The Value of Social Media Anonymity: Evidence from the Stock Market*

Huasheng Gao
Fanhai International School of Finance
Fudan University
huashenggao@fudan.edu.cn

Yongxian Tan
School of Accounting, Economics and Finance
Curtin University
yongxian.tan@curtin.edu.au

Hailong Zhao
School of Accounting
Southwestern University of Finance and Economics
zhaohl@swufe.edu.cn

This Version: April 2024

Abstract: We examine the value implication of social media anonymity in the context of the stock market. Exploiting China's regulatory change that removes anonymity and requires real name registration in social media, we find that such a policy leads to a significant decrease in firm value, especially when investors rely more on social media for stock information. We further show that the treatment leads to lower stock price informativeness, higher stock crash risk, poorer corporate disclosure, and more earnings manipulation. Overall, we provide evidence that removing social media anonymity decreases firm value through the channel of worsening firms' information environment.

Key Words: Firm value; Social media anonymity; Information environment

JEL: G28; G32; G17; K24

* We are grateful for the helpful comments from Jason Huafeng Chen, Zhiguo He, Sheng Huang, Dan Li, Yiqing Lv, Jun Pan, Jun Qian, Jiangping Qiu, Songzhi Song, Nancy Su, Dragon Tang, Jean Tirole, Jun Tu, Frank Yu, Renxuan Wang, Yuxi Wang, Qi Zhang, and the seminar participants from CEIBS, NYU Shanghai, Shanghai Jiaotong University, the First Fanhai-Haitong Finance Workshop and the First Fudan-TSE Economics and Finance workshop. All errors are our own.

1. Introduction

Over the past two decades, there has been an intense debate on whether anonymity in social media should be eliminated. One side argues that anonymity should be eliminated because it facilitates abuse, illegal activity, and the spread of fake news on the internet. The other side feels that anonymity is vital for the protection of freedom of expression and helps to achieve better information dissemination (Lidsky and Cotter. 2007; Moore, 2009; Choi, 2012; Kaminski, 2012; Lee and Liu, 2016; Huang, Li, and Markov, 2020). Despite various possible pros and cons associated with internet anonymity, there is a lack of quantitative evidence identifying its impact on society, possibly because many aspects of its impact are difficult to quantify. In this paper, we shed insight on this issue by focusing on the stock market. Stock market provides a good setting to examine this question because (1) the richer data in this market helps to quantify any value implication in a reasonably accurate way (2) investors are increasingly relying on social media for financial information (Antweiler and Frank, 2004; Das and Chen, 2007; Chen et al., 2014; Ang et al., 2021).

Ex ante, the effect of social media anonymity on firm value could go either way. On one hand, anonymity may encourage the spread of false information because of the difficulty in holding individuals accountable for anonymous posts. False news may make firms' environment opaque and subject investors to greater uncertainty, which could in turn decrease firm value. On the other hand, people are more likely to share important but sensitive information when they are free of concerns over privacy and when they have improved protection from detection, retribution, and embarrassment. Hence, anonymous communication can facilitate the flow of information by protecting privacy and freedom of expression. Free communication of information (especially negative information) in social media could also push firm managers to increase their disclosure and prevent them from hiding and

accumulating negative information (Xu and Zhang, 2013; Cade, 2018; Dube and Zhu, 2021). Thus, anonymity in social medias can enrich firms' information environment faced by investors, which could in turn enhance firm value.

In this study, we identify the effect of social media anonymity on firm value based on the exogenous shock to social media anonymity generated by China's regulatory change toward a real name registration system. In February 2015, the Cyberspace Administration of China (CAC) issued Provisions on the Administration of Account Names of Internet Users (hereafter referred to as the Account Names Provisions). Account Names Provisions require that internet information service providers demand their users to register accounts after undergoing real identity information authentication, effectively imposing real name policies on all internet services in China. Taking advantage of this unique setting, we attempt to provide the first empirical study on the value implication of social media anonymity by examining the impact of removing anonymity in internet stock message boards on firm value.

We focus on China mainly because the regulatory change in China provides us an unique empirical setting for the following reasons. First, real name registration in China was initiated by the government for tightening control of the Internet via a strengthened regulation of Internet account registration. This policy change is not aimed at affecting the stock market; therefore, any potential effect on firm value is likely to be an unintended consequence. Second, because all domestic internet service providers are required to adopt real name polices, Chinese users of internet stock message boards cannot switch to other anonymous platforms to release sensitive information. Third, the Chinese stock market is characterized with high retail investor participation (Carpenter, Lu and Whitelaw, 2015; Liu, Stambaugh and Yuan, 2019). This enhances the power of our tests because retail investors are

more likely to rely on social media for information. It is also worth noting that such a setting is not readily available in many other countries. Taking the U.S. for example, there is no national-wide or state-wide regulation on real name registration. And very few mainstream social media platforms in the U.S. adopt a genuine real name system. Even some particular platforms require real name registration, users can easily switch to other platforms.[1]

We conduct our analysis in a difference-in-differences (DiD) framework using posts data from Eastmoney Guba. As the first and most popular stock message board in China (e.g., Hong et al., 2014; Huang, Qiu and Wu, 2016; Ang et al., 2021), Guba hosts a firm-specific stock message board for each listed firm on the Shanghai and Shenzhen Stock Exchanges. Before February 2015, nearly half of the messages on Guba were posted by anonymous, non-registered users. After the regulatory change, virtually all posts on Guba were made by posters registered under real names. Guba tends to be a more important source of information for firms with more Guba posts. If real name registration has any impact on firm value, firms with more Guba posts before the regulatory change should be affected more. Hence, we calculate the total number of Guba posts about each sample firm in 2014 (the year right before the policy change) and use it as a measure of treatment intensity.

Using a panel of 14,640 firm-year observations from 2010-2019, we find that after real name registration firms with higher treatment intensity experience significantly larger reductions in firm value (measured by Tobin's Q) than firms with lower treatment intensity. On average, an increase in treatment intensity by one standard deviation leads to a 5.44% reduction in firm value, which translates

[1] For example, Facebook and Quora have real name policies, but it is practically impossible for them to force all users to provide identification information. They enforce their real name policies by merely using computer algorithms to flag suspicious accounts. A user who wants to communicate sensitive information anonymously may still sign up accounts with a fake name and an email address. Moreover, a Facebook user can still choose to release sensitive information about a stock anonymously on Twitter.

into RMB 723 million (approximately USD 103 million). Supporting the parallel-trend assumption underlying a causal interpretation of the DiD estimation, we show that the pre-treatment trends of firm value are indeed indistinguishable between the treatment and control groups.

We further investigate the possible channel through which social media anonymity impacts firm value. We first examine the cross-sectional variation in the treatment effect. If real name registration decreases firm value by worsening firms' information environment faced by investors, the impact should be more pronounced in situations where investors rely more on social media for information. Consistent with this conjecture, we find that the treatment effect is indeed more pronounced in firms with poorer analyst coverage, lower institutional ownership, smaller sizes, and lower asset tangibility.

Next, we provide direct evidence that real name registration impairs firms' information environment. First, we demonstrate that the treatment firms experience a larger reduction in Guba posts than the control firms following the treatment. This evidence indicates that the removal of user anonymity discourages the exchange of firm-specific information, thereby reinforcing supports the relevance condition of this policy change. Second, we observe a greater decrease in negative posts compared to positive posts. This evidence is consistent with the view that sharing negative information is potentially riskier and thus anonymity is particularly effective to protect the posters (Froomkin, 1995; Lidsky and Cotter, 2007). Furthermore, we find that the treatment policy leads to lower stock price informativeness, higher stock crash risk, poorer corporate disclosure, greater real earnings management, and more financial restatements. Taken together, these findings support the mechanism by which real name registration decreases firm value — by worsening firms' information environment.

We conduct two robustness checks on our main results. First, there could an alternative explanation: Firms with more Guba posts in the pre-event period were overvalued and thus experience

a greater decline in firm value subsequently in the post-event period. We show that our inference is unchanged when we measure treatment intensity based on the number or fraction of negative posts (rather than the total number of posts). Because firms with more negative posts are less likely to be overvalued (in fact, they are likely undervalued), this evidence alleviates this overvaluation-based explanation. Second, we show that our inference is largely unchanged when we performance DiD analysis using a propensity score matched sample. This result helps to mitigate the concern that our main results are driven by the difference in firm characteristics.

This study contributes to the debate over online anonymity. The dilemma of online anonymity can be traced back to the early days of the Internet. In November 1997, the American Association for the Advancement of Science (AAAS) organized a conference on anonymous communications on the Internet. While conference participants agreed that anonymous communication is a constitutional right protected by the First Amendment, they acknowledged that online anonymity offers both opportunities and risks related to how people exchange information. The existing evidence on the net effect of online anonymity on society is limited, possibly due to the challenge of quantifying these effects in various social contexts. In this study, we shed light on this issue by focusing on the financial market, where we can use stock information to identify possible value implication. To the best of our knowledge, we provide the first empirical evidence that the removal of social media anonymity imposes significant costs in the financial market by inhibiting the free communication of information.

Second, our research also extends the growing field of research that examines the role of social media in financial markets. Previous research shows that social media influences investors' trading behavior (Hong et al., 2014), market volatility (Antweiler and Frank, 2004), future firm performance (Chen et al., 2014; Bartov, Faurel and Mohanram, 2018; Tang, 2018), and corporate information

dissemination (Blankespoor, Miller and White, 2014; Miller and Skinner, 2015). Our study complements the above research by providing evidence that anonymity is a crucial feature for social media to harness the wisdom of crowds and provide valuable information.

2. Institutional Background

The Chinese government's earliest attempt to remove social media anonymity can be traced back to 2003, when it ordered internet cafes to collect customers' identification information. The purported goal of this real-name identification requirement was to fight crime and to protect minors. In the decade that followed, law makers, government agencies and industry associations took more policy initiatives toward a real name system. For example, in 2004, the Ministry of Education required that all university-affiliated message boards strictly enforce real name registration policies. In 2007, internet companies in China signed the China Internet Industry Self-Discipline Pact, which strongly encouraged an end to anonymity for bloggers. In February 2012, in order to comply with a new regulatory requirement by the Beijing Municipal Government, China's major microblog companies, including Sina, Sohu, NetEase, and Tencent, announced that they would implement real name policies. These real name policies required that microblog users register with their real names, but the username that showed on their homepage did not have to be their real legal names.

In December 2012, the Standing Committee of the National People's Congress passed a new law titled Decision on Strengthening Online Information Protection (hereafter referred to as Online Information Decision). Article 6 of the Online Information Decision mandates that "network service providers should require their users to supply true identification information when signing an agreement to provide them with website access, fixed line telephone, mobile phone access, or allow them to post information via the network." While the Online Information Decision made real-name

registration requirement a legal obligation for most online service providers in China, initially neither the government nor the online service providers strictly enforced the real name registration rule. Many internet message boards, such as Eastmoney Guba, still allowed users to post message anonymously.

The institutional reform plan of the State Council, which was issued in 2013, required that the Ministry of Industry and Information Technology, Cyberspace Administration of China (CAC) and the Ministry of Public Security introduce and implement the real name registration system by July 2014. Yet, the plan was delayed. In January 2015, the CAC announced that it would comprehensively promote the real name registration system in 2015 (Li, 2015).

In February 2015, the CAC issued the Account Names Provisions. Article 5 of the Account Name Provisions require that internet information service providers demand internet information service users to register accounts after undergoing real identity information authentication. This effectively imposed real name policies on all internet services, including internet message boards, in China. In April 2015, the CAC announced a new system of regulatory talk, which could be used to supervise online service providers and urge them to rectify their unlawful practices. In response to these regulatory changes, Guba made real name registration mandatory.

The 2015 regulatory change provides a unique opportunity to examine the effect of real name registration on firm value because of its three important features. First, the regulatory change was initiated for reasons unrelated to stock market performance. The stated purposes of the Account Name Provisions include “strengthening the administration of account names of internet users and protecting the lawful rights and interests of citizens, legal persons, and other organizations”. Yang and Kang (2016) analyse Chinese blog entries regarding real name registration and find that freedom, privacy and the deterrence of cybercrime are the three board themes. We search extensive for newspaper

articles on the debate over real name registration in China. None of these articles cites financial markets as a major concern. As such, we conclude that the policy change of the real name registrations generates a plausibly exogenous shock to the information exchange in internet message boards. Second, all internet service providers in China were required to adopt real name polices. A certain social media platform cannot choose whether to comply or not. Therefore, message board users cannot easily switch to anonymous platforms to release sensitive information. This makes the impact of real name registration more detectable. Third, the Chinese stock market is characterized with high retail investor participation. According to the annual report of the Shanghai Stock Exchange in 2016, the retail investors comprise 99.8% of total investor accounts, hold 25.2% of stock value and contribute about 86.9% of the yearly trading volume on the exchange.[2] This enhances the power of our test because retail investors are more likely to rely on stock message boards for information.

3. Hypothesis Development

Information plays a crucial role in allocating resources in a capital market economy. All types of investors, from the sophisticated investors to small investors, struggle to acquire and analyse information in order to evaluate the return potential of investment opportunities and monitor the use of their capital resources once they have been allocated. The explicit information processing costs and implicit opportunity costs affect investors' information choices, trades, and market outcomes (Blankespoor, deHaan and Marinovic, 2020). Due to a lack of ability and resources at the individual level to collect and analyse firm information, small investors usually face information disadvantages as compared to corporate insiders and professional investors (La Porta et al., 2000; Bhattacharya et al.,

[2] On average, the retail shareholder aggregate share ownership is 42.87% of shares outstanding on Shanghai Stock Exchange at the end of 2015. As a comparison, Brav, Cain, and Zytnick (2022) find the corresponding number in the U.S. is 26%.

2007; Malmendier and Shanthikumar, 2007). The growth of the internet stock message board, which is an important form of social media in the stock market, has enabled small investors to exchange information and express opinions effectively (Antweiler and Frank, 2004; Das and Chen, 2007). As a priori, how the anonymity in stock message boards affects firm value can go either way. On the one hand, anonymity in the stock message board can increase firm value by improving the information environment faced by investors for the following three reasons.

First, anonymity plays a vital role in facilitating information exchange on social media, irrespective of whether the information pertains to positive or negative aspects of a company. This phenomenon, as noted by Miller and Skinner (2015), weakens the control that firms typically wield over their information environments. On social media platforms, information is typically exchanged in an interactive and collaborative manner. In this process, the participation of individual contributors is crucial for achieving the "wisdom of crowds" (Surowiecki, 2004). The provision of anonymity enables individuals to share information that they might otherwise hesitate to disclose under their real identity due to social, professional, or cultural constraints. This helps to lower barriers to participation and attract a larger and more diverse group of individuals, such as investors, employees, suppliers and competitors, to stock message boards. Thus, anonymity can potentially increase the volume of discussions, foster a more active and engaged community, and mitigate the biases in the information environment.

Second, anonymity especially helps reveal negative information by shielding whistleblowers or activists who want to expose wrongdoing or raise awareness about the negative aspects of firms without risking personal safety or retaliation. Because corporate disclosure tends to be biased towards good news, the timely dissemination of potentially negative information is particularly valuable to

individual investors (Verrecchia, 1983; Jorgensen and Kirschenheiter, 2003; Acharya, DeMarzo and Kremer, 2011). This view is broadly consistent with findings documented in previous research. For example, Lee and Kim (2020) show that employees are likely to be critics of their companies, especially in an anonymous social media platform. Huang, Li and Markov (2020) find that employee social media disclosures, which can strongly predict bad corporate events, are relatively more important as a source of bad news than as a source of good news. According to Ang et al. (2021), social media criticisms contain incremental value-relevant information that can predict a potential acquirer's subsequent decision to withdraw its acquisition attempt.

Third, the improved information exchange enabled by anonymity can encourage managers to provide better corporate disclosure, which is also an important determinant of corporate information environment. Corporate managers generally have incentives to withhold unfavourable news and their voluntary disclosure decisions are influenced by shareholders' access to information (Beyer et al., 2010). When there are limited ways for outside investors to access firm-specific information, self-interested managers are more inclined to release positive news promptly, while obfuscating and withholding negative news (Healy and Palepu, 2001; Li, 2008; Hutton, Marcus and Tehranian, 2009; Kothari, Shu and Wysocki, 2009) because this strategy buys managers time for more favorable developments or allow them to make adjustments to accounting measures (Dye, 1985; 2001). Active information sharing on social media limits the room for information manipulation and increases the detection risk or litigation risk for managers. This in turn discourages managers from hoarding negative information or delaying its disclosure (Skinner, 1994; 1997).

Taken together, the above discussion leads to our first hypothesis (information environment hypothesis):

Hypothesis 1a. Elimination of anonymity in the stock message board decreases firm value.

On the other hand, anonymity in stock message boards come with risks, and one significant risk is the potential dissemination of misinformation on the platforms. False information, if not promptly corrected by other participants, can mislead investors and significantly diminish the overall reliability of the platform. Anonymous users of stock message boards can avoid reputation damage or legal responsibilities by changing their online identities. The existing literature suggests that anonymous users may post deceptive messages, aiming to gain at the expense of other investors. For example, van Bommel (2003), Eren, Ozsoylev and Street (2009) and Jiang, Mahoney and Mei (2005) find that anonymous fraudsters spread misleading information in stock message boards to "pump" up the price of a stock and then "dump" their shares of the stock at the inflated price. Once the fraudsters dump their shares and stop hyping the stock, the stock price typically falls, causing other investors to lose money. Kogan, Moskowitz and Niessner (2021) find that fraudulent articles on Seeking Alpha (SA) and The Motley Fool leads to abnormal trading volume, accompanied by significant short term increases and subsequent reversals in stock prices. Similarly, Mitts (2020) documents that pseudonymous authors who write articles in SA's "short idea" category generate sharp price drops and reversals in target firms. He argues that this pattern is associated with manipulative stock-option trading by these authors. Dyer and Kim (2021) show that investors discount information in anonymous equity research in comparison to nonanonymous research.

Given the above discussion, anonymity in the stock messages boards may make it easier for individuals to spread false information about companies. These messages can manipulate stock prices and mislead investors. Anticipating such risk, investors require a higher return when making the

investment, leading to a lower firm value. This leads to our second alternative hypothesis (manipulation hypothesis):

Hypothesis 1b. Elimination of anonymity in the stock message board increases firm value.

4. Sample and Descriptive Statistics

We obtain accounting information and stock return data from the China Stock Market & Accounting Research (CSMAR) database and Guba posts data from https://guba.eastmoney.com/ (more details about Guba posts data are provided in Appendix B). Our initial sample consists of all non-financial A-share stocks listed on the Shanghai and Shenzhen stock exchanges from 2010 to 2019. We further require that the sample firms have five years of data before and after the regulatory change. This data requirement allows us to compare the same set of firms between the pre- and post-event periods, effectively using each firm as its own control. The resulting sample consists of 14,640 firm-year observations for 1,464 unique firms.

If real name registration has any impact on firm value, firms previously with more Guba posts should be affected more. Based on this rationale, we calculate the total number of posts about a firm in its firm-specific Guba message board in 2014 (i.e., the year immediately before the real name registration policy) and use this number as a measure of the firm's treatment intensity.

We measure firm value by Tobin's Q, defined the sum of the market value of equity and book value of long-term debt divided by total assets (Roll, Schwartz and Subrahmanyam, 2009; Gurun and Butler, 2012). In our regression analysis, we control for firm characteristics that can potentially affect firm value, including firm size, sales growth, operating performance, R&D, capital expenditure, geographic business segments, firm age, state ownership, CEO and director ownership. These control variables are motivated by prior literature (e.g., La Porta et al., 2002; Masulis and Mobbs, 2011). To

minimize the effect of outliers, we winsorize all continuous variables at the 1st and 99th percentiles. Detailed variable definitions are provided in Appendix A.

Table 1 Panel A reports the descriptive statistics for our sample. *Tobin's Q* has a mean of 1.910 and a median of 1.278. *PreLnPost* has a mean of 1.73 and a standard deviation of 0.689, which indicates substantial variation in message board posts across firms in the year prior to the regulatory change. The average company in our sample has firm size of RMB 14,991 million (or approximately USD 2,141 million), sales growth of 0.119, operating performance of 0.05, R&D intensity of 0.001, CAPEX intensity of 0.072, geographic business segments of 5.2, founder director of 0.126, state ownership of 0.578, and is 15.858 years old since IPO. Table 1 Panel B compares the pre-event firm characteristics between the treated and control groups. Compared to the control group, the treated firms are larger, have a higher sales growth, invest more in R&D and capital expenditure, have more segments and greater CEO ownership, and are more likely to be a state-owned company. As pointed out by Roberts and Whited (2013), it is preferable for the treatment and control groups to be relatively similar across observable dimensions prior to the treatment. However, when this similarity is not achieved, one can directly incorporate control variables in the regression, as we do in our later analysis.

5. Results

5.1. Virtual Illustration and Univariate Test

To visualize the impact of real name registration on firm value, we divide the sample firms into two groups: The treatment (control) group consists of firms with higher (lower) total number of posts in their firm-specific Guba message boards than the sample median in 2014. In Figure 1, we plot the mean cumulative excess returns of the treatment and control group firms from 60 months before the regulatory change (i.e., Month 0) to 60 months afterwards. Before the regulatory change, the 60-month

cumulative excess returns of the two group move in tantum with each other: They are -17.00% for the treatment group and -17.32% for the control group, respectively. In the 24 months following the regulatory change, the treatment group firms have substantially lower stock performance than the control group firms. The difference remains stable in the subsequent 36 months. Over the 60 months after the regulatory change, the cumulative excess returns are -46.70% for the treatment group and -34.11% for the control group.

In Table 2, we run a univariate DiD test to examine how the natural logarithm of Tobin's Q (*LnQ*) of the treatment group firms change relative to the control firms after the regulatory change. Column (1) indicates that the average *LnQ* of the treatment group firms decrease from 0.276 in the pre-event period to 0.171 in the in the post-event period, with this change being statistically significant at the 1% level. In column (2), the average *LnQ* of the control firms are statistically indistinguishable between the pre- and post-event periods (0.389 vs. 0.388). Column (3) reports the DiD estimation of *LnQ*, defined as the average difference in *LnQ* change between the treatment and control group firms. The univariate DiD estimate is negative and statistically significant at the 1% level.

Taken together, Figure 1 and Table 2 indicate that after the regulatory change, the treatment group firms experience a significance decline in firm value relative to the control group firms. These results are consistent with the information environment hypothesis, suggesting that the elimination of anonymity in stock message boards decreases firm value.

5.2. Baseline Results

We perform our baseline regression analysis in a general DiD framework as follows:

$$LnQ_{i,t+1} = \alpha + \beta PreLnPost_i \times After_t + \gamma' Z_{i,t} + \delta_t + \varphi_i + \varepsilon_{i,t}. \quad (1)$$

In Equation (1), i indexes firm, and t indexes year. $PreLnPost_i$ is the natural logarithm of the total number of Guba post about firm i in 2014 and is used to measure the treatment intensity for this firm. $After_t$ is an indicator variable that takes the value of one for the post-treatment period (i.e., the 2015-2019 period) and zero for the pre-treatment period (i.e., the 2010-2014 period). The β coefficient is the DiD estimate that captures the causal effect of real name registration on firm value. Z is a vector of control variables that may affect firm value as indicated by previous research (e.g., La Porta et al., 2002; Masulis and Mobbs, 2011), including firm size, sales growth, operating performance, R&D, capital expenditure, geographic business segments, firm age, state ownership, CEO and director ownership. We include year fixed effects, δ_t, to account for time-specific shocks to a firm's value and firm fixed effects, φ_i, to absorb any time-invariant unobservable firm characteristics that may bias the results. In the presence of the year and firm fixed effects, we do not include *PreLnPost* and *After* in the regression model to avoid collinearity issues. We cluster standard errors by firm in all regressions.

Table 3 presents the regression results. In columns (1)–(3), we only include *PreLnPost*×*After,* firm and year fixed effects in the regression, and implement the estimation using two, six and ten years of window surrounding the regulatory change, respectively. In columns (4)–(6), we re-estimate the models in columns (1)–(3) by additionally including the full set of control variables. In all six columns, the coefficients on *PreLnPost*×*After* are negative and statistically significant at the 1% level. The economic magnitude of the effect is sizable. Taking column (6) for example (the time window is from five years prior to the treatment to five years afterwards), the coefficient on *PreLnPost*×*After* is -0.077, indicating that a one-standard-deviation increase in treatment intensity (0.689) is linked to a 5.44% $(= \mathrm{e}^{0.077\times0.689} - 1)$ decrease in Tobin's Q. Considering that the sample average market capitalization

is RMB 13.3 billion, this reduction in Tobin's Q can be translated into a decrease in market capitalization by RMB 723 million (or equivalently USD 103 million).

Because the validity of the DiD estimate depends critically on the parallel trends assumption, we undertake a diagnostic test to check whether this assumption holds. We construct nine indicator variables to flag years relative to the regulatory change. $Before^{n}$ *(n= 1, 2, 3, 4)* indicates that it is the n^{th} years before the regulatory change. Likewise, $After^{n}$ *(n= 1, 2, 3, 4, 5)* indicates that it is the n^{th} year after the regulatory change. In column (7), we re-estimate Equation (1) after replacing the *After* indicator variable with these nine indicators (the fifth year before the treatment serves as the benchmark year). The coefficients of $PreLnPost \times Before^{1}$, $PreLnPost \times Before^{2}$, $PreLnPost \times Before^{3}$ and $PreLnPost \times Before^{4}$ are especially important, because their significance and magnitudes indicate whether there is any difference in the trend of firm value across firms with different treatment intensity prior to the regulatory change. We find that all these coefficients are statistically insignificant and trivial in magnitude, suggesting that the parallel trends assumption is not violated. In contrast, the coefficients for $PreLnPost \times After^{1}$, $PreLnPost \times After^{2}$, $PreLnPost \times After^{3}$, $PreLnPost \times After^{4}$, and $PreLnPost \times After^{5}$ are significantly negative at or below the 5% level and much larger in magnitude. For example, the coefficient on $PreLnPost \times Before^{1}$ is only -0.001, which is statistically indistinguishable from zero. In contrast, the coefficient of $PreLnPost \times After^{5}$ is -0.062, which is statistically significant at the 1% level. Consistent with the parallel trends assumption, the estimation results in column (7) shows that the firms with different treatment intensity share similar trends in firm value prior to the treatment. The lack of significant lead effects has three noticeable implications. First, it suggests that investors did not fully anticipate the policy change toward real name registration. Second, even if some investors did foresee the policy change, the actual information environment

remained unchanged until the policy took effect. Third, the effect of the real name registration on firm value cannot be attributed to policymakers merely reacting to past stock performance, thus mitigating concerns about reverse causality. This conclusion is also consistent with the institutional background in Section 2, which suggests that the policy is implemented mainly for tightening the control of Internet account registration rather than to exert influence on the stock market).

Overall, Table 3 shows that the real name registration policy leads to a significant decrease in firm value, and that such an effect occurs *after* this policy is enacted, which suggests a causal effect. These results support the information environment hypothesis.

5.3. Channel Test 1: Heterogenous Treatment Effect

To provide evidence on the mechanisms underlying our main findings, we examine heterogeneous treatment effects across firms. By doing so, we can further mitigate the concern that some omitted variables are driving our results, because such omitted variables (if any) would need to be uncorrelated with the control variables in the regression and would also need to explain the cross-sectional variation in the treatment effect (e.g., Claessens and Laeven, 2003; Raddatz, 2006).

If real name registration decreases firm value by exacerbating the information environment faced by investors, the treatment effect should be stronger in situations where investors rely more on social media for information. To this end, we examine how the impact of real name registration vary with analyst coverage, institutional ownership, firm size and asset tangibility. Financial analysts serve as an important traditional information source for investors. When a firm is covered by fewer analysts, investors have greater need to obtain information via alternative sources (such as social media) (Campbell, DeAngelis and Moon, 2019; Hales, Moon and Swenson, 2018). Thus, we expect the treatment effect to be more pronounced for firms with lower analyst coverage. Institutional investors

usually have their own research team, subscribe professional research services and enjoy direct access to managers and equity analysts (Soltes, 2014; Green, Markov and Subasi, 2014; Bushee, Jung and Miller, 2017). They are less likely to rely on message boards for information than retail investors. Hence, we expect the treatment effect to be stronger for firms with lower institutional holdings. We also expect the treatment effect to be greater for smaller firms and firms with fewer tangible assets because these firms are usually less covered by traditional information source and retail investors may particularly rely on social media to obtain information about these firms (Bushee, Matsumoto and Miller, 2003; Campbell, DeAngelis and Moon, 2019).

In columns (1) and (2) of Table 4, we partition our sample based on the number of analyst coverage in 2014, which is the year prior to the policy change. We use analyst information in 2014 to capture the pre-existing level of analyst coverage. Then we re-estimate Equation (1) separately for the terciles of firms with the highest and lowest analyst coverage. For firms with high analyst coverage, the DiD estimate (i.e., the coefficient on *PreLnPost* × *After*) is -0.037, which is statistically significant at the 10% level. In contrast, for firms with fewer analyst coverage, the DiD estimate is -0.082 (significant at the 1% level). The corresponding Wald test indicates that these two coefficients are significantly different at the 5% level. The evidence indicates that the negative effect of real name registration on firm value is more pronounced among firms with fewer pre-existing analyst coverage.

In columns (3) and (4), we partition the sample by the institutional ownership in 2014. The DiD estimate in the tercile of firms with the lowest institutional ownership is -0.082, which is more than two times as large as that for the tercile with highest institutional ownership (-0.039). The difference in these two coefficients are significant at the 5% level. These results indicate that our treatment effect

is stronger for firms with more retail investors, who are more likely to rely on social media for information than institutional investors.

Likewise, we partition our sample by firm size in columns (5)-(6) and by asset tangibility in columns (7)-(8), respectively. In all four columns, the coefficients on *PreLnPost*×*After* are significantly negative. However, the magnitudes of the DiD estimates are substantially greater for smaller firms than larger firms (-0.079 vs. -0.044), indicating a stronger treatment effect for smaller firms. The magnitudes of the DiD estimates are also greater for firms with more intangible assets than for firms with lower intangible assets (-0.084 vs -0.052), indicating a stronger treatment effect for firms with high intangibility.

Taken together, the results in Table 4 provide evidence that the impact of real name registration on firm value is more pronounced when message boards previously played a more important role in information exchange. These results support the view that real name registration policy decreases firm value by worsening the information environment faced by investors.

5.4. Channel Test 2: Impact on Social Media

5.4.1. Number and Structure of Social Media Posts

In this subsection, we examine the impact of the real name registration on the information exchange in social media. In columns (1) and (2) of Table 5 Panel A, the dependent variables are the logarithms of the total number of Guba posts (*LnTotalPosts*) and total number of Guba comments (*LnTotalComments*) about a given firm, respectively. These two variables are direct measures of the information exchange in Guba message boards. The coefficients on *PreLnPost* × *After* are, respectively, -0.161 and -0.122, both statistically significant at the 1% level. The estimation results indicate that an one-standard-deviation increase in treatment intensity (0.689) decreases the numbers

of Guba posts by 12% ($= e^{0.689\times0.161} - 1$) and the number of Guba comments by 9% ($= e^{0.689\times0.122} - 1$). The evidence is consistent with the view that the removal of online anonymity discourage the freedom of expression and information exchange in social media.

Next, we examine how real name registration affects the flow of negative information in Guba. Negative information is particularly important to retail investors because they often hold undiversified portfolios, have lower risk tolerance (Goetzmann and Kumar, 2008), and pay more attention to bad news due to loss aversion and endowment effects (Kahneman, 1979). Furthermore, the retail investors are more sensitive to bad news since bad news spreads faster and wider than good news via the Internet (Hornik et al., 2015). In column (3), the dependent variable is the total number of negative posts minus the total number of positive posts normalized by total number of posts (*GubaNegativeTone*), which measures the proportion of negative messages communicated in social media. Following Antweiler and Frank (2004), a post is classified as negative or positive by using Naive Bayes Classification (NBC) method based on manually labeled training sample with Chinese sentiment dictionary. The coefficient on *PreLnPost* × *After* is -0.386 and statistically significant at the 1% level. These results indicate that the removal of social media anonymity reduces the proportion of negative information in Guba by more than 38 percentage points, relative to the sample mean of -18%. The evidence suggests that real name registration discourages netizens from revealing negative information on social media.

5.4.2. Predicting Earnings Surprise based on Social Media Posts

In Table 5 Panel B, we investigate how real name registration influences the information content of social media. If the real name registration impedes the flow of information (particularly negative information) on social media, we should observe a decline in predictive power of social media on earnings surprise after the regulatory change. Specially, we examine to what extent the negative (or

positive) Guba posts in a given quarter predict the quarterly earnings surprises announced at the quarter end. Since Chinese analysts rarely issue quarterly forecasts (Huyghebaert and Xu, 2016; Ke and Zhang, 2021), we measure the earnings surprise as earnings per share (EPS) in quarter *q* minus EPS in quarter *q-4* (i.e., the same quarter of the previous year), scaled by the absolute value of EPS in quarter *q-4* (Bamber and Cheon, 1995; Bartov, Radhakrishnan and Krinsky, 2000).

In all columns of Panel B, we include firm- and quarter- fixed effects and the same set of control variables as in column (6) of Table 3. Columns (1) and (2) report the results estimated using the firm-quarter observations with negative earnings surprises. Our key variable of interest is *NegativeRatio*, which is defined as the ratio of negative posts to all Guba posts created during the quarter. In the pre-treatment period (column (1)), the coefficient on *NegativeRatio* is -0.023 and statistically significant at the 1% level, indicating that the negative information communicated in the anonymous social media can predict poor firm performance. This finding is broadly consistent with Hales, Moon and Swenson (2018), and Huang, Li and Markov (2020), who show that social media in the U.S. conveys important information to predict corporate earnings. In contrast, in the post-treatment period (column (2)), the coefficient on *NegativeRatio* becomes positive, has a much smaller magnitude (0.012), and is not statistically significant. The corresponding Ward test result indicates that the difference in the coefficient estimates of *NegativeRatio* between the two columns are statistically significant at the 1% level. These results indicate that the negative posts in the social media can predict poor firm performance only when the social media is anonymous and such predictability vanishes once the anonymity is removed.

Columns (3) and (4) report results estimated using firm-quarter observations with positive earnings surprises. The model specifications are the same as in columns (1) and (2), except that we replace

NegativeRatio with *PositiveRatio,* which is defined as the ratio of positive posts to all Guba posts created during the quarter. In the pre-treatment period (column (3)), the coefficient on *PositiveRatio* is 0.015 and significant at the 5% level, indicating that the favourable message in social media predicts positive earnings surprises. In the post-treatment period (column (4)), the coefficient on *PositiveRatio* is also 0.015 and significant at the 10% level. The difference between these two coefficient estimates is statistically indistinguishable from zero.

Overall, Table 5 Panel B indicates that while positive message in social media can predict good firm performance irrespective of anonymity, negative message can predict poor firm performance only when the social media is anonymous. These results provide supporting evidence that real name registration deters the dissemination of negative but valuable information on social media.

5.4.3. Stock Reaction towards Earnings Surprise

If real name registration discourages the disclosure and exchange of negative information among investors, investors should be more surprised when a firm's earnings fall short of expectation (Kothari, Shu and Wysocki, 2009). To test this conjecture, we estimate the market reaction at quarterly earnings announcement in Table 5 Panel C. Following Andrei, Friedman and Ozel (2023), the dependent variable is CAR[-1,+1], the three-day size-adjusted cumulative abnormal return (firm's stock return minus the value-weighted size matched portfolios return) around the announcement date of quarterly earnings.[3] In both columns, we control for the magnitude of earnings surprises and the full set of control variables as in Panel B. The model in column (1) is estimated using firm-quarter observations with negative earnings surprises. The coefficient on *PreLnPost* × *After* is negative and statistically

[3] Our main results on earnings announcement window returns are robust to defining excess daily returns as firm-specific returns adjusted for market returns.

significant at the 1% level, indicating that investors have more negative reactions to negative earnings surprise once social median anonymity is removed. In column (2), we focus on firm-quarter observations with positive earnings surprises. The coefficient on *PreLnPost* × *After* is negligible and statistically indistinguishable from zero. The difference between these two DiD estimates is significant at the 10% level. The findings in Panel C indicate that real name registration policy leads to stronger negative market reactions to negative earnings surprises, while investors' reaction to positive earnings surprises is not affected by the treatment.

5.4.4. Placebo Test on Mainstream Media

A potential concern about our findings is that they might be driven by a change in China's overall media environment rather than the real name registration policy in social media. We investigate this possibility by examining the coverage change and tone change in China's mainstream media. Mainstream media in China refers to a set of central Party-state media (such as People's Daily and Xinhua News Agency) and regional official outlets (such as Shanghai Daily and Beijing Daily). They are regarded as being core to the government's goal of maintaining "correct public opinion guidance". Unlike social media, mainstream media was never anonymous and thus are unaffected the real name policy.

In Table 5 Panel D, we collect the information about each stock's coverage in mainstream media from CNRDS (Chinese Research Data Services Platform) dataset. We re-estimate our baseline regression using proxies for mainstream media coverage and mainstream media tone as dependent variables in columns (1) and (2), respectively. In both columns, the coefficients on *PreLnPost* × *After* are negligible and not significantly different from zero, indicating the mainstream media are unaffected by the real name policy. Overall, these results show that real name policy only affect social media and

has no effect on mainstream media, indicating that the change in social media is likely due to the real name registration policy rather than the changes in China's overall media environment.

Taken together, the results in Table 5 provide supporting evidence for the information environment hypothesis: Real name registration policy decreases firm value by reducing value-relevant information (particularly negative information) in social media.

5.5. Channel Test 3: Impact on Firm Information Environment

In this subsection, we provide further evidence that real name registration indeed worsens the information environment faced by investors. In column (1) of Table 6, we examine the impact of real name registration on firms' stock price informativeness. Higher stock price informativeness indicates more firm-specific information conveyed in firms' stock price and thus better information environment faced by investors (Morck, Yeung and Yu, 2000; Hutton, Miller and Skinner, 2009). Following this strand of literature, we first compute the R^2 obtained from estimating the expanded index model detailed in Appendix A. Then we define a firm's stock price informativeness (*StkInform*) as $\mathrm{Log}(\frac{1-\mathrm{R}^2}{\mathrm{R}^2})$. A higher value of *StkInform* indicates more firm specific information in the firm's stock price and thus greater informativeness. The coefficient of *PreLnPost* × *After* is -0.034 , which is statistically significant at the 5% level. The evidence suggests that real name registrations policy leads to a significant decrease in firms' stock price informativeness.

Previous research shows that informational opaque firms tend to have higher stock price crash risk (Jin and Myers, 2006; Hutton, Marcus and Tehranian, 2009). That is, if negative firm-specific information are not reflected in their stock prices in a timely manner, there will be a higher likelihood of stock price crash when accumulated negative firm-specific information is finally revealed to investors. For this reason, we use the firm's stock price crash risk as another proxy for the information

environment faced by investors. In column (2), we measure crash risk using *NCSKEW*, defined as the negative of the third moment of firm-specific weekly returns in a sample year divided by the standard deviation of firm-specific weekly returns raised to the third power (Chen, Hong and Stein, 2001; Kim, Li and Zhang, 2011; Xu, Xuan and Zheng, 2021).[4] The coefficient on *Treat* × *After* is positive and statistically significant at the 1% level, indicating that firms with greater treatment intensity experience significantly larger increase in the likelihood of having stock price crashes after real name registration.

In column (3), we examine the impact of real name registration on corporate voluntary disclosure. As an important aspect of the information environment faced by investors, corporate voluntary disclosure helps reduce the information asymmetry between corporate insider and outside investors. Similar in spirit to Bryan (1997), we measure voluntary disclosure using the logarithm of the numbers of words about future discussion in the Management Discussion and Analysis (MD&A) section of a company's annual report.[5] The coefficient of *PreLnPost* × *After* is significantly negative, indicating that real name registration policy leads to a significant decrease in corporate voluntary disclosure. Existing literature suggests that corporate managers' voluntary disclosure decisions depend on shareholders' access to information (Healy and Palepu, 2001; Li, 2008; Hutton, Marcus and Tehranian, 2009; Kothari, Shu and Wysocki, 2009). When investors have greater access to information from external sources (such as active information sharing in anonymous social media), managers tend to

[4] In untabulated tests, we also use the down-to-up volatility (*DUVOL*) as an alternative proxy for the stock price crash risk and our inference is largely the same: The coefficient on *Treat* × *After* is 0.053 and significant at the 1% level.

[5] Although the inclusion of MD&A section itself in financial reports is mandatory, the content is largely at managers' discretion (Li, 2008; Bloomfield, 2008; Loughran and McDonald, 2014). Leuz and Schrand (2009) and Bourveau et al. (2018) use the length of MD&A as proxy for the level of value-relevant voluntary disclosure. In general, firms are required to disclose both retrospective and prospective analyses in MD&A. Bryan (1997) finds the prospective disclosures (percentage of content related to the future) are especially helpful in assessing firms' prospects. These evidences suggest that our measurement is a reasonable proxy for voluntary disclosure.

provide more voluntary disclosure to reduce detection risk or litigation risk (Skinner, 1994, 1997; Beyer et al., 2010). Once the removal of anonymity reduces investors' access to information, managers face less pressure for providing more voluntary disclosure. Our results are broadly consistent with this view.

Finally, in a poor information environment, shareholders often do not have sufficient information for monitoring manager's actions. This may encourage managers to engage in earnings management activities (Schipper, 1989; Warfield, Wild and Wild, 1995). We test this conjecture in columns (4) and (5). In column (4), the dependent variable REM measures firms' real earnings management, defined as abnormal production costs minus abnormal cash flow from operations and abnormal discretionary expenses (Kim et al., 2012). The coefficient on *PreLnPost* × *After* is significantly positive at the 5% level, suggesting that the real name registration policy leads to an increase in real earnings management. In column (5), the dependent variable *Restatement* is an indicator variable taking the value of one for firms announcing the correction of misstatement in a subsequent reporting period and zero otherwise. The coefficient on *PreLnPost* × *After* is 0.018 and statistically significant at the 5% level, indicating that the likelihood of financial restatement increases after the treatment policy.[6] The magnitude of the coefficient economically significant. It suggests that a one-standard-deviation increase in treatment intensity results in a 1.24 (=0.689×0.018) percentage points higher likelihood of restatement in

[6] In column (5), we use OLS regression rather than the logistic/probit regression because of the "incidental parameters problem" associated with fixed effects. Greene (2004) and Arellano and Hahn (2007) suggest that nonlinear maximum likelihood models can be inconsistent and biased when fixed effects are included. Ai and Norton (2003) also raise concerns about the coefficient on the interaction term does not represent the marginal effect in logit or probit DiD specifications. Using linear models with discrete outcome variables is common practice (e.g., Gillette, Samuels and Zhou, 2020; Jennings et al., 2023).

comparison to the unconditional probability of 8.09 percentage points. The results in columns (4) and (5) show that managers tend to engage in more earnings management after the regulatory change.

Taken together, the results in Table 6 provide supporting evidence that real name registration decreases firm value by deteriorating the information environment faced by shareholders: This policy shock decreases stock price informativeness, increases stock price crash risk, reduces corporate voluntary disclosure, and boosts earnings manipulation.

6. Robustness Check and Additional Investigation

6.1. Alternative Explanation: Overvaluation and/or Attention

An alternative explanation for our findings is that firms with higher treatment intensities are overvalued in the pre-event period. As a result, these firms attract more Guba posts in the pre-event period and have a greater decline in firm value in the post-event period. This explanation is unlikely to hold because firms with higher treatment intensities actually have lower Tobin's Q in the pre-event period. As shown in Table 2, in the pre-event period, the average *LnQ* of firms with high treatment intensities is only 0.276, which is smaller than that of firms with low treatment intensities (0.389). In addition, Figure 1 shows that the stock return of these two groups is indistinguishable during the 60-month period before the treatment policy.

To further investigate this overvaluation-based explanation, we re-estimate the baseline model in column (6) of Table 3 using three alternative measures of treatment intensity: *PreLnNegPost*, *PreLnNonPosiPost*, and *PreGubaNegativeTone*. *PreLnNegPost* (*PreLnNonPosiPost*) measures the total number of negative (non-positive) Guba posts about a firm in 2014 (the year prior to the regulatory change). *PreGubaNegativeTone* is the difference between the number of negative posts and the number of positive posts scaled by the total number of Guba posts about the company in 2014. A

higher value of *PreLnNegPost*, *PreLnNonPosiPost*, or *PreGubaNegativeTone* indicates more negative tone about a firm. As discussed earlier, real name registration discourage negative posts more than positive posts. Hence, if real name registration has impact on firm value, firms with more negative Guba posts before the regulatory change should be affected more. A key advantage of using these alternative measures of treatment intensity is that negative posts are unlikely to be correlated with stock overvaluation (in fact they are more likely to be correlated with stock undervaluation). We report the estimation results in Table 7 Panel A. In all three columns, the DiD estimates are negative and significant at the 1% level, indicating that the firms with more negative posts in the pre-event period experience larger declines in firm value after the treatment. These results, combined with the evidence from Table 2 and Figure 1, indicate that our main findings are unlikely to be driven by the pre-treatment overvaluation of high-treatment-intensity firms.

Relatedly, another possible alternative explanation is investor attention. That is, real name registration deters the information exchange in social media and thus decreases the investor attention paid to the stock. Considering that investor attention is usually positively associated with stock valuation (Barber and Odean, 2008; Aboody, Lehavy and Trueman, 2010; Hirshleifer, Lim and Teoh, 2009), the drop in investor attention may lead to the decline in firm valuation. To examine this explanation, we use Baidu Searching Index, which is similar to SVI (Searching Volume Index) of Google Trend, to measure the stock level retail investor attention by Baidu Searching Index following Da, Engelberg and Gao (2011), DeHaan, Shevlin and Thornock (2015), and Choi, Gao and Jiang (2020). Baidu Searching Index for the stock covers all the internet search on Baidu with key words related to the stock. To identify the search with the stock, the key words must include either the stock

code, the current/historical stock name, or the current/historical name of the listed company.[7] In Table 7 Panel B column (1), we measure investor attention using *Overall index* , which is calculated based on the searching volumes from all devices. In column (2) and (3), we measure investor attention using *Mobile index* and *PC index*, which represent the searching from mobile phones and personal computers, respectively. We re-estimate our baseline regression model using the attention proxies as dependent variables. In all three columns, the coefficients on *PreLnPost* × *After* are not significantly different from zero, indicating that real name registration does not alter the investor attention to firms with different treatment intensities. Thus, it is unlikely that our main findings can be attributed to changes in investor attention.

6.2. Propensity Score Matching

A potential concern about our main findings is that firms with different treatment intensity may not be comparable due to the differences in firm characteristics. As pointed out by Roberts and Whited (2013), it is preferable for the treatment and control groups to be relatively similar across observable dimensions prior to the treatment. When this similarity is not achieved, one can directly incorporate control variables in the regression specification, as we did in our analysis so far.

To further alleviate this concern, we use a propensity score matching algorism to construct a matched sample of treatment and control firms. Specifically, we first follow the method used in Table 2 to classify all sample firms into treatment and control groups, which consist of firms whose total number of Guba posts is above (below) the sample median in 2014. Then we estimate a probit model using all sample firms in 2014. In the probit model, the dependent variable is *Treat*, which equals one

[7] Information on Baidu Searching Index is obtained from MARK datasets (https://www.macrodatas.cn/article/1147467078). Because this dataset starts in 2011, our sample period for this analysis is from 2011 to 2019.

for treatment firms and zero otherwise. We control for the full set of firm characteristics as in Equations (1). In addition, we control for the average of annual growth rates in Tobin's Q and Guba posts over the prior three years to ensure that the matched pairs have similar pre-existing trends in firm value and Guba posts. We also include industry fixed effects in the probit model.

We present the estimation results in column (1) of Table 8 Panel A. The model captures a significant amount of variation in the choice variable, as indicated by a pseudo-R^2 of 15.5% and a *p*-value from the χ^2 test of the overall model fitness well below 0.1%. Then we match the treatment firms to control firms using one-to-one nearest neighbour propensity score matching, without replacement. That is, we match each treatment firm to a control firm with the closest propensity score (i.e., the predicted probabilities) obtained from the estimation in column (1). We end up with 373 one-to-one pairs of matched firms (746 observations).

Because the validity of propensity score matching depends on finding closely matched firms, we conduct two diagnostic tests. In Panel A column (2), we re-estimate the probit model using the matched sample. All coefficient estimates are statistically insignificant. Panel B report the results of a balance test on the major characteristics between the matched pairs. We show that the pre-event characteristics of the treatment firms and their matched peers are indistinguishable. Overall, the diagnostic test results in Panel A column (2) and Panel B show that the propensity score matching procedure has removed meaningful observable differences in pre-event characteristics between the treatment and control firms (other than the difference in Guba posts in 2014). As a result, the matching procedure increases the likelihood that the observed difference in the change of firm value between the treatment and control firms is caused by real name registration.

We perform DiD analysis using the matched sample and report the estimation results in Panel C column (1). The model in column (1) is the same as the baseline model in Table 3 column (6), except that we now replace *PreLnPost* with *Treat*. The interaction term *Treat* × *After* thus captures the DiD effect. The coefficient of *Treat* × *After* is -0.098, which is statistically significant at the 1% level. The estimation result indicates that the real name registration policy leads to a 9.8% reduction in firm value for treatment firms relative to their propensity-score-matched peers. The model in column (2) is similar to the model in Table 3 column (7), except that we now replace *PreLnPost* with *Treat.* The estimation result indicates that the matched pairs have similar pre-treatment trends in firm value and the effect of real name registration policy shows up only after the treatment.

Taken together, the results in Table 8 are consistent with our baseline results in Table 3. Hence, the analysis based on the propensity-score-matched sample further alleviates the concern that our baseline results might be driven by the differences in firm characteristics between of the treatment and control group firms.

6.3. Taiwan market as control group

China's real name registration is a national policy that affect all firms listed in China. Although firms with more pre-existing Guba posts are affected more than those with fewer pre-existing Guba post, the latter group could still be affected to some extent. This possibility might lead to some underestimation of our treatment effect. As a robustness check, we use public firms in Taiwan as a control group. While Taiwan and Mainland China share cultural and linguistic ties, social media in Taiwan is unaffected by Mainland China's real name registration policy. Therefore, Taiwanese firms could serve as an alternative control group for investigating the value implication of real name registration policy. We collect the information on all public firms in Taiwan from Wind database.

Considering that there are far more public firms listed in Mainland China than in Taiwan (approximately 3,760 vs 942 as of 2019), we match each Taiwanese firm to a firm in Mainland China based on the propensity score method we used in Table 8. We define the $MainlandChina$ indicator taking the value of one for mainland Chinese firms, and zero for Taiwanese firms. In Table 9, we then re-estimate our baseline regression by replacing *PreLnPost* with $MainlandChina$ based on the Taiwan firm and their propensity-score-matched firms in Mainland China. As shown in column (1), the coefficient on *MainlandChina*×*After* is -0.103 and significant at the 1% level, indicating that the real name policy leads to a significant decrease in firm value of Mainland Chinese firms relative to Taiwanese firms. In column (2), we repeat the analysis in column (1) by restricting to the subsample of Mainland Chinese firms headquartered in Fujian and Guangdong provinces, which are the two provinces being most geographically and economically close to Taiwan. The coefficient estimate of *MainlandChina*×*After* is -0.195 and significant at the 5% level, indicating that the real name registration policy leads to a drop in firm value by approximately 20% in Mainland China relative to Taiwan. Overall, these results indicate that our main findings are robust to alternative control groups.

6.4. Instrumental variable approach

In our baseline model, we use the number of Guba post in 2014 to classify the treatment and control groups. However, it is possible that some omitted variables correlated with firms' Guba posts drive their subsequent change in valuation. We apply the two-stage least squares (2SLS) regression with an instrumental variable (IV) to relate the exogenous variation in firms' number of Guba posts and thus help mitigate this concern.

We exploit the proportion of days with inclement weather in 2014 as the IV for the scale of the online discussion in that year. We collect the weather information at the firm's headquarter city from

CNRDS. Inclement weather refers to harsh weather conditions, including heavy rainfall, heavy snow, hail, sandstorm, violent wind (Beaufort scale greater than 7), and extreme temperature (>35 or <-25 degrees Celsius). It is widely documented that individuals are more likely to choose indoor (as opposed to outdoor) activities, such as surfing the Internet and watching TV, when the weather is bad (Starr-McCluer, 2000; Gomez et al., 2007), and that individuals tend to have more information about local firms and exhibit a strong preference for local stocks (Coval and Moskowitz, 1999; Ivkovic and Weisbenner, 2005). Therefore, this IV is likely to be positively associated with the number of Guba posts for a given local firm (satisfying the relevance condition). More importantly, the weather in 2014 is unlikely to affect the change in firm valuation several years later during 2015-2019 (other than through the channel of social media discussion), thus satisfying the exclusion condition.

In the first stage, we run an OLS regression with *LnPost* as the dependent variable. The independent variables include *Badweather* and the firm characteristics used in Equation (1). In the second stage, we re-estimate Equation (1) by using the predicted value of *LnPost* obtained from the first stage. Table 10 column (1) presents the results of the first-stage regression. The coefficient on the IV is positive and significant at the 1% level, indicating that local bad weather increases participation in online postings about local firms. Column (2) present the second-stage regression results. Consistent with the findings from the OLS analysis, the coefficient estimate of *PreLnPost*×*After* is negative and significant at the 1%. Comparing results obtained from the OLS regressions (Table 3 column (6)) with those obtained from the 2SLS regressions, it is interesting to observe that the magnitudes of the 2SLS coefficient estimates (-0.17) are twice as large in magnitude as that of the OLS estimate (-0.077), suggesting that OLS regressions bias the coefficient estimates downward due to endogeneity in investor postings. This finding suggests that some omitted variables (e.g., management talents)

simultaneously make higher firm valuation and more intensively discussed by investors. Once we use the instrument to clean up the correlation between investor postings and the residuals (the firm's unobservable characteristics), the coefficient estimates become larger in magnitude.

Overall, the documented impacts of real name policy on the firm value are robust to addressing the endogeneity associated with investor Guba postings.

7. Conclusion

Despite numerous arguments for and against internet anonymity in the past two decades, quantitatively assessing its societal impact has been challenging. After weighting the potential cost and benefits, what is the "net benefit" of internet anonymity? In this paper, we delve into this issue by examining its effects on firm value in the stock market. This is particularly pertinent given the growing reliance of investors on social media for financial information.

Our study focuses on a unique event – China's transition to a real-name registration system for all users of internet social media in 2015. Using a DiD approach, we observe a substantial decline in firm value for firms more affected by this policy relative to those less affected. Furthermore, we find that the cross-sectional variations in the treatment effect are consistent with the information environment hypothesis. Specifically, we note that the negative effect of real-name registration policy on firm value is more pronounced for firms with poor analyst coverage, firms with more retail investors, smaller firms, and lower-tangibility firms. Finally, we provide evidence that the real-name registration significantly reduces the information exchange in social media, lowers firms' stock price informativeness, elevates firms' stock crash risk, discourages corporate disclosure, and increases earnings manipulation. Collectively, these results indicate that real name registration in social media decreases firm value by worsening the information environment faced by investors.

Several governments have considered or passed laws restricting the rights of internet users to maintain online anonymity. For example, on July 31, 2023, President Putin has signed off on a legal amendment that require all Russian online platforms to verify the identities of new users before providing access to services. In 2021, some British politicians called for an end to online anonymity in the UK's Online Safety Bill, arguing that anonymity on the Internet had contributed to greater toxicity in public discourse. Although the proposal was considered by then-Home Secretary Priti Patel, it faced significant opposition and was not included in the bill's draft. In the U.S., Senator John Kennedy, Republican of Louisiana, announced in 2021 a plan to introduce legislation requiring social media users to verify their legal identities (critics argued that this may have been unconstitutional, and Kennedy has yet to introduce the bill). In the context of the stock market, our study provides the first quantitative analysis on this debate. Our results suggest that eliminating Internet anonymity does more harm than good. This finding is particularly timely and relevant, given the ongoing debate on public policies regarding the termination of online anonymity.

REFERENCES

Aboody, David, Reuven Lehavy, and Brett Trueman. 2010. Limited attention and the earnings announcement returns of past stock market winners. *Review of Accounting Studies* 15: 317-344.

Acharya, Viral V., Peter DeMarzo, and Ilan Kremer. 2011. Endogenous information flows and the clustering of announcements. *American Economic Review* 101 (7): 2955-2979.

Ai, Chunrong, and Edward C. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80 (1): 123-129.

Andrei, Daniel, Henry Friedman, and N. Bugra Ozel. 2023. Economic uncertainty and investor attention. *Journal of Financial Economics* 149 (2): 179-217.

Ang, J. S., C. Hsu, D. Tang, and C. Wu. 2021. The role of social media in corporate governance. *The Accounting Review* 96 (2): 1-32.

Antweiler, Werner, and Murray Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance* 59(3): 1259-1294.

Arellano, Manuel, and Jinyong Hahn. 2007. Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs* 43: 381.

Barber, Brad M., and Terrance Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies* 21(2): 785-818.

Bamber, Linda Smith, and Youngsoon Susan Cheon. 1995. Differential price and volume reactions to accounting earnings announcements. *Accounting Review*: 417-441.

Bartov, Eli, Lucile Faurel, and Partha S. Mohanram. 2018. Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review* 93 (3): 25-57.

Bartov, Eli, Suresh Radhakrishnan, and Itzhak Krinsky. 2000. Investor sophistication and patterns in stock returns after earnings announcements. *The Accounting Review* 75 (1): 43-63.

Beyer, A., Cohen, D. A., Lys, T. Z., and Walther, B. R. 2010. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, 50(2-3), 296-343.

Bhattacharya, N., Black, E. L., Christensen, T. E., and Mergenthaler, R. D. 2007. Who trades on pro forma earnings information? *The Accounting Review* 82(3): 581-619.

Blankespoor, E., G. S. Miller, and H. D. White. 2014. The role of dissemination in market liquidity: Evidence from firms' use of Twitter. *The Accounting Review* 89 (1): 79–112.

Blankespoor, Elizabeth, Ed deHaan, and Ivan Marinovic. 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics* 70 (2-3): 1013-1044.

Bourveau, Thomas, Yun Lou, and Rencheng Wang. 2018. Shareholder litigation and corporate disclosure: Evidence from derivative lawsuits. *Journal of Accounting Research* 56 (3): 797-842.

Brav, Alon, Matthew Cain, and Jonathon Zytnick. 2022. Retail shareholder participation in the proxy process: Monitoring, engagement, and voting. *Journal of Financial Economics* 144 (2): 492-522.

Bryan, Stephen H. 1997. Incremental information content of required disclosures contained in management discussion and analysis. *The Accounting Review,* 285-301.

Bushee, Brian J., Dawn A. Matsumoto, and Gregory S. Miller. 2003. Open versus closed conference calls: The determinants and effects of broadening access to disclosure. *Journal of Accounting and Economics* 34 (1-3): 149-180.

Bushee, Brian J., Michael J. Jung, and Gregory S. Miller. 2017. Do investors benefit from selective access to management? *Journal of Financial Reporting* 2 (1): 31-61.

Cade, N. L. 2018. Corporate social media: How two-way disclosure channels influence investors. *Accounting, Organizations and Society* 68: 63-79.

Campbell, John L., Matthew D. DeAngelis, and James R. Moon. 2019. Skin in the game: Personal stock holdings and investors' response to stock analysis on social media. *Review of Accounting Studies* 24: 731-779.

Carpenter, J. N., F. Lu, and R. F. Whitelaw. 2015. The real value of China's stock market. *Journal of Financial Economics* 139 (3): 679-696.

Chang, Y. C., Hong, H. G., Tiedens, L., Wang, N., and Zhao, B. 2015. Does diversity lead to diverse opinions? Evidence from languages and stock markets. Rock Center for Corporate Governance at Stanford University Working Paper, (168), 13-16.

Chen, H., P. De, Y. J. Hu, and B. H. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies* 27 (5): 1367-1403.

Chen, J., H. Hong, and J. C. Stein. 2001. Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics* 61 (3): 345-381.

Choi, Darwin, Zhenyu Gao, and Wenxi Jiang. 2020. Attention to global warming. *The Review of Financial Studies* 33 (3): 1112-1145.

Choi, Bryan H. 2012. The anonymous internet. *Maryland Law Review* 72: 501-570.

Claessens, Stijn, and Luc Laeven. 2003. Financial development, property rights, and growth. *The Journal of Finance* 58 (6): 2401-2436.

Coval, Joshua, and Tobias Moskowitz, 1999. Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 1-39.

Da, Z., Engelberg, J., and Gao, P. 2011. In search of attention. *The Journal of Finance*, *66*(5), 1461-1499.

Das, Sanjiv R., and Mike Y. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53(9): 1375-1388.

DeHaan, E., Shevlin, T., and Thornock, J. 2015. Market (in) attention and the strategic scheduling. and timing of earnings announcements. *Journal of Accounting and Economics*, *60*(1), 36-55.

Dube, S., and Zhu, C. 2021. The disciplinary effect of social media: Evidence from firms' responses to Glassdoor reviews. *Journal of Accounting Research*, 59(5), 1783-1825.

Dye, Ronald A. 1985. Disclosure of nonproprietary information. *Journal of Accounting Research*: 123-145.

Dye, Ronald A. 2001. An evaluation of "essays on disclosure" and the disclosure literature in accounting. *Journal of Accounting and Economics* 32 (1-3): 181-235.

Dyer, Travis, and Eunjee Kim. 2021. Anonymous equity research. *Journal of Accounting Research* 59 (2): 575-611.

Eren, N., H.N. Ozsoylev and P.E. Street. 2009. Hype and dump manipulation, Working Paper (University of Minnesota).

Froomkin, A. Michael. 1995. Flood control on the information ocean: Living with anonymity, digital cash, and distributed databases. *Journal of Law and Commerce* 15: 395-507.

Gillette, Jacquelyn R., Delphine Samuels, and Frank S. Zhou. 2020. The effect of credit ratings on disclosure: Evidence from the recalibration of Moody's municipal ratings. *Journal of Accounting Research* 58 (3): 693-739.

Goetzmann, William N., and Alok Kumar. 2008. Equity portfolio diversification. *Review of Finance* 12 (3): 433-463.

Green, T. C., Jame, R., Markov, S., and Subasi, M. 2014. Access to management and the informativeness of analyst research. *Journal of Financial Economics*, 114(2), 239-255.

Greene, William. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal* 7 (1): 98-119.

Gurun, Umit G., and Alexander W. Butler. 2012. Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance* 67 (2): 561-598.

Hales, Jeffrey, James R. Moon Jr, and Laura A. Swenson. 2018. A new era of voluntary disclosure? Empirical evidence on how employee postings on social media relate to future corporate disclosures. *Accounting, Organizations and Society* 68: 88-108.

Healy, Paul M., and Krishna G. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31 (1-3): 405-440.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh. 2009. Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance* 64 (5): 2289-2325.

Hong, H., W. Jiang, N. Wang, and B. Zhao. 2014. Trading for status. *Review of Financial Studies* 27 (11): 3171–3212.

Hornik, J., Satchi, R. S., Cesareo, L., and Pastore, A. 2015. Information dissemination via electronic word-of-mouth: Good news travels fast, bad news travels faster! *Computers in Human Behavior* 45, 273-280.

Huang, K., Li, M., and Markov, S. 2020. What do employees know? Evidence from a social media platform. *The Accounting Review*, 95(2): 199-226.

Huang, Y., H. Qiu, and Z. Wu. 2016. Local bias in investor attention: Evidence from China's Internet stock message boards. *Journal of Empirical Finance* 38: 338–354

Hutton, Amy P., Alan J. Marcus, and Hassan Tehranian. 2009. Opaque financial reports, R^2, and crash risk. *Journal of Financial Economics* 94 (1): 67-86.

Hutton, Amy P., Gregory S. Miller, and Douglas J. Skinner. 2003. The role of supplementary statements with management earnings forecasts. *Journal of Accounting Research* 41 (5): 867-890.

Huyghebaert, Nancy, and Weidong Xu. 2016. Bias in the post-IPO earnings forecasts of affiliated analysts: Evidence from a Chinese natural experiment. *Journal of Accounting and Economics* 61 (2-3): 486-505.

Ivkovic, Zoran, and Scott, Weisbenner. 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance* 60: 267-306.

Jennings, J., Kim, J. M., Lee, J., and Taylor, D. 2023. Measurement error, fixed effects, and false positives in accounting research. *Review of Accounting Studies*, 1-37.

Jiang, Guolin, Paul G. Mahoney, and Jianping Mei. 2005. Market manipulation: A comprehensive study of stock pools. *Journal of Financial Economics* 77 (1): 147-170.

Jin, L., and S. C. Myers. 2006. R^2 around the world: New theory and new tests. *Journal of Financial Economics* 79 (2): 257-292.

Jorgensen, Bjorn N., and Michael T. Kirschenheiter. 2003. Discretionary risk disclosures. *The Accounting Review* 78 (2): 449-469.

Kahneman, Daniel. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* 47: 278.

Kaminski, Margot. 2012. Real masks and real name policies: Applying anti-mask case law to anonymous online speech. *Fordham Intellectual Property, Media & Entertainment Law Journal* 23: 815-896.

Kim, J. B., Y. Li, and L. Zhang. 2011. Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics* 100 (3): 639-662.

Kogan, Shimon, Tobias J. Moskowitz, and Marina Niessner. 2021. Social media and financial news manipulation. Available at SSRN 3237763.

Kothari, Sabino P., Susan Shu, and Peter D. Wysocki. 2009. Do managers withhold bad news? *Journal of Accounting Research* 47 (1): 241-276.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. 2002. Investor protection and corporate valuation. *The Journal of Finance*, 57(3): 1147-1170.

Lee, Hyung-Woo, Peter J. Robertson, and Kitai Kim.2020. Determinants of job satisfaction among US federal employees: An investigation of racial and gender differences. *Public Personnel Management* 49 (3): 336-366.

Lee, Jyh-An, and Ching-Yi Liu. 2016. Real-name registration rules and the fading digital anonymity in China. *Washington International Law Journal* 25 (1): 1-34.

Lee, Yeunjae, and Katie Haejung Kim. 2020. De-motivating employees' negative communication behaviors on anonymous social media: The role of public relations. *Public Relations Review* 46 (4): 101955.

Leuz, Christian, and Catherine Schrand. 2009. Disclosure and the cost of capital: Evidence from firms' responses to the Enron shock. No. w14897. *National Bureau of Economic Research*.

Li, D. 2015. CAC: Microblog and posts bars to adopt real name registration this year. *The Beijing Times*.

Li, Feng. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics* 45 (2-3): 221-247.

Lidsky, Lyrissa Barnett, and Thomas F. Cotter. 2007. Authorship, audiences, and anonymous speech. *Notre Dame Law Review* 82: 1537-1603.

Liu, Jianan, Robert F. Stambaugh, and Yu Yuan. 2019. Size and value in China. *Journal of Financial Economics* 134 (1): 48-69.

Loughran, Tim, and Bill McDonald. 2011. Measuring readability in financial disclosures. *The Journal of Finance* 69 (4): 1643-1671.

Lu, Hai, Jee-Eun Shin, and Mingyue Zhang. 2023. Financial reporting and disclosure practices in China. *Journal of Accounting and Economics* (2023): 101598.

Malmendier, Ulrike, and Devin Shanthikumar. 2007. Are small investors naive about incentives? *Journal of Financial Economics* 85 (2): 457-489.

Masulis, Ronald W., and Shawn Mobbs. 2011. Are all inside directors the same? Evidence from the external directorship market. *The Journal of Finance* 66(3): 823-872.

Miller, Gregory S., and Douglas J. Skinner. 2015. The evolving disclosure landscape: How changes in technology, the media, and capital markets are affecting disclosure. *Journal of Accounting Research* 53 (2): 221-239.

Mitts, Joshua. 2020. Short and distort. *The Journal of Legal Studies* 49 (2): 287-334.

Moore, Susanna. 2009. The challenge of internet anonymity: Protecting John Doe on the internet. UIC John Marshall Journal of Information Technology & Privacy Law 26(4): 469-486.

Morck, Randall, Bernard Yeung, and Wayne Yu. 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58 (1-2): 215-260.

Raddatz, Claudio. 2006. Liquidity needs and vulnerability to financial underdevelopment. *Journal of Financial Economics* 80 (3): 677-722.

Roberts, Michael R., and Toni M. Whited. 2013. Endogeneity in empirical corporate finance. *Handbook of the Economics of Finance.* Vol. 2. Elsevier. 493-572.

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam. 2009. Options trading activity and firm valuation. *Journal of Financial Economics* 94 (3): 345-360.

Roychowdhury, Sugata. 2006. Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42 (3): 335-370.

Skinner, Douglas J. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* 32 (1): 38-60.

Skinner, Douglas J. 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics* 23 (3): 249-282.

Soltes, Eugene. 2014. Private interaction between firm management and sell-side analysts. *Journal of Accounting Research* 52 (1): 245-272.

Surowiecki, James. 2004. The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations* 296.5.

Tang, V. W. 2018. Wisdom of crowds: Cross-sectional variation in the informativeness of third-party generated product information on Twitter. *Journal of Accounting Research* 56 (3): 989–1034.

Van Bommel, Jos. 2003. Rumors. *The Journal of Finance* 58 (4): 1499-1520.

Verrecchia, Robert E. 1983. Discretionary disclosure. *Journal of Accounting and Economics* 5: 179-194.

Warfield, Terry D., John J. Wild, and Kenneth L. Wild. 1995. Managerial ownership, accounting choices, and informativeness of earnings. *Journal of Accounting and Economics* 20 (1): 61-91.

Xu, S. X., and Zhang, X. 2013. Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly* 37(4):1043-1068.

Xu, Y., Y. Xuan, and G. Zheng. 2021. Internet searching and stock price crash risk: Evidence from a quasi-natural experiment. *Journal of Financial Economics* 141 (1): 255-275.

Yang, K.C. and Y. Kang. 2016. Real-name registration regulation in China: An examination of Chinese netizens' discussions about censorship, privacy, and political freedom. *In Political Discourse in Emergent, Fragile, and Failed Democracies* IGI Global: 61-87.

Table 1. Summary Statistics

Table 1 reports the summary statistics. The full sample in Panel A consists of 14,640 firm-year observations from 2010 to 2019. We obtain Guba posts information from https://guba.eastmoney.com/ and financial information from CSMAR datasets. Panel B reports the comparison of firm characteristics between treated and control groups in the pre-treatment period. Treatment (control) firms are firms with the number of Guba posts above (below) the sample median in 2014. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles.

Panel A. Summary statistics of the full sample

Variable	N	Mean	P25	Median	P75	SD
TobinQ	14,640	1.910	0.767	1.278	2.224	1.947
PreLnPost	14,640	1.730	1.244	1.725	2.199	0.689
After	14,640	0.500	0.000	0.500	1.000	0.500
Asset (MillionRMB)	14,640	14991.07	1855.62	4312.06	11159.83	37076.93
SaleGrowth	14,640	0.119	-0.008	0.095	0.214	0.244
Cashflow	14,640	0.050	0.002	0.047	0.099	0.096
R&D	14,640	0.001	0.000	0.000	0.000	0.004
Capex	14,640	0.072	0.032	0.060	0.099	0.056
Segments	14,640	5.211	2.000	3.000	7.000	5.443
FounderDirector	14,640	0.126	0.000	0.000	0.000	0.331
BoardOwn	14,640	0.029	0.000	0.000	0.001	0.092
CEOOwn	14,640	0.011	0.000	0.000	0.000	0.050
CertifiedDirector	14,640	0.010	0.000	0.000	0.000	0.045
SOE	14,640	0.578	0.000	1.000	1.000	0.494
FirmAge	14,640	15.858	12.000	16.000	20.000	5.465

Panel B. Comparison of firm characteristics between the treated and control groups in the pre-treatment period

	Treatment		Control		Test of difference	
	Mean (1)	Median (2)	Mean (3)	Median (4)	T-test (1)-(3)	Wilcoxon test (2)-(4)
Tobin'sQ	1.823	1.229	1.982	1.394	-0.159***	-0.165***
TotalPosts	8.854	6.78	3.935	3.374	4.919***	3.406***
Ln(Asset)	22.362	22.199	21.549	21.497	0.813***	0.702***
SalesGrowth	0.159	0.139	0.133	0.114	0.026***	0.025***
CashFlow	0.055	0.048	0.051	0.05	0.004	-0.002
R&D	0.001	0	0.001	0	0.000***	0.000***
Capex	0.083	0.07	0.079	0.068	0.004***	0.002*
Ln(Segments)	1.495	1.386	1.226	1.099	0.269***	0.287***
FounderDirctor	0.128	0	0.143	0	-0.015*	0.000*
BoardOwn	0.032	0	0.033	0	-0.001	0.000***
CEOOwn	0.014	0	0.01	0	0.004***	0.000***
CertifiedDirector	0.010	0	0.008	0	0.002	0.000
SOE	0.607	1	0.578	1	0.029**	0.000***
Ln(FirmAge)	2.418	2.565	2.401	2.565	0.017	0.000

Table 2. Univariate Tests

Table 2 reports the univariate tests that examine the relation between real name regulatory change and firm value. Treatment (control) firms are firms with the number of Guba posts above (below) the sample median in 2014. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. Beneath each univariate test is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Treatment (1)	Control (2)	DiD (1)-(2)
Pre-regulatory *LnQ (a)*	0.276	0.389	
Post-regulatory *LnQ (b)*	0.171	0.388	
Difference=(b)-(a)	-0.105***	-0.001	-0.104***
	(000)	(0.938)	(0.000)

Table 3. Baseline Regression

Table 3 reports the general difference-in-difference tests that examine the relation between real name regulatory change and firm value. The dependent variable is the natural logarithm of Tobin's Q. The variable *PreLnPost* is the natural logarithm of Guba posts number in the year prior to the real name regulatory change. *After* is an indicator variable that takes the value of one for the post-treatment period (i.e., the 2015-2019 period) and zero for the pre-treatment period (i.e., the 2010-2014 period). $Before^n$ ($n \in \{1,2,3,4\}$) is an indicator variable that takes the value of one if it is t year pre to the real name regulatory change and zero otherwise. $After^n$ ($n \in \{1,2,3,4,5\}$) is an indicator variable that takes the value of one if it is n year following the real name regulatory change and zero otherwise. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. All regressions include firm and year fixed effects. Standard errors are robust to heteroscedasticity and adjusted for firm clustering. Beneath each coefficient estimate is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DV=LnQ	(-1,+1)	(-3,+3)	(-5,+5)	(-1,+1)	(-3,+3)	(-5,+5)	(-5,+5)
PreLnPost×After	-0.112***	-0.081***	-0.063***	-0.111***	-0.092***	-0.077***	
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	
PreLnPost×Before^4							-0.004
							(0.712)
PreLnPost×Before^3							-0.003
							(0.820)
PreLnPost×Before^2							0.004
							(0.835)
PreLnPost×Before^1							-0.001
							(0.960)
PreLnPost×After^1							-0.108***
							(0.000)
PreLnPost×After^2							-0.118***
							(0.000)
PreLnPost×After^3							-0.056***
							(0.008)
PreLnPost×After^4							-0.045**
							(0.036)
PreLnPost×After^5							-0.062***
							(0.007)
Ln(Asset)				-0.049	-0.309***	-0.362***	-0.361***
				(0.383)	(0.000)	(0.000)	(0.000)
SalesGrowth				0.051	0.035	0.050**	0.052**
				(0.405)	(0.193)	(0.014)	(0.013)
CashFlow				0.220*	0.481***	0.506***	0.505***
				(0.096)	(0.000)	(0.000)	(0.000)
R&D				-6.053	-2.414	-0.355	-0.448
				(0.235)	(0.204)	(0.826)	(0.781)
Capex				0.066	0.258*	0.262**	0.260**
				(0.826)	(0.066)	(0.025)	(0.026)
Ln(Segments)				-0.021	-0.007	0.004	0.004
				(0.454)	(0.690)	(0.769)	(0.760)
FounderDirctor				0.026	-0.090	-0.095*	-0.096*
				(0.756)	(0.123)	(0.092)	(0.090)
BoardOwn				-0.119	0.029	0.012	0.016
				(0.783)	(0.858)	(0.930)	(0.910)
CEOOwn				-0.428	-0.507**	-0.160	-0.152
				(0.428)	(0.042)	(0.454)	(0.478)
CertifiedDirector				0.173	0.065	0.008	0.000
				(0.379)	(0.586)	(0.939)	(0.998)
SOE				0.004	-0.062	-0.087**	-0.087**
				(0.974)	(0.212)	(0.019)	(0.019)
Ln(FirmAge)				0.281	0.076	-0.103**	-0.101**
				(0.358)	(0.359)	(0.018)	(0.020)
Constant	0.658***	0.445***	0.360***	1.026	7.172***	8.720***	8.698***
	(0.000)	(0.000)	(0.000)	(0.468)	(0.000)	(0.000)	(0.000)
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,928	8,784	14,640	2,928	8,784	14,640	14,640
R-squared	0.878	0.813	0.768	0.877	0.842	0.824	0.825

Table 4. Channel Tests 1: Heterogeneous Treatment Effects based on Information Environments

Table 4 re-estimate column (6) of Table 3 based on various subsamples. The subsample are based on upper tertile compared to lower tertile of *AnalystCoverage, InstitutionalOwnership, Firmsize*, and*AssetTangibility*in the year prior to the real name regulatory change. *AnalystCoverage* is defined as the natural logarithm of the number of analysts following the firm. *InstitutionalOwnership* is defined as the percent of common shares outstanding held by institutional investors. *Firmsize* is defined as the natural logarithm of sales. *AssetTangibility* is defined as the ratio of net property, plant, and equipment over total assets. The Wald test examines the equivalence of the coefficients for *PreLnPost×After* between the high and low groups. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. All regressions include firm and year fixed effects. Standard errors are robust to heteroscedasticity and adjusted for firm clustering. Beneath each coefficient estimate is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

DV=LnQ	AnalystCoverage		InstitutionalOwnership		FirmSize		AssetTangibility	
	High	Low	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PreLnPost×After	-0.037*	-0.082***	-0.039*	-0.082***	-0.044*	-0.079***	-0.052**	-0.084***
	(0.069)	(0.002)	(0.086)	(0.001)	(0.071)	(0.003)	(0.027)	(0.000)
Wald Test	5.78**		5.19**		3.14*		2.85*	
P-value	0.016		0.023		0.077		0.091	
Controls	Same as those in Table 3 column (6)							
Firm_FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,820	5,320	4,880	4,880	4,880	4,880	4,880	4,880
R-squared	0.848	0.811	0.839	0.820	0.790	0.771	0.808	0.849

Table 5. Channel Test 2: Impact on Social Media

Table 5 reports the general difference-in-difference tests to examine the impacts of real name regulatory change on social media. The regression specification follows that in column (6) of Table 3, while the control variables in Panel B and Panel C are at quarter level. Panel A reports the impacts of real name regulatory change on the number and structure of social media posts. Panel B reports the proportion of negative (or positive) posts in Guba posts at quarter q is helpful to predict the negative (or positive) quarterly earnings surprise around the regulatory change. The dependent variable is negative (positive) earnings surprises measured as earnings per share in quarter q minus earnings per share in the same quarter of previous year, scaled by the absolute value of earnings per share in q-4. The *NegativeRatio* (*PositiveRatio*) is the ratio of negative (positive) posts to all posts at quarter q. Panel C reports stock reaction towards the earnings surprise. The dependent variable *CAR[-1, +1]* is the three-day size-adjusted cumulative abnormal return around the announcement date of quarterly earnings. Panel D reports the impacts of real name regulation change on the mainstream media. The mainstream media coverage is defined as the natural logarithm of (1+ # news about the firm) and tone is defined as ratio of # negative news minus # positive news over total # news. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. All regressions include firm and year (quarter) fixed effects. Standard errors are robust to heteroscedasticity and adjusted for firm clustering. Beneath each coefficient estimate is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

Panel A. Number and Structure of Social Media Posts

	LnTotalPosts	LnTotalComments	GubaNegativeTone
	(1)	(2)	(3)
PreLnPost×After	-0.161***	-0.122***	-0.386***
	(0.000)	(0.000)	(0.005)
Controls	Same as those in Table 3 column (6)		
Firm_FE	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes
Observations	14,640	14,640	14,640
R-squared	0.525	0.482	0.426

Panel B. Predicting Earnings Surprise based on Social Media Posts

	Negative Surprises		Positive Surprises	
	(1)	(2)	(3)	(4)
	After=0	After=1	After=0	After=1
NegtiveRatio	-0.023***	0.012		
	(0.004)	(0.350)		
PositiveRatio			0.015**	0.015*
			(0.040)	(0.099)
Wald Test	7.06***		0.00	
P-value	0.000		0.981	
Controls	Same as those at quarter level in Table 3 column (6)			
Firm_FE	Yes	Yes	Yes	Yes
Quarter_FE	Yes	Yes	Yes	Yes
Observations	12,898	12,559	13,552	13,798
R-squared	0.147	0.156	0.110	0.113

Panel C. Stock Reaction towards Earnings Surprise

	CAR[-1,+1]	
	Negative Surprises	Positive Surprises
	(1)	(2)
PreLnPost×After	-0.003**	0.000
	(0.012)	(0.805)
Wald Test	3.21*	
P-value	0.073	
EarningSurprise	0.028*	0.037***
	(0.059)	(0.001)
Controls	Same as those at quarter level in Table 3 column (6)	
Firm_FE	Yes	Yes
Quarter_FE	Yes	Yes
Observations	18,648	20,197
R-squared	0.032	0.044

Panel D. The Impacts on Mainstream Media

	Mainstream Media Coverage	Mainstream Media Tone
	(1)	(2)
PreLnPost×After	-0.002	-0.005
	(0.798)	(0.464)
Controls	Same as those in Table 3 column (6)	
Firm_FE	Yes	Yes
Year_FE	Yes	Yes
Observations	14,640	14,640
R-squared	0.796	0.346

Table 6. Channel Tests 3: Effect on Information Environments

Table 6 reports the general difference-in-difference tests to examine the impacts of real name regulatory change on firms' information environments. The regression specification follows that in column (6) of Table 3. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. All regressions include firm and year fixed effects. Standard errors are robust to heteroscedasticity and adjusted for firm clustering. Beneath each coefficient estimate is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Stock Informativeness	Crash Risk	Voluntary Disclosure	Real Earnings Management	Restatement
	(1)	(2)	(3)	(4)	(5)
PreLnPost×After	-0.034**	0.046***	-0.048***	0.013**	0.018**
	(0.047)	(0.009)	(0.008)	(0.045)	(0.015)
Controls	Same as those in Table 3 column (6)				
Firm_FE	Yes	Yes	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes	Yes	Yes
Observations	14,638	14,634	13,699	13,255	14,640
R-squared	0.565	0.089	0.480	0.358	0.094

Table 7. Alternative Explanation of Overvaluation and/or Attention

Table 7 reports the tests on the alternation explanations of overvaluation and/or investor attention. Panel A re-estimate column (6) of Table 3 by using alternative ways to define treatment intensity. *PreLnNegPost* is the natural logarithm of negative Guba posts number. *PreLnNonPosiPost* is the natural logarithm of non-positive Guba posts number. *PreGubaNegativeTone* is the difference between the number of negative posts and the number of positive posts scaled by the total number of Guba posts about the company. All these three variables are measured in 2014 (the year prior to the real name regulatory change). Panel B reports the variation of investor attention around the regulation change. The Overall Index, Mobile Index and PC index cover the Baidu searching for the stock from all devices, mobile phones, and personal computers. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. Standard errors are robust to heteroscedasticity and adjusted for firm clustering. Beneath each coefficient estimate is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

Panel A. Alternative Definition of Treatment Intensity

DV=LnQ	(1)	(2)	(3)
PreLnNegPost×After	-0.075***		
	(0.000)		
PreLnNonPosiPost×After		-0.077***	
		(0.000)	
PreGubaNegativeTone×After			-0.618***
			(0.002)
Controls	Same as those in Table 3 column (6)		
Firm_FE	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes
Observations	14,640	14,640	14,640
R-squared	0.824	0.824	0.824

Panel B. Investor Attention around the Regulation Change

	Overall index	Mobile Index	PC Index
	(1)	(2)	(3)
PreLnPost×After	0.059	0.042	-0.002
	(0.435)	(0.617)	(0.980)
Controls	Same as those in Table 3 column (6)		
Firm_FE	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes
Observations	13,176	13,176	13,176
R-squared	0.928	0.899	0.923

Table 8. Propensity Score Matching

Table 8 examines the effect of real name regulatory on firm value based on propensity score matching. Panel A reports parameter estimates from the probit model used in estimating the propensity scores for the treatment and control groups. The dependent variable in the probit model is the *Treatment* dummy which equals one for the firms with the number of Guba posts above the sample median in 2014 and 0 otherwise. The "Pre-Matching" column contains the parameter estimates of the probit model estimated using the sample prior to matching. These estimates are then used to generate the propensity scores for matching treatment and control firms. The "Post-Matching" column contains the parameter estimates of the probit model estimated using the subsample of matched treatment-control pairs after matching. Coefficient estimates are reported and p-value are displayed in parentheses below. Panel B reports the balance test results for the pairs of treatment and control firms after matching. Columns (1) and (2) of Panel C re-estimate columns (6) and (7) of Table 3 based on the propensity-score-matched sample. Variable definitions are provided in Appendix A.***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

Panel A. Propensity score regression and diagnostic regression

	Pre-Matching	Post-Matching
	(1)	(2)
Ln(Asset)	0.328***	-0.057
	(0.000)	(0.301)
SalesGrowth	-0.042	0.076
	(0.809)	(0.733)
CashFlow	-0.443	0.065
	(0.276)	(0.904)
R&D	7.133	0.594
	(0.386)	(0.954)
Capex	-0.476	1.577
	(0.533)	(0.120)
Ln(Segments)	0.168***	-0.114
	(0.007)	(0.198)
FounderDirctor	-0.055	0.131
	(0.724)	(0.514)
BoardOwn	-0.613	0.241
	(0.333)	(0.749)
CEOOwn	1.912*	0.011
	(0.061)	(0.994)
CertifiedDirector	0.451	0.741
	(0.590)	(0.544)
SOE	-0.189**	0.178
	(0.028)	(0.122)
Ln(FirmAge)	0.224*	-0.055
	(0.069)	(0.735)
QGrowth	0.132*	0.021
	(0.066)	(0.824)
GubaPostGrowth	0.155***	-0.005
	(0.000)	(0.814)
Constant	-8.018***	1.134
	(0.000)	(0.381)
Industry_FE	Yes	Yes
Observations	1,464	746
Presudo R-Square	0.155	0.024
P-value	0.000	0.757

Panel B. Balance tests

	Treatment	Control	Diff.	T-test	P-value
Ln(Asset)	21.890	21.930	-0.042	0.530	0.599
SalesGrowth	0.100	0.101	0.000	0.030	0.979
CashFlow	0.039	0.037	0.002	0.250	0.799
R&D	0.001	0.001	0.000	0.290	0.773
Capex	0.076	0.070	0.005	1.340	0.182
Ln(Segments)	1.361	1.426	-0.064	1.500	0.135
FounderDirector	0.145	0.131	0.013	0.530	0.596
BoardOwn	0.031	0.030	0.002	0.240	0.813
CEOOwn	0.010	0.009	0.001	0.290	0.771
CertifiedDirector	0.008	0.006	0.002	0.710	0.481
SOE	0.571	0.542	0.029	0.810	0.418
Ln(FirmAge)	2.583	2.598	-0.014	0.510	0.609
QGrowth	0.102	0.095	0.007	0.180	0.856
GubaPostGrowth	-0.227	-0.290	0.063	0.370	0.711

Panel C. Matching Sample Difference-in-Difference Analysis

DV=LnQ	(1)	(2)
Treat×After	-0.098***	
	(0.000)	
Treat×Before^4		0.012
		(0.619)
Treat×Before^3		-0.022
		(0.470)
Treat×Before^2		-0.025
		(0.443)
Treat×Before^1		-0.045
		(0.193)
Treat×After^1		-0.083**
		(0.044)
Treat×After^2		-0.130***
		(0.002)
Treat×After^3		-0.107***
		(0.008)
Treat×After^4		-0.124***
		(0.003)
Treat×After^5		-0.124***
		(0.005)
Controls	Same as those in Table 3 column (6)	
Firm_FE	Yes	Yes
Year_FE	Yes	Yes
Observations	7,460	7,460
R-squared	0.817	0.816

Table 9. Taiwan Markets as Control Groups

Table 9 examines the effect of real name regulatory on firm value by taking Taiwan markets as the control group. We collect the information on all public firms in Taiwan from Wind database. *MainlandChina* is the indicator variable taking the value of one for the firms in Mainland China and zero for Taiwanese firms. In column (1), we focus on Taiwanese firms and their propensity-score-matched peers in Mainland China. In column (2), we focus on Taiwanese firms and their propensity-score-matched peers headquartered in Fujian and Guangdong provinces of Mainland China. Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. All regressions include firm and year fixed effects. Standard errors are robust to heteroscedasticity and adjusted for firm clustering. Beneath each coefficient estimate is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

DV=LnQ	(1)	(2)
$MainlandChina \times After$	-0.103***	-0.195**
	(0.009)	(0.038)
Controls	Same as those in Table 3 column (6)	
Firm_FE	Yes	Yes
Year_FE	Yes	Yes
Observations	4,912	2,719
R-squared	0.770	0.766

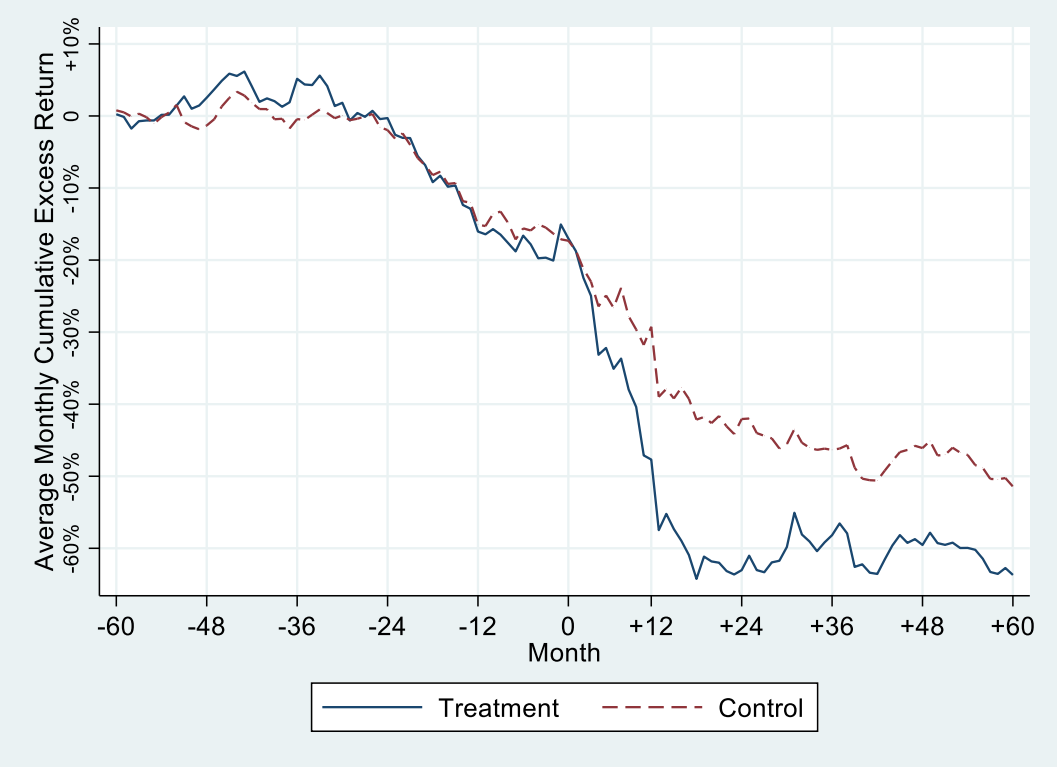
Table 10. The Two-stage Least Squares Regressions

Table 10 reports the two-stage least squares (2SLS) regression results of effect of real name regulatory on firm value. Column (1) reports the first-stage regression with *LnPost* as the dependent variable, and the province-level *Badweather* in 2014 as the instrumental variable. The instrumental variable *Badweather* is the average proportion of days with bad weather across all cities in firm's headquarter province in 2014, where bad weather refers to heavy rainfall, heavy snow, hail, sandstorm, violent wind (Beaufort scale greater than 7), and extreme temperature (>35 or <-25 degrees Celsius). Variable definitions are provided in Appendix A. All continuous variables are winsorized at the 1st and 99th percentiles. The first-stage regression includes industry fixed effects and standard errors are clustered by industry. The second-stage regression include firm and year fixed effects, and standard errors are clustered by firm. Beneath each coefficient estimate is its p-value. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	First stage predicting *LnPost*	Second stage
DV=	*LnPost*	*LnQ*
	(1)	(2)
Badweather	1.461***	
	(0.000)	
PreLnPost×After		-0.170***
(instrumented)		(0.000)
Controls	Same as those in Table 3 column (6)	
Industry_FE	Yes	No
Firm_FE	No	Yes
Year_FE	No	Yes
Observations	1,464	14,640
R-squared	0.189	0.824

Figure 1: Stock Return around the Real Name Regulatory Change

Figure 1 shows the average monthly cumulative excess returns for treatment and control firms. Monthly cumulative excess return is calculated as the sum of monthly return in excess of market return. Treatment (control) firms are firms with the number of Guba posts above (below) the sample median in 2014.



Appendix A: Variable Definitions

Variable	Definition
After	Indicator variable that takes the value of one for the post-treatment period (i.e., the 2015-2019 period) and zero for the pre-treatment period (i.e., the 2010-2014 period).
AnalystCoverage	Natural logarithm of the number of analysts following the firm.
AssetTangibility	The ratio of net property, plant, and equipment over total assets.
BoardOwn	Percent of common shares outstanding held by the board.
Capex	Capital Expense/Assets.
CAR[-1,+1]	The three-day return around the quarterly earnings announcement date minus the value-weighted size matched portfolios return.
Cashflow	The annual cash flow from operations scaled by beginning of year assets.
CEOOwn	Percent of common shares outstanding held by the CEO.
CertifiedDirector	Percentage of all Non-Certified Inside Directors (Certified Inside Directors) on the board.
CrashRisk	Negative of the third moment of firm-specific weekly returns for each year, divided by the standard deviation of firm-specific weekly returns raised to the third power, for a given firm in a year.
EarningSurprise	The difference between actual quarterly earnings per share (EPS) and earnings in the same quarter of previous year scaled by the absolute value of EPS at q-4.
FirmSize	Natural logarithm of sales.
FounderDirector	Indicator variable that takes one if the founder is on the board and is zero otherwise.
GubaPostGrowth	The three-year average annual growth in Guba posts.
GubaNegativeTone	The ratio of Guba # negative posts minus # positive posts over total # posts.
InstitutionalOwnership	Percent of common shares outstanding held by institutional investors.
Ln(Asset)	Natural logarithm of year-end assets.
Ln(Comments)	Natural logarithm of total number of Guba comments (000).
Ln(FirmAge)	Natural logarithm of number of years since the firm's IPO.
Ln(Posts)	Natural logarithm of total number of Guba posts (000).
Ln(Segments)	Natural logarithm of the number of geographic segments.
LnTotalComments	Natural logarithm of Guba comments number (000) in a year.
LnTotalNegativePosts	Natural logarithm of Guba negative posts number (000) in a year.
LnTotalPosts	Natural logarithm of Guba posts number (000) in a year.
NegativeRatio	The ratio of the negative posts to all posts at quarter q
PreLnNegPost	Natural logarithm of negative Guba posts number (000) in the year prior to the real name regulatory change.
PreLnNonPosiPost	Natural logarithm of non-positive Guba posts number (000) in the year prior to the real name regulatory change.
PreLnPost	Natural logarithm of Guba posts number (000) in the year prior to the real name regulatory change.
QGrowth	The three-year average annual growth in Tobin's Q.
R&D	R&D/Assets.
Real Earnings Management	Real activities management is measured as abnormal production costs minus abnormal cash flow from operations and abnormal discretionary expenses.
Restatement	Indicator variable that takes one if firm announces the correction of misstatement in a subsequent reporting period by the end of 2022.
SalesGrowth	The three-year median annual growth in sales.
SOE	Indicator variable takes one for State-Own Enterprises and zero otherwise.
StkInform	The logit transformation of the $(1-R^2)$, where R^2 is obtained from the expanded value-weighted market and value-weighted industry index model using daily stock returns
Tobin's Q	The ratio of the market value of equity and book value of long-term debt all divided by total assets.
Voluntary Disclosure	Natural logarithm of the number of words in MD&A about future.

Appendix B: Data from Social Media

We collect the Guba data from East Money (http://guba.eastmoney.com), which is one of the largest and most comprehensive social media platforms focusing on the capital market in China. The platform includes all stocks that are currently trading or have been traded in the market, with independent sub-forums named according to the stock ticker or name of each company. We develop a web crawler to download all the posts (including posts and reply comments) on the discussion board of each company. Our social media dataset includes 220.26 million posts and 645.44 million comments, covering 4,098 unique Chinese firms from 2010 to 2019.

To construct the measurement of posts tone, we utilize machine learning techniques to measure the tone of Guba posts using Chinese sentiment dictionary from Dalian University of Technology, Tsinghua University, CNKI and Chinese translation of dictionary of Loughran and McDonald. We label tone of 5,000 posts as training set at the post-level because most posts are short and then classify all the posts into positive, negative, or neutral using Naive Bayes Classification (NBC). Since the social media user often using Emoji to express their opinions, we also consider the emoji in our model to define the tone. The in-sample accuracy reaches 85.4%, is larger than that of Antweiler and Frank (2004) (72.3%) using NBC for Yahoo! Finance messages. The out-of-sample validation using 1,000 randomly selected posts shows that the accuracy rate of our model is above 78% in labeling the tone of posts. More importantly, the number of posts classified as the opposite tone (positive posts classified as negative and vice versa) is low (0.4% and 0.2%). The Guba tone is defined as (# Negative Posts - # Positive Posts)/Total # Posts) at firm-year level. Our final sample includes about 124.8 million posts and 396.1 million comments for 1,464 unique firms over the period from 2010 to 2019.