

Do Chinese Social Media Delineate the Optimistic Bias of Traditional Media?

Eric Wang
The Chinese University of Hong Kong, Shenzhen

T.J. Wong
University of Southern California

Tianyu Zhang
The Chinese University of Hong Kong

May 2019

Abstract

This paper examines whether social media delineate the bias of state-controlled traditional media by supplying less optimistically biased information to the market. Using a comprehensive sample of corporate news articles of Chinese newspapers and posts of an online stock forum, East Guba, from 2009 to 2016, we find that East Guba's tone is less positively associated with that of the newspapers for the same firm on the same day when the newspapers are expected to be *more* optimistically biased. This decline in association in tone is significantly larger since the 2015 political shock which suppressed the reporting of negative corporate news by traditional media. Finally, when the tone of the newspapers deviates positively from that of East Guba, the newspaper articles are perceived by the market to be less credible as reflected in their significantly attenuated stock return response.

We appreciate comments from conference participants of "Text Analysis for Asia and Beyond" Conference at USC, and participants of CUHK-Shenzhen SME research workshop.

1. Introduction

Traditional media in autocratic regimes are optimistically biased. The autocrats control the media and bias the news they report as a way to strengthen political power (Lippman, 1953; Noelle-Neumann, 1984; Djankov et al., 2003; Enikolopov et al., 2011). This optimistic bias in reporting is found in corporate news as well since negative news can potentially destabilize the economy and weaken the perceived competency of the government (Stockmann, 2013; Piotroski et al., 2017). However, in these autocratic regimes, many newly developed social media such as online stock forums have started to supply information to the markets. In contrast with traditional media that typically have state-controlled editorial teams, social media are typically private and they rely on wisdom of crowds, with user generated content and user anonymity, to supply information. Users that contribute corporate information to these online platforms are expected to be independent, as they are more likely to be free from direct censorship by the state (Foucault, 1977; Spears and Lea, 1994).

The objective of the paper is to examine if Chinese social media delineate the optimistic bias of traditional media. We first examine if the tone of social media relative to that of traditional media can reflect traditional media's bias when such bias increases. To test this, we regress the tone of social media on the tone of traditional media. The coefficient on the tone of traditional media will capture the relative tone of the two types of media for the same firm on the same day. We posit that when traditional media are expected to be more optimistically biased, the tone of social media will be less positively associated with that of traditional media, as reflected by a significant decrease in the coefficient on the tone of traditional media. This decline in positive association in tone of the two types of media when traditional media's bias increases indicates that social media can help to delineate the bias.

We use whether the tone of traditional media is positive or negative to proxy for their optimistic bias. The media's tone is expected to be *more* optimistically biased when it is positive than when it is negative. We use this proxy for two reasons. First, the overall tone of the articles is an aggregate of the tone of the underlying events and the bias in the reporting of the events. We expect that through a mechanical relation, the articles are likely to be more optimistically biased when their tone is positive than negative. Second, as the underlying events are positive, we expect optimistic bias will increase because the cost to bias the tone positively will decrease.¹ This further enhances the likelihood that the positive tone of the articles can reflect optimistic bias.

Our second analysis examines whether traditional media's bias, as delineated by social media, is perceived by the market to be biased and therefore less useful in trading decisions. We use the difference in tone between traditional and social media for the same firm on the same day to directly capture the traditional media's positive bias. We expect that traditional media's greater positive tone relative to that of social media is an indication that traditional media have a larger optimistic bias. To test whether the difference in tone can indeed capture the optimistic bias, we use the stock price response to the tone of traditional media as a gauge for market's perception of the tone bias. We posit that when traditional media are more optimistically biased, the stock return response to the tone of traditional media will significantly decrease, reflecting that the tone is perceived to be more biased.

This second analysis complements the test of association in tone of the two types of media in two ways. First, the decline in association in tone in the first set of analyses can also be consistent with the interpretation that social media are noisy or reporting on topics different from

¹ A more detailed discussion of why the tone of traditional media is positively associated with the tone bias is presented in subsection 2.3.

those of traditional media when traditional media are more positively biased. By directly measuring whether the tone of traditional media is more or less positive than that of social media, this second analysis provides a directional test on whether social media can actually delineate traditional media's bias that is actually *positive*. Second, this analysis uses the stock prices in the market as an external validation of social media's ability to delineate the bias.

We expect that social media in an autocratic regime can serve this benchmarking role in delineating the bias of traditional media. When traditional media are controlled by the state and the public lack a valid channel to express their opinions, the information disseminated through traditional media will serve as the public opinion, giving rise to the spiral of silence of the majority (Noelle-Neumann, 1984). Computer-mediated communication, such as social media, allows users to express their true mind or authentic selves, feeling less inclined to give in to groups, display slavish conformity and defer to powerful others (Spears and Lea, 1994). The contributors of analyses and opinions to these online stock forums are unlikely to be influenced or controlled by the state as traditional media. Consequently, the business model of these online platforms relies on wisdom of the crowds to provide valuable information to the community (Chen et al., 2014; Bartov et al., 2018). It is easier for the regime to influence the editors and a small set of journalists in traditional media, but it is much harder to control the crowd on social media especially when the contributors remain anonymous. Thus, as social media users post their opinions online, they are providing an independent interpretation of the news that will serve as a check against the bias of traditional media. Though social media sites can be censored by the autocratic regime, the government tends not to police the social media as much because they regard it as a source of grass root information (Qin et al., 2017) and it is unlikely that they find

stock opinions threatening since they are mainly concerned about instigation of collective actions (King et al., 2013, 2014).

On the other hand, social media may not delineate the bias in traditional media. As argued by Spears and Lea (1994), computer-mediated communication may give rise to uninhibited behavior and flaming, and polarized, risky and extreme decisions. Thus, social media posts may simply be noise and have no significant information content, despite they are relatively free from positive bias. As a result, the tone of social media is not positively associated with that of traditional media even when the latter are not biased. Comparing the tone of the two types of media does not allow us to identify the positive bias of traditional media. It is also possible that social media sites are subject to heavy censorship. The government may delete highly critical posts from the sites. Likewise, social media contributors may refrain from delineating the traditional media due to self-censorship. This could be particularly true in politically sensitive periods and after the 2015 stock market intervention, during which the government is concerned about stability and is extremely sensitive to negative news.²

China offers a great setting to study whether the wisdom of crowds of social media can delineate the optimistic bias of traditional media in an autocratic regime. Although the equity market was only established in the early 1990s, it has grown to have more than 3,000 firms listed in its two domestic stock exchanges. The newspaper coverage of these listed firms is vast, with more than 162 state-owned traditional newspapers spanning across most of the provinces in the country in our sample from the Wisenews database. The rich dataset offers us a comprehensive sample to study the reporting bias of corporate news by traditional media. For social media of stock opinions and analyses, we use East Guba internet platform, which is an online stock forum

² The government intervened the 2015 market crash in China by requesting state-owned brokerage firms and fund companies to prop up the market (http://www.xinhuanet.com/fortune/2015-08/06/c_128100347.htm).

allowing contributors to post their analyses and opinions of stocks and share their views with other users. During our sample period from 2009 to 2016, users of East Guba have posted about 146 million posts (about 83,000 posts per trading day), covering over 3,000 listed firms in China. Our final sample comprises more than 970 thousand firm-days with at least one news article from traditional newspapers and three posts from East Guba for each firm on the same day.

Following the machine learning method in Piotroski et al. (2017), we compute the tone of each traditional media article and social media post. We then aggregate the tone of each type of media for each firm on each day for comparison. Our evidence shows that the tone of traditional media is generally positive (77.9% positive) with a mean of 0.3704, while the tone of social media is negative (13% positive) with a mean of -0.2028. The observed tone difference confirms prior research that traditional media are likely to be positively biased, while social media are less positively biased.

To examine whether social media provide less optimistically biased information to the market than traditional media in China, we regress the tone of social media posts on the tone of traditional media's articles of the same firm on the same day. We find that the coefficient on the tone of traditional media is significantly positive at 1% level, suggesting that the tone of the two sources of corporate news is positively correlated. We also find that the coefficient on the tone of traditional media is significantly reduced by 0.0235 when the tone of traditional media is positive, which is 55% of the coefficient when the tone of traditional media is zero or negative. This is consistent with our prediction that when traditional media are likely to be more positively biased (i.e., having a positive tone), their tone will be less positively associated with that of social media.

Next, we exploit two sets of political events as exogenous shocks to test if political incentives affect traditional media's optimistic bias and social media's ability to delineate such bias. The first political event involved the government suppressing traditional media in reporting negative corporate news after it intervened the market during the 2015 stock market crash.³ We find that after the shock, there is a decline in the association of tone between traditional and social media when the former is expected to be more optimistically biased. This suggests that the political shock has a significant effect on traditional media's incentives to bias the tone upwards while social media remains to be less influenced by political pressure.

We use the National Congress Meetings, which are held every five years, as the second set of political events. We find that social media play the benchmarking role in delineating traditional media's bias only during the non-National Congress period and not during period of National Congress Meetings. Additional analysis shows that during the period of the National Congress Meetings (meeting days plus 45 days before and after), the number of posts drops significantly while the number of traditional media articles increases significantly. One possible interpretation is that during politically sensitive periods, even social media are under close watch and refrain from posting critical messages, while traditional media are pressured to report more positive news.⁴

Finally, we study whether traditional media with a higher optimistic bias as delineated by social media is associated with a weaker stock return response. First, we find a significantly positive stock return response to the tone of the traditional media articles at a five- and ten-day

³ There is anecdotal evidence that the government pressures the traditional media to be optimistic in its reporting during the stock market intervention (http://www.chinadaily.com.cn/micro-reading/interface_yidian/2015-06-26/13895012.html).

⁴ This is consistent with the finding in Piotroski et al. (2015) that Chinese firms suppress the release of bad news in the year around the National Congress meetings.

CAR.⁵ This suggests that although the traditional media articles are biased, they still have information value to the market. Second, our evidence shows that when the traditional media are more positively biased as indicated by the divergence in tone between traditional and social media, the stock return response to the traditional media decreases significantly. We find this result in the five, ten and twenty-day CAR. To the extent that the market could see through at least partially the bias and discount the information contained in the positive tone of the traditional media, our results further support that social media serve the benchmarking role in delineating traditional media's reporting bias.

Our study contributes to the literature in the following ways. First, prior research focuses mainly on the information generation and dissemination roles of social media in democratic regimes (e.g. Blankespoor et al., 2013; Chen et al., 2014; Bartov et al., 2018). This is the first paper we know of that studies whether, in an autocratic government, social media delineate the biased corporate news of traditional media by supplying less biased information to the market. Our results show that in China, social media's reliance on wisdom of crowds can shelter themselves from government's direct intervention and provide corporate news that can serve as a benchmark against the positive bias of traditional media.

Second, we provide new evidence that in China the market can discount at least partially the positive bias of traditional media and will discount the information when the deviation in tone of the two types of media indicates that traditional media are positively biased. This extends prior research that newspapers can be optimistically biased for economic or political reasons, in either democratic or autocratic countries (Gurun and Butler, 2012; Solomon, 2012; Stockmann,

⁵ We cumulate the stock returns from the day after (day +1) the news/opinions dissemination by the traditional or social media (day 0) to ensure that our result is not capturing the media's reaction to the stock price movement on day 0 due to reverse causality.

2013; Piotroski et al., 2015). Our paper provides evidence that the market will make adjustments to the bias when responding to the news of traditional media.

Third, we are one of the few studies that find that social media posts have information content (see also Tumarkin and Whitelaw, 2001; Antweiler and Frank, 2004; Tetlock et al.2008, Chen et al., 2014). Different from the results in Chen et al. (2014) that are based on U.S. data, we find that the social media posts have strong positive stock return response in both the shorter (five-day) as well as the longer (twenty-day) windows.

The rest of the paper is organized as follows. Section two provides the discussion of the institutional background and hypothesis development. Data and sample are presented in section three and our results are presented in section four. We provide a number of additional tests in section five and conclude the paper in section six.

2. Institutional Background and Hypothesis Development

2.1 Development of the newspaper industry in China

Ever since the beginning of the economic reforms in 1978, China has to balance between two somewhat conflicting objectives for its newspapers. On the one hand, the newspaper industry plays the role of an important information institution to support its market reforms. On the other hand, it continues to serve as the mouthpiece of the government to maintain the political stability of the regime.

In the pre-reform period, China's newspapers were primarily a propaganda machine with the goal of conveying the government's political narrative to its constituents. To support the market liberalization reforms since 1978, the government began to set up a new category of newspapers called commercialized newspapers. The government began cutting subsidies, forcing certain newspapers to rely solely on advertising and subscription revenues, thus creating a new

set of market-based incentives for these newspapers. As a result of these changes, commercialized newspapers are more responsive to market demands for news in order to maximize advertising and subscription revenue. Together, these commercialization reforms increased the number of newspapers from 186 to 1,943 between 1978 and 2009 (Stockmann, 2012).

Another significant reform in the newspaper industry is the formation of news conglomerates, which combine the Party newspapers that serve primarily as the mouthpiece of the government and the newly established commercialized newspapers that respond to the need of the economic reforms. This reform leads to the formation of newsgroups that are under state control with a common set of senior editors for each newsgroup. Piotroski et al. (2017) document that by combining the two types of newspapers within the same group, the conglomeration reform allows the rich advertising revenues from the commercialized newspapers to be channeled to the Party newspapers. They also find that after the conglomeration, both types of papers continue to perform their distinct roles. More importantly, they find that the commercialized newspapers continue to serve its information role by remaining to be less political and more critical than the Party newspapers in reporting corporate news.

One impetus to the media reforms in China is the establishment of the domestic stock markets in the early 1990's. This has created a strong demand for corporate news of the listed companies in China. Between 2000 to 2010, the sample period of Piotroski et al. (2017), they have collected more than 1.7 million corporate news articles of listed companies from 110 distinct newspapers in China. The number of articles grew from more than 67 thousand articles in 2000 to 270 thousand articles in 2010 in the sample.

Despite these largescale news media reforms, the government retains tight control of the newspapers. None of the newspapers or news groups can be majority owned by non-state entities. With this concentration of state ownership and control, there is no press freedom in China. The government continues to control the country's newspapers through the appointment and dismissal of senior editorial staff. The Propaganda Departments of the Chinese Communist Party (CCP) committees (at every level of government) also exerts influence on the reporting behavior of newspapers to ensure that their reports are consistent with the government's narratives. Finally, the government elevated the news media's regulatory and licensing agency, formerly the Publication Bureau under the Ministry of Culture to ministerial-level status and was renamed the General Administration of Press and Publication (GAPP). These elevations signaled that the government's commitment to retain strict control of the news media in China.

2.2 Development of social media in China

The population of netizens in China surpassed 700 million, among which 92.5% are mobile netizens, by June 2016 (All-China Journalists Association, 2017). The deep internet penetration facilitates the development of social media in China, making the social platform, such as microblog, WeChat, and client app, as main channels to access information by netizens. Most of these social media platforms are privately owned, keeping the sites from the direct control of the state. Also, these platforms allow users to generate their own content, which grant netizens more freedom to express their own opinions, especially when their identity is anonymous. The anonymity enables users to conceal their true identity in the computer-mediated communication without the fear of being chastised or targeted. However, anonymity also allows users to release information or express opinions without sufficient responsibilities, making the information on social platforms lacking credibility. Thus, it is still debatable whether social

media is enhancing information efficiency, especially when it is used for disseminating news and expressing opinions for stock investment purposes.

A number of financial service firms have set up social media platforms to provide financial and investment information as part of the overall services they provide to clients. Quite many of these social media platforms are in the form of bulletin board system (BBS), engaging their clients in providing user-generated content (UGC) information and interacting with each other. The main new media players with a focus on financial information include Hexun (<http://www.hexun.com/index.htm>), Jinrongjia (<http://www.jrj.com.cn>), and StockStar (<http://www.stockstar.com>). EastMoney (www.eastmoney.com) was a newcomer into the market when it was founded in 2005. However, it has become the most successful stock social media platform with a rank of 41 among all China's websites globally and No. 1 among all China's financial websites by 2017 according to ALEXA and iResearch data. East Guba, which is a BBS style platform launched by EastMoney in 2006, becomes the hottest online stock platform in China. For example, the total number of posts on Wanke, a real estate company listed in Shenzhen Stock Exchange, has reached more than 360,000 on East Guba, more than 10 times higher than the second most-popular stock-dedicated social media Hexun.

East Guba operated as a discussion board, which is arranged according to the identity of each firm. Thus, the discussion related to a listed company is posted on the board specifically for the company. The interaction among users is like that on Twitter, but they can voluntarily make a post on the board of a listed company without length limitation. The difference between East Guba and Seeking Alpha, a popular U.S. online stock forum, is that the posts of East Guba are screened by any editor as in Seeking Alpha. However, the posts are subject to the regulation or censorship by the Chinese government. For example, a company operating the social media

platform must install surveillance software, which will monitor the information flow on the platform with algorithmic filtering. The regulatory agents also manually monitor the information on social media by deleting the posts that may cause political stability. Thus, this raises questions on whether social media can serve as the benchmarking role in delineating the positive bias of corporate news provided by state-owned traditional media.

2.3 Hypothesis Development

The Chinese government has the ability and incentives to bias the corporate news of traditional newspapers. As an autocratic regime, the government has never granted press freedom to the newspapers. Through regulations, GAPP continues to exert tight control of the press. The government also maintains direct control of the press through majority ownership of all the newspapers and newsgroups. The Party organizations and GAPP can wield their influence through the appointments of senior editors of all the newspapers in China.

There are strong political incentives to influence corporate news in China because local government leaders' performance is evaluated based on the economic performance of the regions under their control (Li and Zhou, 2005). Piotroski et al. (2015) find that listed firms suppress the release of bad news in the year leading up to the promotion of the political leaders of the province in which the firms operate. Likewise, they find that listed firms refrain from reporting bad news around the National Congress Meetings in order avoid any embarrassment to the central government during this important political event. This bias in corporate information is also found in the news media by Piotroski et al. (2017). They document that the corporate news of the Party newspapers, which serve more as a mouthpiece for the government, have a stronger optimistic bias in tone than that of the commercialized newspapers that are more incentivized to meet the needs of the market.

It is unclear the corporate information being produced and disseminated in social media such as online stock forums share the optimistic bias of traditional newspapers in China. On the one hand, the government cannot exert nearly as much influence on social media because all the large social media platforms in China are privately owned. One key feature that makes the control of social media difficult is that the production of information is not guided by an editorial board, but through the wisdom of crowds. To bias the information production, the government will need to influence the crowds rather than a small group of editors, which is much more difficult. Though there is evidence that the Chinese government censors social media messages, primarily those that are associated with collective actions which can destabilize the government's power (King et al., 2013, 2014), the government even allows political sensitive information to be disseminated through social media because it wants to gather more grass root information that can enhance its ability to control the citizens (Qin et al., 2017).

On the other hand, these online posts of stock opinions and analyses are made by non-professionals. China is famous for its stock market being dominated by unsophisticated retail investors. The information produced by amateur investors in these platforms can simply be noise and will not be able to serve as a benchmark against the reporting bias of the traditional media. In addition, there is an increase in the level of control over the internet by the government in recent years.⁶ The 2015 stock market intervention by the government has led to a tight control against public release of negative information that will destabilize the market. The government's

⁶ Two recent news articles provide discussion on how the government has increased its control of the internet: "The great firewall of China: Xi Jinping's crackdown," The Guardian, June 29, 2018 (<https://www.theguardian.com/news/2018/jun/29/the-great-firewall-of-china-xi-jinpings-internet-shutdown>); "China has launched another crackdown on the internet – but it's different this time," CNBC, October 26, 2017 (<https://www.cnbc.com/2017/10/26/china-internet-censorship-new-crackdowns-and-rules-are-here-to-stay.html>).

tightening of censorship of the social media may lead to an increase in optimistic bias in the corporate information shared in the online stock forums.

Thus, it remains an empirical issue if social media can serve as a check in delineating the optimistic bias of traditional media in China. To test this, we first examine if social media will provide less biased information to the market when traditional media is expected to be more optimistic. Specifically, we examine the tone of corporate articles of the Chinese newspapers relative to that of the posts in East Guba, the online stock forum. We posit that when traditional media are more positively biased, the tone of the two media will deviate more significantly which will attenuate the correlation of the tone as well. Our first hypothesis is as follows:

H1: The tone of social media has a weaker association with that of traditional media for the same firm on the same day when the tone of traditional media is positive than when it is neutral or negative.

In this hypothesis, we assume that there is a stronger positive bias in tone when the tone of the newspaper articles is positive (reporting positive news) than negative (reporting negative news). There are two reasons to support this assumption. First, since the tone of the newspaper articles is the sum of the optimistic bias and the tone of the underlying news events, the tone of the articles and the optimistic bias can be mechanically related. Thus, the articles are likely to have stronger optimistic bias when the tone of the articles is more positive.

Second, the sentiment of the underlying events and optimistic bias are likely to be positively associated, which will further enhance the likelihood of the tone of the articles in capturing their tone bias. The reason is that if the underlying events are positive, the cost to bias the reporting of the events will likely to decrease. This is consistent with the argument that the cost of optimistic bias is lower for reporting good news than bad news because the government wants to avoid creating a crisis (Watts, 2003a, b). That is, shareholders are less likely to protest

when the underlying news is good while the actual positive return turns out to be smaller than expected. However, if the underlying news is bad, introducing a positive bias could create a stock crash, leading to a bigger loss to the investors and the risk of investor protests.

There is another possible reason that the cost of tone bias is lower when the underlying events are good news than bad news. Literature in psychology suggests that in many contexts, people pay much greater attention to negative or unfavorable information than to positive or favorable information (e.g. Cacioppo and Gardner 1999; Baumeister et al. 2001). In economics literature, prospect theory in Kahneman (1979) suggests a stronger investors' attention to negative information because their decision-making processes are influenced by loss aversion and endowment effects. Recent studies on human behaviors show that bad news spreads faster and wider than good news over the internet (e.g. Hornik et al., 2015). Due to investors' heightened attention to the bad news in the market, any unraveling of the government's attempt to cover up the underlying bad news by introducing positive bias to news could cause an even bigger panic or even crisis, compared with the discovery of the government's upward bias of good news.⁷

It is possible that other than the cost of the optimistic bias, the benefit of the optimistic bias may also affect media's reporting incentives. The benefit of the positive bias can be much greater especially when the tone of the underlying events becomes very negative. That is, when the bad news can jeopardize the political capital (e.g. promotion opportunities) of the government officials, they will have strong incentives to suppress the bad news. Working in the

⁷ Another possible reason is that, like earnings management, media would introduce more positive bias when their reported tone is just below zero (neutral) in order to increase the probability of reporting an article with positive tone, in line with the propaganda needs of the government. This additional bias causes a discontinuity around the breakpoint, just like the discontinuity in reported earnings per share well-documented in earnings management literature (e.g. Burgstahler and Dichev, 1997). As a result, when the reported tone is positive, it has more positive bias than when the reported tone is negative.

opposite direction to the cost effect of the optimistic bias, this benefit effect may weaken the positive association between the tone of the articles and optimistic bias.

However, we do not expect that this benefit effect will outweigh the cost of optimistic bias in our analysis. When the state-controlled newspapers face the decision to bias the tone of a negative news event, they are likely to omit reporting the event than manipulating the tone of the reports.⁸ To the extent that these newspapers choose to skip reporting the events, they will be excluded from our analysis because only firm-days with at least one news article are included in our sample. The conjecture that many of the bad news events are unreported by traditional media is consistent with our finding that only 22.1% of the newspaper articles are negative in tone, comparing to East Guba having 87% of the posts reporting in negative tone. Appendix I also shows that when we rank the tone of newspaper articles into deciles, the positive (optimistic) tone bias of newspapers, measured by the positive deviation in tone between newspapers and East Guba, is monotonically increasing in the tone of the newspapers, which supports our assumption that traditional media with more positive tone are more optimistically biased. Notwithstanding these explanations, it remains a joint hypothesis that traditional media's tone is a good proxy for their optimistic bias and social media plays a monitoring role in delineating traditional media's bias.

Next, we examine if social media's role to delineate traditional media's optimistic bias is supported by the market perception of the credibility of traditional media's news articles. To test this, we posit that the market's response to the tone of traditional media will be attenuated when the tone of the traditional media deviates more positively from that of social media. This is a joint hypothesis that social media can serve as a benchmark to identify the tone bias of traditional

⁸ For example, news media in China chose to remain silence especially in the early periods of the milk powder scandal in 2008 and during SARs in 2003.

media, and the market discounts its response to the tone of traditional media articles when it is positively biased. Our second hypothesis is as follows:

H2: The stock return response to the tone of traditional media is significantly reduced when it deviates positively from the tone of social media.

3. Data and Sample

3.1 Data from traditional media

The traditional media data is an updated version of that used in Piotroski et al. (2017) with extended years covered, ambiguity in firm name cleaned, and definition for the tone improved. The original news article covering listed companies are crawled from Newswire with an automatic crawler by searching the stock name throughout the database. We first clean the data to avoid the articles crawled due to the ambiguity of the company name. For example, a large number of articles are crawled for a company named Laobaixing (老百姓), which is also a common term referring to “ordinary people”. We identify a list of companies with ambiguous names and then manually read a set of articles for each name as a training set for identifying whether the articles are talking about that particular company. We then apply the machine learning technology to identify the articles that actually cover each of the firm that has an ambiguous name.

We also apply machine learning technology to construct the measurements of the tone. First, we hire two batches of research assistants to label the tone of each sentence of 50,000 articles randomly picked from our sample as negative, positive, and neutral. Then, we optimize the model within this training sample and extend the model to the full sample. The out-of-sample validation using 10,000 randomly selected sentences shows that the accuracy rate of our model is above 90% in labeling the tone of sentences.

Next, we aggregate sentence-level tone to form measures of article-level tone. The tone of the article is measured by the relative weight of positive sentences to negative sentences in the article. In addition, we also consider the importance of sentences from different positions within an article. That is, we weigh the sentences from the first and last paragraphs as 2, the first and last sentences of first and last paragraph as 3, and other sentences from the article as 1. The tone of the body of the article equals $(\# \text{ of positive sentences} - \# \text{ of negative sentences}) / (\# \text{ of positive sentences} + \# \text{ of negative sentences} + 1)$. The overall tone of the article in the end is defined as $(\text{tone of text body} * 0.7 + \text{tone of title} * 0.3)$.

Our final traditional media news dataset includes about 3.7 million news articles published over the period of 2009 to 2016, by 162 unique newspapers. These newspapers are located in 23 provinces. Among the newspapers, 42 (26%) are party-affiliated and 120 (74%) are non-party-affiliated (commercialized). Party-affiliated newspapers published about 0.64 million (17%) news articles, while non-party-affiliated newspaper published about 3.06 million (83%) news articles.

3.2 Data from Social Media

The Guba platform (aka East Guba) on East Money (<http://guba.eastmoney.com>) provides a separate discussion board for each listed company. East Guba is one of the oldest and most influential social media platforms with a focus on the capital market.⁹ Users can enter and post on the board by searching the stock code or the name of the company. They can also build up a self-defined list of companies by following them and entering the discussion board of the companies on the list by a direct click. Thus, the posts on this platform are well matched with the

⁹ Weibo is the most influential social media platform but it lacks a focus on capital markets. Snowball, another social media platform, was established much later than East Guba and is likely to have more active professionals such as analysts and institutional investors on it.

related companies. We develop a web crawler to download all the main posts on the discussion board of each company.¹⁰ Because the posts on the social media are usually short, we thus label the tone at the post level rather than sentence level. Because the writing style is different from that in traditional media, we use a different machine learning training set for the social media posts. Emoji is often employed by social media users to express their opinions. In our model, we also consider emoji in defining the tone of a post.

We define the tone of social media for a firm as the relative weight of the number of positive posts and negative posts, $(\# \text{ of positive posts} - \# \text{ of negative post}) / (\# \text{ of positive posts} + \# \text{ of negative post} + 1)$, which is analogous to using the tone of sentences in the body of the text to compute the tone of traditional media articles. Our final social media dataset includes about 38.4 million posts, covering 3,011 firms from 2009 to 2016.

3.3 Sample Selection Process and Descriptive Analyses

We start by obtaining all available firm-day daily stock price observations from 2009 to 2016, for companies listed on China's Shanghai and Shenzhen Stock exchanges from the China Stock Market & Accounting Research (CSMAR) database. We first match firm-day stock price observations with firm fundamentals in their last year's annual reports. Next, we match the firm-day sample with our traditional media tone and social media tone as discussed in subsection 3.2. Because our empirical analyses focus on the relationship between social media and traditional media on a firm-day level, we include only firm-day observations with at least one traditional media news articles and at least three social media posts. We require to have at least three social media posts because we want to avoid measurement errors of the social media tone caused by

¹⁰ We ignore the reply comments on the platform.

insufficient posts per day.¹¹ After merging our social media and traditional media file with CSMAR dataset, our final sample includes 970,982 firm-day observations, covering 3,011 unique firms. Detailed sample selection process is reported in Table 1, Panel A.

Table 1, Panel B provides descriptive statistics of traditional media tone and social media tone. On average, a firm in our sample is covered by 3.89 traditional media news articles and 39.49 social media posts per day. The tone of traditional media has a mean (median) of 0.3704 (0.4658), and the tone of social media has a mean (median) of -0.2028 (-0.1778). Also, traditional media tone is positive in 77.9% of the sample, while social media tone is positive in only 12.7% of the sample.¹² Taken together, the descriptive evidence suggests that on average traditional media is more positively biased than social media. When disentangling traditional media based on their affiliation to the CCP as “party media” and “non-party (commercialized) media”, we show that the mean (median) tone of party media is 0.5115 (0.7000) and mean (median) tone of non-party media is 0.1602 (0.0000). This descriptive evidence is consistent with Piotroski et al. (2017) that the positive bias in traditional media is more pronounced in party-media than non-party media.

The correlation between the tone of traditional and social media is shown in Table 2. The tone of the two media is positively correlated at 0.0523, significant at the 5% level. When the traditional media tone is positive, the correlation drops to 0.0459, consistent with our conjecture that social media is maintaining its independence when traditional media becomes more positively biased. The correlation between the tone and CARs is also significantly positive for

¹¹ Results are robust when we increase the threshold to five posts per day and ten posts per day, despite a reduced sample size due to increased requirement.

¹² Our measurement of social media tone using the Chinese “Guba” is comparable to the measurements using US social media the “StockTwits”. For example, Byard and Wang (2018) reports a mean(median) tone of -0.326 (-0.405) when applying similar textual analysis method to StockTwits dataset. Consistent with sample distribution of Byard and Wang (2018), we also show that the average tone of Chinese social media is negative.

the both types of media. However, the magnitude of CARs' correlation with social media is three times larger than that of traditional media, suggesting that the former has more information content and is likely to be less biased. The sample distribution of the firm-day observations by year is reported in Table 3 Panel A. The number of firm-days are slightly smaller in 2015 and 2016, but the distribution is generally quite even over the sample period. The sample distribution by industry is presented in Table 3 Panel B. Most of the industries are well represented in our sample.

Figure 1 provides descriptive evidence on the relationship between the tone of social media and traditional media. The slope coefficient between social media and traditional media is higher when traditional media is negative and it decreases when traditional media is positive.

4. Empirical Analyses

4.1 Traditional Media Bias and Social Media's Benchmarking Role

Our primary empirical analyses examine the relationship between the tone of social media and the tone of traditional media covering the same firm on the same day, and how this relationship varies when traditional media tone is more positive than neutral or negative. We estimate the following model with OLS regression using firm-day data:

$$(1) \quad \text{Social Media Tone}_{it} = \alpha + \beta_1 \text{Traditional Media Tone}_{it} + \beta_2 \text{Traditional Media Tone}_{it} \times \text{TM Positive}_{it} + \beta_3 \text{TM Positive}_{it} + \beta_4 \text{Size}_{it} + \beta_5 \text{ROA}_{it} + \beta_6 \text{Market to Book}_{it} + \beta_7 \text{Leverage}_{it} + \text{Firm Dummy} + \text{Year Month Dummy} + \varepsilon_{it}$$

In these estimations, TM Positive_{it} is a dummy variable equal to one (and zero otherwise) if $\text{Traditional Media Tone}_{it}$ is greater than zero, indicating that the average tone of

all social media articles covering a firm is positive. Coefficient β_1 captures the unconditional association between the tone of social media and the tone of traditional media. Coefficient β_2 of the interaction term $Traditional\ Media\ Tone_{it} \times TM\ Positive_{it}$ captures the reduction in the association between the tone of social media and traditional media when the tone of traditional media is positive. Because both social media and traditional media provide relevant information of a company's fundamentals (Bartov et al. 2017), we expect β_1 to be positive and significant. We expect β_2 to be negative and significant because when traditional media are *more* optimistically biased (as proxied by the positive tone), the tone of social media will be *less* positively biased relative to traditional media.

We estimate the model with a group of firm-level control variables including *Size*, *ROA*, *Market to Book* and *Leverage* to capture the variation in firm characteristics that could affect the bias in traditional media, and the relationship between the tone of social media and the tone of traditional media. All variable definitions are presented in Appendix II. However, even with these control variables, there is still concern that our results could be driven by correlated omitted variables. To alleviate this concern, we include firm fixed effect and year-month fixed effect to absorb further unobservable firm characteristics and time-variant characteristics that potentially affect the tone of both social media tone and traditional media. In all the regressions, we report in parentheses t-statistics with robust standard errors two-way clustered at the firm level and year-month level.

The regression results of model 1 are reported in Table 4. In columns 1 and 2, we show a positive association between the tone of traditional media and social media, indicating that on average traditional media tone is positively associated with social media tone. In columns 3 to 5, the coefficient estimates on traditional media tone continue to be positive and significant, and the

coefficient estimates on $Traditional\ Media\ Tone_{it} \times TM\ Positive_{it}$ are negative and significant across all model specifications. The negative coefficients on the interaction term suggest that the positive association between traditional media tone and social media tone is reduced when traditional media is positive. In terms of economic significance, in column 5 the coefficient on the tone of traditional media is significantly reduced by 0.0235 when the tone of the traditional media is positive, which is 55% of the coefficient when the tone of the traditional media is zero or negative.¹³

Overall, our baseline regression results show a positive association between traditional tone and social media tone of the same firm on the same day, and this association is attenuated when traditional media tone is positive, consistent with our first hypothesis that social media serves as a benchmark to delineate the optimistic bias of traditional media.¹⁴

4.2 Influences of Political Incentives on the Benchmarking Role of Social Media

Prior literature shows that traditional media bias varies with political incentives of the government (Piotroski et al. 2015; Gentzkow et al. 2006). In this subsection, we exploit China's recent political events as exogenous shocks to test if the changes in the association of the tone of traditional and social media reflect how the two media respond to changes in political incentives.

4.2.1 Media Intervention in 2015 Stock Market Crash

¹³ The coefficient on Traditional Media Tone is low (0.0426 in column 5), which is likely due to noise caused by the two types of media reporting on different topics. When we increase the requirement of having more than one article by traditional media on each firm-day, it increases the chances that the two types of media are reporting on the same topics. We find that the coefficient on Traditional Media Tone (untabulated) increases to 0.134 (0.156) when we require more than three (five) articles per firm-day.

¹⁴ A possible alternative explanation is that when traditional media tone is positive, social media are reporting topics different from those of traditional media, which causes the decline association and the negative coefficient on $Traditional\ Media\ Tone_{it} \times TM\ Positive_{it}$. However, when we increase the chances that the two types of media are covering similar topics by requiring each firm-day to have more than three (five) articles (see footnote 13) or to be on the three days around the earnings announcement dates, the coefficient on $Traditional\ Media\ Tone_{it} \times TM\ Positive_{it}$ remains significantly negative.

First, we use the media intervention in 2015 as an exogenous shock for our tests. In 2015, the Chinese stock market experienced a historical crash with a third of the market value in the Shanghai Stock Exchange being lost in the month of June. This was followed by 1,400 companies (more than half of all the listed firms) filed for a trading halt in early July in an attempt to avoid further losses. To stabilize the turbulence in the stock market, the National Bureau of Television, Broadcast, and Newspaper (a bureau under GAPP) issued an authoritative order to all press in China, demanding them to decrease the coverage of stock market-related news, and stop using negative words such as “tumble” or “crash” in the news. The Chinese press responded to this authoritative order by reducing the coverage of negative corporate news, resulting in an increased level of bias.¹⁵

To test the impact of such intervention formally, we partition the full sample into pre-intervention period which includes all the sample data before or on July 23rd, 2015, and post-intervention period after this date. We estimate the baseline regression (Model 1) separately in these two sub-samples and test for the difference between regression coefficients using a Chow test. Results are reported in Table 5. The association between traditional media tone and social media tone is not statistically different in the pre-intervention period and post-intervention period, but the reduction in the association when traditional media is positive has significantly increased in the post-intervention period. Specifically, in the pre-intervention period, the reduction in association between traditional media tone and social media tone is about 49% (0.0209/0.0425) when traditional media tone is positive, compared to a reduction of about 97% (0.0352/0.0360) in the post-intervention period. Taken together, the results suggest that the 2015 media intervention, as an exogenous political shock, has increased the positive bias in

¹⁵ The mean tone of the newspapers in 2015 is 0.4343 in the first 6 months and 0.4282 in the second 6 months.

traditional media's coverage of company news. Also, social media are able to maintain their benchmarking role in delineating traditional media's bias by sheltering themselves from the shock.

4.2.2 National Congress Meeting Period versus Non-National Congress Meeting Period

Another exogenous political shock we exploit is the CCP National Congress Meeting, which is an important political event held every five years in Beijing. During this period, the government intensifies its monitoring over traditional media to maintain a stable social environment. Our untabulated evidence shows that around the 18th CCP National Congress Meeting, traditional media tone has a positive deviation from that of social media, with a range of 0.4455 to 0.5316, from the 3 months prior to the meeting during the month of the meeting, and remained positive for the 3 months after the meeting. Given the heightened bias in traditional media during this politically sensitive period, we are interested in whether social media, given it is under more loosed control and monitoring from the regulator than traditional media, can continue to delineate traditional media's optimistic bias.

We estimate the same regression model in our baseline analysis (Model 1) separately in the two sample partitions: congress meeting sample which includes a window of -45 to +45 days relative to the opening day of the 18th CCP National Congress Meeting, and non-congress meeting sample which includes the rest of the firm-day observations. Column 1 of Table 6 shows that there is no significant association between the tone of social media and the tone of traditional media around the congress meeting sample. In addition, the attenuation in the association between the tone social media and traditional media also disappeared during this same period. However, column 2 shows similar results to our main finding in Table 4, suggesting our baseline results mainly concentrate in the non-congress meeting period. One possible

interpretation of the results is that social media can delineate the bias in traditional media during the non-congress meeting periods but not during congressional meeting periods.

Next, we explore the potential causes for the loss of social media's ability in delineating traditional media's bias during congress meeting period. In Figure 2, we show univariate evidence that the number of social media posts experienced a gradual decline starting from about 6 months ahead of the congress meeting, reached the lowest level during the meeting and started a slow recovery to reach its normal level in about three months following the meeting. However, during the entire time the number of news articles published by traditional media for a firm-day remains largely unchanged.

Table 7 provides results on the level of activities for traditional media and social media during the Congress Meeting period. We estimate a pooled OLS regression model to determine the impact of the Congress Meeting on the level of activities proxied by the number of social media posts and the number of new articles. The model controls for firm characteristics (*Size*, *ROA*, *Market to Book*, *Leverage*), daily returns (stock return and index return), news from other channels (number of news, and number of social media posts as control variables) and a firm dummy to capture observed firm-level characteristics. Our results show that the number of corporate news covering on a firm-day by traditional media increases by an average of 0.3 (7.7% increase) articles during Congress Meeting period, while the number of social media posts decrease by an average of 12.35 posts (31.3% decrease) during the same period. This finding provides possible explanations for the disappearance of social media's ability in delineating the bias – during politically sensitive periods such as the National Congress Meetings, even social media users refrain from posting critical (negative) messages while the traditional media is pressured to report more positive news.

4.3 The Tone of Social Media and Traditional Media and Stock Returns

Prior studies have investigated the stock return responses to traditional media tone and social media tone: Tetlock et al. (2008) focus on traditional media and find that the fraction of negative words in firm-specific news articles is associated with the same day and following day abnormal returns. Chen et al. (2014) find that opinion transmitted through seeking alpha is associated with abnormal stock returns from 3 to 60 days after the post. Bartov et al. (2018) show that Twitter contains original information that is useful in predicting announcement period returns. However, to the best of our knowledge, there is little evidence on whether there are significant stock return responses to traditional media and social media in China. Given China's unique media system (Piotroski et al. 2017; Qin et al. 2018) and institutions, empirical findings using U.S. data may not be generalizable to the Chinese setting.

In addition, existing literature examines the information content of social media and traditional media *independently*. We add to existing literature by jointly analyzing the relative association between the tone of the articles of social media and traditional media and the corresponding future abnormal stock returns. More specifically, we want to use the relative stock return response to traditional and social media as a way to gauge if social media can serve as a benchmark against the positive tone bias of traditional media. We expect that the stock return will discount its response to the tone of traditional media when it deviates positively in tone from that of social media, which provides support to our conjecture that social media can serve as a benchmark in delineating the positive bias of traditional media.

4.3.1 Descriptive Evidence

To study the information content of social media and traditional media in predicting returns, we first adopt a portfolio analysis method. Each week, stocks are ranked based on their

average tone of traditional media (social media) in the previous week (from last Monday to last Friday) into ten deciles. Equal-weighted portfolios are formed by buying stocks that fall into the 10th and the 1st deciles of last week's traditional media tone (social media tone), these two portfolios are labeled TM10 (SM10) and TM1 (SM1), respectively.¹⁶ Portfolios are adjusted weekly, and we impose a hypothetical transaction cost of 5 bps on each weekly adjustment.

Figure 3 presents the net value of these weekly-adjusted portfolios from 2009 to 2016, with a beginning value of 100, in comparison with China Stock Index 300 (CS300). Both TM10 and SM10 outperform the index, and both TM1 and SM1 underperform the index, suggesting traditional media and social media both provide information useful in predicting stock returns. When looking at positive news and negative news separately, we show that SM10 yields a raw return of 446% over the course of eight years, almost doubling TM10's return for the same period; however, TM1 yield a more negative return than SM1 (-51% vs. -31%). Taken together, this descriptive evidence suggests that 1) overall tone of social media is useful in predicting returns than tone of traditional media, and 2) tone of traditional media is less useful than tone of social media in predicting returns when the news is positive (i.e. more biased), but traditional media is more useful than social media in predicting returns when the news is negative (i.e. less biased). However, these descriptive findings should be interpreted with caution, because they have not considered many factors that limit the ability to form the portfolio, such as liquidity, restriction to buy/sell and price impact.

4.3.2 Regression Analyses

¹⁶ Short-selling is strictly controlled by the China Security Regulation Committee (CSRC), and it is banned for the majority time in our sample. To approximate the reality, we only calculate returns on the long side during our sample period.

In this subsection, we explore whether Chinese social media and traditional media contain information that is useful to predict future stock returns using a more rigorous regression analysis that controls for other factors. Specifically, we test 1) whether the tone of traditional media and tone of social media are associated with future abnormal returns at the firm level, 2) whether the association between the tone of traditional media and future abnormal returns is attenuated when it deviates from the tone of social media (i.e., more positively biased).

We estimate the following regressions models with firm fixed effect and year-month fixed effect. CAR is the value-weighted market-adjusted cumulative abnormal return; we use three different measurement windows to capture the short-term and longer-term impact news and social media. To reduce the likelihood of reverse causality that news or social media is reacting to extreme abnormal returns, all three CAR windows start from day +1 relative to day 0, where traditional media and social media are measured. *TM Bias* is a dummy variable that equals one if the difference between traditional media tone and social media tone for the same firm on the same day falls in the top quintile, representing the highest level of optimistic bias in traditional media.¹⁷ Following Tetlock et al. (2008), we include CAR of the last week to control for momentum in price movement, and standard errors are double clustered at the firm and year-month level. Our model is as follows:

$$(2) \quad CAR_{it} = \alpha + \beta_1 Traditional\ Media\ Tone_{it} + \beta_2 Traditional\ Media\ Tone_{it} \times \\ TM\ Bias_{it} + \beta_3 TM\ Bias_{it} + \beta_4 Social\ Media\ Tone_{it} + \beta_5 CAR[-5, -1]_{it} + \\ \gamma' Control\ Variables_{it} + Firm\ Dummy + Year\ Month\ Dummy + \varepsilon_{it}$$

¹⁷ We calculate the difference between the tone of traditional media and the tone of social media as follows: $Diff_{tm-sm} = Traditional\ Media\ Tone_{it} - Social\ Media\ Tone_{it}$, we avoid using a continuous measure of the difference between TM and SM to alleviate the concern of multi-collinearity. In robustness check, we define *TM Bias* as top quartile and top tercile, and the results remains unchanged.

Table 8 presents the regression results. In all three CAR measurement windows, the tone of social media is positively associated with CAR, and the coefficients are both statistically and economically significant. An increase in the tone of social media from zero (neutral) to one (most positive) is associated with an increase of 56 basis points in cumulative abnormal returns over the following five trading days. Moreover, this association does not reverse in the 10-day or 20-day windows.¹⁸ Consistent with findings of Tetlock et al. (2008) and Chen et al. (2014), our results suggest that traditional media also have positive associations with the subsequent CAR in the 5-day and 10-day windows, but the associations are significantly attenuated when the tone of traditional media deviates from the tone of social media. Comparing the coefficients of traditional media and social media, the results suggest that the tone of social media has much stronger predictive power of future abnormal returns than traditional media.

In summary, the regression analysis provides evidence that social media and traditional media both provide useful information to predict future abnormal returns, and traditional media's information content is reduced when its tone deviates from the tone of social media. These findings support our conjecture that social media serve as a check in delineating the positive bias in traditional media.¹⁹

5. Additional Analysis and Robustness Tests

5.1 Social Media's delineation of bias for State-owned Firms versus Non-State-Owned Firms

In this subsection, we focus on whether state ownership of a listed firm influences the degree of social media's ability to delineate the reporting bias of traditional media. Piotroski et

¹⁸ This further supports that the information provided by social media is not noise but value relevant.

¹⁹ We conclude that the stock price results support our conjecture that social media supplies less biased information that serves as a benchmark for delineating the traditional media's optimistic bias. However, we do not claim that it is the social media information that *causes* the market to discount its response to the biased information reported by the traditional media. The price discount could simply be reflecting the market's response to the unbiased underlying information that is captured by the social media.

al. (2017) document that Chinese newspapers are more likely to optimistically bias the news of state-owned enterprises (SOEs) firms because negative news of SOEs will impose a higher cost to the politicians than non-SOEs. If social media maintain their independence and not bias one type of firms more than the other, we expect social media's postings are able to delineate a higher optimistic bias by traditional media for SOEs than non-SOEs. We use the following cross-sectional model to examine whether social media can delineate the difference in traditional media's optimistic bias between SOEs and non-SOEs:

$$(3) \quad \begin{aligned} \text{Social Media Tone}_{it} = & \alpha + \beta_1 \text{Traditional Media Tone}_{it} \times \text{TM Positive}_{it} \times \text{SOE}_{it} + \\ & \beta_2 \text{TM Positive}_{it} \times \text{SOE}_{it} + \beta_3 \text{Traditional Media Tone}_{it} \times \text{TM Positive}_{it} + \\ & \beta_4 \text{Traditional Media Tone}_{it} \times \text{SOE}_{it} + \beta_5 \text{TM Positive}_{it} + \beta_6 \text{SOE}_{it} + \\ & \beta_7 \text{Traditional Media Tone}_{it} + \gamma' \text{Control Variables}_{it} + \text{Industry Dummy} + \\ & \text{Year Month Dummy} + \varepsilon_{it} \end{aligned}$$

Table 9 presents the result of OLS regression with industry and year-month fixed effects to control for unobservable correlated omitted variables. The results in column 1 suggest that social media can delineate a stronger traditional media's optimistic bias for SOEs than non-SOEs, but only the period prior to the 2015 media intervention (pre-intervention period). In such a period, the reduction in the association between traditional media tone and social media tone when traditional media tone is positive is about 69% for SOEs and 40% for non-SOEs.²⁰ We do not find similar results in the post-intervention sample or the full-sample. This is likely because as a response to the strict media intervention order from the government, traditional media increased the bias in corporate news upward for *all* firms regardless of their state ownership in

²⁰ The reduction in association between traditional media tone and social media tone when traditional media tone is positive (i.e. Social media's delineation of Traditional Media Bias) is calculated as follow: for SOEs $(\beta_1 + \beta_3)/(\beta_7 + \beta_4)$, for non-SOEs β_3/β_7

the post-intervention period. The results suggest that the traditional media bias for SOEs and non-SOEs has been leveled following the intervention.²¹

5.2 The Information Content of Social Media and Traditional Media: SOEs Versus Non-SOEs

Our findings in the previous subsection that social media can delineate a different level of bias across SOEs and non-SOEs. In this section we explore whether the information content of traditional media tone and social media tone varies across SOEs and non-SOEs. We estimate Model 3 separately in the SOE subsample and non-SOE subsample with firm fixed effect and year-month fixed effect, and compare the coefficients across-sample using Chow test.

Results are reported in Table 10. In all three measurement windows, the tone of social media is positively associated with future abnormal returns for both SOE and non-SOE subsamples, and the coefficients of social media are significantly higher than coefficients of traditional media, suggesting that social media has better information content in predicting future returns in both the SOE and non-SOE subsamples. The tone of traditional media fails to predict future abnormal returns in two out of three measurement windows for SOEs but still exhibit an ability to predict future abnormal returns for non-SOEs. Even in the [1,5] days window where traditional media are able to predict abnormal returns for both types of companies, the coefficient on *Traditional Media Tone* is significantly greater for non-SOEs than for SOEs (difference is significant at 0.01 level). This evidence suggests that the bias in the tone of traditional media associated with SOEs has weakened its information content. Our results also suggest that social media – due to its benchmarking role in delineating the traditional media bias – can serve as a

²¹ We find that in the pre-intervention period, tone of traditional media is significantly more positive for SOEs than for no-SOEs, but this difference has disappeared in the post-intervention period. The results (untabulated) are robust when controlling for firm characteristics and including industry fixed effect and year-month fixed effect.

substitute in providing more useful information to the market when the bias in traditional media is more prominent.

5.3 Social Media's delineation of bias for Party Papers versus Non-Party Papers

Piotroski et al. (2017) show that party papers in China are more positively biased than non-party papers for political reasons. Consistent with this finding, the mean tone of party media is 0.5115, notably more positive compared to the mean tone of non-party (commercialized) media of 0.1602, in our sample. We then explore whether social media differentiate the degree of positive bias in party papers and in non-party papers. We estimate our baseline model separately by first replacing *Traditional Media Tone* with *Party Media Tone* and then with *Non-Party Media Tone*. Because party papers publish corporate news less frequently than non-party media, a large number of firm-day observations are missing due to the lack of party media coverage.

Results are reported in Table 11. The association between the tone of social media and tone of party media is significantly smaller (Chi-square statistics =8.57), accounting for only about 33% (0.0192/0.0586) of the association between the tone of social media and the tone of non-party media. However, the coefficients on the interaction term which are to capture the extent to which social media delineate traditional media's positive bias are not statistically different across the two specifications. This is likely due to the already weak association between the tone of social media and the tone of party media. In general, the tone of social media exhibits significantly lower association with the tone of party media than with tone of the non-party media, due to the relatively high positive bias in party media.

5.4 The Information Content of Social Media and Traditional Media during the National Congress Meeting Period

In the previous subsection we find that social media fail to delineate the bias in traditional media in the presence of strong political incentive from the government to curb negative news, we extend this finding by investigating whether social media and traditional media provide useful information in predicting future abnormal returns.

We estimate Model 3 separately for National Congress Meeting period and Non-National Congress Meeting period. Our (untabulated) results indicate that neither traditional media tone nor social media tone is associated with future abnormal returns during the National Congress Meeting period. It is possible that the intensified monitoring over both traditional media and social media during the politically sensitive period has diminished the information content of both media. However, both traditional and social media tone is significantly associated with future abnormal returns in the non-congress meeting period, except for the twenty-day CAR and the traditional media tone.

5.5 Robustness Tests

5.5.1 *Alternative Measures of Traditional Media Bias*

Our results of social media's benchmarking role primarily rely on using an indicator variable whether traditional media tone is positive (i.e., tone of traditional media > 0). From the descriptive statistics, we show that traditional media is positive in about 77.9% of the firm-day observations. One may be concerned that this indicator variable has insufficient variation in capturing the bias in tone. As a robustness check, we replace the *TM Positive* indicator variable with *TM Above Median*, which is an indicator variable set to one if tone of traditional media is above the median tone of all firms on a given day. We repeat our baseline regressions and find that our results (untabulated) are statistically and economically robust to this alternative measure.

5.5.2 *Alternative Measures of the Dependent Variable - Social Media Tone*

In our sample, we require at least three social media posts about the same company for a firm-day to be included in our sample, in order to avoid measurement errors due to a few extreme and unrepresented posts. One may be concerned that limiting the sample to three posts minimum would be insufficient in addressing this issue. As a robustness check, we tried the following restrictions in constructing our dependent variable – social media tone: 1) include only samples with a minimum of five posts per day 2) include only samples with a minimum of ten posts per day 3) winsorize the most positive and most negative social media posts in 1) & 2). Our results (untabulated) show that the baseline results largely remain unchanged in both sign and significance after applying these three alternative measures of social media tone to our main results.

6. Conclusions

This paper examines whether China's social media supply less positively biased information to the market than traditional media. Using a comprehensive sample from 2009 to 2016 of corporate news of newspapers and the opinions shared by investors in East Guba, an online stock forum, we document that there is a positive association in the tone of the newspapers and East Guba for the same firm on the same day. However, we find that the tone of East Guba is less positively associated with that of the newspapers when the tone of the newspapers is positive. Consistent with the conjecture that East Guba plays the benchmarking role in delineating the bias of state-owned media, these results suggest that the tone of East Guba's posts is less optimistic than that of the newspapers when the latter are more likely to be positively biased.

Exploiting the 2015 media intervention as an exogenous political shock, we find that this political intervention has increased the newspapers' optimistic bias and the deviation in tone

between the newspapers and East Guba. This suggests the political influence on East Guba is smaller, which allows it to continue playing the benchmarking role to delineate the newspapers' optimistic bias. Finally, the positive stock return response to the tone of the newspapers' articles is significantly attenuated when it deviates positively from that of East Guba, further supporting that East Guba can serve as a benchmark against newspapers' reporting bias.

Reference

- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of finance*, 59(3), 1259-1294.
- Bartov, E., Faurel, L., & Mohanram, P. S. (2017). Can Twitter help predict firm-level earnings and stock returns?. *The Accounting Review*, 93(3), 25-57.
- Baumeister, Roy F., et al. "Bad is stronger than good." *Review of general psychology* 5.4 (2001): 323.
- Blankespoor, E., Miller, G. S., & White, H. D. (2013). The role of dissemination in market liquidity: Evidence from firms' use of Twitter™. *The Accounting Review*, 89(1), 79-112.
- Burgstahler, David, and Ilia Dichev. "Earnings management to avoid earnings decreases and losses." *Journal of accounting and economics* 24.1 (1997): 99-126
- Byard, Donal, and Eric Wang. "The Impact of Public Disclosure on Information Asymmetry between Sophisticated and Unsophisticated Investors: Evidence from an Investor Social Media Network." (2016).
- Cacioppo, John T., and Wendi L. Gardner. "Emotion." *Annual review of psychology* 50.1 (1999): 191-214.
- Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5), 1367-1403.
- Djankov, S., McLiesh, C., Nenova, T., & Shleifer, A. (2003). Who owns the media?. *The Journal of Law and Economics*, 46(2), 341-382.
- Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and political persuasion: Evidence from Russia. *American Economic Review*, 101(7), 3253-85.
- Foucault, M.. 1977. Discipline and punish. Harmondsworth: Penguin.
- Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of political Economy*, 114(2), 280-316.
- Gurun, Umit G., and Alexander W. Butler. 2012. Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value. *Journal of Finance* 67:561-97.
- Hornik, Jacob, et al. "Information dissemination via electronic word-of-mouth: Good news travels fast, bad news travels faster!." *Computers in Human Behavior* 45 (2015): 273-280.
- Kahneman, Daniel. "Prospect theory: An analysis of decisions under risk." *Econometrica* 47 (1979): 278.

- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Anomalies: The endowment effect, loss aversion, and status quo bias." *Journal of Economic perspectives* 5.1 (1991): 193-206.
- King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, 107(2), 326-343.
- King, G., Pan, J., & Roberts, M. E. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science*, 345(6199), 1251722.
- Li, H., & Zhou, L. A. (2005). Political turnover and economic performance: the incentive role of personnel control in China. *Journal of public economics*, 89(9-10), 1743-1762.
- Noelle-Neumann, Elisabeth, 1984. *The Spiral of Silence: Public Opinion-Our Social Skin*. The University of Chicago Press.
- Piotroski, J. D., Wong, T. J., & Zhang, T. (2015). Political incentives to suppress negative information: evidence from Chinese listed firms. *Journal of Accounting Research*, 53(2), 405-459.
- Piotroski, J. D., Wong, T. J., & Zhang, T. (2017). Political Bias in Corporate News: The Role of Conglomerate Reform in China. *The Journal of Law and Economics*, 60(1), 173-207.
- Qin, B., Strömberg, D. and Wu, Y., 2017. Why does China allow freer social media? Protests versus surveillance and propaganda. *Journal of Economic Perspectives*, 31(1), pp.117-40.
- Solomon, D. H. (2012). Selective publicity and stock prices. *The Journal of Finance*, 67(2), 599-638.
- Spears R. and Lea M.. 1994. Panacea or Panopticon? The hidden power in computer-mediated communication. *Communication Research* 21(4), 247-459.
- Stockmann, D. (2013). *Media commercialization and authoritarian rule in China*. Cambridge University Press.
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437-1467.
- Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57(3), 41-51.
- Tversky, Amos, and Daniel Kahneman. "Loss aversion in riskless choice: A reference-dependent model." *The quarterly journal of economics* 106.4 (1991): 1039-1061.
- Watts, R.L., 2003a. Conservatism in accounting part I: Explanations and implications. *Accounting horizons*, 17(3), pp.207-221.

Watts, R.L., 2003b. Conservatism in accounting part II: Evidence and research opportunities. *Accounting horizons*, 17(4), pp.287-301.

Table 1- Panel A Sample Selection Process

	Firms	Firm-day
CSMAR-Financial Statement File Merged with Stock Price File <i>Less firm-days with less than 3 posts</i>	3,171 (132)	4,421,222 (136,019)
Social Media with no less than 3 posts/day <i>Less firm-days with no news articles</i>	3,039 -	4,285,203 (3,273,687)
Traditional Media with no less than 1 news/day <i>Less firm-days with missing control variables</i>	3,039 (28)	1,011,516 (40,534)
Final Sample	3,011	970,982

This table presents the sample selection process of our firm-day observations used in the main tests. We require each firm-day observations to have at least three social media posts to avoid measurement errors of the social media tone caused by insufficient posts per day. To assess the tone of traditional media on the same day, we require at least one traditional media news article for a firm-day to be included in our sample. Results are robust when we increase the threshold to five social media posts per day, despite a reduced sample size. Results are also robust when we increase the threshold on the daily traditional media news articles to 2,3, and 4 articles per day.

Table 1- Panel B Descriptive Statistics

	<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>P25</i>	<i>Median</i>	<i>P75</i>
Traditional Media	<i>Number of news</i>	970,982	3.8953	9.2842	1.0000	1.0000	3.0000
	<i>Traditional media tone</i>	970,982	0.3704	0.4947	0.0742	0.4658	0.7958
	<i>Party media tone</i>	124,616	0.5115	0.4998	0.2682	0.7000	0.9182
	<i>Non-party media tone</i>	970,982	0.1602	0.4021	0.0000	0.0000	0.3979
	<i>Traditional media positive</i>	970,982	0.7790	0.4149	1.0000	1.0000	1.0000
Social Media	<i>Social media tone</i>	970,982	-0.2028	0.3145	-0.4118	-0.1778	0.0000
	<i>Number of posts</i>	970,982	39.4893	94.8157	3.0000	16.0000	43.0000
Stock Return	<i>Raw return</i>	970,982	0.0014	0.0339	-0.0157	0.0007	0.0169
	<i>Abnormal return</i>	970,982	0.0012	0.0275	-0.0130	-0.0017	0.0115
	<i>CAR[1,5]</i>	970,982	0.0006	0.0600	-0.0301	-0.0041	0.0253
	<i>CAR[1,10]</i>	970,982	0.0016	0.0821	-0.0416	-0.0054	0.0368
	<i>CAR[1,20]</i>	970,982	0.0036	0.1122	-0.0587	-0.0062	0.0551
Firm Fundamentals	<i>SOE</i>	18,499	0.4278	0.4948	0.0000	0.0000	1.0000
	<i>Size</i>	18,499	21.9302	1.3771	20.9622	21.7418	22.6726
	<i>ROA</i>	18,499	0.0539	0.0809	0.0142	0.0422	0.0834
	<i>Market to Book</i>	18,499	3.8148	3.8706	1.7172	2.7480	4.5508
	<i>Leverage</i>	18,499	0.2125	0.3827	-0.0000	0.0342	0.2606

This table present descriptive statistics on the measures of social media, traditional media and daily stock returns for the sample of 970,982 firm-day observations from the period of July 2009 to 2016. Each firm-day observation must have 1) no less than three social media posts and 2) no less than one traditional media news article. Party media tone has smaller number of observations because party-affiliated news media publish less frequently than non-party-affiliated news media, resulting in observations with no news articles by party-affiliated media. Firm fundamentals are firm-year measures of firm characteristics at each fiscal year end.

Table 2: Correlation Table

	1	2	3	4	5	6	7	8	9	10	11
1 <i>number of news</i>											
2 <i>traditional media tone</i>	-0.0184*										
3 <i>party media tone</i>	-0.0048*	0.2826*									
4 <i>non-party media tone</i>	-0.0199*	0.5905*	-0.1036*								
5 <i>traditional media positive</i>	0.0699*	0.8217*	0.2093*	0.4903*							
6 <i>social media tone</i>	-0.0113*	0.0523*	0.0042*	0.0446*	0.0459*						
7 <i>number of posts</i>	0.1523*	-0.0470*	-0.0301*	-0.0155*	-0.0221*	-0.0382*					
8 <i>raw return</i>	-0.0029*	0.0460*	-0.0009	0.0492*	0.0408*	0.2115*	-0.0125*				
9 <i>abnormal return</i>	-0.0028*	0.0524*	-0.0025	0.0582*	0.0474*	0.2183*	-0.0089*	0.8226*			
10 <i>CAR[1,5]</i>	-0.0059*	0.0109*	0.0041*	0.0067*	0.0087*	0.0324*	-0.0439*	0.0593*	0.0639*		
11 <i>CAR[1,10]</i>	-0.0083*	0.0098*	0.0044*	0.0049*	0.0084*	0.0305*	-0.0505*	0.0468*	0.0530*	0.7280*	
12 <i>CAR[1,20]</i>	-0.0109*	0.0079*	0.0051*	0.0043*	0.0052*	0.0274*	-0.0602*	0.0311*	0.0357*	0.5338*	0.7257*

This table presents the Pearson's correlation coefficients of variables used in the regressions. *denote that correlation coefficients are significant at 0.05 level.

Table 3: Sample Distribution

Panel A- by Year

Sample Distribution - By Year

Year	Freq.	Percent	Cum.
2009	121,729	12.54%	12.54%
2010	127,999	13.18%	25.72%
2011	129,894	13.38%	39.10%
2012	135,332	13.94%	53.03%
2013	135,203	13.92%	66.96%
2014	117,543	12.11%	79.06%
2015	101,934	10.50%	89.56%
2016	101,347	10.44%	100.00%

Total	970,982	100%	100%
-------	---------	------	------

Table 3: Sample Distribution

Panel B- by Industry			
Industry	Frequency	Percentage	Cumulative
Computer and Communications	66,710	6.87	6.87
Real Estate	64,572	6.65	13.52
Pharmaceutical	51,478	5.30	18.82
Electrical Manufacture	45,168	4.65	23.47
Automotive	36,948	3.81	27.28
Chemical Products	36,397	3.75	31.03
Retails	36,032	3.71	34.74
Financial Service	32,730	3.37	38.11
Specialized Equipment Manufacture	32,630	3.36	41.47
Business Service	31,977	3.29	44.76
Software and Information Technology	29,581	3.05	47.81
Alcoholic Beverage, Non-alcoholic Beverage and Tea	27,605	2.84	50.65
Wholesale	24,947	2.57	53.22
Construction	24,101	2.48	55.70
Electricity and Heat Supply	24,088	2.48	58.18
Non-metallic Mineral	23,029	2.37	60.56
General Equipment Manufacture	20,766	2.14	62.70
Ferrous Metal Smelting	20,560	2.12	64.81
Non-Ferrous Metal Smelting	19,740	2.03	66.85
Water Transportation	17,568	1.81	68.65
Aero Transportation	17,493	1.80	70.46
Transportation Equipment Manufacture	16,846	1.73	72.19
Business Service	15,746	1.62	73.81
Road Transportation	14,345	1.48	75.29
Coal Mining and Washing	13,477	1.39	76.68
Internet Service	12,649	1.30	77.98
Food Manufacture	10,856	1.12	79.10
Metallic Product Manufacture	10,286	1.06	80.16
Agriculture	9,859	1.02	81.17
Others	182,798	18.82	100
Total	970,982	100	

Table 4: Social Media’s Delineation of Traditional Media Bias

Dependent Variable: Social Media Tone

<i>Independent Variables</i>	1	2	3	4	5
Traditional Media Tone	0.0331*** (51.43)	0.0306*** (43.68)	0.0711*** (23.62)	0.0415*** (16.67)	0.0426*** (17.11)
Traditional Media Tone x TM_Positive			-0.0676*** (-19.99)	-0.0224*** (-7.84)	-0.0235*** (-8.23)
TM_Positive			0.0145*** (8.32)	0.0056*** (3.97)	0.0056*** (3.95)
Size					-0.0327*** (-24.78)
ROA					0.0429*** (5.84)
Market to Book					-0.0025*** (-13.64)
Leverage					0.0262*** (12.52)
Adj-R2	0.003	0.092	0.005	0.092	0.093
N	970,982	970,982	970,982	970,982	970,982
Firm-Fixed Effect	NO	YES	NO	YES	YES
Year-Month-Fixed Effect	NO	YES	NO	YES	YES

This table presents the results of OLS regressions of the tone of social media on the tone of traditional media of the same firm on the same day using Model (1) – Section 4.1. The dependent variable is the daily Social Media Tone, which is measured using textual analysis of the posts on East Money Guba, a major social media platform dedicated to the discussion of financial news related to a firm. Traditional Media Tone is the mean tone of traditional media news articles extracted using textual analysis. TM_Positive is a dummy variable set to 1 if Traditional Media Tone is >0. Standard errors are estimated by two-way clustering at firm and year-month level with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Social Media’s Delineation of Traditional Media Bias: Pre and Post Media Intervention

Dependent Variable: Social Media Tone

	(1)	(2)	Difference
<i>Independent Variables</i>	Pre-Intervention	Post-Intervention	(2)-(1)
Traditional Media Tone	0.0425*** (16.26)	0.0360*** (5.99)	-0.0065 (-1.02)
Traditional Media Tone x TM_Positive	-0.0209*** (-7.00)	-0.0352*** (-5.15)	-0.0143** (-3.96)
TM_Positive	0.0052*** (3.44)	0.0098*** (2.75)	
Size	-0.0204*** (-16.96)	-0.0805*** (-14.61)	
ROA	0.0096 (1.55)	0.0404 (1.49)	
Market to Book	-0.0010*** (-6.09)	-0.0059*** (-14.13)	
Leverage	0.0227*** (12.62)	0.0516*** (7.99)	
Adj-R2	0.084	0.095	
N	802,523	168,459	
Firm-Fixed Effect	YES	YES	
Year-Month-Fixed Effect	YES	YES	

This table presents the result of OLS regressions of the tone of social media on the tone of traditional media of the same firm on the same day in two sub-sample period: pre and post media intervention in which traditional media are required to decrease the coverage of stock market-related news, and stop using negative words such as “tumble” or “crash” in the news. Pre-intervention period is before or on July 23rd, 2015, and post-intervention period after this date. We use Chow Test and report Chi-square Statistics to test the equality of two coefficients across sample. Standard errors are estimated by two-way clustering at firm and year-month level with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Social Media's Delineation of Traditional Media Bias: National Congress Meeting Period vs Non-National Congress Meeting Period

Dependent Variable: Social Media Tone		
	1	2
<i>Independent Variables</i>	Congress Meeting	Non-Congress Meeting
Traditional Media Tone	0.0206 (1.55)	0.0431*** (17.02)
Traditional Media Tone x TM_Positive	0.0094 (0.62)	-0.0248*** (-8.51)
TM_Positive	0.0021 (0.29)	0.0058*** (4.03)
Size	-0.0436 (-1.48)	-0.0331*** (-14.86)
ROA	-0.0252 (-0.01)	0.0468*** (6.26)
Market to Book	-0.0738 (-1.11)	-0.0025*** (-13.41)
Leverage	0.3566*** (3.29)	0.0266*** (12.59)
Adj-R2	0.196	0.092
N	41,110	929,872
Firm-Fixed Effect	YES	YES
Year-Month-Fixed Effect	YES	YES

This table presents the result of OLS regressions of the tone of social media on the tone of traditional media of the same firm on the same day in two sub-sample periods: Congress Meeting Period and Non-Congress Meeting Period. Congress meeting sample includes firm-days in a window of -45 to +45 days relative to the opening day of the 18th CCP National Congress Meeting (Nov.8th, 2012), and non-congress meeting sample includes the rest of the firm-day observations. Standard errors are estimated by two-way clustering at firm and year-month level with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Number of Social Media Post/ Traditional Media News Articles during Congress Meeting Period

	1	2
	Dependent Variables	
<i>Independent Variables</i>	Number of Social Media Posts	Number of News Articles
Congress Meeting	-12.3525*** (-7.87)	0.3003*** (4.09)
Number of News	1.5869*** (4.19)	
Number of Social Media Posts		0.0150*** (7.21)
Social Media Tone	-0.1680 (-0.31)	-0.1510*** (-2.66)
Traditional Media Tone	-3.2516*** (-5.20)	-0.3299*** (-6.13)
Stock Return	-27.0297 (-1.50)	1.7709** (2.48)
Index Return	32.4363 (1.31)	2.1961 (0.97)
Size	-17.2684* (-1.89)	0.1506 (0.52)
ROA	17.7713 (0.41)	1.8110** (1.99)
Market to Book	-2.1499* (-1.88)	0.0292 (1.59)
Leverage	8.6415* (1.67)	0.0605 (0.26)
Adj-R2	0.412	0.791
N	970,982	970,982
Firm-Fixed Effect	YES	YES

This table present the result of regression of the relationship between level of activities for traditional media and social media and congress meeting. The dependent variable is the number of social media posts/ traditional media news articles. We estimate a pooled OLS regression model to determine the impact of the Congress Meeting on the level of activities proxied by the number of social media posts and the number of new articles. Standard errors clustered at firm with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 8: Market Responses to Traditional Media’s Bias as Delineated by Social Media

	1	2	3
	CAR(1,5)	CAR(1,10)	CAR(1,20)
<i>Independent Variables</i>			
Traditional Media Tone	0.0009*** (5.23)	0.0007*** (2.82)	-0.0001 (-0.21)
Traditional Media Tone x TM Bias	-0.0042*** (-5.47)	-0.0047*** (-4.41)	-0.0067*** (-4.43)
TM Bias	0.0035*** (5.58)	0.0045*** (5.14)	0.0075*** (5.99)
Social Media Tone	0.0056*** (21.24)	0.0069*** (18.47)	0.0080*** (15.13)
Size	-0.0052*** (-14.17)	-0.0099*** (-16.05)	-0.0189*** (-19.44)
ROA	-0.0026 (-1.25)	-0.0042 (-1.21)	-0.0124** (-2.21)
Market to Book	-0.0006*** (-10.63)	-0.0011*** (-12.42)	-0.0022*** (-15.11)
Leverage	0.0033*** (6.03)	0.0062*** (6.76)	0.0126*** (8.46)
CAR [-5,-1]	0.0031* (1.91)	0.0044** (2.13)	-0.0024*** (3.79)
Adj- R2	0.0143	0.0223	0.0345
N	970,982	970,982	970,982
Firm Fixed Effect	YES	YES	YES
Year-Month Fixed Effect	YES	YES	YES

This table reports the OLS regression estimates of traditional media tone of Day 0 on the cumulative abnormal return (CAR) of three windows. Traditional Media Tone is the mean tone of all traditional media news articles covering the same firm on day 0; TM Bias is an indicator variable set to 1 if the distance between traditional media and social media calculated as (Traditional Media Tone – Social Media Tone) is in the 1st quintile of all firms on the same day. Social Media Tone is the mean tone of all social media articles covering the same firm on day 0. Standard errors are estimated by two-way clustering at firm and year-month level with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Social Media’s Delineation of Traditional Media Bias- SOE vs Non-SOE: Pre and Post Media Intervention

Dependent Variable: Social Media Tone

<i>Independent Variables</i>	Pre-Intervention	Post - Intervention	Full Sample
Traditional Media x TM_Positive x SOE	-0.0114** (-2.12)	0.0186 (0.95)	0.0022 (0.47)
TM_Positive x SOE	0.0076*** (2.76)	-0.0120** (-2.08)	0.0021 (0.83)
TM_Positive x Traditional Media	-0.0222*** (-5.65)	-0.0538*** (-7.36)	-0.0327*** (-9.44)
SOE x Traditional Media	-0.0067 (-1.40)	-0.0098 (-1.00)	-0.0090** (-2.09)
TM_Positive	0.0027 (1.30)	0.0231*** (5.78)	0.0076*** (4.11)
SOE	-0.0343*** (-14.91)	0.0016 (0.34)	-0.0280*** (-13.48)
Traditional Media	0.0552*** (16.11)	0.0500*** (7.72)	0.0548*** (18.07)
Firm-Level Control Variables	YES	YES	YES
Adj-R2	0.051	0.019	0.077
N	802,523	168,459	970,982
Industry-Fixed Effect	YES	YES	YES
Year-Month-Fixed Effect	YES	YES	YES

This table presents the result of OLS regressions of the relationship between the tone of social media and the tone of traditional media of the same firm on the same day in two sub-sample period, and across SOE and non-SOE: pre and post media intervention in which traditional media are required to decrease the coverage of stock market-related news, and stop using negative words such as “tumble” or “crash” in the news. Pre-intervention period is before or on July 23rd, 2015, and post-intervention period after this date. SOE is a dummy variable set to 1 if the firm is ultimately controlled by either the central or local government. TM_Positive is a dummy variable set to 1 if Traditional Media Tone is >0. All regression includes the following control variables: Size, Market to Book, Leverage, ROA. Standard errors are estimated by two-way clustering at firm and year-month level with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10: Market Responses to Traditional Media and Social Media: SOE and Non-SOE

	1		2		3		4		5		6	
	CAR(1,5)		CAR(1,5)		CAR(1,10)		CAR(1,10)		CAR(1,20)		CAR(1,20)	
<i>Independent Variables</i>	SOE		Non-SOE		SOE		Non-SOE		SOE		Non-SOE	
Traditional Media Tone	0.0004**	0.0012***	0.0008***		0.0004	0.0012***	0.0008**		-0.0004	0.0013***	0.0017***	
	(2.20)	(6.30)	(10.44)		(1.51)	(4.47)	(5.44)		(-1.30)	(3.61)	(12.49)	
Social Media Tone	0.0042***	0.0063***	0.0021***		0.0050***	0.0074***	0.0024***		0.0050***	0.0077***	0.0027***	
	(16.55)	(19.50)	(26.50)		(14.36)	(16.83)	(19.61)		(10.64)	(13.00)	(13.29)	
Size	-0.0053***	-0.0059***			-0.0105***	-0.0106***			-0.0194***	-0.0204***		
	(-18.71)	(-19.46)			(-27.42)	(-25.71)			(-37.24)	(-36.71)		
ROA	-0.0018	-0.0038**			-0.0044*	-0.0059***			-0.0181***	-0.0119***		
	(-1.03)	(-2.46)			(-1.90)	(-2.77)			(-5.68)	(-4.17)		
Market to Book	-0.0007***	-0.0007***			-0.0013***	-0.0013***			-0.0024***	-0.0024***		
	(-16.25)	(-17.27)			(-22.96)	(-23.67)			(-31.59)	(-33.38)		
Leverage	0.0041***	0.0027***			0.0080***	0.0045***			0.0155***	0.0096***		
	(11.12)	(4.66)			(16.19)	(5.79)			(23.01)	(9.16)		
Adj-R2	0.0114	0.0218			0.0205	0.0309			0.0335	0.0469		
N	528,234	442,748			528,234	442,748			528,234	442,748		
Firm Fixed Effect	YES	YES			YES	YES			YES	YES		
Year-Month Fixed Effect	YES	YES			YES	YES			YES	YES		

This table reports the OLS regression estimates of Traditional Media Tone and Social Media Tone of Day 0 on the cumulative abnormal return (CAR) of three windows for SOE and non-SOE sample. Traditional Media Tone is the mean tone of all traditional media news articles covering the same firm on day 0; Social Media Tone is the mean tone of all social media articles covering the same firm on day 0. We use Chow Test and report Chi-square Statistics to test the equality of two coefficients across SOE and non-SOE sample. All regressions include firm fixed effect and year-month fixed effects. Standard errors are estimated by two-way clustering at firm and year-month level with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 11: Social Media’s Delineation of Traditional Media Bias: Party Media and Non-Party Media

Dependent Variable: Social Media Tone Day 0

	1	2	Diff.
<i>Independent Variables</i>	Party Media	Non-Party Media	(2) – (1)
Traditional Media Tone	0.0192** (2.50)	0.0586*** (17.70)	0.0394*** (8.57)
Traditional Media Tone x	-0.0145*	-0.0392***	-0.0247
TM_Positive	(-1.70)	(-11.33)	(0.974)
TM_Positive	0.0038 (0.82)	0.0054*** (3.19)	
Size	-0.0443*** (-14.08)	-0.0326*** (-24.65)	
ROA	0.0623*** (3.39)	0.0443*** (6.03)	
Market to Book	-0.0036*** (-7.84)	-0.0025*** (-13.55)	
Leverage	0.0318*** (6.49)	0.0261*** (12.49)	
Adj-R2	0.115	0.093	
N	124,616	970,982	
Firm-Fixed Effect	YES	YES	
Year-Month-Fixed Effect	YES	YES	

This table presents the result of OLS regressions of the relationship between the tone of social media and the tone of Party-affiliated/Non-Party affiliated traditional media of the same firm on the same day. TM_Positive is a dummy variable set to 1 if Traditional Media Tone (Party-Affiliated or Non-Party Affiliated) is >0. The list of Party-Affiliated Media can be found in Pitroski et al. (2017). We use Chow Test and report Chi-square Statistics to test the equality of two coefficients across the two sample. Standard errors are estimated by two-way clustering at firm and year-month level with t-statistics presented in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix I

Traditional Media Tone and Traditional Media – Social Media Tone Difference

Traditional Media Tone Decile	Mean of TM Bias (Traditional Media Tone- Social Media Tone)	Standard Deviation	Frequency	
1	-0.3794	0.3760	98,421	▲
2	0.0287	0.3535	97,206	TM Negative
3	0.2889	0.3410	97,943	TM Positive
4	0.4654	0.3279	94,736	▼
5	0.6061	0.3314	96,345	
6	0.7229	0.3316	97,796	
7	0.8421	0.3340	97,993	
8	0.9648	0.3337	98,644	
9	1.0683	0.3323	97,264	
10	1.1415	0.3294	94,634	
Total	0.5732	0.5722	970,982	

Appendix II Variable Definitions

Abnormal Return = Index adjusted return, equals to the raw return, which equals to (closing price – opening price)/opening price, of a firm minus the index return on the same day (Shanghai-index, Shenzhen-index, Small Cap-index, and GEM-index), which can be expressed as

$$AbRet_{i,t} = Ret_{i,t} - Ret_{M,t}$$

CAR[x,y] = Cumulated abnormal returns of a stock from day x to day y , calculated as $CAR(x,y) = \sum_{t=x}^{t=y} AbRet_{i,t}$

Congress Meeting = Indicator variable set to 1 if a firm day is within a window of -45 to +45 days relative to the opening day of the 18th CCP National Congress Meeting

Index Return = The raw return of Shanghai Stock Exchange daily Index

Leverage = Ratio of the firm's long-term debt to shareholders' equity at the end of year t

Market to Book = Ratio of the firm's market value of equity to book value of equity at the end of year t

Number of News = Number of traditional media news articles covering a firm on a given day

Number of Posts = The total number of social media articles related to a firm posted on a given day

Non-Party Media Tone = Mean of all news articles published by a traditional media that is not under direct control of the CCP party.

Party Media Tone = Mean of all news articles published by a traditional media that is under direct control of the CCP party, for a complete list of party-affiliated media, please see Piotroski (2017) for detail.

ROA = Firm's return on assets, measured as net income in year t scaled by total assets at the end of fiscal year t

SIZE = Natural Logarithm of the firm's total market value of equity at the end of fiscal year t

Stock Return = The raw return of a stock calculated as (closing price – opening price) / opening price

SOE = Indicator variable equal to one if the company is ultimately controlled by either the central or local government; a firm is considered state controlled if the state owns more than 20 percent of its equity in year t

Social Media Tone = for a firm day, social media tone is calculated as (Number of positive social media posts – Number of negative social media posts) / Total Number of social media posts.

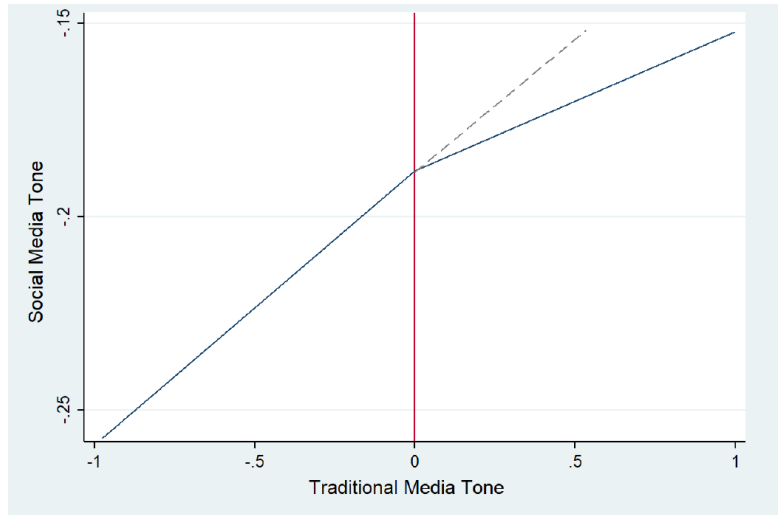
TM_Bias = Indicator variable set to 1 if the distance between traditional media and social media calculated as (Traditional Media Tone – Social Media Tone) is in the 1st quintile of all firms on the same day

Traditional Media Tone = Mean of the tone of all traditional media news articles covering a firm on a given day. For measurement of article tone using textual analysis, please see section 3.1 for detail.

Traditional Media Positive = A dummy variable set to 1 if Traditional Media Tone is greater than 0; set to 0 otherwise.

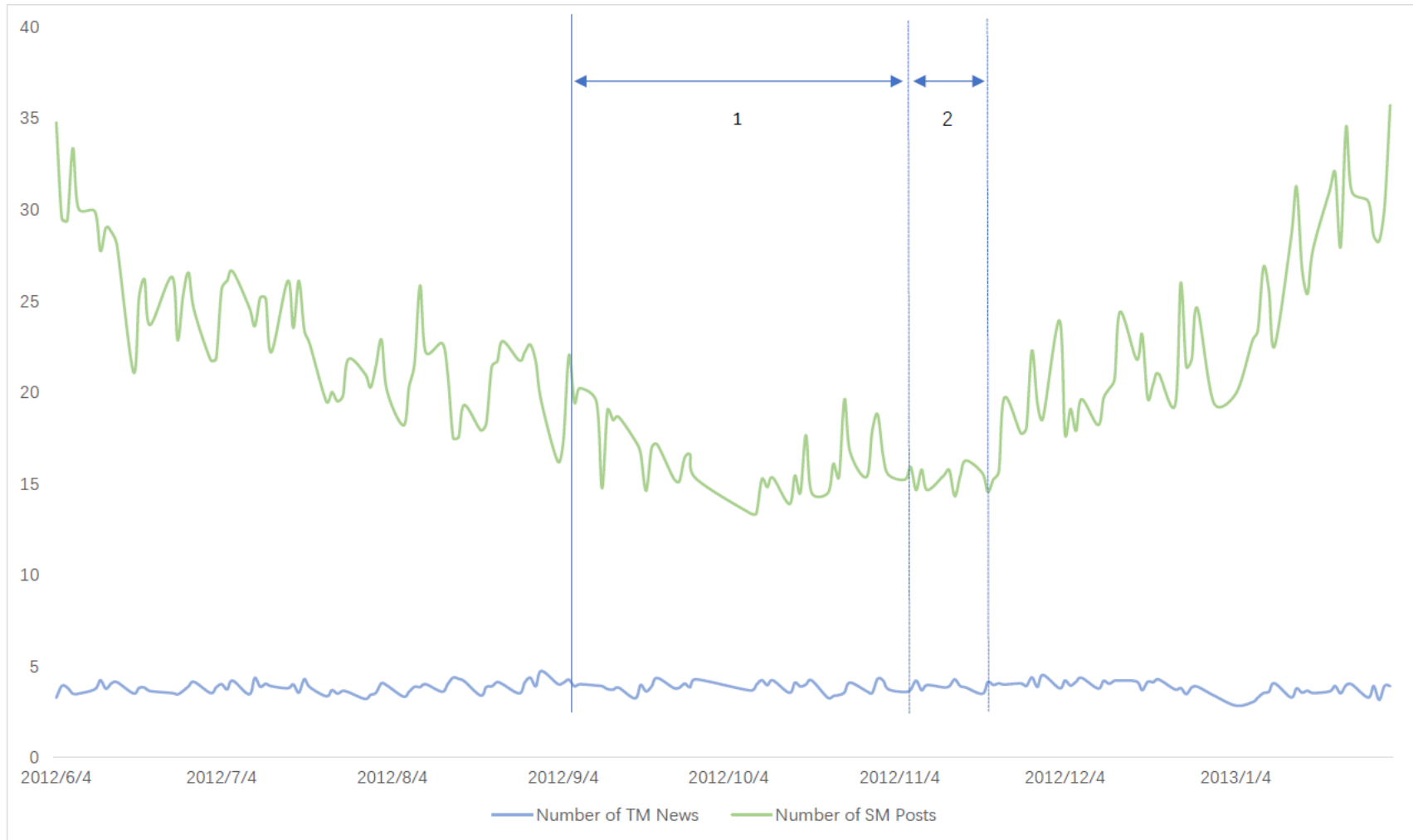
Figure 1

Linear Prediction Plot of Association Between the Tone of Social Media and the Tone of Traditional Media*



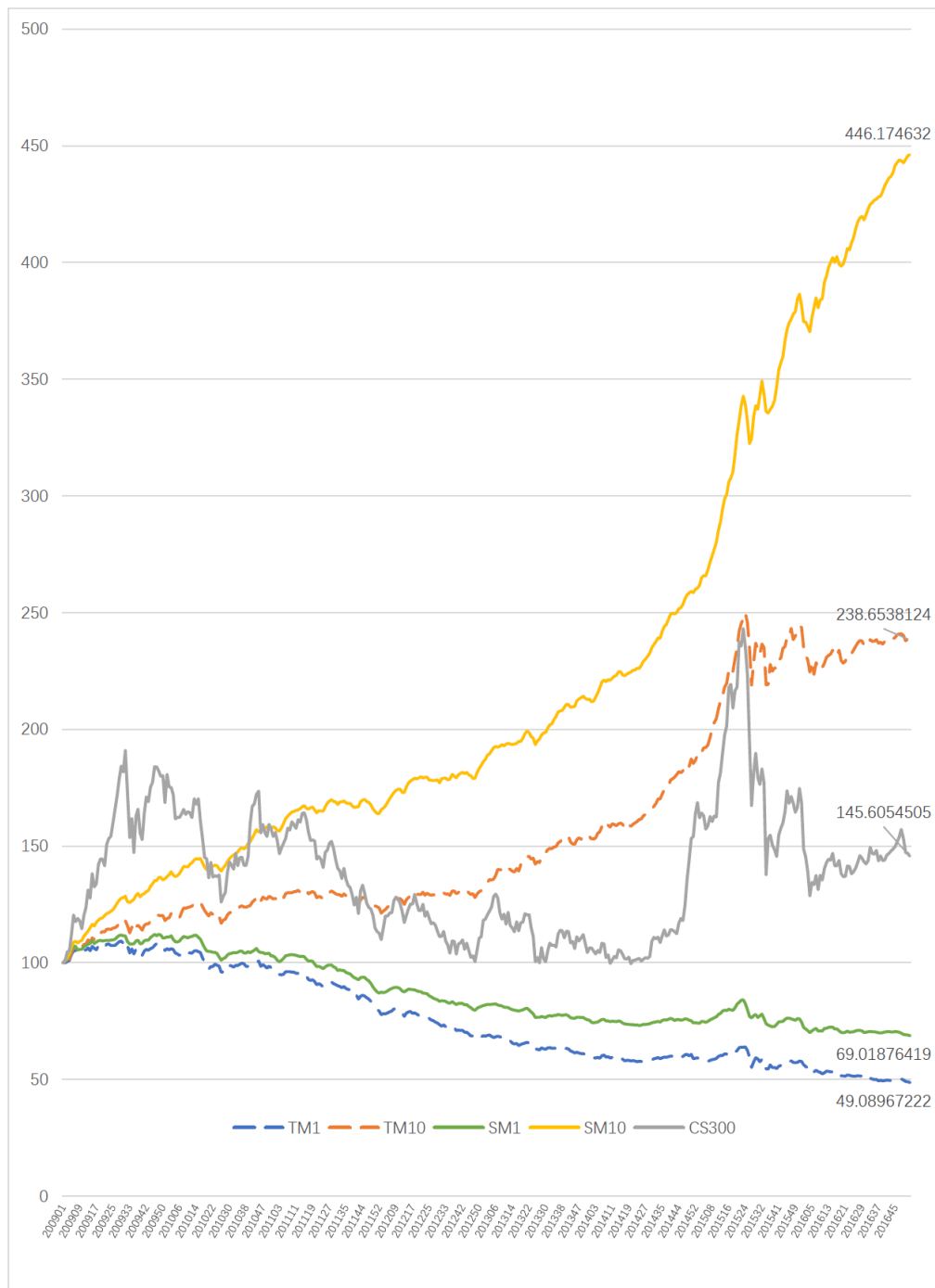
*This figure presents the linear prediction for Social Media Tone from a linear regression of Social Media Tone on Traditional Media Tone and plots the resulting line with 95% confidence interval.

Figure 2: The Daily Number of News Articles and Social Media Posted Around 18th CCP National Congress Meeting



This figure presents the number of social media posts and number of traditional media news articles per firm-day before and after the 18th CPC national congress meeting period. Period 1 represents the pre-meeting period (60 days prior to the opening date); Period 2 represents the meeting period.

Figure 3: Net Value of Weekly-Adjusted Stock Portfolios Formed Based on Tone of Traditional Media and Tone of Social Media



This figure presents the net value of weekly-adjusted portfolios from 2009 to 2016, with a beginning value of 100, in comparison with China Stock Index 300 (CS300). Each week, stocks are ranked based on their average tone of traditional media (social media) in the previous week (from last Monday to last Friday) into ten deciles. Equal-weighted portfolios are formed by buying stocks that fall into the 10th and the 1st deciles of last week's traditional media tone (social media tone), these two portfolios are labeled TM10 (SM10) and TM1 (SM1), respectively. Portfolios are adjusted weekly, and we impose a hypothetical transaction cost of 5 bps on each weekly adjustment.