. (2019), proving that both transitions of opinions from homogeneous to heterogeneous and from heterogeneous to homogeneous can generate abnormal trading volumes.

We further examine the relationship between heterogeneous opinions and stock returns during the earnings guidance period. For PRE_A, the group with homogeneous views before the announcement, the cumulative abnormal return on the announcement date and the next date is 1.4%, while for PRE_D it is −0.06%, with a large difference that persists until 10 days after the announcement. The coefficient of the PRE_OP variable in the multivariate regression can also prove the significant difference in cumulative abnormal returns between the two groups. This finding is consistent with Miller (1977) theory that disagreement causes optimists to price stocks, leading to decreased future returns.

By comparing the cumulative abnormal returns of groups AD, AA, DA, and DD, according to Miller (1977) theory, the AD group's returns should experience an increase, because when opinions become more heterogeneous, stocks are priced by optimists and prices rise. This paper finds that this is not obvious, only after the 9th day after opinion release does the AD group rise, a possible explanation is that there is a certain lag effect from opinion heterogeneity to impact on stock prices. For the DA group, it is consistent with Miller (1977) theory, because when opinions become more consensus, the cumulative abnormal return of the DA group does not decrease but rises instead. And in the DA group, the announcement itself is mostly positive news, which also provides an explanation from an alternative vantage.

These findings provide new insights into explaining trading volume and stock returns from the perspective of heterogeneous opinions. This paper adopts a relatively novel direct method of measuring investor sentiment to test the theories and hypotheses in historical literature, finding that both homogenization and heterogenization of investor sentiment lead to increased anomalous trading volume, which is consistent with the theories of Kim and Verrecchia (1991a&b) and Kandel and Pearson (1995). In addition, we find that Miller (1977) theory on stock returns and heterogeneous opinions is confirmed, that is, heterogeneous opinions lead to stock prices being set by optimists and overestimated, resulting in declining future stock returns. When divided into four groups before and after the announcement, the DA group has higher returns than the AD group, which is also consistent with Miller (1977) theory.

The heterogeneous and homogeneous opinion metrics used in this paper are also correlated to some extent with other existing indicators. For individual investors, other easily accessible indicators can be used instead of obtaining a large amount of investor sentiment from social networks. For example, to determine whether opinions before the announcement release are heterogeneous or homogeneous, investors can refer to historical returns, which have a strong correlation with the heterogeneity and homogeneity of investor opinions, and the higher the historical returns, the more likely it is homogeneous opinions. There is also a significant relationship between heterogeneous/homogeneous opinions and the inverse of turnover rate and PB ratio. The higher the turnover rate and the lower the PB ratio, the more likely it is heterogeneous opinions.

Section 2 provides literature reviews. Section 3 provides hypotheses development. Section 4 will describe data and calculation of direct measure of opinions. Section 5 includes empirical analysis and results. Section 6 concludes.

2. Literature review

2.1. Investors' disagreement and high trading volume

The release of corporate pre-earnings announcement often causes investors to change their opinions. Therefore, in the historical literature, there are many studies that use this event as the background and employ heterogeneous opinions to explain trading volumes that are higher than hedging needs in the market. There are two main theories.

Kim and Verrecchia (1991a&b) argue that high trading volumes are due to news about corporate earnings causing investors to re-examine and change their opinions. Those who think the stock is overvalued will sell, while those who think it is undervalued will buy. This manifests as investor opinions gradually converging. This view is confirmed by other studies. For example, Atiase and Bamber (1994) find that before news is released, investor heterogeneous opinions are highly correlated with high trading volumes.

Kandel and Pearson (1995) argue that high trading volumes are because investors have different understandings of the news itself, causing investors who interpret it as good news to buy and vice versa. This manifests as investor opinions diverging. This view is also confirmed in the literature. Bamber et al. (1999) use analysts' different forecasts during the earnings announcement period to prove that the news itself caused people to form different opinions, thus generating trading volume.

Measurement of investor heterogeneous opinions is not easy, and usually adopted indirect methods: such as analyst forecast dispersion, historical trading volume, return volatility, company age, etc. (Berkman, Dimitrov, Jain, Koch, & Tice, 2009). Among them, although analyst forecast dispersion more directly measured heterogeneous opinions, it only represented a small portion of market participants (Atiase and Bamber, 1994; Bamber et al., 2011). In addition, some factors cannot change rapidly, especially those related to the company itself, such as company age. To verify the above two explanations required dynamically viewing whether investor opinions changed from heterogeneous to homogeneous or vice versa during pre-earnings announcement releases. Giannini et al. (2019) directly used comments on stocks at the Stocktwits

website along with corporate news, manually classifying positive and negative sentiments to train algorithms to automatically classify large amounts of comments and news, then evaluated the consensus between investor comments and news. Giannini et al. (2019) also found that consensus between investor comments and news was especially greater for firms with more diversified shareholders, and heterogeneous opinions were more likely when firm specific risk is higher. Giannini et al. (2019) used this factor to show both previous explanations were reasonable. Tetlock (2007) recorded that news sentiment reflected institutional investor trading activity and also reflect the overall state of investor heterogeneous opinions.

2.2. Investors' disagreement and stock returns

Heterogeneous opinions will lead to changes in asset prices. There are two strands of literature on the operating mechanism:

Miller (1977) proposes a model in which under dissenting opinions and short-sale constraints, stock prices are determined by optimists, thus overvalued, leading to lower future returns. This view is supported by Berkman et al. (2009). Giannini et al. (2019) also found relevant evidence: when investor opinions diverged after news releases, stock prices rose and future returns decreased. Varian (1985) considers heterogeneous opinions as an additional risk factor, leading to increased stock prices and possible lower future returns.

2.3. Literature on text mining in social media

Tumarkin and Robert (2001) criticized social media messaging platforms for having a lot of noise, but recently Chen, De, Hu, and Hwang (2014) proved that using text mining to extract investment information from new generation social media is valuable. Giannini et al. (2019) directly measured investor heterogeneous opinions using Twitter messages, achieving accuracy and timeliness, and found the results were close to those of previous indirect opinion measurement methods. Bollen et al. (2011) use information on Twitter to analyze changes in public sentiment with two models, one categorizing sentiment as positive and negative and the other as sentiment such as calm, alert, sure, vital, kind and happy. The study finds a clear correlation between public sentiment and the Dow Jones index.

Different types of social media means different indicator construction methods. For forums, there are often discussions on post quality and whether comment sentiments match the original post (Chen et al., 2014). For Twitter-style social media, number of followers can be extracted. Company's number of official followers represents a corporate attention metric. The number of Individual followers can be used to weight their opinions (Giannini et al., 2019).

3. Hypothesis development

3.1. Disagreement in investors' opinions and trading volume

Using a rational expectation model, Kim and Verrecchia (1991a) proposed that when the accuracy of information collected by individual investors prior to announcements differs, they will give different weights to the announcement and change their opinions to varying weights. The difference in such opinions will lead to abnormally high trading volume. Consistent with this theory, previous studies have shown that investors' heterogeneous opinions before earnings announcements are related to the number of views investors express (Atiase and Bamber, 1994; Bamber, Barron, & Stober, 1997). Thus, we test whether the heterogeneous opinion measurement method in this paper's social network can validate Hypothesis 1.

Hypothesis 1. Heterogeneous investor opinions before earnings pre-announcement are associated with higher trading volumes.

Banerjee and Kremer (2010) provided a more nuanced model, arguing that trading volume reflects changes in the degree of disagreement among people's opinions. In their model, trading volume can be explained by both dispersion and convergence of opinions. When people agree on the interpretation of common good and bad signals, trading volume depends on the degree of heterogeneous opinions before the signal - the greater the degree of heterogeneous opinions, the easier it is to converge opinions and generate trading volume. On the other hand, different understandings of earnings announcement can also lead to continued heterogeneous opinions and generate trading volume from opinion divergence after announcement. Although both opinion convergence and divergence trades are reasonable in Banerjee and Kremer (2010) model, it is not easy to detect the combination of both trades. In this paper, we test the dynamics of trading volume during pre-earnings announcement, where the direct flexible opinion measurement method derived from investors' posts on a financial media makes it possible.

Hypothesis 2. Either convergence or divergence of investor opinions during earnings pre-announcement periods are both accompanied by higher trading volumes.

3.2. Disagreement in investors' opinions and stock returns

Miller (1977) proposes a model where people held homogeneous or heterogeneous views before earnings announcements. A key assumption in Miller's model is the limited short-selling, so pessimistic investors either short-sell or cannot invest in the stock. Thus, investors' heterogeneous opinions will cause optimists to set asset prices, leading to overpricing and lower future returns. Under the short-selling constraints in the Chinese stock market, we test Hypothesis 3 implied by Miller (1977) theory.

Hypothesis 3. Disagreement in investor opinions before earnings pre-announcement leads to negative stock returns after the announcement.

According to Miller (1977)'s theory, information disclosure often leads to convergence of opinions, as it can reduce dissent. In fact, earnings announcements may also cause investor disagreements due to different understandings of earnings news. In the latter case, large short-sale constraints lead to stock prices set by optimists, while without constraints there is no partial pricing situation. Given the short-sale constraints in the Chinese stock market, we propose to test Hypothesis 4.

Hypothesis 4. If earnings pre-announcements make investor opinions converge, returns on the announcement date are negative; if pre-announcements make opinions diverge, returns are positive.

4. Data and opinion measurement

4.1. Data on Pre-earnings announcement

According to SEC regulations, the listed companies in China must regularly disclose reports, including annual reports, interim reports and quarterly reports, and are required to issue earnings announcement before report releases under certain conditions. In general, listed companies in the Shanghai and Shenzhen markets need to provide pre-earnings announcements when they expect the following situations in the current reporting period: (1) net profit is negative; (2) net profit increases or decreases by more than 50% from the same period last year; (3) turning losses into gains. For annual reports, over 400 listed companies often issue pre-earnings announcements, while for quarterly reports, over 160 companies issue announcements. In order to test hypothesis about investor opinions driving abnormal trading volume and returns. Announcements before the formal reports tend to contain more negative or surprising news, which can better motivate investors to change their opinions.

We crawl pre-earnings announcement data in 2021 from Sina Finance platform, collecting stock codes, stock names, announcement

types, announcement dates, reporting periods, earnings guidance highlights and earnings per share from the same period last year, a total of 1345 company announcements.

4.2. Data on individual Investors' posts on Eastmoney platform

Eastmoney platform is operated by Eastmoney security company that integrates functions such as financial transactions, market quotes, news browsing, community communication and financial data support. The company's web and app platforms are widely used domestically. People browse financial information and even engage asset transactions on these platforms, which have had a profound impact on various aspects of domestic investment activities.

Eastmoney platform has separate forums for discussing every stock. According to our statistics, popular individual stock forums often have over 400 posts on trading days expressing investors' views and emotions about stock ups and downs. Even less popular forums have over 20 posts. The number of posts is sufficient to form the daily homogeneous and heterogeneous opinion metrics for a stock.

We use Python web crawling technology to crawl opinions in the individual stock forums of companies that released 2021 annual earnings guidance, 5 days before and after the release date, 11 days in total. After removing advertisements, reposts, and official consultations, I obtained a total of 124,601 posts, 1345 companies, with each record mainly containing the post content and time. The posts are not long, often just one sentence, for example: "Everyone run away, this stock will hit the limit down tomorrow!" which directly expresses an opinion without analysis. Some posts have slightly more analysis, such as "This stock covers several popular sectors like lithium batteries and e-sports, it must go up!" Some questioning posts are marked as neutral by this paper, such as "The uptrend continued after the opening in the afternoon session, have the good news run out?"

4.3. Classification of bullish/bearish sentiment

In this paper, we categorize the viewpoints in the posts into bullish, bearish, and neutral. Sentimental analysis is a field of natural language processing that involves using text-mining methods to identify and extract subjective information from the original material. The purpose is usually to find out the author's point of view and attitude; this attitude may be the author's individual judgment or their emotional state at that time. This research applies sentimental analysis to financial news headlines using a Naïve Bayesian classifier, which is as follows.

$$P(A|B) = \frac{P(B|A) \bullet P(A)}{P(B)}$$

The Bayesian theorem formula states that given event B has occurred, the probability of event A occurring can be derived from the formula on the right. Suppose there is a post saying "This is good news." Event A is the post having a bullish view, and event B is the post "This is good news". According to the formula, to infer the probability that this post has a bullish view, we need to find out $P(B|A)$, $P(A)$, and $P(B)$. The probability of its occurrence $P(B)$ can be neglected since only the rank of probability for each emotional tendency is of interest. In the nominators, $P(A)$ is priori probability, which can be obtained from previous experience or from training set. Assuming the independence of each word in a sentence, which is why the model is naïve, $P(\textit{This is good news}|A)$ can be expressed as $P(\textit{This}|A)\times P(\textit{is}|A)\times P(\textit{good}|A)\times P(\textit{news}|A)$. Therefore, the basic idea of naive Bayesian sentiment classification is to first annotate part of the corpus as a training set, then train the model to obtain the probability of words appearing in each sentiment classification.

We use the above method to classify the 124,601 posts obtained. First, we randomly selected and manually categorized 2000 comments as bullish, bearish, and neutral. Next, due to the difference between Chinese and English, the Scikit-learn Python library could not be directly used for training and classification of posts, because there are spaces between words in English but not in Chinese. Therefore, we use the Jieba library for word segmentation on the posts. The Jieba library can efficiently scan the word graph based on a prefix dictionary to generate a directed acyclic graph (DAG) of all possible word formation in a sentence, and uses dynamic programming to find the maximum probability path and obtain the maximum combination based on word frequency. At the same time, this paper uses the precise mode in Jieba library word segmentation types, which is most suitable for textual analysis.

The specific data form of each post in this paper is as follows: All the words in the manually categorized posts are formed into a vector, where the number of occurrences of each word in the post is reflected in the corresponding position in the vector. One occurrence is 1, two occurrences is 2, and not occurring is 0. Thus each post's data point is a vector, such as (0,1,1,0,0,0,2, …,0). Calling Scikit-learn's Naive Bayes learner and inputting the training dataset, we can use the Naive Bayes algorithm to categorize the remaining post data - bearish sentiments are marked as −1, neutral as 0, and bullish as 1. Based on the test data, the categorization accuracy reached over 80%, similar to literature on recognizing sentiment of text opinions.

4.4. Classification of opinion convergence/divergence

For each company, since the readership of each post is similar, for simplicity we do not weight the posts by readership. Instead, we directly sum the sentiment labels of all posts each day to obtain the daily opinion tendency (opinion_sum). The larger the value, the stronger the bullish sentiment. The smaller the value, the stronger the bearish sentiment. Values close to 0 indicate more pronounced heterogeneous opinions.

Furthermore, we divide the observation window into before and after pre-earnings announcement, as well as daily frequency, and try to determine whether the opinions on each stock are heterogeneous or homogeneous. For all comments on a stock before the announcement, we calculate the standard deviation of the −1, 1, 0 sentiment values, so each stock has an announcement sentiment standard deviation. If the sentiment standard deviation of this stock falls below the median of the standard deviations of the entire sample, then this stock is labeled as the PRE_A (previously in agreement) group, recorded as 0, i.e. homogeneous opinions before the announcement; otherwise it is the PRE_D (previously in disagreement) group, recorded as 1, i.e. the heterogeneous group. This variable is called PRE_OP. Based on the opinion status before and after the announcement, stocks can also be divided into four groups according to the dynamic changes: 1) AA group, i.e. homogeneous opinions both before and after the announcement date; 2) AD group, i.e. homogeneous opinions before the announcement date and heterogeneous opinions after; 3) DA group, i.e. heterogeneous opinions before the announcement date and homogeneous opinions after; 4) DD group, i.e. heterogeneous opinions both before and after the announcement date. In addition, we also take a cross section of each day's situation, and labels whether the stock has homogeneous or heterogeneous opinions that day, recording this variable as OP_DAILY.

4.5. Description on indirect measurements of opinion disagreement

According to the literature, commonly used indicators for measuring heterogeneous opinions include: the inverse of company age (AGE); turnover rate (TURN), taking the daily average over the 45-day window from 56 days to 11 days before the earnings announcement; profit volatility (INCVOL), i.e. the standard deviation of profit changes over the past 20 quarters. In light of Giannini et al. (2019), we also develop three indicators for measuring heterogeneous opinions based on abnormal trading volume: one is the standard unexplained volume (SUV), which is the ratio of the difference between the actual and predicted trading volume to its own standard deviation, calculating the sum of the ratios for the announcement date and the next date, divided by the predicted trading volume for the announcement date obtained from linearly regressing trading volume on the 50-day window return rate

from 54 days to 5 days before the announcement; the second indicator is the market-adjusted turnover (MATO), which is the mean difference between the stock turnover rate on the announcement date and the previous date, and the market turnover rate; the third indicator is the change in market-adjusted turnover, which is calculated by taking the 50-day window mean of market-adjusted turnover from 54 days to 5 days before the announcement, and taking the difference between the announcement date turnover rate and the mean. These variables are also described in detail in Appendix. Most of the variables were obtained from the Joinquant platform,[1] while the market turnover rate variable was downloaded from the CSMAR database.

5. Empirical analysis and results

5.1. Summary statistics and disagreement measurement

We summarize descriptive statistics for the direct indices for heterogeneous opinion in this paper, indirect indicators from previous literatures, and control variables. Table 1 presents summary statistics for the 1345 sample pre-earnings announcements during the sample period of Jan, 2021 to Dec, 2021. ME is the announcing firm's market capitalization. CAR [0,1] is the cumulative abnormal return for the announcement date and the next date used later, where 0 is the earnings announcement day. Daily abnormal return is based on the four-factor model estimated in the 90-day window [−120,−31] ending 31 days prior to the earnings announcement. PRE_OP is the dummy variable of investor disagreement constructed using forum posts prior to the pre-earnings announcement. We also report summary statistics for the commonly used measures of investor disagreement. INCVOL is the standard deviation of the quarterly operating income. RETVOL is the standard deviation of the announcing firm's daily excess returns. Log(1/AGE): take the inverse of company age, then the natural logarithm. TURN is average daily turnover from 45 days to 11 days before the announcement. Earnings Persistence: earnings stickiness, the coefficient of the lagged year variable when regressing operating profits of the past 8 years with each lagged by 1 year. Ln(ME): take natural logarithm of the market cap of the company last year. Ln(B/M): take inverse then inverse of market-to-book ratio. Ret[−12,−2]: stock returns from 12 months to 2 months before the announcement month. MATO (market adjusted turnover) is the average daily abnormal turnover in the [0,1] window. ΔTO is calculated as MATO minus the announcing firm's average daily abnormal turnover on the window [−45,-5] before the announcement. Standardized unexpected volume, SUV, is the abnormal volume in the [0,1] window[2]. These variables are described in detail in Appendix A.

Table 1
Descriptive statistics.

Variable	Obs	Mean	Std.Dev.	Min	Max
PRE_OP	1345	0.493	0.501	0	1
CAR[0,1]	1296	0.007	0.055	−0.235	0.195
INCVOL	1339	0.050	0.306	0.002	5.564
RETVOL	1339	0.028	0.012	0	0.105
AGE	1345	13	8.033	0	31
TURN	1344	2.658	4.163	0.225	52.333
Earnings Persistence	1345	0.287	0.630	−3.244	5.644
ln(ME)	1345	4.001	1.125	0	8.223
ln(B/M)	1339	0.951	0.859	−1.243	4.165
Ret[−12,-2]	1331	0.121	0.530	−0.701	3.787
MATO	1345	1.303	3.742	−1.266	40.016
ΔTO	1343	−0.158	3.312	−30.002	28.348
SUV	1345	0.332	1.727	−2.943	5.920

It can be seen from PRE_OP that the numbers of companies with homogeneous and heterogeneous opinions are similar, and the variable ranges are similar compared to previous literature. The mean of CAR [0,1] is 0.7% and the standard deviation is 5.5%, indicating both positive and negative abnormal returns on the announcement date and the next date, with the mean close to zero and a relatively large standard deviation of 5.5%. The minimum and maximum are −24% and 20% respectively, indicating a potentially large absolute value of abnormal returns. INCVOL is the volatility of the operating profit ratio (standard deviation). It can be seen that the average quarterly standard deviation of the operating profit to asset ratio is around 5%, with a minimum volatility of 0.2% and a maximum standard deviation of 500%. RETVOL shows the historical return volatility, and the average standard deviation of returns is 2.8%. AGE is the company age since listing, showing a large range from less than 1 year to 31 years, with an average of 13 years. This indicates our sample is quite comprehensive. It should be noted that some variables require the company to be listed for a certain number of years, such as the previous variable INCVOL which needs 20 quarters of financial data, and RET[−12,-2] which needs 1 year of return data, so some sample loss is inevitable. TURN is the average turnover rate over window [−45,-11] in percentage points. It can be seen the average is about 2.7% with a standard deviation of 4.2%, indicating turnover is not high but has some fluctuation. Earnings persistence shows for ¥1 increase in last year's profit, this year's profit increases by ¥0.29. ME is market cap in 100 million yuan, and the average of Ln(ME) is around 4. Looking at Ret[−12,-2], the average return over the 10 months before 2 months prior to the announcement reaches 12%, with a standard deviation of 53% and maximum of 379%. So these stocks have relatively good historical returns with large differences. MATO is the average market-adjusted turnover on the announcement date and next date. It can be seen the average drops after market adjustment, and more importantly volatility also decreases. ΔTO is MATO minus average historical market-adjusted turnover, and its mean is −0.15%, indicating historical turnover is almost the same as MATO, and average turnover during the announcement period does not increase significantly. From SUV it can be seen that trading volume on the announcement date and next date is still higher than expected, since its mean is positive, from the perspective of predicting trading volume solely based on returns.

5.1.1. Comparison between PRE_OP and indirect measures of disagreement

To better understand the variables of heterogeneous opinion status, we regress the PRE_OP variable on indicators from previous literature to explore their relationships, including profit volatility INCVOL, return volatility RETVOL, company age AGE, and turnover TURN. Additionally, some common stock characteristics are included: ln(ME) - natural logarithm of market cap, ln(B/M) - natural logarithm of the inverse of market-to-book ratio, and stock returns RET[−12,-2] over the window from 12 months to 2 months before the announcement.

We also regress PRE_OP on variables describing information quality, including two variables describing earnings quality: The first is earnings persistence, which is the coefficient when regressing earnings from the past 8 years against the corresponding previous year's earnings before the announcement; The second is the absolute accrual amount of the company, |Accrual|. In addition, the number of institutional

[1] Joinquan platform is a financial database on Chinese stock market. Please check the website, www.joinquant.com, for detailed information.

[2] SUV: use return rates on window [−54,-5] before the announcement date to regress trading volume, use the resulting formula to predict trading volume on the announcement date and next date, take the average after comparing the difference between actual and predicted trading volume with standard deviation.

shareholders can also reflect information quality. We take the logarithm after adding 1 to this number to form the variable Ln(1 + #inst.). We first regress using common stock characteristics, and then group the regressions by existing indicators reflecting heterogeneous opinions. Models (2), (3), and (4) are the regressions using indicators reflecting volatility, market sentiment, and information quality respectively. Model (5) uses all variables for regression. The regression results are shown in Table 2.

It can be seen from the regression that historical returns are consistently negatively correlated with heterogeneous opinions, that is, the higher the historical returns, the lower the likelihood of heterogeneous opinions, while other indicators make positive contributions to heterogeneous opinions, except for market value which varies. In addition, it can be seen that among the volatility indicators, the volatility of stock returns is more significantly related to heterogeneous opinions than the volatility of operating profits. At the same time, the average historical turnover of stocks is also significant, making a positive contribution to heterogeneous opinions. The total accrual amount of listed companies also has a significant contributing effect. Overall, historical returns, historical return volatility, historical turnover, and accruals can better reflect the degree of opinion heterogeneity compared to the opinions used in this paper.

5.1.2. Convergence/divergence of opinion

In Section 3.4, firms were categorized into four clusters: progressively heterogenizing, homogenizing, retaining heterogeneity, and maintaining homogeneity. In this section, we explore the relationship between these four dynamic types around pre-earnings announcement and other common indicators. Bamber et al. (1999) proposed that low returns and high trading volumes often accompany the heterogeneity of opinions. We examine this theory and whether other indicators can predict the heterogeneity of opinions.

First, we select the sample with homogeneous opinions before and after the announcement and the sample with opinions changing from homogeneous to heterogeneous, and establishes a dummy variable DIV, with the AD group coded as 1 and the AA group as 0. Next, a dummy variable LowretHighvol is established to test Bamber et al. (1999) theory. If the cumulative abnormal return of the company over the [−1,1] window is in the lowest quintile of the sample, and the cumulative abnormal turnover is in the higher half, then LowretHighvol = 1, otherwise LowretHighvol = 0. If DIV is significantly positively correlated with LowretHighvol, then high-volume low-return guidance leads to opinion heterogeneity, validating Bamber et al. (1999) view.

We also include other independent variables such as company size, PB ratio, historical returns, etc. The three variables describing information environment quality mentioned in the previous section are also included: number of institutional investors, earnings persistence, and total accruals. The more institutional investors, the more shareholders with considerable weights have their own unique inside information about the company. With the same pre-earnings announcements, it is more likely for them to react differently, leading to heterogeneous opinions.

Panel A of Table 3 shows the regression results of opinion divergence. It is found that when LowretHighvol alone is regressed on DIV or turnover change is controlled, the significance of LowretHighvol is higher, while the significance decreases after controlling other common indicators. But the regression coefficients are positive for all three regressions, so it can be seen that Bamber et al. (1999) view makes sense that low returns plus high trading volumes do accompany opinion divergence. In addition, it can be seen that the inverse PB ratio has higher significance, and the lower the PB ratio, the easier it is to cause opinion divergence. For historical returns, the lower the returns, the easier it is to cause opinion divergence. In addition, earnings persistence and number of institutional holdings reflecting company information quality are negatively correlated with opinion divergence, indicating the greater the earnings persistence and the more institutional holdings, the less likely opinion divergence is, information is more stable with higher quality.

Similarly, we take DD as reference and regresses DD and DA groups, with DD coded as 0 and DA as 1 (i.e. constructing dummy CON = 1,0). Other variables are the same as previous group. It can be seen that except for historical returns, the signs of other variables are opposite to the previous group, indicating these variables contribute in the opposite way to homogenization. Notably here the LowretHighvol variable is highly significant, contributing negatively to opinion convergence. Change in turnover is also negatively correlated with convergence, that is, if turnover increases after guidance, the likelihood of opinion convergence decreases. For variables related to information quality, it can be seen the more institutional investors, the more likely opinion convergence occurs. The greater earnings persistence, the easier convergence occurs after guidance. Accruals are significant and negatively correlated, indicating that more accruals reflect lower information quality, making it difficult for people to reach consensus.

In summary, three insights emerge from regressing divergence and convergence on extant variables: Firstly, Bamber et al. (1999) hypothesis proves valid, with the extracted low return high volume trait positively associated with divergence and negatively with convergence, especially significantly for the convergence panel. Secondly, turnover changes demonstrate weaker significance versus the former table, implying superior explanatory powers of historical turnover for announcement homogeneity and heterogeneity, alongside weaker associations between turnover shifts and opinion changes. Thirdly,

Table 2
Regressions of PRE_OP on existing disagreement measures and firm characteristics. This table presents logistic regressions of the PRE_OP measure. The main dependent variables include the existing disagreement measures. INCVOL is the standard deviation of the quarterly operating income. RETVOL is the standard deviation of the announcing firm's daily excess returns. Log(1/AGE): Natural logarithm of inverse of firm age. TURN is average daily turnover measures. We also examine firm characteristics that can potentially affect investor disagreement. ln(1 + #inst) is natural log of 1 plus the number of institutional shareholders of the announcing firm. Earnings persistence is the coefficient in regressions of annual earnings on lagged annual earnings in the eight years ending in the year before the earnings announcement. |Accrual| is absolute value of the announcing firm's total accrual of the year before the earnings announcement. We further control for commonly used firm characteristics, where ln(ME) is natural log of the market capitalization of the announcing firm; ln(B/M) is natural log of the book-to-market ratio of the announcing firm; Ret [−12,−2] is the buy-and-hold stock return of the announcing firm from month 12 to month 2, where month 0 is the month of announcement.. ***, **, * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variable: PRE_OP				
	(1)	(2)	(3)	(4)	(5)
INCVOL		0.132			0.372
		(0.20)			(0.52)
RETVOL		7.812***			3.598*
		(3.16)			(1.92)
log(1/AGE)			0.019		0.042
			(0.43)		(0.91)
TURN			0.043***		0.028*
			(3.18)		(1.78)
Ln(1 + #inst.)				0.029	0.017
				(0.51)	(0.30)
Earnings Persistence				0.020	0.028
				(0.43)	(0.61)
|Accrual|				0.000*	0.000
				(1.75)	(1.58)
ln(ME)	0.017	0.018	0.035	−0.032	−0.003
	(0.56)	(0.60)	(1.14)	(−0.60)	(−0.06)
ln(B/M)	0.043	0.032	0.029	0.065	0.036
	(1.15)	(0.87)	(0.70)	(1.65)	(0.84)
Ret[−12,-2]	−0.060	−0.108**	−0.087*	−0.055	−0.086
	(−0.94)	(−1.96)	(−1.74)	(−0.85)	(−1.28)
Pseudo-R^2	0.038	0.058	0.031	0.049	0.079

Table 3
Determinants of divergence and convergence of opinion. This table shows logistic regressions on factors that may cause the divergence of opinion (Panel A) or convergence of opinion (Panel B). Panel A includes the AA and AD groups, the sample consists of earnings announcements in the "Agree / Agree" group and the "Agree / Disagree" group. The dependent variable is a dummy variable, DIV (Divergence) that equals 1 if an event is in the "Agree / Disagree" group, and 0 if an event is in the "Agree / Agree" group. Panel B includes DD and DA groups, the sample consists of earnings announcements in the "Disagree / Agree" group and the "Disagree / Disagree" group. The dependent variable, CON (Convergence) is a dummy variable that equals 1 if an event is in the "Disagree / Agree" group, and 0 if an event is in the "Disagree/ Disagree" group. The independent variables in Panels A and B include Low|Ret|/HighVol, which is a dummy variable for earnings announcement with low absolute earnings-announcement return and high earnings-announcement trading volume (CAV[−1,1]), and 0 otherwise. Δ TURN is the change in turnover (TURN) around the earnings announcement. ΔRETVOL: Change in return volatility before and after guidance. Ln(ME): Natural log of market cap last fiscal year-end. Ln(B/M): Natural log of inverse of market-to-book ratio. Ret[−12,-2]: Returns from month −12 to −2 before the announcement month. Ln(1 + #inst.): Number of institutional investors holding the stock. Earnings Persistence: Coefficient estimate from regressing past 8 years earnings on previous year earnings before announcement. |Accrual|: Absolute accrual amount of most recent fiscal year. See Appendix A for variable definitions. ***, **, * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Panel A:Divergence of Opinions Dependent Variable:DIV			Panel B:Convergence of Opinions Dependent Variable:CON		
	(1)	(2)	(3)	(1)	(2)	(3)
Lowret	0.234**	0.240**	0.159*	−0.243**	−0.251**	−0.270**
Highvol	(1.97)	(1.93)	(1.90)	(−1.98)	(−2.03)	(−2.15)
ΔTURN		0.514	0.366		−0.776	−0.705
		(0.27)	(0.19)		(−0.54)	(−0.48)
Ln(ME)			0.094			−0.039
			(1.06)			(−0.45)
Ln(B/M)			0.082			0.018
			(1.30)			(0.28)
Ret[−12,-2]			−0.116			−0.043
			(−1.12)			(−0.34)
Ln(1 + #inst.)			−0.081*			0.027
			(−1.84)			(0.31)
Earnings Persistence			−0.051			0.052
			(−0.57)			(0.84)
|Accrual|			0.000			0.000
			(0.98)			(−1.50)
Pseudo-R^2	0.021	0.021	0.043	0.014	0.014	0.029

fundamentals and information quality variables exhibit generally feeble significance, despite opposite correlations across panels. Overall, the directly social media-extracted DIV and CON variables encompass appreciable explanatory capacity from pertinent academic factors, relatively accurately representing divergence and convergence. This further validates social media's information harvesting capabilities.

5.2. Disagreement and trading volume

In this section, we attempt to explore through event studies whether heterogeneous opinions before announcement are accompanied by high trading volumes during the announcement period, and whether both opinion divergence and convergence can explain high trading volumes, namely Hypothesis 1 and Hypothesis 2.

Abnormal trading volume in this paper is measured as follows: taking the turnover of the stock and the market from 245 days to 45 days before the announcement date, regressing the market turnover on the stock turnover, then using the obtained linear regression formula to predict the stock turnover in the [−5,10] window during the announcement, and taking the difference between the predicted and actual turnover to obtain the abnormal turnover (may be referred to as abnormal trading volume or announcement trading volume below). Expressed in formulas:

$$AV_{it} = V_{it} - (\alpha_i + \beta_i V_{mt}) \quad (1)$$

V_{it} and AV_{it} are the turnover rate and abnormal turnover of the stock on that day, i.e. Abnormal Volume, V_{mt} is the market turnover rate that day, and α_i and β_i are the parameters obtained from regressing the stock turnover rate against the market turnover rate over the [−245,-45] window.

Firstly, we examine cumulative abnormal volumes for stocks with announcement homogeneous and heterogeneous opinions. The total sample is divided into homogeneous PRE_A and heterogeneous PRE_D groups per the pre-5-day sentiment standard deviation magnitudes. Mean [−5,10] window AV is tallied and plotted. As Fig. 1 illustrates, [−5,10] window AV changes appear analogous between the homogeneous and heterogeneous stocks, with pronounced elevation for both on the announcement and succeeding dates, alongside slight 7th day rebounds after two-day announcement drops. However, heterogeneous opinions persistently maintain positive abnormal volumes while homogeneous ones exhibit negative volumes, validating Hypothesis 1 that heterogeneous cohorts engender higher announcement trading volumes. Additionally, it can be observed that changes in abnormal trading volumes of stocks with heterogeneous opinions are more drastic, with higher volatility than stocks with homogeneous opinions.

As shown in Fig. 2, it shows the change in cumulative abnormal volume, CAV. In this chart, the difference between the homogeneous group before the announcement and the heterogeneous group before the announcement can be more clearly seen. It can be seen that the heterogeneous group before the announcement remained above the homogeneous group, validating that heterogeneous opinions lead to abnormally high trading volume during announcements. The cumulative abnormal turnover of the heterogeneous group kept rising, reaching around 2% on the announcement day, possibly reflecting that the heterogeneous group had many institutions with different information sources, obtaining relevant information in advance before the announcement and starting to trade. The cumulative abnormal trading volume of the homogeneous group experienced a fall then rise, being exactly 0 on the announcement day. The cumulative abnormal turnover of both groups continued to grow slowly 2 days after the announcement.

Panel A of Table 4 shows the abnormal trading volume data at different stages within the [−5,10] window: it is found that the cumulative abnormal trading volume is higher for the heterogeneous group at each stage, and the Welch's *t*-test on their difference is statistically significant. The difference in cumulative abnormal trading volume on the announcement date and the next date is the most significant. This indicates that homogeneous and heterogeneous opinions have a relatively large impact on trading volumes on and after the announcement

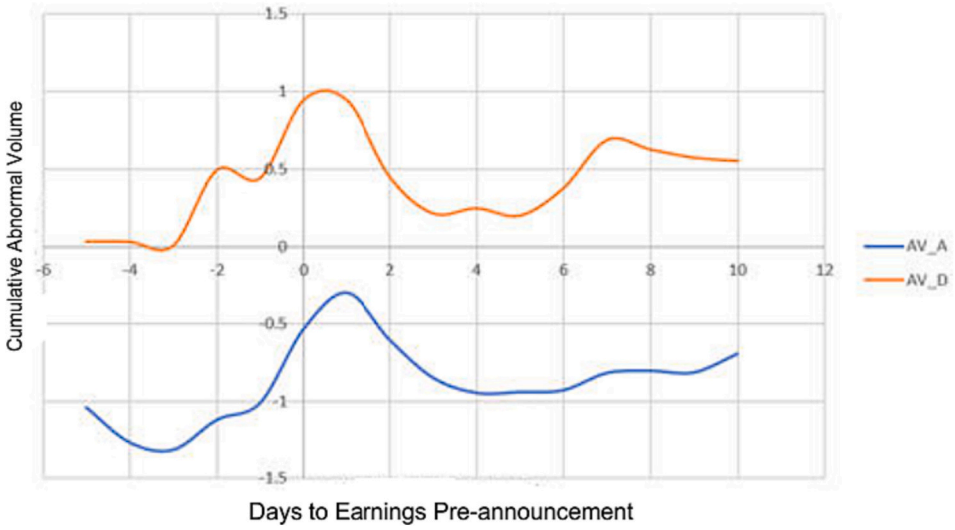

Fig. 1. Changes in Abnormal Trading Volumes of Homogeneous and Heterogeneous Opinion Groups around Pre-earnings announcement. This figure depicts the daily abnormal turnover changes of homogeneous and heterogeneous opinion groups within the [−5,10] window around guidance releases, where day 0 is the announcement date. The abnormal turnover of this window is obtained by subtracting expected turnover from realized turnover, and the expected turnover is calculated using a regression formula of market turnover on stock turnover over the [−245, −45] window.

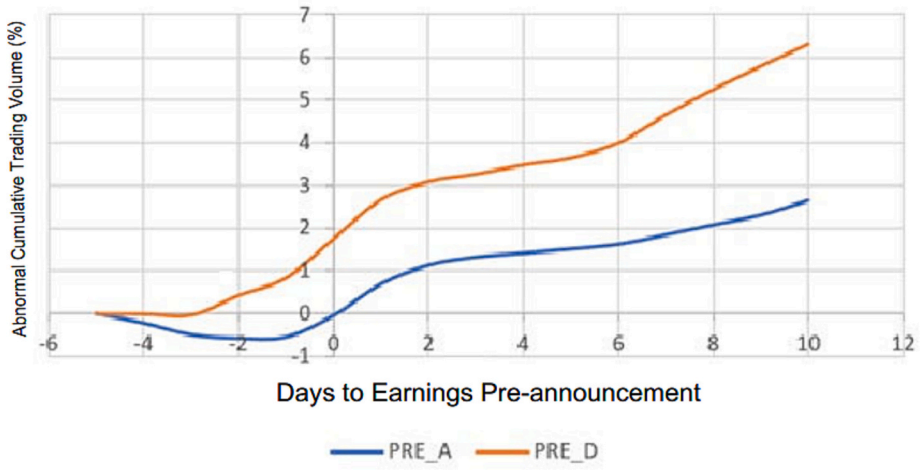

Fig. 2. Cumulative Abnormal Trading Volume Changes of Homogeneous and Heterogeneous Opinion Groups during Pre-earnings announcement Period. This figure depicts the cumulative abnormal turnover changes of the homogeneous opinion group and the heterogeneous opinion group over the [−5,10] window around pre-earnings announcement, where day 0 is the announcement date. The abnormal turnover for this window is obtained by subtracting expected turnover from realized turnover, where expected turnover is calculated by regressing market turnover on stock turnover over the [−245,-45] window. Cumulative abnormal turnover sums up the daily abnormal turnovers. For ease of observation, day −5 abnormal turnover is set to 0.

date.

Next, this paper calculates the cumulative abnormal trading volumes of the aforementioned 4 groups with different opinion change dynamics over the [−5,10] window: AD, AA, DD, DA. As shown in Fig. 3, it can be seen that within the [−5,10] window around the announcement date, each group exhibits roughly similar but slightly different patterns.

It can be seen the abnormal turnover patterns of the four groups are similar, with significant increases on the announcement date and the next date, followed by declines starting the second day after. The abnormal trading volume of the group with heterogeneous opinions both before and after the announcement remained positive and highest. The abnormal trading volumes of groups DA, AD, AA followed in turn. Notably, compared to group AA, group AD's abnormal turnover was lower than AA's before 3 days prior to the announcement and after 9 days, but exceeded and remained above AA over the [−2,8] window. This shows that if pre-earnings announcement causes investor opinions to become more heterogeneous, it will lead to increased abnormal trading volume during the announcement period, but will become similar to group AA about a week later. For groups DD and DA, DD remained higher than DA. This validates the rationality of Hypothesis 2, because in the core announcement period AD is greater than AA, opinion divergence leads to increased abnormal trading volume. Although DA is below DD, when opinions become more homogeneous, it still brings a significant increase in abnormal trading volume, thus strongly proving Hypothesis 2 that opinion convergence also increases trading volume.

As shown in Fig. 4, the differences of cumulative abnormal trading volume between groups become more distinct. AD notably exceeds DA during the announcement, DD leads other groups by a wide margin,

Table 4
Cumulative abnormal trading volume across groups of disagreement. Panel A presents cumulative abnormal trading volumes (CAV) around pre-earnings announcements across the groups of investor disagreement. We first sort announcements into two groups based on the value of the PRE_OP measure. We then report the averages of the cumulative abnormal volumes (CAVs) in the [0,1], [2,10], and [0,10] windows for the two groups, where day 0 is the earnings announcement day. For an announcement, we use the methodology of Campbell and Wasley (1996) to daily abnormal trading volume where the coefficients are estimated in the 200-day window ending 45 days prior to the earnings announcement [−245,-45]. We also report the volume differences and the associated two-sample t-statistics assuming unequal variances. Panel B presents CAVs for the four groups of the changes in investor agreement. We calculate the PRE_OP measures after earnings announcements. The "Agree / Agree" group contains the announcements where the PRE_OPs equal 0 in both pre- and post-announcement periods. The "Agree / Disagree" group contains the announcements where PRE_OP equals 0 in the pre- announcement period but 1 in the post-announcement period. The "Disagree / Agree" group contains the announcements where PRE_OP equals 1 in the pre- announcement period but 0 in the post-announcement period. The "Disagree / Disagree" group contains the announcements where the PRE_OPs equal 1 in both pre- and post-announcement periods. We also report the differences in CAV between the "Agree / Agree" group and the other three groups and the associated t-statistics. In addition, the CAV differences between groups AD, DA, DD and group AA are calculated using Welch's *t*-test. ***, **, * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Observations	CAV [0,1]	CAV [2,10]	CAV [0,10]
Panel A:cumulative abnormal volume (CAV) across Groups of Disagreement (%)				
PRE_A	634	1.244	1.956	3.200
PRE_D	623	1.824	3.640	5.464
PRE_D-PRE_A		0.580*	1.684*	2.264*
t-stat		1.858	1.575	1.653
Panel B:cumulative abnormal volume across Groups of AA, AD, DA, DD				
AA	383	0.935	1.933	2.868
AD	251	1.752	2.073	3.825
AD – AA		0.817**	0.140**	0.957**
t-stat		1.989	2.047	2.261
DA	246	1.313	2.120	3.433
DA - AA		0.378*	0.187**	0.565**
t-stat		1.801	2.062	2.156
DD	377	2.130	4.524	9.478
DD – AA		1.195**	2.591**	6.610**
t-stat		2.101	2.170	2.213

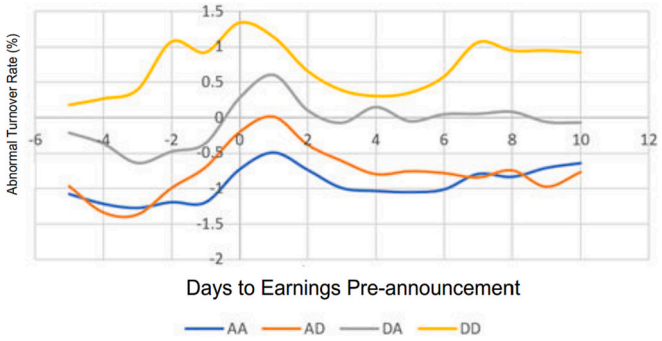

Fig. 3. Daily Abnormal Turnover of Four Groups with Different Opinion. This figure depicts the different abnormal turnover changes of groups AA, AD, DA, DD over the [−5,10] window. Day 0 is the announcement date. Abnormal turnover is obtained by subtracting expected turnover from realized turnover, where expected turnover is calculated by regressing market turnover on stock turnover over the [−245, −45] window.

while DA has a sharp trading volume increase during the announcement. Similarly, Panel B of Table 4 also shows abnormal trading volume data of the 4 groups over the [−5,10] window. Taking the differences of AD, DA, DD with AA, all differences are positive. This indicates all three situations can generate abnormal trading volume. Over the time dimension, the differences in cumulative abnormal trading volume between groups AD, DA, DD and AA are least significant over [2,10]. They are most significant over [0,1], the announcement date and the next date. This shows the effect of opinion changes on increased trading volume is short-term. Over the dimension of different opinion change situations, the significance of trading volume difference between DD and AA is greatest, followed by AD and DA is lowest. This indicates that during the announcement period, continuously heterogeneous opinions generate the highest trading volume, while opinion divergence contributes more to high trading volume than convergence. (See Fig. 5.)

Next, we use multivariate linear regression to examine the impact of heterogeneous opinions on cumulative abnormal trading volume. The independent variables use include a dummy variable for homogeneous/heterogeneous opinions before the announcement PRE_OP (heterogeneous = 1, homogeneous = 0), dummy variables for the DD, AD, DA types (1 for stocks of that type, else 0), and OP the sum of opinions over the 5 days before the announcement. Different windows around the announcement are used for the regressions. In addition, a new variable is constructed for the sentiment of the earnings guidance itself, with ANNOUNCE = 1 for predicted profit growth or profit, else 0. Table 5 shows the regression results on cumulative abnormal trading volume.

In Table 5, model (1) regresses cumulative abnormal trading volume on PRE_OP. It is found correlation with announcement heterogeneous opinions is statistically significant at 10% level, and opinion heterogeneity can bring about 0.5% higher turnover, comparable to the difference between heterogeneous and homogeneous groups in the previous table. In model (2), PRE_OP is replaced with dummy variables for AD, DA and DD. It can be seen their contributions to cumulative abnormal returns relative to group AA are also close the differences between groups in the previous table. Also consistent with Panel B of the previous table, DA's contribution to abnormal turnover is less than AD's, and DD's contribution is the largest and most significant. In model (3), PRE_OP is used together with PRE_OP_SUM and ANNOUNCE variables. It is found cumulative abnormal trading volume has significant positive correlation with PRE_OP_SUM, meaning higher bullish sentiment leads to higher abnormal trading volume. It is also positively correlated with ANNOUNCE, that predicted profit growth increases cumulative abnormal trading volume. In model (4) using AD, DA, DD variables together with PRE_OP_SUM and ANNOUNCE, PRE_OP_SUM retains the highest significance. Similarly, models (5) and (6) use the variables in (4) to regress CAV[2,10] and CAV[0,10]. Consistent with the previous table, the explanatory power of all variables weakens for CAV[2,10] compared to [0,1], but except for AD, others remain positively correlated with the dependent variable, and PRE_OP_SUM retains the highest significance. This reflects that not only announcement heterogeneous/homogeneous opinions, but also overall opinion polarity contributes greatly to trading volume, with bullish opinions promoting and bearish opinions inhibiting trading volume. The intensity of bullish/bearish sentiment is also reflected in trading volume.

In summary, this section explores the rationality of Hypotheses 1 and 2. First, the author divides the sample into a homogeneous opinion group before the announcement and a heterogeneous opinion group, examining their abnormal trading volumes and cumulative abnormal trading volumes during the announcement period. From graphs and tables over the [−5,10] window, it can be seen that both groups experience increased trading volume on the announcement day, but the heterogeneous group's abnormal trading volume remained above the homogeneous group. According to the table, the heterogeneous group has on average 0.58% higher abnormal trading volume over [0,1] compared to the homogeneous group. This proves Hypothesis 1 that announcement heterogeneous opinions lead to higher trading volume during the announcement. Next, this paper divides the announcement into groups transitioning from homogeneous to heterogeneous, from heterogeneous to homogeneous, remaining homogeneous and remaining heterogeneous, making graphs and tables of cumulative abnormal

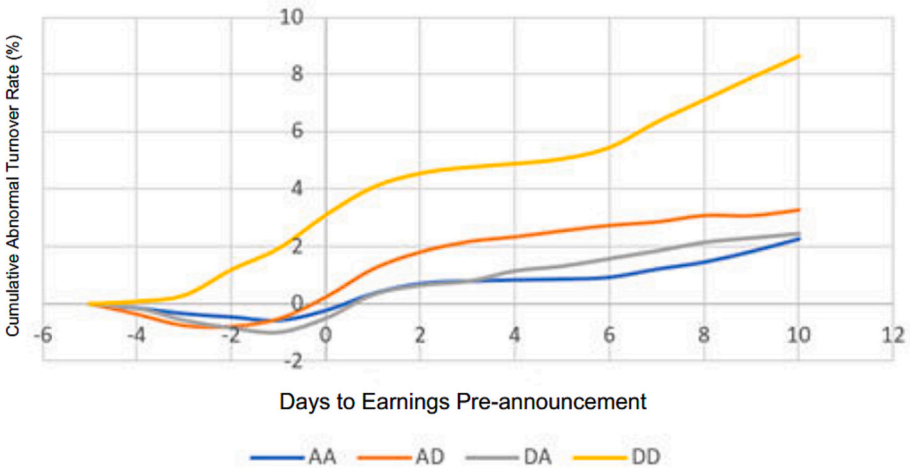

Fig. 4. Changes in Cumulative Abnormal Trading Volume of Groups with Different Opinion. This figure depicts the cumulative abnormal turnover changes of groups AA, AD, DA, DD over the [−5,10] window around announcements, where day 0 is the announcement date. Abnormal turnover is obtained by subtracting expected turnover from realized turnover, where expected turnover is calculated by regressing market turnover on stock turnover over the [−245, −45] window. Cumulative abnormal turnover sums the daily abnormal turnovers. For ease of observation, day −5 abnormal turnover is set as 0.

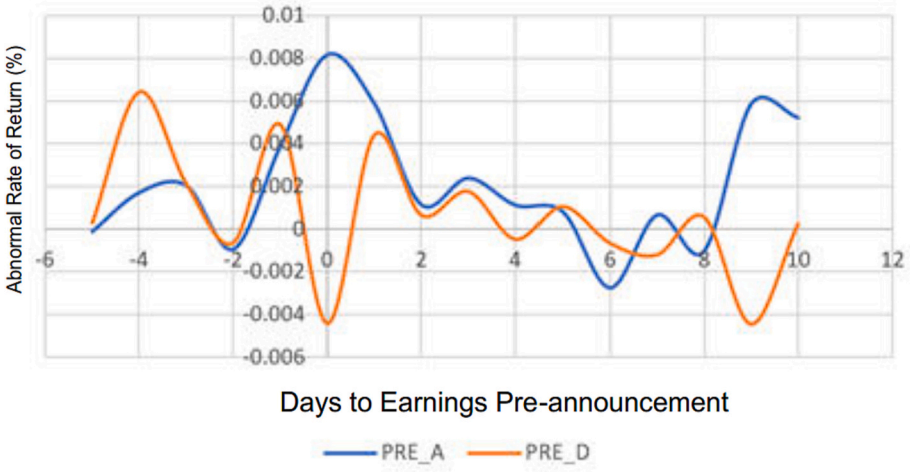

Fig. 5. Daily Abnormal Returns of Homogeneous and Heterogeneous Groups during Pre-earnings announcement Period. This figure depicts the daily average return changes of the group with homogeneous opinions over the 5 days before pre-announcement and the group with heterogeneous opinions over [−5,10]. Day 0 indicates the pre-announcement date. Returns here are the difference between predicted returns from the Carhart four factor model and realized returns.

turnover over [−5,10]. Comparing the homogeneous group with the group transitioning from homogeneous to heterogeneous shows the latter's cumulative abnormal turnover exceeded and remained above the former's on the announcement date, consistent with the hypothesis that opinion divergence leads to abnormal trading volume. Comparing heterogeneous and heterogeneous-to-homogeneous groups, the heterogeneous group has higher abnormal trading volume, but the heterogeneous-to-homogeneous opinion group exhibits a steeper increase in abnormal trading volume on the announcement date, and is higher than the stable homogeneous group. This also proves the hypothesis that opinion convergence leads to higher trading volume. To further examine these hypotheses and conclusions, we convert announcement homogeneous/heterogeneous opinions and the four opinion change situations into dummy variables, and adds the announcement opinion index sum and the sentiment orientation of the guidance itself to regress different stages of cumulative abnormal returns. These regressions validate the differences in abnormal turnover between groups found earlier through summary statistics, and also discover significant positive correlation between the announcement opinion index sum, as well as increased abnormal turnover when guidance predicts profit growth. These variables contribute more to explaining abnormal turnover over [0,1] compared to [2,10].

5.3. Disagreement in opinions and stock returns

In Hypothesis 3, based on Miller (1977)'s theory, we infer that if investors hold divergent opinions before pre-earnings announcement, due to short-selling constraints, stock prices will be set solely by optimists and be overvalued. After the announcement, updated information brings in a fuller set of investors, which will lead stock prices to drop. Therefore prior to the announcement heterogeneous opinions lead to lower returns during the announcement period. We test whether the abnormal returns of homogeneous and heterogeneous groups during the pre-earnings announcement fit this hypothesis.

First, abnormal returns around announcement are measured as follows: Over the 90-day window [−120,-31] before earnings announcement, the Carhart four factor model is constructed.

$$R_{expect,it} = \widehat{\alpha_i} + \widehat{\beta_{1i}}\left(R_{mt} - r_f\right) + \widehat{\beta_{2i}}SMB_t + \widehat{\beta_{3i}}HML_t + \widehat{\beta_{4i}}UMD_t \quad (2)$$

α_i, β_{1i} are coefficients obtained by regression over the window of

Table 5
Regression on Cumulative Abnormal Turnover. This table shows regressions on cumulative abnormal turnover over different pre-earnings announcement windows using prior homogeneous/heterogeneous opinion measures, opinion change dynamics during announcements, sum of opinion sentiment before announcements, and positivity/negativity of the announcement itself. PRE_OP: Dummy variable indicating homogeneous/heterogeneous opinions 5 days before announcement date, heterogeneous = 1. AD: Dummy variable, stocks with homogeneous opinions before and heterogeneous opinions after guidance are coded 1, others 0. DA: Dummy variable, stocks with heterogeneous opinions before and homogeneous opinions after are coded 1, others 0. DD: Dummy variable, stocks with heterogeneous opinions before and after are coded 1, others 0. PRE_OP_SUM: Summation of opinion sentiment over 5 days before announcements. ANNOUNCE: Positivity/negativity of the announcement, profit growth/profit with coded 1, profit decline/loss with coded 0. Variables are elaborated in Appendix A.

	Dependent variable					
	CAV [0,1]				CAV [2,10]	CAV [0,10]
Indep. Var.	(1)	(2)	(3)	(4)	(5)	(6)
PRE_OP	0.522*		0.329			
	(1.79)		(0.51)			
AD		0.872*		0.707	−0.071	0.636
		(1.92)		(0.77)	(−0.02)	(0.13)
DA		0.488		0.440	0.342	0.781
		(0.50)		(0.46)	(0.08)	(0.16)
DD		1.058**		0.690	0.146	0.836
		(1.97)		(0.85)	(0.04)	(0.20)
PRE_OP_SUM			0.054***	0.053***	0.139**	0.192***
			(3.97)	(3.89)	(2.31)	(2.70)
ANNOUNCE			0.511	0.484	1.432	1.916
			(0.72)	(0.68)	(0.46)	(0.52)
Pseudo-R^2	0.022	0.141	0.143	0.214	0.314	0.329

[−120,-31] , $R_{mt} - r_f$ is the market factor, R_{mt} the market return at time t, r_f is the risk free rate; SMB_t is the size factor (small minus big), the return on the small market cap portfolio minus that on the big cap portfolio at time t; HML is the value factor (high minus low), the return on the low PB portfolio minus that on the high PB portfolio at time t; UMD is the momentum factor, the return on past winner portfolio minus past loser portfolio at time t. Then the abnormal return of stock i at time t, $AR_{it} = R_{it} - R_{expect.it}$ is obtained.

According to the above formula, the abnormal returns AR over [−5,10] for each stock can be obtained. The two groups with homogeneous and heterogeneous opinions before pre-earnings announcement are examined. As shown in Fig. 6, the dynamic patterns in abnormal returns of the two groups are roughly similar, but very different on the announcement date. The heterogeneous group's returns becomes negative while the homogeneous group reaches 0.8%. The two groups had similar returns for a week after, then another large difference emerged on the 9th day, again with the homogeneous group ahead. It can be inferred that for the heterogeneous group, some investors change their opinions after the announcment and participate in price-setting, leading to the originally overvalued stock price dropping, hence negative returns. For the homogeneous group, the opposite may have happened, where pessimists leaving led to overvaluation and return increases.

In Fig. 6, we plot the dynamic variation in cumulative abnormal returns, which shows the differences between the two groups more clearly. Before the announcement, the heterogeneous opinion group had higher cumulative abnormal returns. But from the announcement date, the cumulative abnormal returns of the homogeneous group kept rising while the heterogeneous group started dropping and meandering, keeping the returns of the homogeneous group above. By the 8th day after the announcement, the returns of the homogeneous group began rising rapidly while the heterogeneous group experienced a decline. Therefore, we find the supporting evidence to Hypothesis 3.

As abnormal returns changed notably over the few days around the announcement, Panel A of Table 6 shows the cumulative abnormal returns over [0,1] and [−1,1]. It can be seen that the homogeneous group has higher cumulative abnormal returns over both windows, more significantly in [0,1], where the difference between groups is also more

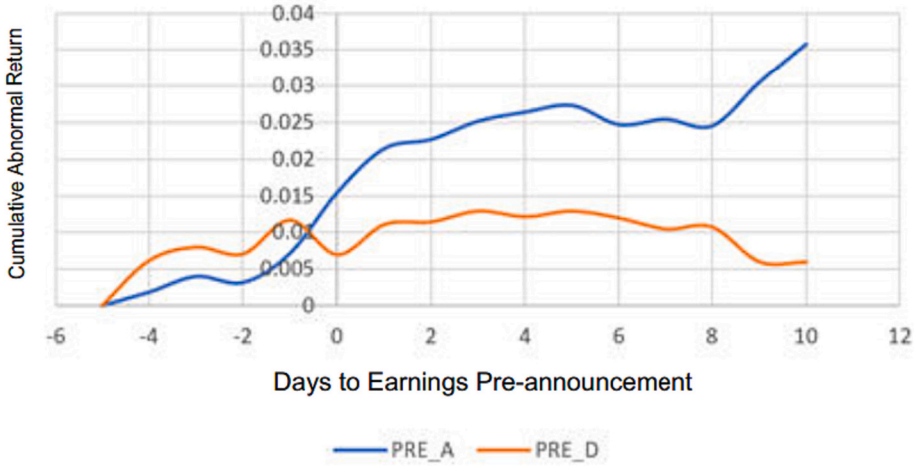

Fig. 6. Cumulative Abnormal Returns of Homogeneous and Heterogeneous Groups during announcement Period. This figure depicts the cumulative average return changes of the group with homogeneous opinions over the 5 days before announcement and the group with heterogeneous opinions over [−5,10]. Day 0 indicates the announcement date. Returns here are the difference between predicted returns from the Carhart four factor model and realized returns, accumulated from 5 days before announcment. For ease of observation, abnormal returns on day −5 before announcement is set as 0.

Table 6
Earnings announcement returns across disagreement groups. Panel A presents cumulative abnormal trading returns (CARs) around pre-earnings announcements across the groups of investor disagreement. We first sort announcements into two groups based on the value of the PRE_OP measure. We then report the averages of the cumulative abnormal returns in the [0,1], [2,10], and [0,10] windows for the two groups, where day 0 is the earnings announcement day. For an announcement, we use the methodology of Campbell and Wasley (1996) to daily abnormal return where the coefficients are estimated in the 200-day window ending 45 days prior to the earnings announcement [−245,-45]. "Full Sample" includes all earnings announcements in our sample. The low institutional ownership subsample and the high institutional ownership subsample are classified according to institutional ownerships of the announcing firms at the end of the quarter prior to the earnings announcement. Panel B presents CARs for the four groups of the changes in investor agreement. We calculate the PRE_OP measures after earnings announcements. The "Agree / Agree" group contains the announcements where the PRE_OPs equal 0 in both pre- and post-announcement periods. The "Agree / Disagree" group contains the announcements where PRE_OP equals 0 in the pre- announcement period but 1 in the post-announcement period. The "Disagree / Agree" group contains the announcements where PRE_OP equals 1 in the pre- announcement period but 0 in the post-announcement period. The "Disagree / Disagree" group contains the announcements where the PRE_OPs equal 1 in both pre- and post-announcement periods. We also report the differences in CAV between the "Agree / Agree" group and the other three groups and the associated t-statistics. ***, **, * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	observations	CAR[0,1]	CAR[−1,1]
Panel A: Cumulative Abnormal Returns (CAR) across Groups of Prior Disagreement (%)			
Full Sample			
PRE_A	649	0.014	0.018
PRE_D	647	−0.0006	0.004
PRE_D-PRE_A		−0.0146**	−0.014**
t-stat		−1.960	−1.967
High Institutional Ownership			
PRE_A	268	0.012	0.019
PRE_D	381	−0.006	−0.000
PRE_D-PRE_A		−0.018**	−0.019**
t-stat		−2.1510	−2.1200
Low Institutional Ownership			
PRE_A	381	0.016	0.018
PRE_D	266	0.006	0.009
PRE_D-PRE_A		−0.010*	−0.009
t-stat		−1.693	−0.444
Panel B:cumulative abnormal return across Groups of AA, AD, DA, DD			
AD	258	−0.004	0.0003
AA	391	0.026	0.030
DD	391	0.005	0.013
DA	256	−0.009	−0.010
DA-AD		−0.005*	−0.0103**
t-stat		−1.933	−2.022

significant. This further shows that for the group of the heterogeneous opinions, after pre-earnings announcements, some other investors quickly participate in stock market, leading prices to drop, while pessimists left the homogeneous group market, allowing prices to be set by optimists and rise, increasing returns. This is consistent with Hypothesis 3.

Miller's theory assumes the existence of short-sale constraints, which prevent pessimists from participating in the market, leading to overvaluation. Therefore, higher short-sale constraints should make the return difference between homogeneous and heterogeneous groups more significant. In light of Nagel (2005), institutional ownership is used to measure the degree of short-sale constraints. Lower institutional ownership implies higher short-sale constraints. Therefore, in Panel A of Table 6, this paper calculates CAR for different periods separately for subsamples with high and low institutional ownership. It is found that the return difference has higher significance for the high institutional ownership sample for the Chinese stock market, which is contrary to Miller's theory. There may be other effects overriding the short-sale constraint effect. For example, companies with high institutional ownership are likely to have a large number of institutional investors, each with proprietary information, easily leading to a diversity of opinions in the market, i.e. high extent of heterogeneous opinions.[3] This extremely high opinion heterogeneity acts as a risk factor, biasing prices higher. When the announcement reduces the extent of heterogeneous opinions, prices drop more dramatically, causing lower returns.

After analyzing the homogeneous and heterogeneous groups before pre-earnings announcement, we divide the sample into groups AD, AA, DD, DA by opinion dynamics, and plots the abnormal returns respectively in Fig. 7. It can be seen that overall the patterns are similar, but only group AA had return increases on the announcement date, while other groups' returns dropped. In addition, group AA remained almost at the top afterwards, followed by DD, AD and DA. After announcement, the two groups with opinion convergence, AA and DA, behaved similarly, while DD and AD behaved differently. According to Miller (1977)'s theory, opinion convergence leads to trading volume. For group DA, returns did significantly drop on the announcement date. For group AD, there was a decreasing trend for a period after announcement. But this effect diminished quickly, with AD's returns exceeding DA's from the 5th day onwards. From the cumulative perspective in Fig. 8, the returns of the four groups from highest to lowest were: AA, DD, AD, DA, with the first two positive and the latter two negative. Overall, Miller's theory manifested to some extent, validating Hypothesis 4, since group DA experienced a large drop on the announcement date, meaning opinion convergence leads to lower returns. But the duration was still short, and the effect was not obvious in cumulative terms, with groups AA and DD remaining higher. As shown in Panel B of Table 6, which displays the performance of the four groups in different announcement scenarios, it can be observed that group AA performed the best, with the highest cumulative returns over both [0,1] and [−1,1] windows. DA-AD were negative, meaning the convergence groups had lower returns than the divergence groups, consistent with Miller (1977)'s theory.

Similarly, we regress CAR over window [0,1] on the heterogeneous opinion metrics constructed in this paper and those borrowed from the previous literature, as shown in Table 7. It can be seen that the variable PRE_OP constructed in this paper depicting announcement homogeneous/heterogeneous opinions has strong explanatory power on CAR [0,1]. Announcement heterogeneous opinions lead to a 1.6% decrease in cumulative returns on the announcement date and the next date. PRE_OP_SUM and ANNOUNCE are insignificant in this regression, suggesting returns around guidance are not sensitive to intensity of positive/negative sentiment or positive/negative sentiment of the guidance itself. But their positive coefficients also suggest positive sentiment and profit guidance may promote return increases. Among the existing heterogeneous opinion metrics from the literature, return volatility and historical turnover are more significant, and are significantly negatively correlated, consistent with the heterogeneous opinion metrics constructed in this paper. That is, heterogeneous opinions lead to lower returns around announcements. For example, every 1% increase in historical turnover is associated with a 0.4% decrease in CAR[0,1]. In comparison, the explanatory powers of profit volatility and company age on CAR[0,1] are weaker.

In this section we explore the impact of heterogeneous opinions and opinion changes on returns during pre-earnings announcement periods. First, we construct a Carhart four-factor model over the guidance period [−120,−31] and used it to predict and difference returns over [−5,10] to obtain abnormal returns as the dependent variable. The sample was divided into homogeneous and heterogeneous groups before guidance, graphs and tables are constructed to test the significance of Miller (1977)'s view. It is found that the homogeneous group had significantly

[3] This can be seen from the larger number of heterogeneous companies than homogeneous

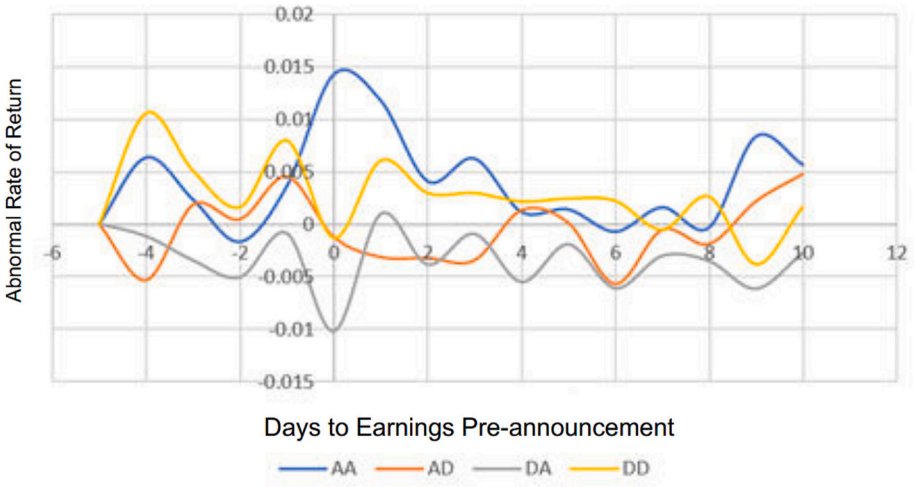

Fig. 7. Abnormal Returns of Groups with Different Opinion during announcement Period. This figure depicts the abnormal returns of groups AA, AD, DA, DD over the [−5,10] window. Day 0 is the announcement date. Abnormal returns are calculated by subtracting predicted returns from realized returns. Predicted returns are from the Carhart four factor model using [−120,-31] window data.

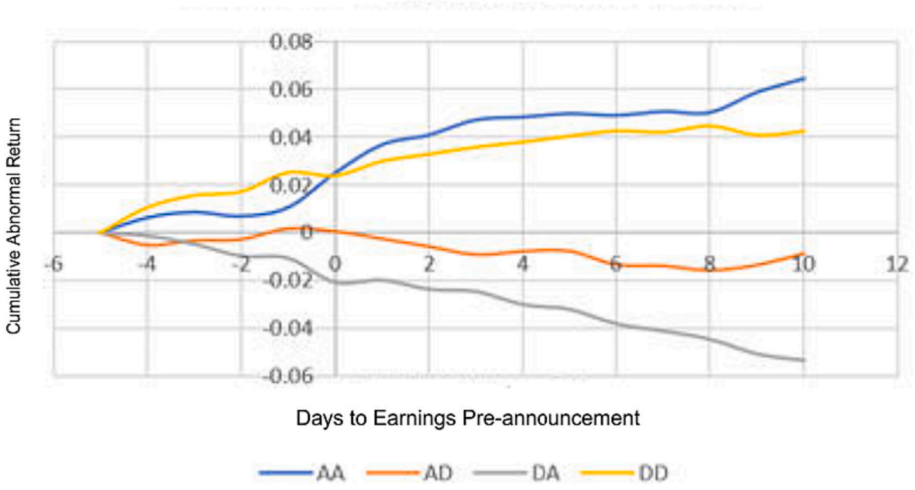

Fig. 8. Cumulative Abnormal Returns of Groups with Different Opinion during announcement Period. This figure depicts the cumulative abnormal returns of groups AA, AD, DA, DD over the [−5,10] window. Day 0 is the announcement date. Abnormal returns are calculated by subtracting predicted returns from realized returns and accumulating over time. Predicted returns are from the Carhart four factor model using [−120,-31] window data. For ease of observation, returns on day −5 are set as 0.

higher returns than the heterogeneous group after announcment, proving the assumption that stocks with heterogeneous opinions before announcement were priced solely by optimists and thus overvalued. In addition, since Miller (1977)'s theory is based on the assumption of short-sale constraints, and a stock's institutional ownership level is often used as a measure of short-sale constraints, with lower institutional ownership implying higher short-sale constraints, this section also divided the full sample into subsamples with high and low institutional ownership levels and tested the homogeneous and heterogeneous groups again. The result is that the difference between the two groups is more pronounced for higher institutional ownership, contrary to expectations. We argue that this may be because the high institutional ownership subsample followed Varian (1985)'s view that heterogeneous opinions are a risk factor when short-sale constraints are low. When guidance reduces the degree of heterogeneity, prices drop. Stocks with high institutional ownership are likely to have a high degree of heterogeneous opinions before guidance due to the large amount of proprietary information and opinions. Therefore, the effect is more pronounced.

We further divide the sample into four groups: homogeneous both before and after pre-earnings announcement, heterogeneous before and after, transition from homogeneous to heterogeneous, and transition from heterogeneous to homogeneous. [−5,10] window abnormal return charts are plotted and cumulative returns over [0,1] and [−1,1] windows for each group tested to scrutinize Miller (1977)'s theory in more detail. It is found that the cumulative returns of the homogeneous to heterogeneous transition group are indeed greater than the heterogeneous to homogeneous group, indicating a process of pessimists leaving and re-entering the market. Finally, we regress the [0,1] window cumulative returns against the heterogeneous opinion variables constructed in this paper and those from the literature. It is found the homogeneous/heterogeneous group variable is very significant, again proving the validity of the variable constructed in this paper and that social media can be used to uncover information. Also, the stock return volatility and historical turnover from the literature are highly significant, and likewise contributed negatively to returns as do heterogeneous opinions.

Table 7
Regressions of earnings announcement returns. This table presents the regressions of earnings announcement returns on the investor disagreement measures. The dependent variables are the two-day cumulative abnormal returns in the [0,1] window with respect to earnings announcement, where day 0 is the announcement day. For an earnings announcement, we estimate daily abnormal returns using the four-factor model where the coefficients are estimated in the 90-day window [−120,-31] ending 31 days prior to the earnings announcement. We include PRE_OP as the key explanatory variable. We also control for PRE_OP_SUM, which measures the sum of the changes of opinions in two weeks prior to pre-earnings announcement; ANNOUNCEMENT, which measures the direction (positive/negative) of pre-earnings announcement. We also include disagreement proxies in Berkman et al. (2009), including INCVOL, RETVOL, LN(1/AGE) and TURN. INCVOL is the standard deviation of the quarterly operating income. RETVOL is the standard deviation of the announcing firm's daily excess returns. AGE is the number of years the announcing firm has been covered by CSMAR. TURN is average daily turnover measures over a 45-day period ending 10 days before the earnings announcement day. We report in parentheses the Driscoll and Aart (1998) robust t-statistics, which control for time series and cross-sectional correlation.

Dependent Variable: CAR[0,1]						
	(1)	(2)	(3)	(4)	(5)	(6)
PRE_OP	−0.016**	−0.016**	−0.016**	−0.013**	−0.016**	−0.014**
	(−2.47)	(−2.52)	(−2.48)	(−2.03)	(−2.50)	(−2.19)
PRE_OP_SUM		0.000	0.000	0.000	0.000	0.000
		(0.65)	(0.65)	(1.44)	(0.60)	(1.20)
ANNOUNCE		0.003	0.003	0.005	0.003	0.005
		(0.46)	(0.41)	(0.72)	(0.48)	(0.66)
INCVOL			0.000			
			(−0.45)			
RETVOL				−0.947***		
				(−3.43)		
LN(1/AGE)					−0.001	
					(−0.17)	
TURN						−0.004**
						(−2.13)
Pseudo-R^2	0.021	0.021	0.043	0.014	0.014	0.029

6. Conclusions

In this paper, we validat the relationship between heterogeneous opinions and abnormal trading volume and returns during pre-earnings announcement periods by directly obtaining individual investor opinions from social media. First, posts in individual stock forums during announcement are obtained, and machine learning methods are used to classify the sentiment of these posts. The standard deviation of this sentiment index is used as the basis for dividing homogeneous and heterogeneous opinions. It is found this directly obtained heterogeneous opinion metric is highly correlated with many metrics from the literature. Next, according to Kim and Verrecchia (1991a&b), Kim and Verrecchia (1994), announcement heterogeneous opinions, divergence of opinions after announcement, and convergence of opinions after announcement all lead to abnormal trading volume. Therefore, we divide the sample into subsamples by announcement opinion heterogeneity and opinion dynamics during the announcement period. The results are consistent with the hypotheses, namely that heterogeneous opinions, opinion convergence and opinion divergence can all lead to increased trading volume. Finally, according to Miller (1977)'s theory, under short-sale constraints heterogeneous opinions cause pessimists to exit the market leading to overvaluation. Pre-earnings announcements may cause pessimists to re-enter the market, lowering prices and returns. We also divide the sample into subsamples by announcement opinion heterogeneity and opinion dynamics. It is found the homogeneous opinion group did have higher returns after announcement, and the convergence group has lower returns than the divergence group. But in the discussion of short-sale constraints, it is found the proxy variable institutional ownership is not consistent with Miller's view. It is conjectured institutional ownership may greatly affect the degree of heterogeneous opinions, enhancing the Varian (1985) view that heterogeneous opinions are another risk factor, resulting in the phenomenon of lower returns for the heterogeneous group under low short-sale constraints.

The largely literature-consistent results indicate social media can uncover useful information for investors. In terms of future research, further expanding the investigation of trading volume and return differences between groups over longer periods after the announcement may yield additional meaningful discoveries. In addition, there are many machine learning methods for text sentiment classification, for example, SVM, neural networks and etc. Using other methods may also help prove the robustness of our results.

Data availability

Data will be made available on request.

Appendix A. Definition of variables of interest

1. Log(1/AGE): take the inverse of company age, then the natural logarithm.
2. TURN, turnover: is average daily turnover from 45 days to 11 days before the announcement.
3. Earnings Persistence, earnings stickiness: the coefficient of the lagged year variable when regressing operating profits of the past 8 years with each lagged by 1 year.
4. Ln(ME): take natural logarithm of the market cap of the company last year.
5. Ln(B/M): take inverse then inverse of market-to-book ratio.
6. Ret[−12,-2]: stock returns from 12 months to 2 months before the announcement month.
7. MATO (market-adjusted turnover): is the average daily abnormal turnover in the [0,1] window.
8. ΔTO: is calculated as MATO minus the announcing firm's average daily abnormal turnover on the window [−45,-5] before the announcement.
9. SUV (standard unexpected volume), $SUV = \frac{1}{\sqrt{2}}\sum_{t=0}^{t=1}\frac{UV_{it}}{\sigma_{it}}$, take the average after comparing the difference between actual and predicted trading volume with standard deviation.
10. RETVOL, return volatility: is the standard deviation of the announcing firm's daily excess returns.

11. Ln(1 + #inst.), is natural log of 1 plus the number of institutional shareholders of the announcing firm.
12. |Accrual| is absolute value of the announcing firm's total accrual of the year before the earnings announcement.
13. ΔTURN: the difference between daily average turnover over windows [−25,-11] and [10,25].
14. ΔRETVOL: change in return volatility over the pre-earnings announcement windows [−25,-11] and [10,25].
15. AD: a dummy variable, stocks with homogeneous opinions 5 days before and heterogeneous opinions 5 days after pre-earnings announcement are coded 1 and 0, respectively.
16. DA: a dummy variable, stocks with heterogeneous opinions 5 days before the announcement, 1 and homogeneous opinions 5 days after, 0.
17. DD: a dummy variable, stocks with heterogeneous opinions 5 days both before and after the announcement, 1 and others 0.
18. PRE_OP_SUM: summation of opinion sentiment values over the 5 days before guidance.
19. ANNOUNCE: Positivity/negativity dummy variable for the pre-earnings announcement itself, profit growth/profit outcome coded 1, profit decline/loss guidance coded 0.
20. R_{mt}: market return at period *t*.
21. r_f: risk-free rate at period *t*.
22. SMB_t (small minus big): size factor, return on the small market cap portfolio minus return on the big market cap portfolio at time t, rebalanced monthly, with the small and big portfolios as the bottom and top 10% by market cap ranking of A-shares last month.
23. HML_t (high minus low): value factor, return on the low PB portfolio minus return on the high PB portfolio at time t, rebalanced monthly, with the low and high PB portfolios as the bottom and top 10% by PB ranking of A-shares last month.
24. UMD_t (up minus down): momentum factor, return on past winner portfolio minus return on past loser portfolio at time t, rebalanced monthly, with winner and loser portfolios as bottom and top 10% by return ranking of A-shares last month.

References

Atiase, Rowland K., & Bamber, Linda Smith (1994). Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry. *Journal of Accounting and Economics, 17*(3), 309–329. https://doi.org/10.1016/0165-4101(94)90031-0. ISSN 0165-4101.

Bamber, L., Barron, O., & Stober, T. (1999). Differential interpretations and trading volume. *Journal of Financial and Quantitative Analysis, 34*(3), 369–386.

Bamber, L. S., Barron, O. E., & Stevens, D. E. (2011). *Trading Volume Around Earnings Announcements and Other Financial Reports: Theory, Research Design, Empirical Evidence, and Directions for Future Research. Contemporary Accounting Research, 28*, 431–471. https://doi.org/10.1111/j.1911-3846.2010.01061.x

Banerjee, S., & Kremer, I. (2010). Disagreement and learning: Dynamic patterns of trade. *The Journal of Finance, 65*(4), 1269–1302.

Berkman, H., Dimitrov, V., Jain, P., Koch, P., & Tice, S. (2009). Sell on the news: Differences of opinion, short-sales constraints and returns around earnings announcements. *Journal of Financial Economics, 92*, 376–389.

Bamber, L., Barron, O., & Stober, T. (1997). Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review, 72*(4):), 575–597.

Bollen, Johan, Mao, Huina, & Zeng, Xiaojun (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007. ISSN 1877-7503.

Campbell, C. J., & Wasley, C. E. (1996). Measuring abnormal daily trading volume for samples of NYSE/ASE and NASDAQ securities using parametric and nonparametric test statistics. *Rev Quant Finan Acc, 6*, 309–326. https://doi.org/10.1007/BF00245187

Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies, 27*, 1367–1403.

Driscoll, John C., & Aart, C. Kraay. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *The Review of Economics and Statistics, 80*(4), 549–560. http://www.jstor.org/stable/2646837.

Giannini, Robert, Irvine, Paul, & Shu, Tao (2019). The convergence and divergence of investors' opinions around earnings news: Evidence from a social network. *Journal of Financial Markets, Volume, 42*, 94–120. https://doi.org/10.1016/j.finmar.2018.12.003

Kandel, E., & Pearson, N. (1995). Differential interpretations or public signals and trade in speculative markets. *Journal of Political Economy, 103*(4), 831–872.

Kim, O., & Verrecchia, R. (1991a). Trading volume and price reactions to public announcements. *Journal of Accounting Research, 29*(2), 302–321.

Kim, O., & Verrecchia, R. (1991b). Market reaction to anticipated announcements. *Journal of Financial Economics, 30*(2), 273–309.

Kim, O., & Verrecchia, R. (1994). Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics, 17*, 41–67.

Miller, E. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance, 32*, 1151–1168.

Nagel, S. (2005). Short sales, institutional investors and the cross-section of stock returns. *Journal of Financial Economics, 78*, 277–309.

Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139–1168.

Tumarkin, Robert, & Robert, F. Whitelaw (2001). News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal, 57*(3), 41–51. http://www.jstor.org/stable/4480315.

Varian, H. (1985). Divergence of opinion in complete markets: A note. *The Journal of Finance, 40*, 309–317.