Local Information Advantage and Stock Returns: Evidence from Social Media*

Yuqin Huang, Feng Li, Tong Li, and Tse-Chun Lin

April 2023

Abstract

We examine the information asymmetry between local and nonlocal investors with a large dataset of stock message board postings. We document that abnormal relative postings of a firm, i.e., unusual changes in the volume of postings from local versus nonlocal investors, capture locals' information advantage. This measure positively predicts firms' short-term stock returns as well as those of peer firms in the same city. Posting activities primarily reflect good news, potentially due to social transmission bias and short-sales constraints. We identify the information driving return predictability through content-based analysis. Abnormal relative postings also lead analysts' forecast revisions. Overall, investors' interactions on social media contain valuable geography-based private information.

JEL Classifications: D82, G12, G14.

Keywords: Local Information Advantage, Social Interactions, Topical Analysis, Stock Message Boards, Return Predictability.

*Yuqin Huang (huangyuqin@cufe.edu.cn) is at the School of Finance, Central University of Finance and Economics; Feng Li (feng.li@cufe.edu.cn) is at the School of Statistics and Mathematics, Central University of Finance and Economics; Tong Li (litong17@hku.hk) and Tse-Chun Lin (tsechunlin@hku.hk) are at the HKU Business School, the University of Hong Kong. We gratefully acknowledge helpful comments from Francois Brochet (the editor), two anonymous referees, as well as Justin Birru, Rui Chen, Lauren Cohen, Stefanos Delikouras, Xi Dong, Wenxuan Hou, Fuwei Jiang, Chen Lin, Yan Luo, Dragon Yongjun Tang, Shoufu Xu, Johan Walden, Ying Wang, Deniz Yavuz, Bohui Zhang, Xueyong Zhang, Guofu Zhou, and participants at 2016 AFA Annual Meeting, 2015 FMA Annual Meeting, 12th Chinese Finance Annual Meeting, and 4th Chinese Capital Markets Conference. All errors are our own. The paper was previously circulated under the title "Local Information Advantage, Investor Attention and Stock Returns." Huang gratefully acknowledges the financial support from the National Natural Science Foundation of China (Project No. 71673318) and the Major Program of National Fund of Philosophy and Social Science of China (No. 19ZDA098).

1 Introduction

Geography-based information asymmetry is a prominent phenomenon in financial markets. In theory, two views prevail. On one hand, local investors who are geographically close to a firm may have better access to private information about the stock. On the other hand, sophisticated foreign institutional investors may have a particular advantage in local markets over domestic investors because of their superior capabilities to gather and process information. Existing empirical evidence mostly supports the first view that local investors earn superior returns thanks to their information advantage.[1] In this paper, we approach the geography-based information asymmetry from a different angle. We examine the content of social media interactions as a function of users' location and study the consequent asset pricing implications. Said differently, based on the notion that investors are not isolated individuals but instead observe and communicate with each other (Hirshleifer, 2020), we examine the local information advantage revealed on Internet stock message boards and its association with local firms' future performance.

Local investors may enjoy an information advantage by gaining access to information earlier than distant investors. After receiving information about a firm, local investors may want to communicate more with others about this particular stock. Several reasons can explain such information-sharing behavior. First, investors have incentives to disseminate their private information if they have established a position in the stock and hope to move the price accordingly. Given that people tend to overweight the opinions of those with whom they interact, one has the motivation to share information and to become well-connected in the network and thus influential (DeMarzo, Vayanos, and Zwiebel, 2003). Second, even if they do not hold a position in the particular stock and therefore cannot realize capital gains, investors still have the incentive to share information because they could derive utility from information-sharing activities (Jame, Johnston, Markov, and Wolfe, 2016) and social recognition after the stock market subsequently confirms their opinions (Chen, De, Hu, and Hwang, 2014).[2] Consequently, local investors are likely to post more messages about the stock before the information reaches nonlocal investors. Hence, it is likely that the relative strength of investors' posting activities conveys local investors' information advantage.

[1] For example, Coval and Moskowitz (2001), Choe, Kho, and Stulz (2005), Dvořák (2005), Ivković and Weisbenner (2005), Kalev, Nguyen, and Oh (2008), Baik, Kang, and Kim (2010), Shive (2012), Sulaeman (2014), and Ferreira, Matos, Pereira, and Pires (2017) show that local investors possess an information advantage. The few exceptions are Chen, Johnson, Lin, and Liu (2009) and Chhaochharia, Kumar, and Niessen-Ruenzi (2012).

[2] Evidence from the psychology (Berger, 2014) and crowdsourcing (Estellés-Arolas and González-Ladrón-De-Guevara, 2012) literature suggests that individuals may be willing to share information to serve other functions such as impression management, emotion regulation, information acquisition, self-esteem, and skill development.

Although the information received by locals could be either good news or bad news, posting activities may not reflect positive and negative information symmetrically. As suggested by Han, Hirshleifer, and Walden (2022), investors suffer from self-enhancing transmission bias and are more willing to talk about their excellent investments than their poor investment performance. Moreover, investors tend to selectively pay more attention to their investments after good news arrives (Sicherman, Loewenstein, Seppi, and Utkus, 2015). This behavior, known as the ostrich effect, may arise from the preference for optimism, the desire to maintain a positive self-image as a good investor, or the incentive to avoid unpleasure caused by bad news. In addition, retail investors are disincentivized to post negative information because short-sales constraints make it difficult for them to profit from bad news. Finally, investors tend to avoid publicizing negative information since spreading bad news could be risky and may lead to severe punishment.[3] All arguments above suggest that local investors are more likely to post messages after acquiring positive information. Therefore, we expect stock message postings to capture good news better than bad news.

If local investors have superior access to value-relevant information, which has not reached distant investors when locals start posting more messages, it may take time for such information to be fully incorporated into stock prices. As information gradually transmits from local investors to nonlocal investors through social interactions, stock prices slowly adjust. Thus, the relative volume of postings from locals and nonlocals should help predict future stock returns. Moreover, a higher relative posting volume is likely associated with higher stock returns if message postings are more likely to convey positive information.

Our message posting data come from Guba Eastmoney, the most popular stock message board in China. Unlike other message boards which identify posters with their username, Guba Eastmoney allows users to read and post messages without logging into a registered account, and it identifies these users with a non-confidential IP address.[4] This unique feature enables us to distinguish local posters from nonlocal posters and investigate local information advantage with message postings. Our analysis covers more than 300 million postings on 2,239 listed firms in Chinese A-share markets. With this comprehensive dataset, we construct abnormal relative postings (ARP) to capture unusual patterns in posting activities by local relative to nonlocal investors, where local (nonlocal) investors refer to those who are (are not) located in the same city as the focal firm. Specifically, we compute

[3] For anecdotal evidence, we refer the reader to news articles at https://www.wsj.com/articles/SB123178875223174403 and https://www.bbc.com/news/world-asia-china-32964606.

[4] Our data was collected in June 2013. Guba Eastmoney has changed its disclosure policy and required a registered account for participants to post a message.

the ARP measure as the difference between current relative postings and the normal level of relative postings in a few preceding periods, where relative postings are the log difference between the number of messages from local posters and the number of messages from nonlocal posters.

We find strong and robust evidence that higher ARP predicts higher future stock returns. A portfolio that longs high-ARP stocks and shorts low-ARP stocks earns an excess return of 0.24% per week (equivalent to 12.5% per annum), suggesting that the information embedded in posting activities has significant economic value. Moreover, a firm's ARP measure positively predicts subsequent stock returns for firms headquartered in the same city, though with a smaller magnitude. The cross-firm return predictability indicates that local investors receive not only firm-specific information but also location-specific information. The results further corroborate our interpretation that geography-based information can be revealed by social media. Our results are robust to both equal-weighted and value-weighted returns, abnormal returns from alternative factor models, and various construction methods for the ARP measure. Further tests confirm that the variation in our ARP measure is driven by the variation in local information rather than that in local Internet traffic. The positive ARP-return relation confirms that abnormal relative postings are likely to convey good rather than bad news. Consistent with this view, abnormal returns of the high-ARP portfolio are positive and large in magnitude, while alphas of the low-ARP portfolio are close to zero.

We also run Fama and MacBeth (1973) regressions and find that the cross-sectional stock return predictability of the ARP measure cannot be explained by a list of well-known determinants of expected stock returns. In particular, the positive ARP-return relation holds after controlling for firm size, book-to-market ratio, momentum, reversal, asset growth, profitability, idiosyncratic volatility, illiquidity, institutional ownership, insider trading, local media coverage, and variables related to the degree of local bias. Further analysis shows that the return predictability lasts for four weeks. More importantly, we find that stock prices will not reverse beyond the one-month horizon, suggesting that our results are unlikely driven by temporary price pressure.

If local investors' information advantage drives the return predictability of our ARP measure, one may expect that the ARP measure better predicts the returns of firms whose investors suffer from more severe information asymmetry. Our evidence supports this conjecture. First, the positive ARP-return is more pronounced among firms located in under-developed and less transparent regions. Second, the predictive power of ARP for stock returns is stronger among firms with worse information environments and higher limits to arbitrage.

We then analyze the information content of posts on stock message boards to better understand what information drives the return predictability. Topical analysis based on the Latent Dirichlet Allocation (LDA) method reveals that posts from local investors usually contain discussions about firm fundamentals, stock trading activities, and unsystematic noises. We further show that messages about fundamentals and trading are informative about future stock returns. Additional tests suggest that local investors tend to be better informed about firms' investment activities.

Given that abnormal relative postings are informative about firms' future fundamentals, one may expect that they would lead analysts' forecast revisions. Consistent with this conjecture, we find that abnormal relative postings positively predict revisions in analysts' earnings forecasts and recommendations. Specifically, a higher value of ARP is associated with a higher (lower) portion of analysts who revise their earnings forecasts upward (downward) and analysts who upgrade (downgrade) their recommendations. These results again support the idea that local postings mainly convey positive private information in capital markets.

Our paper contributes to the literature on geography-based information asymmetry. Existing studies based on both the U.S. (e.g., Bernile, Kumar, and Sulaeman, 2015) and other markets (e.g., Dvořák, 2005; Choe, Kho, and Stulz, 2005) show that investors who are geographically close to the firm earn superior returns. Both local institutional investors and local retail investors enjoy such information advantage (Ivković and Weisbenner, 2005).[5] Given this widespread perception, distant investors are eager to close the gap by following what locals do (Sulaeman, 2014). However, this is a difficult task since local investors' transaction records are not accessible to the public. In contrast to prior studies relying on confidential information from investors' investment accounts, we establish that publicly-available message postings on Internet stock message boards capture locals' superior information. More importantly, we document that the information revealed by abnormal relative postings has significant economic value. Compared with prior research, our findings provide more meaningful guidance for practitioners. Recent work by Bernile, Kumar, Sulaeman, and Wang (2019) documents that, although the long-horizon information advantage of local institutional investors in the U.S. has been eroded by the availability of more efficient communication technologies and the improvement in firms' information environment, these investors continue to make superior profits from their intraquarter trading activities. Our evidence from China differs from and complements

[5] Consistent with the view that local investors are informed, prior studies (e.g., Shive, 2012; Brown, Stice, and White, 2015) show that the adverse selection effect in stock prices becomes less severe when local investors' trading activities are constrained.

their findings by showing that local retail investors also still enjoy an information advantage (revealed by their posts on message boards) in the short horizon.

Our work is related to Chi and Shanthikumar (2017) who investigate how local bias in Internet search affects the market response to earnings information. Both papers share the common theme of investors' local bias and Internet activities. However, while their focus is on information search on the Internet, i.e., the *demand* for information, ours is on information sharing, i.e., the *supply* of information. Hence, the economic mechanisms of how private information held by local investors is incorporated into stock prices are fundamentally different. More importantly, their study examines variables that do not specify the direction of private information such as bid-ask spread, trading volume, and earnings response coefficients, but we show that our measure predicts stock returns and analyst revisions. Our results thus provide direct evidence of local information advantage, supporting the supply of information mechanism.

Similar to Chi and Shanthikumar (2017), several other papers study local bias in information demand. By examining Internet search volume around insider sales, Omer, Roulstone, Sharp, and Twedt (2012) provide suggestive evidence for local information leakage before the news of insider selling becomes public. Cziraki, Mondria, and Wu (2021) demonstrate that the asymmetric attention allocation by local investors relative to nonlocals as proxied by the Google search volume is positively associated with one-month-ahead stock returns. Chen (2022) studies whether the dispersion in locations of the requests for firms' EDGAR filings induces more trading volume around the earnings announcements and a stronger market response to earnings news. Our paper differs from these studies as we provide evidence that the information supplied by local investors in stock messages could predict local firms' performance. However, given the nature of the intertwined relation between information demand and supply, i.e., those who supply information may also demand information to verify their private signal, we do not rule out that our local investors also demand information.

Our paper is also related to Baik, Cao, Choi, and Kim (2016) who document that the negative tone of local tweets predicts future stock returns and subsequent earnings announcement returns. Although both papers suggest that local social media contain value-relevant information, our paper differs from their study in several distinct dimensions. First, they focus on the United States while our study analyzes China, where retail trading is more prominent. Second, we conduct topical analysis to identify the information that drives the return predictability. Third, in addition to the stock returns of the focal firm, we show that abnormal relative postings can predict peer firms'

returns and analysts' forecast revisions. Fourth, different from their findings, our evidence suggests that local postings mainly reflect good news, which is consistent with social transmission bias and the existence of short-sales constraints. Finally, their sample contains only 645 unique firms for a nine-month period, whereas our sample is from a large Internet message board, which is designed for investors to discuss stocks, spans six years and covers more than 300 million messages for 2,239 firms.

Our paper also adds to the literature on social interactions in financial markets. Earlier investigations focus on how social interactions affect individual behavior (e.g., Hong, Kubik, and Stein, 2004) and corporate outcomes (e.g., Bizjak, Lemmon, and Whitby, 2009). Recent research on social networks (e.g., Ozsoylev and Walden, 2011; Han and Yang, 2013; Walden, 2019) suggests that information diffuses gradually through networks among investors and that information networks have important implications for assets prices, trading volume, liquidity, as well as traders' trading profits and welfare. Moreover, peer pressure associated with social interactions is shown to contribute to investors' disposition effect (Heimer, 2016). In particular, investors tend to sell winning assets and hold onto losers in order to maintain a positive self-image within social networks. The impression management motives also lead to biases in social transmission (e.g., Han, Hirshleifer, and Walden, 2022; Simon and Heimer, 2015). Our research suggests that local investors convey value-relevant information via social media, especially when the information is positive. These results facilitate our understanding of investors' information-sharing behavior via social media.

Last but not least, our paper extends the emerging literature attempting to better understand individuals' behavior and its implications for financial markets using the information they share online. Researchers have analyzed information contained in various sources available online, including tweets (e.g., Bartov, Faurel, and Mohanram, 2018; Crowley, Huang, and Lu, 2022), customer product reviews (Huang, 2018), crowdfunding campaigns for individual entrepreneurs' projects (Lin and Pursiainen, 2022, 2021), information about accessing or registering on a website (Berg, Burg, Gombović, and Puri, 2020), and more closely related to our paper, messages from online investment forums (e.g., Antweiler and Frank, 2004; Chen, De, Hu, and Hwang, 2014).[6] Existing studies on stock message boards investigate the aggregate posting volume but do not distinguish between local

[6] Other examples of recent accounting and finance studies that utilize data on social media include Johnston, Kang, and Wolfe (2013), Blankespoor, Miller, and White (2014), Huang, Qiu, and Wu (2016), Jame, Johnston, Markov, and Wolfe (2016), Chawla, Da, Xu, and Ye (2021), Drake, Thornock, and Twedt (2017), Tang (2018), Blankespoor, Dehaan, Wertz, and Zhu (2019), Campbell, DeAngelis, and Moon (2019), Jiang, Liu, and Yang (2019), Cookson and Niessner (2020), Hsu, Ang, Tang, and Wu (2021), Huang, Li, and Markov (2020), Lerman (2020), and Drake, Moon, Twedt, and Warren (2022). See, e.g., Blankespoor, deHaan, and Marinovic (2020) for a review.

postings and nonlocal postings.[7] The unique data from Guba Eastmoney allow us to conduct such an analysis, echoing Teoh (2018) on the promise of new datasets for accounting and finance research.

2 Data and Measures

2.1 Data Sources

Our stock message data are from Guba Eastmoney.[8] According to statistics of unique visitors and page views, Guba Eastmoney is the most popular stock message board in China. This site provides a unique message board for every stock. Huang, Qiu, and Wu (2016) and Jiang, Liu, and Yang (2019) employ the Guba Eastmoney data to explore investor attention and stock return comovement, respectively.

On Guba Eastmoney, users are free to decide whether to create an account or not; and they are allowed to read, post, or reply to stock messages even without logging into a registered account. Posters who log in to an account are identified by their account ID, while other participants (guest users) are identified by their IP address. Figure 1 provides a screenshot of the message board for Shanghai Airport (Ticker: 600009). As it shows, the majority of authors are identified by their IP address. In our sample, about 86% of the postings are from guest users.

We use a "Web-scraper" program to download the message posting information. Each piece of posting information includes (1) a unique ID number; (2) the number of clicks; (3) the number of replies; (4) title of the original post; (5) author of the post (either the registered account ID or the IP address); (6) content of the post; (7) date of the post; and (8) time of the post.

We employ the IP-address geolocation technology (the QQ database) to locate each guest user with their IP address. We only retain messages with IP addresses that indicate a location in mainland China. Consequently, each message poster in our final sample is matched with cities in the 31 provincial-level administrative divisions in mainland China.[9] Following prior research, we regard a firm's headquarters city as its home city. In our paper, a firm's "local investors" ("nonlocal investors") refer to posters who are (are not) located in the same city as the firm.

[7] Hao, Dong, and Wu (2019) examine how aggregate posts affect the market reaction to earnings news using data from the same stock message board as ours. Based on data on 19 firms, they find that more online posts after earnings announcements facilitate the incorporation of earnings surprises into stock prices.

[8] The website is `http://guba.eastmoney.com`.

[9] The hierarchy of administrative divisions in China is as follows: provincial level (province, autonomous region, and municipality), prefectural level (city), county level, township level, and village level.

Figure 1
Screenshot from Guba Eastmoney

This figure is a screenshot from Guba Eastmoney, the most popular Internet stock message board in China. The site provides a unique message board for every stock. It identifies posters who log in to a registered account by their account ID and labels guest users with their IP addresses. Each piece of posting information includes (1) a unique ID number; (2) the number of clicks; (3) the number of replies; (4) title of the original post; (5) author of the post (either the registered account ID or the IP address); (6) content of the post; (7) date of the post; and (8) time of the post.

Table 1 presents examples of messages posted by local investors.[10] The message contents suggest that local investors indeed have access to nonpublic information, which they could obtain through several channels. For example, they may work at the company or know someone who is familiar with the company (see, e.g., Message #4: "*I also learned from employees of Zhongtai that good news would be released*"). Alternatively, they may learn the information from certain local events that are related to the firm (see, e.g., Message #2: "*Jizhong Energy is holding a party at the Jinniu Hotel to celebrate its good start in the first quarter*"). Note that although locals appear to be better informed than nonlocals, most local postings do not explicitly reveal specific private information about the firm. Instead, the messages often convey vague signals about the information possessed by the poster. This is reasonable because divulging and spreading insider information is strictly prohibited in China.[11]

[Insert Table 1 Here]

We obtain firm location information, accounting information, stock trading data, institutional

[10] Additional examples are provided in the Internet Appendix (Table IA.1).
[11] See Article 53 of the Securities Law of the People's Republic of China at https://www.lawinfochina.com/display.aspx?id=31925&lib=law&EncodingName=big5.

ownership, analyst forecasts, earnings announcement, and media coverage information from the China Stock Market & Accounting Research (CSMAR) database. In addition, the information on city-level GDP, population, area, and employment is collected from the China City Statistical Yearbook. Our sample period is from June 2007 to May 2013. After excluding financial firms, our final dataset covers 2,239 A-share listed companies.

2.2 Abnormal Posting Measure

In this paper, we use relative postings (RP) to measure the relative strength of posting activities by locals and nonlocals. For firm i headquartered in city c, its relative postings measure in week t is calculated as:

$$RP_{i_c,t} = \ln\left(1 + P^{c}_{i_c,t}\right) - \ln\left(1 + P^{-c}_{i_c,t}\right) \tag{1}$$

where $P^{c}_{i_c,t}$ ($P^{-c}_{i_c,t}$) is local (nonlocal) postings, that is, the total number of messages posted in week t by investors in (outside) city c.

To measure unusual changes in relative postings, we construct abnormal relative postings (ARP), which is computed as the difference between current relative postings and the normal level of relative postings in a few preceding periods. Specifically, we define the ARP measure for firm i in week t as its RP in week t minus the median of its RP in the previous ten weeks. That is,

$$ARP_{i_c,t} = RP_{i_c,t} - \text{median}\left(RP_{i_c,t-1}, RP_{i_c,t-2}, \cdots, RP_{i_c,t-10}\right) \tag{2}$$

The ARP measure captures abnormal variations in message postings for a stock by locals relative to nonlocals. A higher value of ARP means that compared with their respective normal levels, local investors' postings are much higher than those from nonlocals. Table 2 shows that the mean and median of ARP are almost zero, but there are significant variations in ARP.

[Insert Table 2 Here]

3 Postings and Stock Returns

3.1 Portfolio Sorts

If the information conveyed by investors' posting activities diffuses gradually through information networks, then abnormal relative postings should help predict future stock returns. To investigate

the relation between abnormal relative postings and firms' subsequent stock returns, we first perform univariate portfolio sorts. Specifically, we sort all stocks into quintiles based on their ARP measure in week t. We then track the performance of each portfolio over week $t+1$. The portfolio performance is presented in Panel A of Table 3. For each portfolio, we report the weekly average equal-weighted excess returns, value-weighted excess returns, abnormal returns (alphas) from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model.[12] We also report the performance of a long-short portfolio which longs stocks in the highest ARP quintile and shorts stocks in the lowest ARP quintile.

[Insert Table 3 Here]

The results show that firms with higher ARP earn higher returns in the subsequent week. The average excess return increases monotonically from –0.19% to 0.05% from the lowest to the highest equal-weighted ARP quintile portfolios. Consequently, the equal-weighted long-short portfolio earns an excess return of 0.24% per week (equivalent to 12.5% per annum), which is statistically significant at the 1% level ($t = 5.72$). This increasing pattern also holds for value-weighted excess returns and abnormal returns from alternative factor models. For example, abnormal returns from the Carhart (1997) four-factor model increase from –0.02% in the low-ARP portfolio to 0.22% in the high-ARP portfolio, resulting in a weekly alpha of 0.24% ($t = 5.55$) for the long-short portfolio. The evidence that high-ARP firms earn positive alphas is consistent with our view that high values of ARP indicate positive private information possessed by local investors. Moreover, the result that abnormal returns of the high-ARP portfolio are large in magnitude but alphas of the low-ARP portfolio are close to zero further confirms our view that posting activities capture more good news than bad news.[13]

Panel B of Table 3 presents the cross-sectional average firm or city characteristics for stocks in each ARP-sorted portfolio. Specifically, we report average firm size, book-to-market ratio, asset-growth rate, return-on-assets, idiosyncratic volatility, Amihud (2002) illiquidity, institutional ownership, insider trading activity, abnormal local media coverage, local population density, the logarithm of local GDP per capita, and share of industry employees. The results show that stocks in the extreme ARP portfolios do not differ significantly in these dimensions, suggesting that the outperformance of high-ARP stocks is unlikely driven by these firm or city characteristics.

[12] As shown in Table IA.2 in the Internet Appendix, we obtain similar results using industry-adjusted returns and returns adjusted by Daniel, Grinblatt, Titman, and Wermers (1997) characteristic-based benchmarks.

[13] The long-short portfolio based on our ARP measure earns positive excess returns in all subsamples classified by industries (Figure IA.1 in the Internet Appendix).

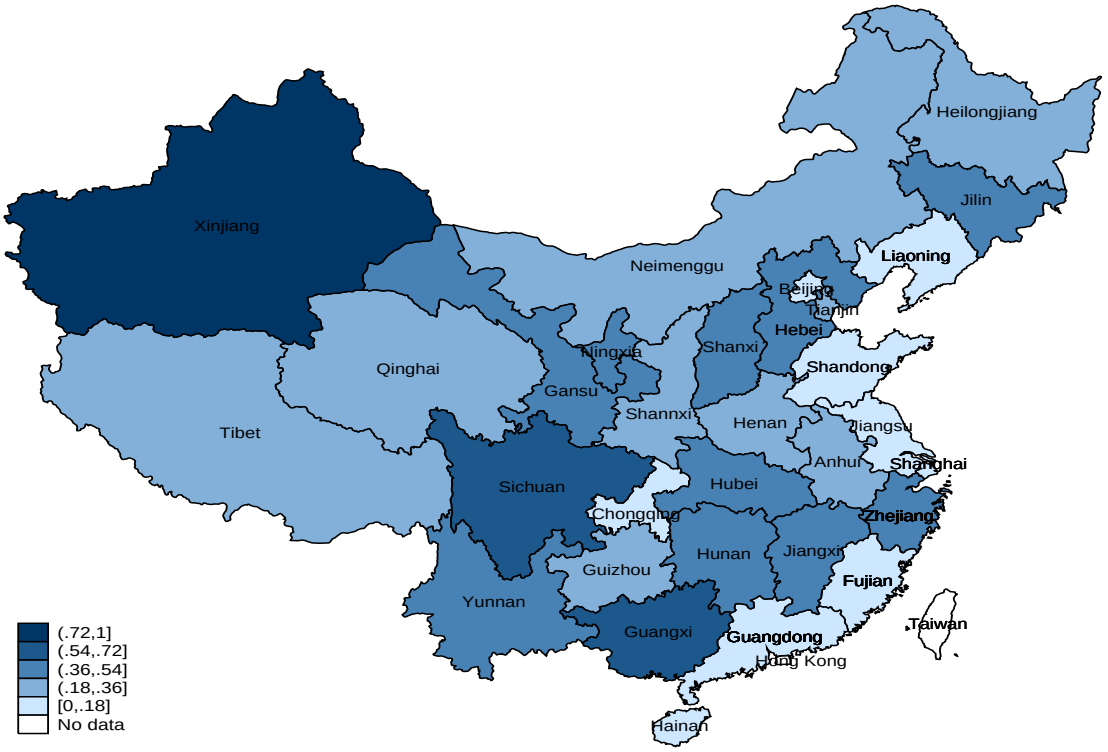


Figure 2
Profits of Long-short Strategy by Administrative Divisions

This figure plots profits of the ARP-sorted long-short strategy for the thirty-one provincial administrative divisions (provinces, autonomous regions, and municipalities) in mainland China. Each week, firms in every administrative division are sorted into quintile portfolios based on their ARP measure, and portfolios are rebalanced every week. The ARP measure is constructed as relative postings for a firm in one week minus the median value of its relative postings in the previous ten weeks where the relative posting measure is the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. Average weekly equal-weighted excess returns (in percentage) of the high-minus-low ARP portfolios are plotted for each administrative division, with darker colors indicating higher returns. The sample period is from June 2007 to May 2013.

To better illustrate geographical variations in the return predictability of our ARP measure, we plot the performance of ARP-sorted long-short portfolios formed with stocks in different locations. Each week, firms in every provincial administrative region are sorted into quintile portfolios based on their ARP measure. Figure 2 displays the return difference between equal-weighted excess returns of high-ARP portfolios and low-ARP portfolios. It shows that the long-short portfolio based on our ARP measure generates positive excess returns in all geographical areas. Interestingly, it also demonstrates that the ARP-based trading strategy is more profitable in under-developed inland regions where firms are relatively opaque than in well-developed coastal regions where firms are relatively transparent. This pattern is consistent with the return predictability of abnormal relative postings arising from the ability of ARP to capture information asymmetry between locals and nonlocals.

Prior studies (Pirinsky and Wang, 2006; Parsons, Sabbatucci, and Titman, 2020; Huang, Lin, and Xiang, 2021) find that stock returns of co-headquartered firms can be driven by common fundamental shocks that are location-specific. The private information held by local investors could be firm-specific. However, it could also be location-specific. If the private information local investors hold is location-specific, we expect that abnormal relative postings also predict stock returns of firms in the same location. Hence, to test this conjecture, we investigate the cross-firm return predictability of ARP. Specifically, we sort firms into quintile portfolios every week based on the median value of peer firms' ARP measure, in which peer firms refer to firms that share the same headquarters city with the focal firm.[14] We then track portfolio performance in the following week. As shown in Table 4, our ARP measure exhibits significant predictive power for peer firms' stock returns. This finding corroborates our interpretation that local investors have better access to the location-specific information contained in social media.

[Insert Table 4 Here]

3.2 Fama-MacBeth Regressions

In addition to portfolio sorts, we estimate Fama and MacBeth (1973) regressions that allow us to control for a list of determinants of expected returns. The model specification is as follows:

$$R_{i,t+1} = \alpha + \beta ARP_{i,t} + \delta \mathbf{X_{i,t}} + \epsilon_{i,t+1} \tag{3}$$

where $R_{i,t+1}$ is the excess return on stock i in week $t+1$; $ARP_{i,t}$ is abnormal relative postings of firm i in week t; and $\mathbf{X_{i,t}}$ is a vector of firm-level characteristics. Each week we estimate a cross-sectional regression of weekly stock returns on lagged abnormal relative postings as well as control variables, and the time-series average of coefficient estimates is taken as the final estimate. β in Equation (3) is our coefficient of interest.

Table 5 reports coefficient estimates from Fama and MacBeth (1973) regressions together with t-statistics based on Newey and West (1987) adjusted standard errors with three lags. In Column (1), stock returns are regressed on the ARP measure only. The coefficient on ARP is positive (0.91) and significant at the 1% level ($t = 5.49$). In Column (2), we include standard variables related to expected stock returns, namely, firm size (Size), book-to-market ratio (BM), reversal ($\text{Ret}_{\text{t–4:t–1}}$),

[14] We focus on firms with more than fifteen peer firms. This restriction leaves us with 1,589 (out of 2,239) firms which are located in around forty cities.

and momentum ($\text{Ret}_{t-52:t-5}$). Specifically, Size is the logarithm of market capitalization calculated as shares outstanding multiplied by the closing price at the end of the previous week; BM is the ratio of the book value of equity at the end of the previous fiscal year divided by market value of equity at the end of the previous week; $\text{Ret}_{t-4:t-1}$ is the stock return over the previous four weeks; $\text{Ret}_{t-52:t-5}$ is the stock return over 48 weeks preceding the previous month. We find that the coefficient on ARP reduces to 0.86 but remains statistically significant ($t = 5.57$).

Column (3) in Table 5 introduces additional firm characteristics that affect stock returns. In particular, we include annual asset growth rate (Cooper, Gulen, and Schill, 2008, AG), return-on-assets (Haugen, Baker, et al., 1996, ROA), idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang, 2006, IVOL), illiquidity (Amihud, 2002, ILLIQ). These variables are defined as follows: AG is the annual growth rate of total assets; ROA is net income divided by total assets; IVOL is the standard deviation of residuals from the Carhart (1997) four-factor model with daily stock returns in the past three months in which at least fifteen observations are required in the calculation; ILLIQ measure is computed as the weekly average of the ratio of absolute daily price change to daily trading volume.

Additionally, we incorporate several variables that are related to the transmission of local information and investors' local bias. First, we consider two variables that are related to firms' information environment. That is, institutional ownership (IO), defined as the percentage of shares outstanding owned by institutional investors; and insiders' net purchase ratio (NPR), defined as the number of purchases minus the number of sales divided by the total number of transactions by managers and large shareholders of a firm in a given week (Lakonishok and Lee, 2001). Second, one could argue that local posters might simply repeat the information disseminated by local media. To distinguish between the effect of message postings and that of media coverage, we consider abnormal local media coverage (ALMedia), which is computed as local media coverage on a firm in a given week (i.e., the logarithm of one plus the number of news reports issued by news agencies located in the firm's headquarters city within a week) minus the median value of its local media coverage in the previous ten weeks. Third, one may worry that abnormal relative postings capture the local bias of individuals' investment portfolios. To alleviate this concern, we incorporate important regional characteristics that are related to local bias in China (e.g., Hong, Jiang, Wang, and Zhao, 2014; Shao and Wang, 2021) such as the logarithm of local GDP per capita (Log(GDP)) and local population

density (PopDensity).[15] Finally, if the economy in a city heavily relies on one industry, abnormal relative postings might capture industry information rather than local information. Therefore, we control the industry-level employment share in a city (EmpShare), which is the total number of employees in an industry in a given city scaled by the total number of employees in the city.

In this model specification, the coefficient on ARP is 0.81, with a t-statistic of 5.39. The positive coefficient on ARP is consistent with our portfolio sorting results and supports our view that higher ARP predicts higher future stock returns. The result also suggests that well-known determinants of cross-sectional stock returns cannot explain the positive ARP-return relation.

[Insert Table 5 Here]

If message postings mainly capture positive information received by local investors, we expect that the positive ARP-return relation concentrates in the sample where the ARP measure is positive (i.e., locals post abnormally more messages compared with nonlocals). To test this conjecture, we rerun Equation (3) using subsamples with positive and negative ARP separately. Consistent with this asymmetric effect, Table IA.3 in the Internet Appendix shows that positive coefficients on ARP are statistically significant only in the subsample with positive abnormal relative postings.

We also examine the ability of our ARP measure to predict stock returns beyond one week. In Column (4) of Table 5, we regress two-week-ahead stock returns on the ARP measure as well as control variables. We find that, although the coefficient on ARP reduces in magnitude to 0.39, it remains statistically significant ($t = 2.88$).[16] As shown in Columns (5) to (8), the predictability of ARP for stock returns lasts for four weeks. More importantly, we do not find any reversals up to twelve weeks ahead. Our results suggest that the return predictability of abnormal relative postings arises from the slow diffusion of value-relevant information rather than temporary price pressure associated with limited investor attention (Barber and Odean, 2008) or a risk premium associated with firms' geographic concentration (Garcia and Norli, 2012).

[15] Prior studies document that as investors in the US and other countries, investors in mainland China exhibit significant local bias in their portfolio choices. On average, Chinese investors over-weight stocks of firms from their province of residence by approximately 9% relative to the market portfolio (Feng and Seasholes, 2008). The magnitude of local bias is similar among men and women. Shao and Wang (2021) show that local bias is higher in provinces with more population and lower GDP per capita.

[16] In an untabulated analysis, we employ a bootstrapping method to test whether the difference in coefficients on ARP (i.e., the coefficient on ARP in Column (3) minus the coefficient on ARP in Columns (4) and (8)) is statistically significant. We find the bootstrap-based 95% confidence intervals for the difference in coefficients do not contain zero, suggesting the differences are statistically different from zero at the 5% level.

3.3 Information Environment and Limits to Arbitrage

If abnormal relative postings capture local investors' information advantage relative to nonlocal investors, one may expect that the predictability of our ARP measure for future stock returns is stronger among firms with less transparent information environments and higher limits to arbitrage. To test this conjecture, we investigate the variation in returns across ARP-sorted portfolios conditional on proxies for firms' informational efficiency and limits to arbitrage. Specifically, each week we first split stocks in our sample into tercile groups based on one proxy and then sort stocks in each group into quintile portfolios based on their ARP measure.

Table 6 reports the performance of ARP-sorted long-short portfolios in subsamples with extreme values of information environment proxies and the measure of limits to arbitrage. As in Panel A of Table 3, we report the weekly average equal-weighted excess returns, value-weighted excess returns, abnormal returns (alphas) from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model. In Panel A, firms are split based on their market capitalization. The predictive power of ARP for future stock returns is stronger among small firms than large firms. For example, the equal-weighted long-short portfolio based on ARP earns an excess return of 0.42% (0.13%) in the small-cap (large-cap) group. The difference in the return gaps between small and large firms (0.29%) is statistically significant at the 1% level. The results are consistent with our expectation since small firms usually have worse information environments.

[Insert Table 6 Here]

We also consider analyst coverage, computed as the number of analysts providing earnings forecasts for a firm in the previous fiscal year. Higher analyst coverage tends to be associated with higher transparency (Lang and Lundholm, 1996). As expected, Panel B in Table 6 shows that the return predictability of ARP is concentrated among firms with low analyst coverage. For example, abnormal returns from the Carhart (1997) four-factor model are 0.30% (t = 4.22) and 0.06% (t = 0.71) in the low-analyst-coverage and high-analyst-coverage group, respectively. The difference in these return gaps (0.24%) is statistically significant (t = 2.40).

We then divide our sample based on institutional ownership. Since institutional investors could enhance corporate transparency by influencing firms' disclosure policy (Ajinkya, Bhojraj, and Sengupta, 2005; Boone and White, 2015), we expect stronger return predictability of ARP among firms with low institutional ownership. This expectation is supported by our results (Panel C

in Table 6). Regardless of which performance measure is used, the outperformance of high-ARP firms relative to low-ARP firms is more pronounced in the low-institutional-ownership group. The differences in return gaps between low-institutional-ownership and high-institutional-ownership firms are statistically significant, except with value-weighted returns.

Finally, we examine the role of stock return idiosyncratic volatility, which is commonly used as a proxy for arbitrage risks (e.g., Stambaugh, Yu, and Yuan, 2015). Higher idiosyncratic volatility is associated with higher arbitrage risks and limits to arbitrage. We thus anticipate stronger results for stocks with higher idiosyncratic volatility. Panel D in Table 6 shows that the positive ARP-return relation is indeed concentrated among stocks with high idiosyncratic volatility.

Overall, results in Table 6 show that the positive ARP-return relation is stronger among firms with opaque information environment, including small firms, firms with low analyst coverage, and firms with low institutional ownership. The relation is also more pronounced among firms with higher limits to arbitrage. The evidence further supports the idea that abnormal relative postings reflect information asymmetry between local and nonlocal investors.

3.4 Local Internet Traffic

Our ARP measure captures the abnormal change in the relative volume of postings from locals and nonlocals. A higher value of ARP suggests local investors are posting abnormally more messages about local firms relative to nonlocal investors. As we have explained, an increase in abnormal relative postings indicates that local investors are accessing value-relevant information about local firms that has not yet reached distant nonlocal investors. However, one could argue that an increase in the ARP measure might reflect an increase in the overall Internet traffic in a given region rather than an increase in local information. To put it differently, we would observe higher abnormal relative postings if investors in a city are posting abnormally more messages in general (i.e., for both local and nonlocal firms) compared to investors in other cities.

To address the potential effects of the change in local Internet traffic, we construct a measure of standardized abnormal relative postings (ARP^{s}). For firm i headquartered in city c, the standardized abnormal relative postings in week t is calculated as follows:

$$ARP^{s}_{i_c,t} = ARP_{i_c,t} - ARP_{-c,t},$$

where $ARP_{-c,t} = RP_{-c,t} - median\left(RP_{-c,t-1}, RP_{-c,t-2}, \cdots, RP_{-c,t-10}\right)$ is the abnormal relative

postings of firms not headquartered in city c in week t. $RP_{-c,t}$ is the relative postings of firms not headquartered in city c in week t, and it captures the number of postings from investors in city c for nonlocal firms (i.e., firms not headquartered in city c) relative to the number of postings from investors outside city c for these firms. That is, $RP_{-c,t} = \ln(1 + \sum_{j_d, d \neq c} P^{c}_{j_d,t}) - \ln(1 + \sum_{j_d, d \neq c} P^{-c}_{j_d,t})$, where $P^{c}_{j_d,t}$ ($P^{-c}_{j_d,t}$) represents the number of messages posted in week t for firm j headquartered in city d by investors located in (outside) city c.

Table IA.5 in the Internet Appendix reports the relation between the standardized abnormal relative postings and stock returns. Panel A shows that the average excess returns and abnormal returns of the high-minus-low ARP^{s} portfolios are positive and statistically significant, which is consistent with our baseline findings. Moreover, the magnitude of profits from the long-short portfolios is comparable to that reported in Table 3. In addition, as shown in Panel B, the standardized measure positively predicts peer firms' stock returns. These results suggest that the variation in our abnormal relative postings measure is unlikely to be solely driven by the variation in local Internet traffic.

3.5 Alternative Posting Measures

Our analysis so far classifies posters as local posters for a firm if they are from the firm's headquarters city. In addition to this natural definition of locality based on city boundaries, prior research adopts alternative classification standards such as geographic distance between a firm and investors (e.g., Coval and Moskowitz, 2001; Ivković and Weisbenner, 2005) and boundaries across states (e.g., Baik, Kang, and Kim, 2010; Bernile, Kumar, and Sulaeman, 2015). In this subsection, we present regression results with the ARP measure constructed based on alternative definitions of investors' locality. First, we consider investors located within a certain radius of a firm's headquarters city as local investors for that company. Column (1) of Table IA.6 in the Internet Appendix reports results with a radius equal to 100 KM.[17] The results show that the coefficient on ARP is positive (0.45) and statistically significant at the 1% level ($t = 3.26$). Second, we define local posters as those from the province (autonomous region or municipality) in which a firm's headquarters is located. As shown in Column (2) of Table IA.6, this definition does not change our conclusions.

Moreover, our baseline ARP measure treats all message postings equally. One may argue that some postings are more important than others and thus should be given a higher weight. For example, as suggested in Antweiler and Frank (2004), longer messages are likely to contain more information

[17] Our results are qualitatively and quantitatively similar when using other cutoffs such as 50 KM or 200 KM.

and thus are more important than shorter ones, implying that the length of a message is a potential measure of its importance. To take into account the heterogeneity in the importance of messages, we construct an alternative ARP measure in which the posting volume is weighted by message length, that is, the number of characters contained in a message posting. Column (3) in Table IA.6 shows that the length-weighted ARP measure still positively predicts firms' future stock returns. Overall, the results in Table IA.6 suggest that the positive ARP-return relation is robust to alternative measures of abnormal relative postings. Our conclusions also remain the same if the ARP measure is constructed with alternative definitions of normal relative postings or if relative postings are calculated as the difference between the number of local postings and nonlocal postings divided by the number of total postings. These results are also presented in the Internet Appendix (Table IA.7).

4 Information Content of Local Postings

Our results so far have shown that the abnormal online posts from local investors relative to those from nonlocal investors are informative for firms' stock returns. To better understand this return predictability, we analyze the contents of stock messages. We aim to figure out what topics investors tend to discuss on the Internet stock message board and which topics predict returns. Therefore, we first perform a topical analysis for local posts based on the topic model proposed by Blei, Ng, and Jordan (2003), keeping in mind the complexity of analyzing text from social media.[18] The implementation of this analysis involves the following steps. First, we segment local posts into tokens that contain single words and adjacent words. Next, we exclude stopwords from our analysis. We then employ the Latent Dirichlet Allocation (LDA) method to identify topics in the messages. Furthermore, we use a topic model to generate the distribution of these topics, ensuring that it fits our data accurately.[19]

Figure 3 plots the results from the topical analysis. Subfigure (a) displays a global perspective of the topics projected into two-dimensional principal component analysis (PCA) space with the distance map. The circle size for each topic shows inter-topic differences computed by the Jensen-Shannon divergence. Subfigures (b) to (d), respectively, present the top ten most relevant terms (in English translation) for the three topics obtained from the LDA model. The first topic contains several generic terms such as "company" and "shares", but it also includes terms related

[18] Machine learning techniques have been successfully applied to identify topics in 10-K filings (Dyer, Lang, and Stice-Lawrence, 2017; Brown, Crowley, and Elliott, 2020), analyst reports (Huang, Lehavy, Zang, and Zheng, 2018), and SEC comment letters (Ryans, 2021). However, text from social media tends to be noisier and less structured than professional reports and corporate filings.

[19] Due to the limitation of computation power, we conduct this analysis with 100,000 randomly-drawn posts.

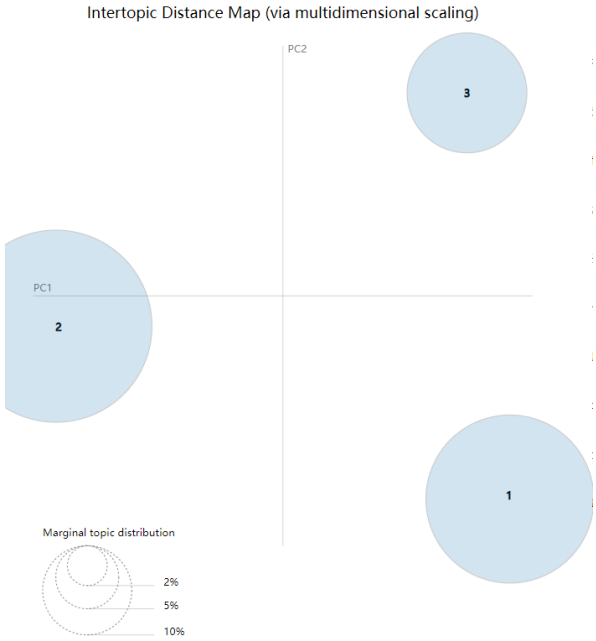
to firm fundamentals such as "performance", "restructuring", and "announcement". Therefore, we label Topic 1 as "Fundamentals". Most terms embedded in the second topic are related to trading activities, including "buy", "sell", and "trend". Thus, we label Topic 2 as "Trading". Unlike the first two topics, it is difficult to identify the intrinsic connection among the frequent terms that appear in the third topic. As a result, Topic 3 is labeled as "Noises".

We also conduct the topical analysis on nonlocal posts. Figure IA.2 in the Internet Appendix presents the results. We do not find systematic differences between posts from local investors and those from nonlocals. In each topic, nonlocal posts share some frequent terms with local posts. This could be due to multiple reasons. For example, investors who write about the equity market are likely to choose similar words, regardless of their locations. It is also possible for investors to repeat others' opinions in their own posts. In addition, it is highly plausible that nonlocal investors will continue the discussion on the same topic when they reply to messages posted by locals.

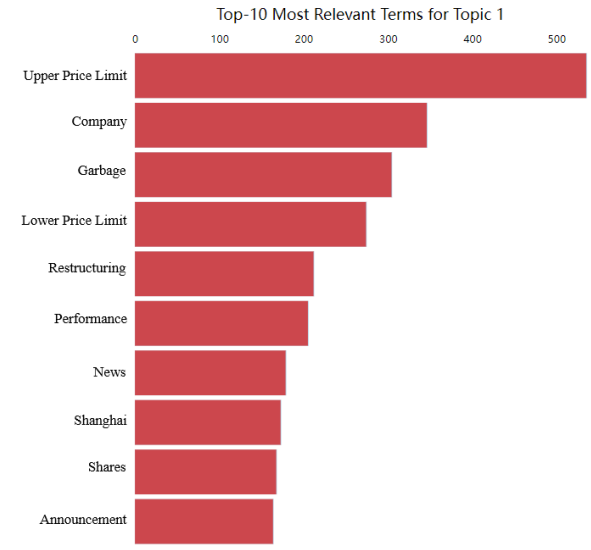
We now investigate which topics discussed in stock message boards predict future stock returns. Specifically, we construct our abnormal relative posting measure based on messages that belong to one of the three topics extracted from local posts, respectively. Posts are classified into a particular topic if they contain three or more of the top ten most frequent terms in that topic. Table 7 reports the results from Fama and MacBeth (1973) regressions of stock returns on these ARP measures. Columns (1) to (3) present results for the ARP measure based on messages with the topic of "Fundamentals", "Trading", and "Noises", respectively. We find that the ARP measure based on messages with the "Fundamentals" topic has strong predictive power for future stock returns, with a coefficient estimate of 1.18 ($t = 5.96$). The coefficient on ARP in Column (2) is also positive (0.68) and statistically significant ($t = 3.08$), suggesting that messages with the "Trading" topic are informative for future stock returns. However, the ARP measure based on posts with the "Noises" topic does not predict stock returns in the next week.

[Insert Table 7 Here]

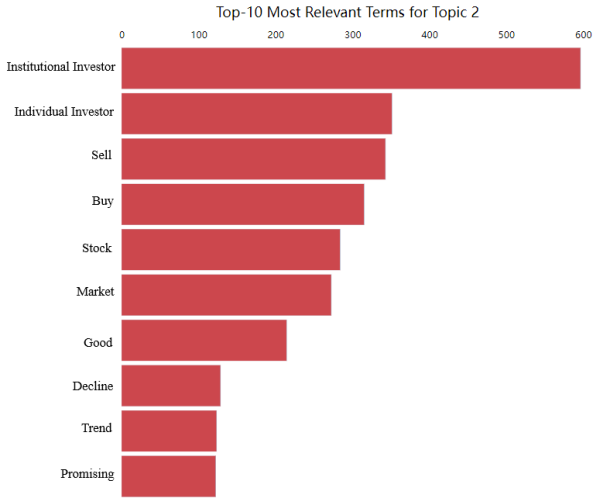
Furthermore, we investigate the return predictability of posts that are related to insider knowledge, which we label as the "Insider" topic. We identify these messages using the following keywords: internal, inside, board, employee, and family. These posts are not identified by our LDA model as a separate topic potentially because investors may avoid insider-related terms in their discussions to reduce the risk of litigation. That is, even though these keywords may appear occasionally,



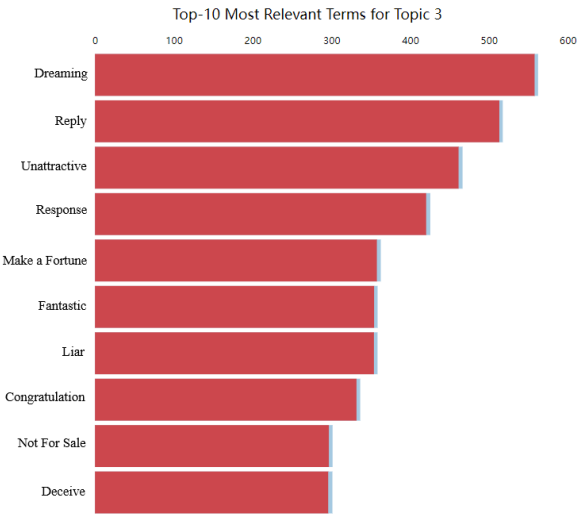
(a) Intertopic Distance Map



(b) Top Terms in Topic 1



(c) Top Terms in Topic 2



(d) Top Terms in Topic 3

Figure 3

Topical Analysis of Local Posts

This figure illustrates the results from the topical analysis of local posts using the Latent Dirichlet Allocation (LDA) model. Subfigure (a) plots a global perspective of the topics projected into two-dimensional principal component analysis (PCA) space with the distance map. The circle size for each topic shows inter-topic differences computed by the Jensen-Shannon divergence. Subfigures (b) to (d) demonstrate the top ten most relevant terms for the topics 1 to 3, respectively. The length of the blue bar indicates the overall term frequency, and the length of the red bar indicates the estimated term frequency within a given topic.

litigation concerns could prevent posters from explicitly discussing their privileged knowledge about the firm. As a result, the number of postings containing insider-related terms can be noisy and therefore cannot predict stock returns. Consistent with the litigation story, the ARP measure based on messages with the "Insider" topic is not predictive of stock returns (Column (4)). Overall, results from our content analysis suggest that the return predictability of ARP is mainly concentrated in messages that discuss firm fundamentals and trading activities.[20]

We also provide additional support to the view that our ARP measure captures a local information advantage by examining posting activities ahead of earnings announcements, which are important informational corporate events. If local investors are better informed about firms' earnings, then we would expect the information asymmetry between local and nonlocal investors to be more pronounced when firms are about to announce earnings news. Accordingly, to the extent local investors are inclined to post during this information-sensitive window, we expect abnormal relative postings to spike ahead of earnings announcements. However, constraints such as litigation risk may temper local investors' willingness to post private information during that window. Indeed, our findings support this conjecture. We find suggestive evidence that locals tend to post more messages before earnings announcements than nonlocals, although the magnitude of the increase is not large (Figure IA.3 in the Internet Appendix). These local posts also receive slightly more clicks and replies before earnings announcements (Figure IA.4 in the Internet Appendix). In addition, there is an upward trend in the percentage of local posts containing "quarterly report" or "annual report" before earnings announcements (Figure IA.5 in the Internet Appendix). The evidence indicates that local investors are posting information related to firms' earnings announcements before the announcement dates, consistent with the idea that messages about firm fundamentals are informative.

5 Postings and Analyst Forecasts

Financial analysts are important information intermediaries in equity markets (e.g., Lang, Lins, and Miller, 2004). If abnormal local postings are informative about firms' future fundamentals, one may expect that abnormal relative postings would lead revisions in analysts' earnings forecasts and recommendations. To test this conjecture, we estimate regressions with the following specification:

[20] Additional analysis confirms that local investors appear to be better informed about firms' fundamentals, and especially about their investment activities. Specifically, we find that the positive ARP-return relation is statistically significant when the messages contain "profit", "investment", "industry", "product", or "acquisition" (Appendix Table IA.4 in the Internet Appendix). However, the return predictability does not exist for messages containing "board", "customer", or "financing".

$$Revision_{i,t+1} = \alpha + \beta ARP_{i,t} + \delta \mathbf{X_{i,t}} + Year + Firm + \epsilon_{i,t+1} \quad (4)$$

where $Revision_{i,t+1}$ represents revisions in analysts' forecasts on earnings per share (EPS) or revisions in analysts' recommendations for firm i in week $t+1$; $ARP_{i,t}$ is abnormal relative postings of firm i in week t, which is defined as before; $\mathbf{X_{i,t}}$ is a vector of control variables. The control variables include firm characteristics such as firm size, book-to-market ratio, asset growth, return-on-assets, idiosyncratic volatility, Amihud (2002) illiquidity, institutional ownership, net purchase ratio of insiders, abnormal local media coverage. We also incorporate city- and industry-level control variables, including population density of the firm's headquarters city, logarithm of GDP per capita in the firm's headquarters city, and share of industry employees. In addition, we control for analyst coverage (Log(Analysts)) following prior studies on analyst forecast revisions. Analyst coverage of a firm is defined as the logarithm of one plus the number of analysts covering the firm in a given week. We also include firm and year fixed effects. Standard errors are clustered at the firm-year level.

[Insert Table 8 Here]

Table 8 reports the results. In Columns (1) and (2), the dependent variables are the percentage of analysts who revise their forecasts of earnings per share (EPS) upward and downward, respectively. The results show that a higher ARP is associated with a higher (lower) portion of analysts who revise their earnings forecasts upward (downward), consistent with our arguments that local investors have access to superior information and that postings mainly reflect positive information.

Columns (3) and (4) consider analysts' recommendation revisions, defined as the percentage of analysts who upgrade and downgrade their recommendations respectively. There are five categories of recommendations (from the least favorable to the most favorable): Sell, Underperform, Neutral, Outperform, and Buy. An upgrade (downgrade) revision refers to a change towards a more (less) favorable category. In line with the results from earnings forecast revisions, a higher ARP is associated with a higher (lower) portion of analysts who upgrade (downgrade) their recommendations. Overall, the results in Table 8 show that our ARP measure has predictive power for future revisions in analysts' forecasts and recommendations, which corroborates our view that investors' posting behavior is informative about firms' future fundamentals.

6 Conclusion

This paper investigates the information asymmetry between local and nonlocal investors with a comprehensive dataset on Internet stock message boards in China. We construct abnormal relative postings to measure unusual changes in the number of messages posted by locals relative to nonlocals. We document strong evidence that our ARP measure positively predicts short-term stock returns. High-ARP stocks outperform low-ARP stocks by 0.24% per week (equivalent to 12.5% per annum), implying that local investors' information advantage has significant economic value. The return predictability lasts for four weeks, and there are no reversals in the longer term. Moreover, a firm's ARP measure positively predicts stock returns of co-headquartered firms, suggesting that local investors' information advantage contains a location-specific component. The return predictability of our ARP measure cannot be absorbed by well-known determinants of expected returns. Furthermore, our results are more pronounced among firms with a more opaque information environment and higher limits to arbitrage, which supports our view that the return predictability of ARP stems from geography-based information asymmetry. Topical analysis suggests that the return predictability is primarily driven by messages about firm fundamentals and trading activities. We also find that abnormal relative postings positively predict analysts' revisions in their earnings forecasts.

Our paper examines the value of local information advantage from a novel angle and provides evidence suggesting that local investors' private information is revealed via social media. We hope our results can help practitioners make better investment decisions in their nonlocal holdings and inspire researchers to examine other implications of investors' interactions on social media.

References

Ajinkya, Bipin, Sanjeev Bhojraj, and Partha Sengupta, 2005, The association between outside directors, institutional investors and the properties of management earnings forecasts, *Journal of Accounting Research* 43, 343–376.

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

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

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

Baik, Bok, Qing Cao, Sunhwa Choi, and Jin-Mo Kim, 2016, Local Twitter activity and stock returns, Working paper, Seoul National University.

Baik, Bok, Jun-Koo Kang, and Jin-Mo Kim, 2010, Local institutional investors, information asymmetries, and equity returns, *Journal of Financial Economics* 97, 81–106.

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, *Review of Financial Studies* 21, 785–818.

Bartov, Eli, Lucile Faurel, and Partha S Mohanram, 2018, Can twitter help predict firm-level earnings and stock returns?, *Accounting Review* 93, 25–57.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri, 2020, On the rise of FinTechs: Credit scoring using digital footprints, *Review of Financial Studies* 33, 2845–2897.

Berger, Jonah, 2014, Word of mouth and interpersonal communication: A review and directions for future research, *Journal of Consumer Psychology* 24, 586–607.

Bernile, Gennaro, Alok Kumar, and Johan Sulaeman, 2015, Home away from home: Geography of information and local investors, *Review of Financial Studies* 28, 2009–2049.

Bernile, Gennaro, Alok Kumar, Johan Sulaeman, and Qin Wang, 2019, Has local informational advantage disappeared?, *Review of Financial Economics* 37, 38–60.

Bizjak, John, Michael Lemmon, and Ryan Whitby, 2009, Option backdating and board interlocks, *Review of Financial Studies* 22, 4821–4847.

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, 101344.

Blankespoor, Elizabeth, Ed Dehaan, John Wertz, and Christina Zhu, 2019, Why do individual investors disregard accounting information? The roles of information awareness and acquisition costs, *Journal of Accounting Research* 57, 53–84.

Blankespoor, Elizabeth, Gregory S Miller, and Hal D White, 2014, The role of dissemination in market liquidity: Evidence from firms' use of twitter, *Accounting Review* 89, 79–112.

Blei, David M, Andrew Y Ng, and Michael I Jordan, 2003, Latent dirichlet allocation, *Journal of Machine Learning Research* 3, 993–1022.

Boone, Audra L, and Joshua T White, 2015, The effect of institutional ownership on firm transparency and information production, *Journal of Financial Economics* 117, 508–533.

Brown, Nerissa C, Richard M Crowley, and W Brooke Elliott, 2020, What are you saying? using topic to detect financial misreporting, *Journal of Accounting Research* 58, 237–291.

Brown, Nerissa C, Han Stice, and Roger M White, 2015, Mobile communication and local information flow: Evidence from distracted driving laws, *Journal of Accounting Research* 53, 275–329.

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.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chawla, Nitesh, Zhi Da, Jian Xu, and Mao Ye, 2021, Information diffusion on social media: Does it affect trading, return, and liquidity?, Working paper, University of Notre Dame.

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of crowds: The value of stock opinions transmitted through social media, *Review of Financial Studies* 27, 1367–1403.

Chen, Jason V, 2022, Wisdom of crowds and the market's response to earnings news: Evidence from the geographic dispersion of investors, *Journal of Accounting and Economics, forthcoming*.

Chen, Li-Wen, Shane A Johnson, Ji-Chai Lin, and Yu-Jane Liu, 2009, Information, sophistication, and foreign versus domestic investors performance, *Journal of Banking & Finance* 33, 1636–1651.

Chhaochharia, Vidhi, Alok Kumar, and Alexandra Niessen-Ruenzi, 2012, Local investors and corporate governance, *Journal of Accounting and Economics* 54, 42–67.

Chi, Sabrina S, and Devin M Shanthikumar, 2017, Local bias in google search and the market response around earnings announcements, *Accounting Review* 92, 115–143.

Choe, Hyuk, Bong-Chan Kho, and René M Stulz, 2005, Do domestic investors have an edge? The trading experience of foreign investors in Korea, *Review of Financial Studies* 18, 795–829.

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

Cooper, Michael J, Huseyin Gulen, and Michael J Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1651.

Coval, Joshua D, and Tobias J Moskowitz, 2001, The geography of investment: Informed trading and asset prices, *Journal of Political Economy* 109, 811–841.

Crowley, Richard M, Wenli Huang, and Hai Lu, 2022, Executive tweets, Working paper, Singapore Management University.

Cziraki, Peter, Jordi Mondria, and Thomas Wu, 2021, Asymmetric attention and stock returns, *Management Science* 67, 48–71.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel, 2003, Persuasion bias, social influence, and unidimensional opinions, *Quarterly Journal of Economics* 118, 909–968.

Drake, Michael S, James R Moon, Brady J Twedt, and James D Warren, 2022, Social media analysts and sell-side analyst research, *Review of Accounting Studies* 1–36.

Drake, Michael S, Jacob R Thornock, and Brady J Twedt, 2017, The Internet as an information intermediary, *Review of Accounting Studies* 22, 543–576.

Dvořák, Tomáš, 2005, Do domestic investors have an information advantage? Evidence from Indonesia, *Journal of Finance* 60, 817–839.

Dyer, Travis, Mark Lang, and Lorien Stice-Lawrence, 2017, The evolution of 10-k textual disclosure: Evidence from latent dirichlet allocation, *Journal of Accounting and Economics* 64, 221–245.

Estellés-Arolas, Enrique, and Fernando González-Ladrón-De-Guevara, 2012, Towards an integrated crowdsourcing definition, *Journal of Information Science* 38, 189–200.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Feng, Lei, and Mark S Seasholes, 2008, Individual investors and gender similarities in an emerging stock market, *Pacific-Basin Finance Journal* 16, 44–60.

Ferreira, Miguel A, Pedro Matos, João Pedro Pereira, and Pedro Pires, 2017, Do locals know better? A comparison of the performance of local and foreign institutional investors, *Journal of Banking & Finance* 82, 151–164.

Garcia, Diego, and Øyvind Norli, 2012, Geographic dispersion and stock returns, *Journal of Financial Economics* 106, 547–565.

Han, Bing, David Hirshleifer, and Johan Walden, 2022, Social transmission bias and investor behavior, *Journal of Financial and Quantitative Analysis* 57, 390–412.

Han, Bing, and Liyan Yang, 2013, Social networks, information acquisition, and asset prices, *Management Science* 59, 1444–1457.

Hao, Qian, Dayong Dong, and Keke Wu, 2019, Online investment forum and the market response around earnings announcement in the Chinese stock markets, *International Journal of Accounting & Information Management* 27, 615–631.

Haugen, Robert A, Nardin L Baker, et al., 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

Heimer, Rawley Z, 2016, Peer pressure: Social interaction and the disposition effect, *Review of Financial Studies* 29, 3177–3209.

Hirshleifer, David, 2020, Presidential address: Social transmission bias in economics and finance, *Journal of Finance* 75, 1779–1831.

Hong, Harrison, Wenxi Jiang, Na Wang, and Bin Zhao, 2014, Trading for status, *Review of Financial Studies* 27, 3171–3212.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, 2004, Social interaction and stock-market participation, *Journal of Finance* 59, 137–163.

Hsu, Charles, James Ang, Di Tang, and Chaopeng Wu, 2021, The role of social media in corporate governance: Social media and corporate governance, *Accounting Review* 96, 1–32.

Huang, Allen H, Reuven Lehavy, Amy Y Zang, and Rong Zheng, 2018, Analyst information discovery and interpretation roles: A topic modeling approach, *Management Science* 64, 2833–2855.

Huang, Jiekun, 2018, The customer knows best: The investment value of consumer opinions, *Journal of Financial Economics* 128, 164–182.

Huang, Kelly, Meng Li, and Stanimir Markov, 2020, What do employees know? Evidence from a social media platform, *Accounting Review* 95, 199–226.

Huang, Shiyang, Tse-Chun Lin, and Hong Xiang, 2021, Psychological barrier and cross-firm return predictability, *Journal of Financial Economics* 142, 338–356.

Huang, Yuqin, Huiyan Qiu, and Zhiguo Wu, 2016, Local bias in investor attention: Evidence from China's internet stock message boards, *Journal of Empirical Finance* 38, 338–354.

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

Jame, Russell, Rick Johnston, Stanimir Markov, and Michael C Wolfe, 2016, The value of crowd-sourced earnings forecasts, *Journal of Accounting Research* 54, 1077–1110.

Jiang, Lei, Jinyu Liu, and Baozhong Yang, 2019, Communication and comovement: Evidence from online stock forums, *Financial Management* 48, 805–847.

Johnston, Rick, Tong Kang, and Michael Wolfe, 2013, Crowdsourcing forecasts: Competition for sell-side analysts?, *Journal of Accounting Research* 493–522.

Kalev, Petko S, Anh H Nguyen, and Natalie Y Oh, 2008, Foreign versus local investors: Who knows more? Who makes more?, *Journal of Banking & Finance* 32, 2376–2389.

Lakonishok, Josef, and Inmoo Lee, 2001, Are insider trades informative?, *Review of Financial Studies* 14, 79–111.

Lang, Mark H, Karl V Lins, and Darius P Miller, 2004, Concentrated control, analyst following, and valuation: do analysts matter most when investors are protected least?, *Journal of Accounting Research* 42, 589–623.

Lang, Mark H, and Russell J Lundholm, 1996, Corporate disclosure policy and analyst behavior, *Accounting Review* 467–492.

Lerman, Alina, 2020, Individual investors' attention to accounting information: Evidence from online financial communities, *Contemporary Accounting Research* 37, 2020–2057.

Lin, Tse-Chun, and Vesa Pursiainen, 2021, The round number heuristic and entrepreneur crowdfunding performance, *Journal of Corporate Finance* 68, 101894.

Lin, Tse-Chun, and Vesa Pursiainen, 2022, Gender differences in reward-based crowdfunding, *Journal of Financial Intermediation* 101001.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-fefinite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 703–708.

Omer, Thomas C, Darren T Roulstone, Nathan Y Sharp, and Brady J Twedt, 2012, The people in your neighborhood: Evidence of local information leakage around corporate headquarters when insiders trade, Working paper, Texas A&M University.

Ozsoylev, Han N, and Johan Walden, 2011, Asset pricing in large information networks, *Journal of Economic Theory* 146, 2252–2280.

Parsons, Christopher A, Riccardo Sabbatucci, and Sheridan Titman, 2020, Geographic lead-lag effects, *Review of Financial Studies* 33, 4721–4770.

Pirinsky, Christo, and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, *Journal of Finance* 61, 1991–2015.

Ryans, James P, 2021, Textual classification of sec comment letters, *Review of Accounting Studies* 26, 37–80.

Shao, Ran, and Na Wang, 2021, Trust and local bias of individual investors, *Journal of Banking & Finance* 133, 106273.

Shive, Sophie, 2012, Local investors, price discovery, and market efficiency, *Journal of Financial Economics* 104, 145–161.

Sicherman, Nachum, George Loewenstein, Duane J Seppi, and Stephen P Utkus, 2015, Financial attention, *Review of Financial Studies* 29, 863–897.

Simon, David, and Rawley Heimer, 2015, Facebook finance: How social interaction propagates active investing, Working paper, Boston College.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.

Sulaeman, Johan, 2014, Do local investors know more? Evidence from mutual fund location and investments, *Quarterly Journal of Finance* 4, 1450010.

Tang, Vicki Wei, 2018, Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on twitter, *Journal of Accounting Research* 56, 989–1034.

Teoh, Siew Hong, 2018, The promise and challenges of new datasets for accounting research, *Accounting, Organizations and Society* 68, 109–117.

Walden, Johan, 2019, Trading, profits, and volatility in a dynamic information network model, *Review of Economic Studies* 86, 2248–2283.

Table A.1
Variable Definitions

Variable	Definition
Posting Variables	
RP	Relative postings, defined as the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters
ARP	Abnormal relative postings, defined as relative postings for a firm in one week minus the median value of its relative postings in the previous ten weeks
Other Variables	
AG	Asset growth, defined as the annual growth rate of total assets
ALMedia	Abnormal local media coverage, defined as local media coverage on a firm in a given week (i.e., the logarithm of one plus the number of news reports issued by news agencies located in the firm's headquarters city within a week) minus the median value of its local media coverage in the previous ten weeks
BM	Book-to-market ratio, defined as the book value of equity divided by market value of equity
EmpShare	Share of industry employees, defined as the total number of employees in an industry in a given city divided by the total number of employees in the city
ILLIQ	Amihud (2002) illiquidity measure, defined as the weekly average of the ratio of absolute daily price change to daily trading volume
IO	Institutional ownership, defined as percentage of shares outstanding owned by institutional investors
IVOL	Idiosyncratic volatility, defined as the standard deviation of residuals from the Carhart (1997) four-factor model with daily stock returns in the past three months, where at least fifteen observations are required in the calculation
Log(Analysts)	Analyst coverage, defined as logarithm of one plus the number of analysts covering the firm in a given week
Log(GDP)	Logarithm of annual GDP per capita (RMB) of a city
NPR	Net purchase ratio, defined as the number of purchases minus the number of sales divided by the total number of transactions by managers and large shareholders of a firm in a given week
PopDensity	Population density of the firm's headquarters city, defined as total population (in 10 thousand) in a city divided by the total area (in squared KM) of the city
$\text{Ret}_{t-4:t-1}$	Cumulative return from week $t-4$ to week $t-1$
$\text{Ret}_{t-52:t-5}$	Cumulative return from week $t-52$ to week $t-5$
ROA	Return on assets, defined as net income divided by total assets
Size	Firm size, defined as the logarithm of market capitalization

Table 1

Example Messages from Local Posters

This table presents examples of messages from local posters. Local posters for a firm refer to posters that are located in the headquarters city of the firm. The table includes the ticker, name, and headquarters city of the company associated with the message. The table also shows the IP address of the poster and the posting time of the message. The last column contains the English translation of the message content.

No.	Ticker	Company Name	Company Location	Poster IP	Date and Time	Content (English Translation)
1	000680	SHANTUI CONSTRUCTION MACHINERY CO., LTD.	Jining, Shandong	123.132.47.*	2008/10/13 23:01:34	I am in Jining City, the headquarters city of Shantui. Shantui's orders have suddenly decreased since May this year. A few days ago, a rumor went out that the company was facing large claims for overseas orders. Shantui is tight inside and loose outside, which is known to all locals.
2	000937	JIZHONG ENERGY RESOURCES CO., LTD.	Xingtai, Hebei	222.223.124.*	2011/04/07 19:34:50	Jizhong Energy is holding a party at the Jinniu Hotel to celebrate its good start in the first quarter.
3	002001	ZHEJIANG NHU CO., LTD.	Shaoxing, Zhejiang	124.91.14.*	2009/07/24 22:04:30	The new factory was already going through trial production procedures with the safety supervision department by the end of June.
4	002092	XINJIANG ZHONGTAI CHEMICAL CO., LTD.	Urumqi, Xinjiang	222.82.25.*	2012/09/17 19:33:49	I also learned from employees of Zhongtai that good news would be released. However, I do not know what it exactly is.
5	002481	YANTAI SHUANGTA FOOD CO., LTD.	Yantai, Shandong	219.146.143.*	2012/06/11 11:10:11	The price will reach 50 RMB based on inside information.
6	300041	HUBEI HUITIAN NEW MATERIALS CO., LTD.	Xiangyang, Hubei	27.22.161.*	2012/11/06 22:26:47	For months, employees could not work at full capacity and not get paid. You said it was outstanding. Aren't you a shill?
7	300131	SHENZHEN YITOA INTELLIGENT CONTROL CO., LTD.	Shenzhen, Guangdong	183.14.3.*	2012/08/17 10:26:02	One can talk about it only after checking by himself. I went to the factory myself. Delivery of many orders was delayed from the first half of the year to the third quarter. I believe you did not know it.
8	600252	GUANGXI WUZHOU ZHONGHENG GROUP CO., LTD.	Wuzhou, Guangxi	116.9.142.*	2009/09/15 00:03:01	My friend on the 310th floor, thank you for your praise. As a person from Wuzhou, I do know some information. But I cannot disclose it flagrantly. Excuse me.
9	600639	SHANGHAI JINQIAO EXPORT PROCESSING ZONE DEVELOPMENT CO., LTD.	Shanghai	124.77.101.*	2012/04/07 10:46:28	Here is a message from Shanghai: I am optimistic about Jinqiao.
10	600665	TANDE CO., LTD.	Shanghai	221.133.233.*	2008/08/05 09:58:21	There is an inside story about Tande's revenue.

Table 2
Summary Statistics

This table reports the summary statistics. Panel A and Panel B report descriptive statistics and correlation coefficients of major variables, respectively. Posting variables include relative postings (RP) and abnormal relative postings (ARP). Firm characteristics include firm size (Size), book-to-market ratio (BM), asset growth (AG), return-on-assets (ROA), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), institutional ownership (IO), net purchase ratio of insiders (NPR), abnormal local media coverage (ALMedia), local population density (PopDensity), logarithm of local GDP per capita (Log(GDP)), and share of industry employees (EmpShare). Detailed definitions are provided in Table A.1. Correlations that are significant at the 5% level are in bold-faced type. The sample period is from June 2007 to May 2013.

Panel A: Descriptive Statistics

	Mean	Std	Q1	Median	Q3
Posting Variables					
RP	–0.33	0.13	–0.41	–0.32	–0.24
ARP	0.00	0.08	–0.05	0.00	0.05
Firm Characteristics					
Size	8.02	1.05	7.31	7.92	8.64
BM	0.63	0.56	0.27	0.46	0.80
$Ret_{t-4:t-1}$	0.18	3.98	0.03	2.17	4.40
$Ret_{t-52:t-5}$	0.18	1.62	–0.74	–0.25	0.59
AG	0.22	0.49	0.01	0.11	0.26
ROA	0.04	0.06	0.01	0.04	0.07
IVOL	2.18	1.02	1.53	1.99	2.55
ILLIQ	0.78	1.17	0.18	0.39	0.86
IO	0.34	0.23	0.14	0.32	0.52
NPR	–0.01	0.11	0.00	0.00	0.00
ALMedia	0.02	0.13	0.00	0.00	0.00
PopDensity	0.09	0.06	0.04	0.08	0.10
Log(GDP)	11.13	0.83	10.54	11.19	11.65
EmpShare	0.06	0.09	0.01	0.03	0.07

Panel B: Correlation Coefficients

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Ret_{t+1}	**1.00**														
2 ARP	**0.02**														
3 Size	**–0.04**	0.00													
4 BM	**0.03**	0.00	**–0.15**												
5 $Ret_{t-4:t-1}$	0.00	**–0.07**	**0.01**	**–0.09**											
6 $Ret_{t-52:t-5}$	**–0.02**	**0.01**	**0.13**	**–0.25**	**–0.01**										
7 AG	**–0.02**	0.00	**0.06**	**0.12**	**–0.04**	**0.02**									
8 ROA	**–0.02**	0.00	**0.23**	**0.04**	**–0.04**	**–0.04**	**0.19**								
9 IVOL	**–0.01**	**–0.01**	**–0.13**	**–0.15**	**–0.02**	**0.19**	**0.08**	**–0.02**							
10 ILLIQ	**0.04**	**0.03**	**–0.47**	**0.15**	**–0.22**	**–0.16**	**–0.03**	**–0.11**	**0.02**						
11 IO	**–0.01**	0.00	**0.53**	**–0.17**	**–0.03**	**–0.08**	**0.02**	**0.16**	**–0.16**	**–0.13**					
12 NPR	**0.01**	**0.01**	**0.02**	**0.04**	**–0.04**	**–0.02**	**0.01**	**0.02**	**–0.04**	**0.02**	**0.01**				
13 ALMedia	**0.00**	**0.00**	**0.06**	**0.02**	**0.01**	0.00	**0.01**	**0.01**	**0.00**	**0.00**	**0.03**	**–0.01**			
14 PopDensity	0.00	0.00	**0.02**	**0.05**	**–0.01**	**–0.02**	**–0.02**	**0.03**	**–0.04**	**0.03**	**0.06**	0.00	**0.14**		
15 Log(GDP)	0.00	**0.01**	**0.01**	**0.09**	**–0.02**	**–0.11**	0.00	**0.06**	**–0.10**	**0.02**	**0.11**	**0.01**	**0.18**	**0.50**	
16 EmpShare	0.00	0.00	**–0.09**	**0.02**	**–0.01**	**–0.04**	**–0.02**	**0.04**	**0.01**	**0.04**	**–0.01**	**–0.01**	**0.12**	**0.21**	**0.51**

Table 3

Abnormal Relative Postings and Stock Returns: Univariate Sorts

This table reports weekly excess returns and abnormal returns (in percentage) of quintile portfolios formed based on one-week-lagged abnormal relative postings (ARP), which is constructed as relative postings for a firm in one week minus the median value of its relative postings in the previous ten weeks where the relative posting measure is the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. Each week, firms are sorted into quintile portfolios based on their ARP measure, and portfolios are rebalanced every week. Panels A reports average weekly equal-weighted excess returns, value-weighted excess returns, abnormal returns from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model for quintile portfolios and the high-minus-low ARP portfolios. Panel B reports the average portfolio characteristics that are defined in Table A.1. The sample period is from June 2007 to May 2013. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A

	Low	2	3	4	High	High – Low	t statistic
EW	–0.19	–0.16	–0.13	–0.07	0.05	0.24***	(5.72)
VW	–0.30	–0.28	–0.27	–0.23	–0.11	0.20***	(2.60)
Carhart four-factor alpha	–0.02	–0.00	0.04	0.09	0.22	0.24***	(5.55)
FF five-factor alpha	–0.00	0.01	0.06	0.12	0.23	0.24***	(5.35)

Panel B

	Low	2	3	4	High	High – Low	t statistic
Size	7.78	7.96	8.02	7.95	7.76	–0.02	(–0.78)
BM	0.65	0.66	0.67	0.67	0.67	0.02	(1.10)
AG	0.22	0.22	0.23	0.23	0.22	0.00	(–0.22)
ROA	0.04	0.04	0.04	0.04	0.04	0.00	(0.57)
IVOL	2.37	2.34	2.34	2.33	2.34	–0.03	(–0.87)
ILLIQ	0.86	0.79	0.78	0.83	0.92	0.06	(0.99)
IO	0.29	0.29	0.30	0.30	0.28	0.00	(–0.73)
NPR	–0.01	–0.01	–0.01	–0.01	–0.00	0.00	(1.22)
ALMedia	0.02	0.03	0.03	0.03	0.02	0.00	(0.67)
PopDensity	0.08	0.09	0.09	0.09	0.08	0.00	(–0.60)
Log(GDP)	11.03	11.11	11.15	11.12	11.03	0.00	(0.00)
EmpShare	0.06	0.06	0.06	0.06	0.06	0.00	(–0.27)

Table 4

Abnormal Relative Postings and Stock Returns: Cross-firm Return Predictability

This table reports weekly excess returns and abnormal returns (in percentage) of quintile portfolios formed based on peer firms' one-week-lagged abnormal relative posting (ARP) measure. Peer firms are defined as firms that share the same headquarters city with a focal firm. The ARP measure is constructed as relative postings for a firm in one week minus the median value of its relative postings in the previous ten weeks. The relative posting measure is the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. Each week, firms are sorted into quintile portfolios based on the median value of peer firms' ARP measure, and portfolios are rebalanced every week. Firms with less than fifteen peer firms are excluded from the sample. Each panel reports average weekly equal-weighted excess returns, value-weighted excess returns, abnormal returns from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model for quintile portfolios and the high-minus-low ARP portfolios. The sample period is from June 2007 to May 2013. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Low	2	3	4	High	High – Low	t statistic
EW	–0.22	–0.11	–0.05	–0.09	–0.08	0.14**	(2.38)
VW	–0.43	–0.32	–0.18	–0.19	–0.24	0.19*	(1.92)
Carhart four-factor alpha	–0.08	0.06	0.11	0.05	0.10	0.19***	(3.09)
FF five-factor alpha	–0.09	0.07	0.12	0.07	0.12	0.21***	(3.31)

Table 5

Abnormal Relative Postings and Stock Returns: Fama-MacBeth Regression

This table reports results from Fama and MacBeth (1973) cross-sectional regressions of weekly stock returns (in percentage) on lagged abnormal relative postings (ARP) as well as control variables. The ARP measure is constructed as relative postings for a firm in week t minus the median value of its relative postings in the previous ten weeks where relative postings are the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. In Columns (1) to (3), the dependent variable is stock returns in week $t+1$. In Columns (4) to (8), the dependent variables are stock returns in weeks $t+2$, $t+4$, $t+6$, $t+8$, and $t+12$, respectively. The control variables include firm size (Size), book-to-market ratio (BM), reversal ($\text{Ret}_{t-4:t-1}$), momentum ($\text{Ret}_{t-52:t-5}$), asset growth (AG), return-on-assets (ROA), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), institutional ownership (IO), net purchase ratio of insiders (NPR), abnormal local media coverage (ALMedia), local population density (PopDensity), logarithm of local GDP per capita (Log(GDP)), and share of industry employees (EmpShare). Detailed definitions are provided in Table A.1. The sample period is from June 2007 to May 2013. t-statistics based on Newey-West adjusted standard errors with three lags are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Ret_{t+1}			Ret_{t+2}	Ret_{t+4}	Ret_{t+6}	Ret_{t+8}	Ret_{t+12}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ARP	0.91***	0.86***	0.81***	0.39***	0.38**	0.06	0.13	0.05
	(5.51)	(5.61)	(5.39)	(2.88)	(2.57)	(0.36)	(0.94)	(0.32)
Size		–0.14***	–0.05	–0.09*	–0.09*	–0.05	–0.08	–0.06
		(–3.11)	(–1.01)	(–1.72)	(–1.77)	(–1.12)	(–1.65)	(–1.33)
BM		0.06	0.04	–0.02	0.02	–0.00	–0.01	0.00
		(0.92)	(0.75)	(–0.29)	(0.34)	(–0.07)	(–0.14)	(0.04)
$\text{Ret}_{t-4:t-1}$		–0.05***	–0.03***	–0.05***	–0.04***	–0.02**	–0.02**	–0.01
		(–3.99)	(–2.95)	(–4.39)	(–3.87)	(–2.57)	(–2.18)	(–0.89)
$\text{Ret}_{t-52:t-5}$		–0.07***	–0.04	–0.04*	–0.05*	–0.04	–0.06**	–0.06*
		(–2.72)	(–1.49)	(–1.67)	(–1.71)	(–1.53)	(–2.12)	(–1.92)
AG			–0.06	–0.06	–0.11**	–0.15***	–0.16***	–0.14***
			(–1.18)	(–1.51)	(–2.56)	(–3.37)	(–4.13)	(–3.32)
ROA			0.10	–0.06	–0.62	–0.68	–0.40	–0.45
			(0.22)	(–0.13)	(–1.43)	(–1.46)	(–0.91)	(–0.99)
IVOL			–0.13***	–0.08**	–0.03	–0.03	–0.01	–0.03
			(–3.57)	(–2.22)	(–0.99)	(–0.81)	(–0.24)	(–0.85)
ILLIQ			0.38***	0.18***	0.11**	0.16***	0.07**	0.14***
			(8.09)	(4.80)	(2.42)	(3.62)	(1.98)	(3.93)
IO			0.04	0.06	–0.02	0.03	–0.09	–0.14
			(0.29)	(0.46)	(–0.18)	(0.27)	(–0.63)	(–0.96)
NPR			0.35**	0.04	0.10	0.30	–0.08	0.16
			(2.03)	(0.21)	(0.59)	(1.63)	(–0.55)	(0.84)
ALMedia			–0.07	–0.07	0.07	–0.13	–0.10	0.01
			(–0.70)	(-0.70)	(0.55)	(–1.39)	(–1.03)	(0.07)
PopDensity			0.18	0.18	0.26	0.09	–0.05	0.07
			(0.48)	(0.48)	(0.75)	(0.27)	(–0.14)	(0.17)
Log(GDP)			–0.02	–0.02	–0.04	–0.01	–0.02	–0.03
			(–0.54)	(–0.54)	(–1.26)	(–0.41)	(–0.59)	(–1.12)
EmpShare			–0.29	–0.11	0.10	0.07	0.04	0.14
			(–1.11)	(–0.41)	(0.36)	(0.25)	(0.13)	(0.42)
Intercept	–0.06	0.91	0.44	0.82	1.06	0.47	0.86	0.80
	(–0.18)	(1.63)	(0.67)	(1.16)	(1.47)	(0.69)	(1.23)	(1.17)
Obs	303,361	303,361	303,361	293,425	279,472	275,838	272,375	265,509
Adj. R^2	0.05%	3.60%	6.39%	5.90%	5.67%	5.27%	5.07%	5.00%

Table 6

Portfolio Returns: Information Environment and Limits to Arbitrage

This table reports weekly excess returns and abnormal returns (in percentage) of quintile portfolios formed based on abnormal relative postings (ARP) and market capitalization (stock price times the number of shares outstanding), analyst coverage (number of analysts following the firm), institutional ownership (percentage of shares owned by institutions), or idiosyncratic volatility (the standard deviation of residuals from the Carhart (1997) four-factor model). The ARP measure is constructed as relative postings for a firm in week t minus the median value of its relative postings in the previous ten weeks where relative postings are the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. Each week, we split our sample into tercile groups based on market capitalization, analyst coverage, institutional ownership, and idiosyncratic volatility. Then we sort firms within each group into quintile portfolios based on their ARP measure, and portfolios are rebalanced every week. The average weekly equal-weighted excess returns, value-weighted excess returns, abnormal returns from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model for the high-minus-low ARP portfolios are presented in each panel. The second last row of each panel reports differences in the return gaps between portfolios with high and low levels of market capitalization, analyst coverage, institutional ownership, or idiosyncratic volatility. The sample period is from June 2007 to May 2013. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: Market Capitalization				
	EW	VW	Carhart four-factor alpha	FF five-factor alpha
Small	0.42***	0.43***	0.45***	0.44***
	(5.83)	(5.66)	(5.94)	(5.72)
Large	0.13**	0.12	0.12*	0.10
	(1.99)	(1.31)	(1.79)	(1.39)
Small – Large	0.29***	0.30**	0.33***	0.34***
	(2.99)	(2.39)	(3.23)	(3.35)
Panel B: Analyst Coverage				
	EW	VW	Carhart four-factor alpha	FF five-factor alpha
Low	0.29***	0.23**	0.30***	0.27***
	(4.22)	(2.31)	(4.22)	(3.64)
High	0.07	0.21*	0.06	0.11
	(0.89)	(1.66)	(0.71)	(1.24)
Low – High	0.22**	0.02	0.24**	0.16*
	(1.97)	(0.35)	(2.40)	(1.65)
Panel C: Institutional Ownership				
	EW	VW	Carhart four-factor alpha	FF five-factor alpha
Low	0.36***	0.26***	0.37***	0.38***
	(4.75)	(2.74)	(4.79)	(4.69)
High	0.15**	0.21*	0.14*	0.14*
	(2.14)	(1.93)	(1.93)	(1.80)
Low – High	0.20*	0.05	0.23**	0.24**
	(1.94)	(0.29)	(2.09)	(2.14)
Panel D: Idiosyncratic Volatility				
	EW	VW	Carhart four-factor alpha	FF five-factor alpha
Low	0.08	0.04	0.09	0.08
	(1.27)	(0.36)	(1.35)	(1.30)
High	0.26***	0.30**	0.26***	0.25***
	(3.44)	(2.09)	(3.28)	(3.05)
High – Low	0.19**	0.26*	0.17**	0.16**
	(2.33)	(1.77)	(2.15)	(2.10)

Table 7

Abnormal Relative Postings and Stock Returns: Messages with Different Topics

This table reports results from Fama and MacBeth (1973) cross-sectional regressions of weekly stock returns (in percentage) on lagged abnormal relative postings (ARP), which is defined as before except that it is constructed based on messages with different topics (listed in the first row). Columns (1) to (3) present results for topics extracted from the LDA model. In each column, the ARP measure is constructed based on messages containing three or more of the top ten most frequent terms in a given topic (shown in Figure 3). Column (4) presents results for the ARP measure constructed based on messages that are related to insider knowledge. Posts are classified into this topic if they contain "internal", "inside", "board", "employee", or "family". The control variables include firm size (Size), book-to-market ratio (BM), reversal ($\text{Ret}_{t-4:t-1}$), momentum ($\text{Ret}_{t-52:t-5}$), asset growth (AG), return-on-assets (ROA), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), institutional ownership (IO), net purchase ratio of insiders (NPR), abnormal local media coverage (ALMedia), local population density (PopDensity), the logarithm of local GDP per capita (Log(GDP)), and share of industry employees (EmpShare). Detailed definitions are provided in Table A.1. The sample period is from June 2007 to May 2013. *t*-statistics based on Newey-West adjusted standard errors with three lags are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Topic	Fundamentals	Trading	Noises	Insider
	(1)	(2)	(3)	(4)
ARP	1.18***	0.68***	0.80	0.33
	(5.96)	(3.08)	(0.19)	(1.47)
Size	-0.05	-0.05	-0.05	-0.05
	(-1.05)	(-1.08)	(-0.99)	(-1.03)
BM	0.05	0.05	0.04	0.05
	(0.84)	(0.82)	(0.79)	(0.83)
$\text{Ret}_{t-4:t-1}$	-0.03***	-0.03***	-0.04***	-0.03***
	(-2.84)	(-2.81)	(-3.05)	(-2.95)
$\text{Ret}_{t-52:t-5}$	-0.04	-0.04	-0.04	-0.04
	(-1.60)	(-1.57)	(-1.51)	(-1.56)
AG	-0.06	-0.06	-0.06	-0.06
	(-1.21)	(-1.25)	(-1.16)	(-1.12)
ROA	0.12	0.11	0.12	0.10
	(0.26)	(0.25)	(0.25)	(0.23)
IVOL	-0.13***	-0.13***	-0.13***	-0.13***
	(-3.57)	(-3.53)	(-3.56)	(-3.52)
ILLIQ	0.38***	0.38***	0.38***	0.38***
	(8.17)	(8.16)	(8.23)	(8.15)
IO	0.04	0.04	0.04	0.04
	(0.27)	(0.29)	(0.25)	(0.24)
NPR	0.33*	0.35**	0.36**	0.36**
	(1.91)	(2.03)	(2.09)	(2.12)
ALMedia	-0.06	-0.06	-0.07	-0.07
	(-0.54)	(-0.57)	(-0.69)	(-0.66)
PopDensity	0.15	0.18	0.16	0.18
	(0.42)	(0.49)	(0.44)	(0.49)
Log(GDP)	-0.01	-0.01	-0.02	-0.01
	(-0.43)	(-0.44)	(-0.51)	(-0.45)
EmpShare	-0.30	-0.31	-0.31	-0.31
	(-1.15)	(-1.19)	(-1.20)	(-1.19)
Intercept	0.42	0.44	0.43	0.42
	(0.65)	(0.67)	(0.65)	(0.64)
Obs	303,361	303,361	303,361	303,361
Adj. R^2	6.45%	6.45%	6.39%	6.39%

Table 8

Abnormal Relative Postings and Analyst Forecast Revisions

This table reports the relation between abnormal relative postings and future analyst forecast revisions. The first two columns report results for earnings forecast revisions, while the last two columns present results for recommendation revisions. In Columns (1) and (2), the dependent variable is the percentage of analysts who revise their forecasts of earnings per share (EPS) upward and downward, respectively. In Columns (3) and (4), the dependent variable is the percentage of analysts who upgrade and downgrade their recommendations, respectively. There are five categories of recommendations (from the least favorable to the most favorable): Sell, Underperform, Neutral, Outperform, and Buy. An upgrade (downgrade) revision refers to a change towards a more (less) favorable category. The independent variable is a firm's abnormal relative postings in a given week, which is constructed as relative postings for the firm in the current week minus the median value of its relative postings in the previous ten weeks. The relative posting measure is the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. The control variables include firm size (Size), book-to-market ratio (BM), asset growth (AG), return-on-assets (ROA), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), institutional ownership (IO), net purchase ratio of insiders (NPR), abnormal local media coverage (ALMedia), population density of the firm's headquarters city (PopDensity), logarithm of GDP per capita (Log(GDP)), share of industry employees (EmpShare), and analyst coverage (Log(Analysts)). Detailed definitions are provided in Table A.1. Abnormal relative postings and control variables are one-week lagged. The sample period is from June 2007 to May 2013. Firm and year fixed effects are included in the regressions. *t*-statistics based on standard errors clustered at the firm-year level are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	EPS Forecast Revision		Recommendation Revision	
	Upward (1)	Downward (2)	Upgrade (3)	Downgrade (4)
ARP	0.02**	–0.08***	0.03**	–0.04**
	(2.51)	(–2.68)	(2.53)	(–2.56)
Size	0.01***	0.01	0.00	–0.01
	(3.20)	(1.08)	(0.76)	(–1.26)
BM	–0.00	–0.00	–0.01*	0.02***
	(–0.49)	(–0.08)	(–1.95)	(3.41)
AG	0.00**	0.01	0.00	0.00
	(2.03)	(0.91)	(0.86)	(0.18)
ROA	0.05**	0.16**	0.00	0.13***
	(2.57)	(2.44)	(0.15)	(4.01)
IVOL	–0.00***	–0.01***	–0.00	0.00
	(–3.04)	(–3.69)	(–1.37)	(0.30)
ILLIQ	–0.00***	–0.02***	–0.00**	–0.01***
	(–2.95)	(–5.11)	(–2.14)	(–3.47)
IO	–0.02***	0.01	0.02**	0.05***
	(–2.94)	(0.45)	(2.36)	(4.64)
NPR	0.01	0.01	0.01	0.00
	(1.63)	(0.62)	(1.24)	(0.50)
ALMedia	–0.00	–0.00	–0.01	–0.00
	(–0.40)	(–0.03)	(–1.03)	(–0.20)
PopDensity	–0.56***	–1.07	–0.59*	–0.64
	(–2.59)	(–1.33)	(–1.77)	(–1.46)
Log(GDP)	–0.03**	–0.09**	–0.01	-0.02
	(–1.99)	(–2.22)	(–0.40)	(–0.94)
EmpShare	–0.15	–0.06	0.04	–0.15
	(–1.58)	(–0.20)	(0.41)	(–1.11)
Log(Analysts)	0.01***	0.04***	0.00	0.01***
	(4.45)	(8.30)	(0.58)	(3.11)
Intercept	0.26*	0.92*	0.14	0.36
	(1.69)	(1.95)	(0.66)	(1.27)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs	229,182	229,182	277,779	277,779
Adj. R^2	0.50%	1.65%	0.33%	0.31%

Internet Appendix:

Local Information Advantage and Stock Returns: Evidence from Social Media

Not for Publication

This Internet Appendix consists of additional results. Table IA.1 presents additional examples of messages from local posters. Table IA.2 reports the relation between abnormal relative postings and alternative return measures, including Daniel, Grinblatt, Titman, and Wermers (1997) characteristic-based benchmark-adjusted returns and industry-adjusted returns. Table IA.3 reports Fama and MacBeth (1973) regression results for subsamples with positive and negative abnormal relative postings separately. Table IA.4 presents the results from Fama and MacBeth (1973) regressions using ARP measures constructed based on stock messages containing certain words. Table IA.5 reports the relation between the standardized abnormal relative postings and stock returns. Table IA.6 presents the return predictability of the distance-based abnormal relative posting measure, the province-level abnormal relative posting measure, and the length-weighted abnormal relative posting measure. Table IA.7 reports the relation between stock returns and abnormal relative postings calculated based on alternative definitions for the normal level of relative postings.

Figure IA.1 plots profits of the ARP-sorted long-short strategy for each industry. Figure IA.2 illustrates the results from the topical analysis of nonlocal posts using the Latent Dirichlet Allocation (LDA) model. Figure IA.3 plots the percentage of local posts, the percentage of clicks on local posts, and the percentage of replies to local posts around firms' earnings announcements. Figure IA.4 plots the percentage of local posts among messages with a high number of clicks or with a high number of replies around firms' earnings announcements. Figure IA.5 plots the percentage of local posts that contain the terms "quarterly report" or "annual report" around firms' earnings announcements.

Table IA.1
Additional Examples of Messages from Local Posters

This table presents additional examples of messages from local posters. Local posters for a firm refer to posters that are located in the headquarters city of the firm. The table includes the ticker, name, and headquarters city of the company associated with the message. The table also shows the IP address of the poster and the posting time of the message. The last column contains the English translation of the message content.

Ticker	Company Name	Company Location	Poster IP	Date and Time	Content (English Translation)
000005	SHENZHEN FOUNTAIN CORP.	Shenzhen, Guangdong	116.25.34.*	2009/04/17 20:24:01	Believe it or not, the chance of closing with a mild increase next Monday is 90%.
000410	SHENYANG MACHINE TOOL CO., LTD.	Shenyang, Liaoning	59.44.39.*	2010/03/04 10:24:34	If you know the situation in Shenyang Machine Tool, you will never buy its stock. Everything is fake. It is comparable to or even worse than Changhong.
000425	XCMG CONSTRUCTION MACHINERY CO., LTD.	Xuzhou, Jiangsu	218.3.210.*	2011/03/24 15:52:22	Hah-hah, there will be more news coming out of this. Hold the stock firmly. Good news will be released in a couple of days.
000488	SHANDONG CHENMING PAPER HOLDINGS LTD.	Weifang, Shandong	113.120.252.*	2010/04/19 20:35:16	What I mentioned was just released today by Voice of The Morning at Shouguang (author's note: Shouguang is a county in Weifang, Shandong) TV. You can search shouguang.com and Voice of The Morning, but there is no video playback. It can only be viewed from local TV. I'm sorry I didn't give you a link.
000536	CPT TECHNOLOGY (GROUP) CO., LTD.	Fuzhou, Fujian	218.85.32.*	2010/04/01 10:36:03	Please go to Mawei (author's note: Mawei is a district of Fuzhou) and check it. What a broken factory. I am from Fuzhou. Hot money was pushing up the price today.
000635	NINGXIA YOUNGLIGHT CHEMICALS CO., LTD.	Shizuishan, Ningxia	119.60.135.*	2010/02/24 08:41:27	Young Light will beat the market recently for sure.
000950	CHONGQING JIANFENG CHEMICAL CO., LTD.	Chongqing	222.180.172.*	2012/02/08 09:10:38	Profits will be prosperous if you buy the stock at a low price recently. No matter whether you believe it or not, I do believe and have bought the shares.
002199	ZHEJIANG EAST CRYSTAL ELECTRONIC CO., LTD.	Jinhua, Zhejiang	122.226.246.*	2010/07/22 12:43:02	I work at 02199, and I know some information about it.
002247	ZHEJIANG DILONG NEW MATERIAL CO., LTD.	Hangzhou, Zhejiang	125.122.11.*	2009/05/07 17:01:10	The target price for Dilong is 25 RMB.
002269	SHANGHAI METERSBONWE FASHION & ACCESSORIES CO., LTD.	Shanghai	222.69.31.*	2009/04/14 22:10:08	Be patient.
300140	XI'AN QIYUAN MECHANICAL AND ELECTRICAL EQUIPMENT CO., LTD.	Xi'an, Shannxi	1.80.224.*	2011/04/07 18:41:27	Hello, Thread-starter. Please do not spread rumors. I myself am an employee of Qiyuan. The company is currently operating well. It is impossible to change the chairman, let alone appoint Zhang Jingtao as the chairman.
600019	BAOSHAN IRON & STEEL CO., LTD.	Shanghai	221.130.187.*	2012/03/09 20:45:24	No hurry. Surprises are coming next week.

Continued on the next page

Ticker	Company Name	Company Location	Poster IP	Date and Time	Content (English Translation)
600037	BEIJING GEHUA CATV NETWORK CO., LTD.	Beijing	219.239.202.*	2010/11/02 14:25:43	Hold it to next week if you believe me.
600062	CHINA RESOURCES DOUBLE-CRANE PHARMACEUTICAL CO., LTD.	Beijing	124.200.52.*	2012/03/26 22:33:26	The unorganized salary system at Double-Crane makes employees fail to perform their own tasks well.
600071	PHENIX OPTICAL CO., LTD.	Shangrao, Jiangxi	113.195.117.*	2011/09/28 15:12:01	I am from Shangrao, Jiangxi Province. The new factory of Phoenix Optics has been built very quickly.
600098	GUANGZHOU DEVELOPMENT GROUP INC.	Guangzhou, Guangdong	113.109.203.*	2009/10/29 17:59:22	The company will benefit from the LPG bus incident in Guangzhou due to its product related to dimethyl ether.
600106	CHONGQING ROAD & BRIDGE CO., LTD.	Chongqing	125.81.254.*	2011/04/01 11:31:10	The good news for 600106 will come out continuously. I will wait and see.
600108	GANSU YASHENG INDUSTRIAL (GROUP) CO., LTD.	Lanzhou, Gansu	115.85.201.*	2010/10/14 13:10:47	The most important message is that local farmers have received subsidies for potatoes. I am not sure why there has been no announcement on such a huge issue as receiving fiscal subsidies. Yasheng received the potato subsidy and, more importantly, the subsidy to breed. No.1 document of central government stipulates that the subsidy is available from both local and central governments.
601899	ZIJIN MINING GROUP CO. LTD.	Longyan, Fujian	220.162.152.*	2010/10/08 08:50:17	The local authorities are stepping up their investigations into false compensation and fraudulent claims! The locals know about it. If you don't believe it, you can ask.
601999	NORTHERN UNITED PUBLISHING & MEDIA (GROUP) CO. LTD.	Shenyang, Liaoning	218.25.42.*	2010/11/19 12:39:51	I am going to buy this stock next week.

Table IA.2

Abnormal Relative Postings and Alternative Stock Return Measures

This table reports weekly excess returns and abnormal returns (in percentage) of quintile portfolios formed based on a one-week-lagged abnormal relative posting (ARP) measure. Panel A reports results using the Daniel, Grinblatt, Titman, and Wermers (1997) characteristic-based benchmarks to calculate excess returns, while Panel B reports results based on returns adjusted by the average returns of industry peers where industries are defined using sectors in the *Guidance for Industry Classification of Listed Companies* released by the CSRC. The ARP measure is constructed as relative postings for a firm in one week minus the median value of its relative postings in the previous ten weeks. The relative posting measure is the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. Each week, firms are sorted into quintile portfolios based on their ARP measure, and portfolios are rebalanced every week. Each panel reports average weekly equal-weighted excess returns, value-weighted excess returns, abnormal returns from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model for quintile portfolios and the high-minus-low ARP portfolios. The sample period is from June 2007 to May 2013. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: DGTW-adjusted Returns

	Low	2	3	4	High	High – Low	t statistic
EW	–0.15	–0.16	–0.11	–0.06	0.05	0.20***	(5.16)
VW	–0.32	–0.31	–0.27	–0.20	-0.16	0.17**	(2.47)
Carhart four-factor alpha	–0.15	–0.17	–0.10	–0.08	0.06	0.21***	(5.15)
FF five-factor alpha	–0.14	–0.17	–0.11	–0.07	0.06	0.20***	(4.73)

Panel B: Industry-adjusted Returns

	Low	2	3	4	High	High – Low	t statistic
EW	–0.20	–0.16	–0.11	–0.07	0.05	0.25***	(6.05)
VW	–0.30	–0.28	–0.26	–0.19	–0.11	0.19**	(2.58)
Carhart four-factor alpha	–0.19	–0.16	–0.10	–0.08	0.06	0.25***	(5.80)
FF five-factor alpha	–0.19	–0.16	–0.10	–0.06	0.06	0.24***	(5.63)

Table IA.3
Abnormal Relative Postings and Stock Returns: Asymmetric Effect

This table reports results from Fama and MacBeth (1973) cross-sectional regressions of weekly stock returns (in percentage) on one-week-lagged abnormal relative posting (ARP) measure as well as control variables. The ARP measure is constructed as relative postings for a firm in week t minus the median value of its relative postings in the previous ten weeks where relative postings are the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. Columns (1) to (3) present results for the subsample with positive ARP measure, while Columns (4) to (6) present results for the subsample with negative ARP. The control variables include firm size (Size), book-to-market ratio (BM), reversal ($\text{Ret}_{t-4:t-1}$), momentum ($\text{Ret}_{t-52:t-5}$), asset growth (AG), return-on-assets (ROA), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), institutional ownership (IO), net purchase ratio of insiders (NPR), abnormal local media coverage (ALMedia), local population density (PopDensity), logarithm of local GDP per capita (Log(GDP)), and share of industry employees (EmpShare). Detailed definitions are provided in Table A.1. The sample period is from June 2007 to May 2013. t-statistics based on Newey-West adjusted standard errors with three lags are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Positive ARP			Negative ARP		
	(1)	(2)	(3)	(4)	(5)	(6)
ARP	1.47***	1.08***	0.91***	0.09	0.40	0.33
	(5.22)	(3.90)	(3.34)	(0.25)	(1.04)	(0.90)
Size		–0.15***	–0.06		–0.13**	–0.03
		(–3.30)	(–1.23)		(–2.55)	(–0.55)
BM		0.04	0.02		0.09	0.07
		(0.59)	(0.39)		(1.27)	(1.11)
$\text{Ret}_{t-4:t-1}$		–0.05***	–0.04**		–0.04***	–0.03**
		(–3.39)	(–2.57)		(–3.53)	(–2.28)
$\text{Ret}_{t-52:t-5}$		–0.06**	–0.03		–0.11***	–0.08**
		(–2.08)	(–1.08)		(–3.25)	(–2.20)
AG			–0.05			–0.10*
			(–0.78)			(–1.72)
ROA			–0.08			0.18
			(–0.16)			(0.37)
IVOL			–0.13***			–0.14***
			(–3.26)			(–3.53)
ILLIQ			0.35***			0.44***
			(5.47)			(6.69)
IO			0.08			0.02
			(0.46)			(0.10)
NPR			0.25			0.55**
			(1.38)			(2.44)
ALMedia			0.02			–0.13
			(0.13)			(–1.00)
PopDensity			0.31			0.10
			(0.76)			(0.23)
Log(GDP)			–0.02			–0.01
			(–0.55)			(–0.26)
EmpShare			–0.41			–0.40
			(–1.25)			(–1.22)
Intercept	–0.09	0.95*	0.55	–0.11	0.76	0.21
	(–0.28)	(1.69)	(0.80)	(–0.35)	(1.28)	(0.27)
Obs	147,611	147,611	147,611	155,750	155,750	155,750
Adj. R^2	0.01%	3.64%	6.53%	0.05%	3.81%	6.77%

Table IA.4

Abnormal Relative Postings and Stock Returns: Messages with Keywords

This table reports results from Fama and MacBeth (1973) cross-sectional regressions of weekly stock returns (in percentage) on lagged abnormal relative postings (ARP) as well as control variables. The ARP measure is constructed as relative postings for a firm in week t minus the median value of its relative postings in the previous ten weeks, where relative postings are the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. In each column, the ARP measure is constructed based on messages containing a certain word (listed in the first row). The dependent variable is stock returns in week $t+1$. The control variables include firm size (Size), book-to-market ratio (BM), reversal ($\text{Ret}_{t-4:t-1}$), momentum ($\text{Ret}_{t-52:t-5}$), asset growth (AG), return-on-assets (ROA), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), institutional ownership (IO), net purchase ratio of insiders (NPR), abnormal local media coverage (ALMedia), local population density (PopDensity), the logarithm of local GDP per capita (Log(GDP)), and share of industry employees (EmpShare). Detailed definitions are provided in Table A.1. The sample period is from June 2007 to May 2013. t-statistics based on Newey-West adjusted standard errors with three lags are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Keyword	Profit	Investment	Industry	Product	Acquisition	Board	Customer	Financing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ARP	1.07***	0.80***	0.63***	0.54**	0.48*	0.46	0.34	0.07
	(5.51)	(4.77)	(2.80)	(2.33)	(1.86)	(1.45)	(0.72)	(0.23)
Size	–0.05	–0.05	–0.05	–0.05	–0.05	–0.04	–0.05	–0.05
	(–1.08)	(–1.08)	(–1.01)	(–0.98)	(–1.03)	(–0.90)	(–1.00)	(–1.00)
BM	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	(0.81)	(0.86)	(0.85)	(0.85)	(0.87)	(0.82)	(0.81)	(0.80)
$\text{Ret}_{t-4:t-1}$	–0.03***	–0.03***	–0.03***	–0.04***	–0.03***	–0.04***	–0.04***	–0.04***
	(–2.87)	(–2.86)	(–2.95)	(–3.02)	(–3.00)	(–3.00)	(–3.03)	(–3.09)
$\text{Ret}_{t-52:t-5}$	–0.04	–0.04	–0.04	–0.04	–0.04	–0.04	–0.04	–0.04
	(–1.57)	(–1.52)	(–1.51)	(–1.49)	(–1.51)	(–1.53)	(–1.55)	(–1.49)
AG	–0.06	–0.06	–0.06	–0.06	–0.06	–0.06	–0.06	–0.06
	(–1.10)	(–1.14)	(–1.17)	(–1.15)	(–1.15)	(–1.10)	(–1.20)	(–1.18)
ROA	0.11	0.13	0.10	0.09	0.12	0.09	0.10	0.11
	(0.23)	(0.27)	(0.21)	(0.20)	(0.26)	(0.19)	(0.22)	(0.23)
IVOL	–0.13***	–0.13***	–0.13***	–0.13***	–0.13***	–0.13***	–0.13***	–0.13***
	(–3.60)	(–3.59)	(–3.53)	(–3.58)	(–3.59)	(–3.56)	(–3.50)	(–3.58)
ILLIQ	0.38***	0.38***	0.38***	0.38***	0.38***	0.38***	0.38***	0.38***
	(8.18)	(8.17)	(8.14)	(8.11)	(8.16)	(8.19)	(8.06)	(8.20)
IO	0.04	0.04	0.04	0.03	0.04	0.02	0.02	0.04
	(0.29)	(0.29)	(0.25)	(0.18)	(0.24)	(0.11)	(0.15)	(0.26)
NPR	0.34**	0.34**	0.37**	0.34**	0.35**	0.36**	0.35**	0.36**
	(2.00)	(2.03)	(2.11)	(2.03)	(2.03)	(2.11)	(2.01)	(2.09)
ALMedia	–0.05	–0.06	–0.07	–0.08	–0.07	–0.07	–0.07	–0.08
	(–0.50)	(–0.58)	(–0.68)	(–0.77)	(–0.67)	(–0.65)	(–0.71)	(–0.80)
PopDensity	0.18	0.17	0.19	0.18	0.16	0.16	0.17	0.18
	(0.49)	(0.46)	(0.52)	(0.50)	(0.44)	(0.44)	(0.46)	(0.50)
Log(GDP)	–0.01	–0.01	–0.01	–0.02	–0.01	–0.02	–0.02	–0.02
	(–0.43)	(–0.41)	(–0.47)	(–0.52)	(–0.42)	(–0.53)	(–0.48)	(–0.49)
EmpShare	–0.31	–0.32	–0.32	–0.31	–0.32	–0.30	–0.29	–0.29
	(–1.21)	(–1.24)	(–1.24)	(–1.19)	(–1.24)	(–1.17)	(–1.12)	(–1.11)
Intercept	0.43	0.43	0.41	0.43	0.42	0.40	0.43	0.42
	(0.66)	(0.65)	(0.64)	(0.66)	(0.65)	(0.62)	(0.66)	(0.64)
Obs	303,361	303,361	303,361	303,361	303,361	303,361	303,361	303,361
Adj. R^2	6.44%	6.43%	6.45%	6.39%	6.39%	6.39%	6.42%	6.38%

Table IA.5
Standardized Abnormal Relative Postings and Stock Returns

This table reports the relation between standardized abnormal relative postings and future stock returns. The standardized abnormal relative posting (ARP^s) measure is the difference between abnormal relative postings of one firm and abnormal relative postings of firms not located in the focal firm's headquarters city. For firm i headquartered in city c, its abnormal relative posting measure in week t is constructed as relative postings for the firm in week t minus the median value of its relative postings in the previous ten weeks. Relative postings of firm i in week t are the logarithm of one plus the number of messages posted for firm i by posters in city c minus the logarithm of one plus the number of messages posted for firm i by posters outside city c. Abnormal relative postings of firms not located in city c are relative postings of firms not headquartered in city c minus the median value of the relative postings in the previous ten weeks. Relative postings of firms not headquartered in city c are the logarithm of one plus the number of messages from posters in city c for all firms not headquartered in the city minus the logarithm of one plus the number of messages from posters outside city c for all firms not headquartered in city c. In Panel A, firms are sorted into quintile portfolios each week based on their own ARP^s measure. In Panel B, firms are sorted into quintile portfolios each week based on the median value of peer firms' ARP^s measure. Peer firms are defined as firms that share the same headquarters city with a focal firm. Firms with less than fifteen peer firms are excluded from the sample. Both panels report average weekly equal-weighted excess returns, value-weighted excess returns, abnormal returns from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model for quintile portfolios and the high-minus-low ARP^s portfolios. The sample period is from June 2007 to May 2013. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: Sorting on ARP^s							
	Low	2	3	4	High	High – Low	t statistic
EW	–0.19	–0.17	–0.10	–0.07	0.04	0.22***	(5.49)
VW	–0.30	–0.24	–0.28	–0.23	–0.13	0.18**	(2.45)
Carhart four-factor alpha	–0.02	–0.01	0.07	0.09	0.20	0.22***	(5.15)
FF five-factor alpha	0.00	0.01	0.08	0.12	0.22	0.22***	(4.96)
Panel B: Sorting on Peer Firms' ARP^s							
	Low	2	3	4	High	High – Low	t statistic
EW	–0.26	–0.18	–0.14	–0.21	–0.11	0.15***	(2.85)
VW	–0.39	–0.39	–0.31	–0.28	–0.22	0.17*	(1.80)
Carhart four-factor alpha	–0.08	–0.02	0.02	–0.04	0.08	0.17***	(2.99)
FF five-factor alpha	–0.03	0.03	0.05	0.02	0.14	0.17***	(2.95)

Table IA.6

Distance-based, Province-level, and Length-weighted Abnormal Relative Postings and Stock Returns

This table reports results from Fama and MacBeth (1973) cross-sectional regressions of weekly stock returns (in percentage) on one-week lagged alternative abnormal relative postings (ARP) as well as control variables. In Column (1), posters from any city that is within 100 KM away from a firm's headquarters city are classified as local posters. In Column (2), posters from any city in the province (autonomous region or municipality) where a firm's headquarters city is located are classified as local posters. In Column (3), the number of messages is weighted by message length (number of characters in a message). The ARP measure is constructed as relative postings for a firm in one week minus the median value of its relative postings in the previous ten weeks where relative postings are the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. The control variables include firm size (Size), book-to-market ratio (BM), reversal ($Ret_{t-4:t-1}$), momentum ($Ret_{t-52:t-5}$), asset growth (AG), return-on-assets (ROA), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), institutional ownership (IO), net purchase ratio of insiders (NPR), abnormal local media coverage (ALMedia), local population density (PopDensity), logarithm of local GDP per capita (Log(GDP)), and share of industry employees (EmpShare). Detailed definitions are provided in Table A.1. The sample period is from June 2007 to May 2013. *t*-statistics based on Newey-West adjusted standard errors with three lags are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Distance-based Measure (1)	Province-level Measure (2)	Length-weighted Measure (3)
ARP	0.45***	0.44***	0.32**
	(3.26)	(2.91)	(2.23)
Size	–0.05	–0.06	–0.05
	(–1.13)	(–1.25)	(–0.96)
BM	0.05	0.05	0.05
	(0.94)	(0.89)	(0.79)
$Ret_{t-4:t-1}$	–0.04***	–0.04***	–0.04***
	(–3.10)	(–3.22)	(–3.07)
$Ret_{t-52:t-5}$	–0.04	–0.04	–0.04
	(–1.47)	(–1.49)	(–1.50)
AG	–0.06	–0.05	–0.06
	(–1.40)	(–1.06)	(–1.17)
ROA	0.19	0.13	0.08
	(0.44)	(0.29)	(0.18)
IVOL	–0.12***	–0.13***	–0.13***
	(–3.46)	(–3.53)	(–3.53)
ILLIQ	0.38***	0.36***	0.37***
	(8.78)	(8.44)	(8.04)
IO	0.05	0.05	0.03
	(0.35)	(0.37)	(0.21)
NPR	0.13	0.26	0.36**
	(0.89)	(1.64)	(2.09)
ALMedia	–0.07	–0.07	–0.07
	(–0.71)	(–0.69)	(–0.71)
PopDensity	–0.05	0.09	0.19
	(–0.15)	(0.26)	(0.53)
Log(GDP)	–0.02	–0.02	–0.01
	(–0.58)	(–0.62)	(–0.46)
EmpShare	–0.22	–0.28	–0.29
	(–0.89)	(–1.17)	(–1.12)
Intercept	0.48	0.53	0.39
	(0.76)	(0.85)	(0.59)
Obs	303,361	303,361	303,361
Adj. R^2	6.15%	6.17%	6.36%

Table IA.7

Abnormal Relative Postings and Stock Returns: Alternative Definitions for Normal Relative Postings

This table reports weekly excess returns and abnormal returns (in percentage) of quintile portfolios formed based on one-week-lagged abnormal relative postings (ARP) constructed with alternative approaches. In Panel A, the ARP measure is constructed as relative postings for a firm in one week minus the median value of its relative postings in the previous twelve weeks. The relative posting measure is the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. In Panel B, the ARP measure is constructed as relative postings for a firm in one week minus the mean value of its relative postings in the previous ten weeks. In Panel C, the ARP measure is constructed as relative postings for a firm in one week minus the weighted average of its relative postings in the previous ten weeks with decreasing weights. Specifically, the weight assigned to k-week-lagged relative posting is $\frac{10-k+1}{1+2+\cdots+10}$, with $k = 1, 2, \cdots, 10$. In Panel D, the ARP measure is constructed with an alternative measure for relative postings (RP'), which is defined as the difference between the number of local postings and nonlocal postings divided by the number of total postings. Each week, firms are sorted into quintile portfolios based on their ARP measure, and portfolios are rebalanced every week. Each panel reports average weekly equal-weighted excess returns, value-weighted excess returns, abnormal returns from the Carhart (1997) four-factor model, and abnormal returns from the Fama and French (2015) five-factor model for quintile portfolios and the high-minus-low ARP portfolios. The sample period is from June 2007 to May 2013. t-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Panel A: $ARP_{i_c,t} = RP_{i_c,t} -$ median$(RP_{i_c,t-1}, RP_{i_c,t-2}, \cdots, RP_{i_c,t-12})$

	Low	2	3	4	High	High – Low	t statistic
EW	–0.19	–0.16	–0.13	–0.06	0.05	0.24***	(5.66)
VW	–0.30	–0.27	–0.25	–0.24	–0.15	0.15**	(2.03)
Carhart four-factor alpha	–0.02	0.00	0.03	0.09	0.22	0.24***	(5.50)
FF five-factor alpha	–0.00	0.02	0.05	0.12	0.23	0.23***	(5.20)

Panel B: $ARP_{i_c,t} = RP_{i_c,t} -$ mean$(RP_{i_c,t-1}, RP_{i_c,t-2}, \cdots, RP_{i_c,t-10})$

	Low	2	3	4	High	High – Low	t statistic
EW	–0.17	–0.12	–0.14	–0.08	0.02	0.19***	(4.50)
VW	–0.27	–0.25	–0.32	–0.23	–0.10	0.17**	(2.11)
Carhart alpha	–0.01	0.05	0.03	0.08	0.18	0.19***	(4.35)
Five-factor alpha	0.00	0.07	0.04	0.10	0.20	0.20***	(4.42)

Panel C: $ARP_{i_c,t} = RP_{i_c,t} - \sum_{k=1}^{10} \frac{10-k+1}{1+2+\cdots+10} ARP_{i_c,t-k}$

	Low	2	3	4	High	High – Low	t statistic
EW	–0.20	–0.14	–0.13	–0.08	0.05	0.25***	(5.99)
VW	–0.26	–0.28	–0.27	–0.27	–0.08	0.17**	(2.13)
Carhart alpha	–0.04	0.03	0.03	0.08	0.22	0.25***	(5.89)
Five-factor alpha	–0.02	0.05	0.05	0.09	0.24	0.26***	(5.91)

Panel D: $ARP_{i_c,t} = RP'_{i_c,t} -$ median$(RP'_{i_c,t-1}, RP'_{i_c,t-2}, \cdots, RP'_{i_c,t-10})$

	Low	2	3	4	High	High – Low	t statistic
EW	–0.11	–0.18	–0.15	–0.09	0.05	0.16***	(3.77)
VW	–0.28	–0.24	–0.30	–0.24	–0.11	0.17**	(2.16)
Carhart alpha	0.04	–0.02	0.02	0.06	0.21	0.17***	(3.85)
Five-factor alpha	0.05	0.03	0.04	0.09	0.22	0.18***	(3.85)

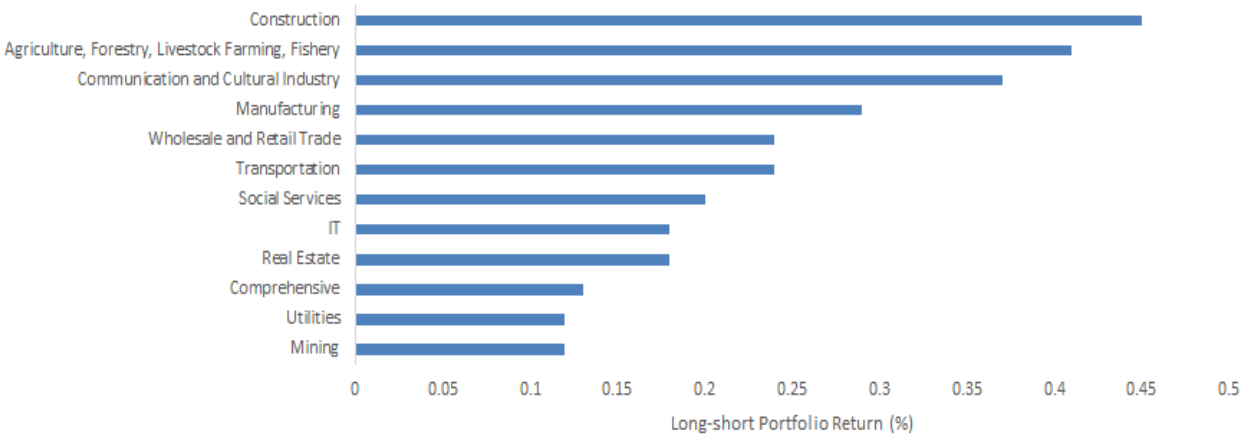
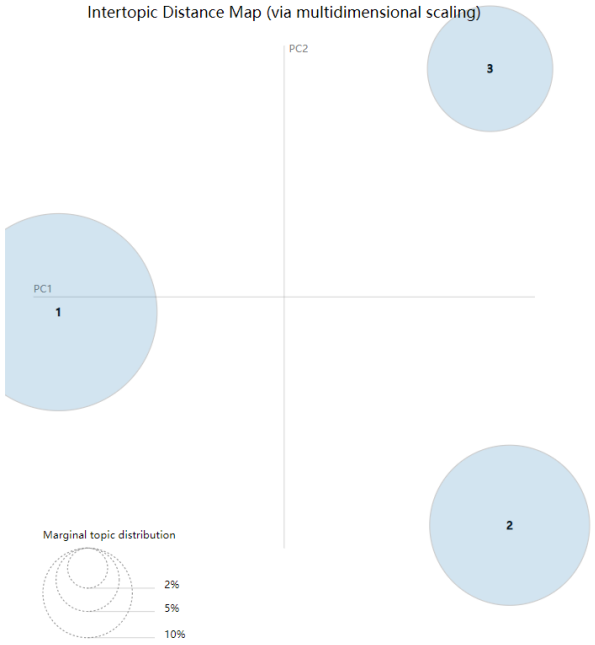
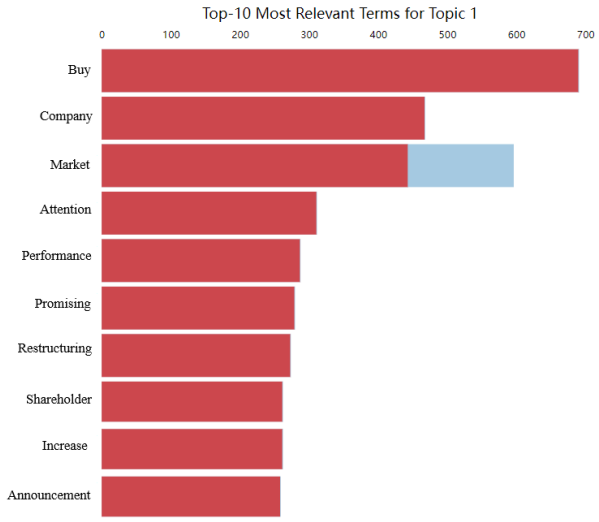


Figure IA.1
Profits of Long-short Strategy by Industry

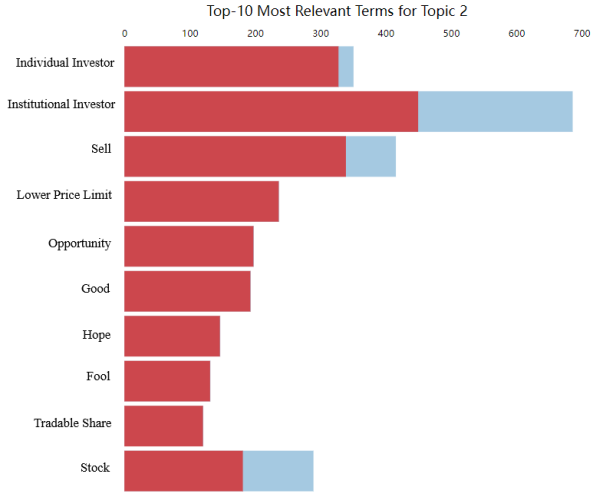
This figure plots profits of the ARP-sorted long-short strategy for each industry. Industries are defined using sectors in the *Guidance for Industry Classification of Listed Companies* released by the CSRC. The ARP measure is constructed as relative postings for a firm in one week minus the median value of its relative postings in the previous ten weeks where relative postings are the logarithm of one plus the number of messages from local posters minus the logarithm of one plus the number of messages from nonlocal posters. Each week, firms in each industry are sorted into quintile portfolios based on their ARP measure, and portfolios are rebalanced every week. Each bar represents average weekly equal-weighted excess returns of the high-minus-low ARP portfolios. The sample period is from June 2007 to May 2013.



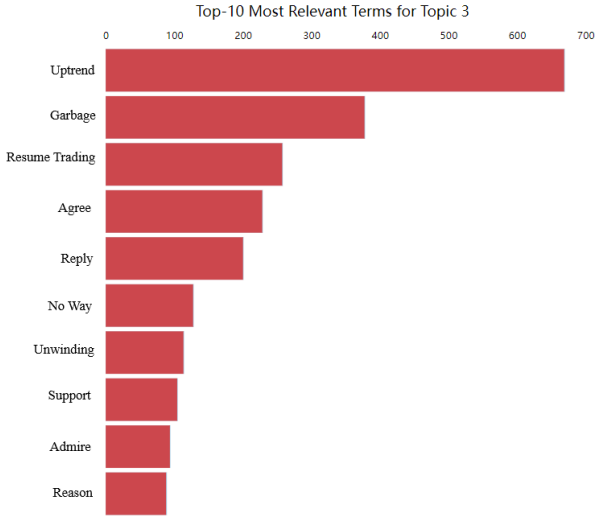
(a) Intertopic Distance Map



(b) Top Terms in Topic 1



(c) Top Terms in Topic 2

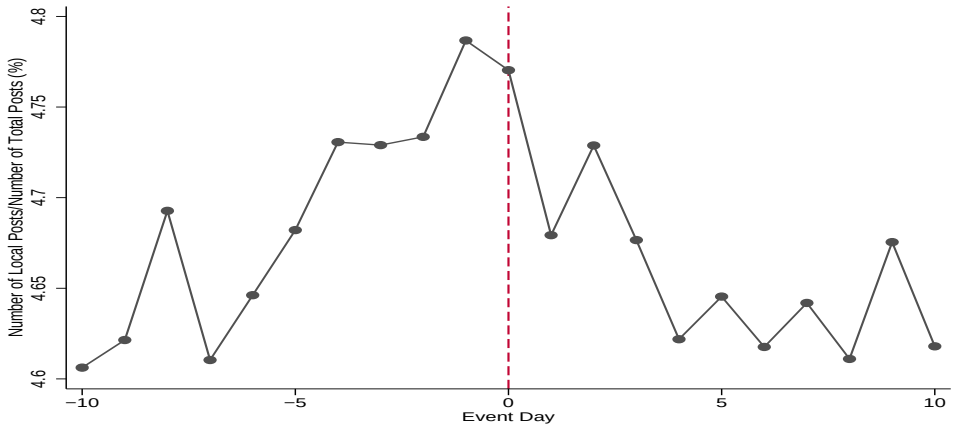


(d) Top Terms in Topic 3

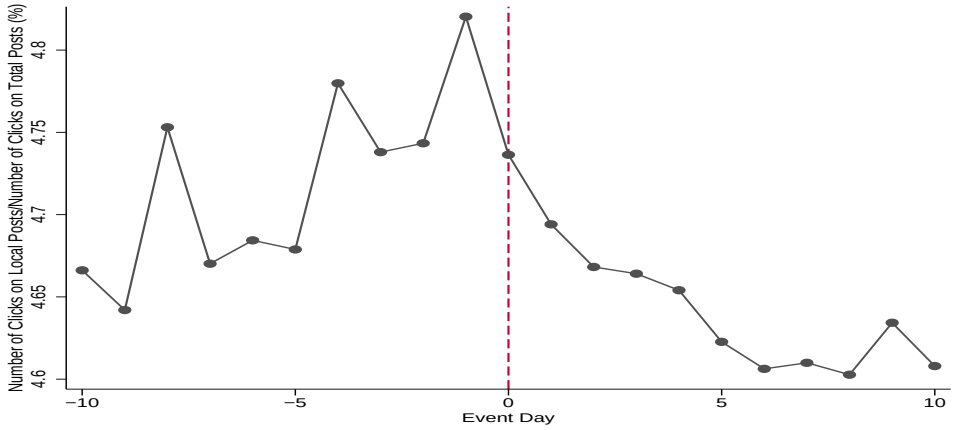
Figure IA.2
Topical Analysis of Nonlocal Posts

This figure illustrates the results from the topical analysis of nonlocal posts using the Latent Dirichlet Allocation (LDA) model. Subfigure (a) plots a global perspective of the topics projected into two-dimensional principal component analysis (PCA) space with the distance map. The circle size for each topic shows inter-topic differences computed by the Jensen-Shannon divergence. Subfigures (b) to (d) demonstrate the top ten most relevant terms for the topics 1 to 3, respectively. The length of the blue bar indicates the overall term frequency, and the length of the red bar indicates the estimated term frequency within a given topic.

(a) Percentage of Local Posts around Earnings Announcements



(b) Percentage of Clicks on Local Posts around Earnings Announcements



(c) Percentage of Replies to Local Posts around Earnings Announcements

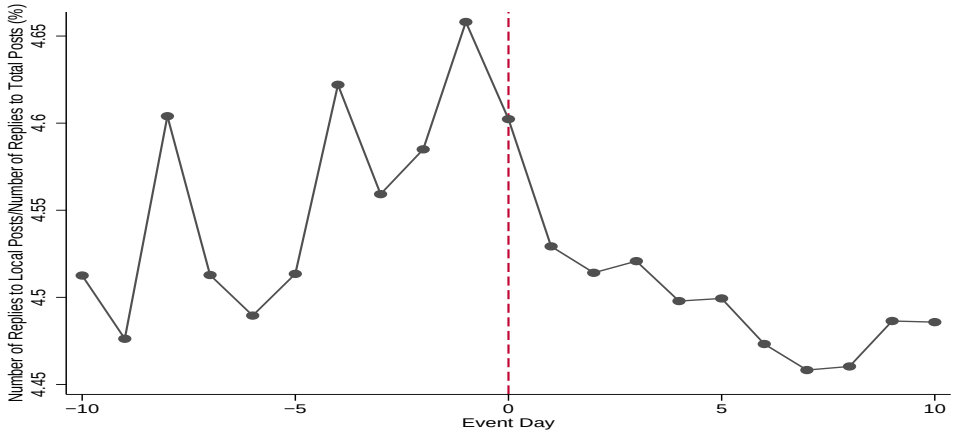
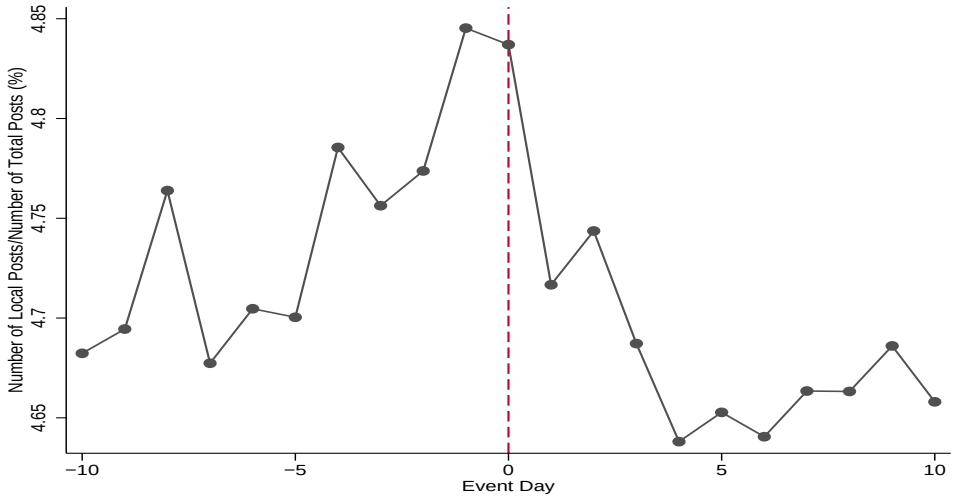


Figure IA.3
Characteristics of Local Posts Around Earnings Announcements

This figure plots the percentage of local posts (subfigure (a)), the percentage of clicks on local posts (subfigure (b)), and the percentage of replies to local posts (subfigure (c)) around firms' earnings announcements. The x-axis is the number of trading days relative to the quarterly earnings announcement date. The sample period is from June 2007 to May 2013.

(a) Percentage of Local Posts Among Messages with High Number of Clicks



(b) Percentage of Local Posts Among Messages with High Number of Replies

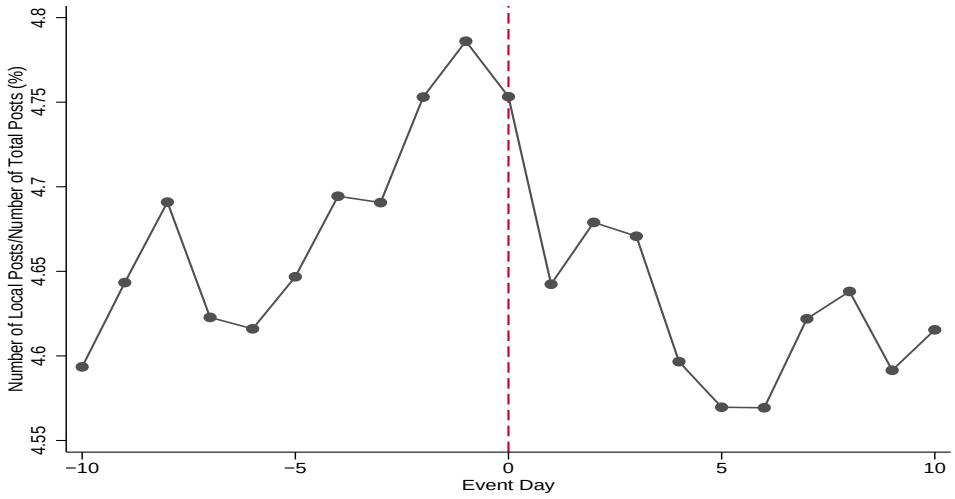


Figure IA.4
Percentage of Local Posts Among High-click or High-reply Messages Around Earnings Announcements

This figure plots the percentage of local posts among messages with a high number of clicks (subfigure (a)) or with a high number of replies (subfigure (b)) around firms' earnings announcements. The x-axis is the number of trading days relative to the quarterly earnings announcement date. A high-click (high-reply) message is one that receives more clicks (replies) than the median value for a given firm on a given day. The sample period is from June 2007 to May 2013.

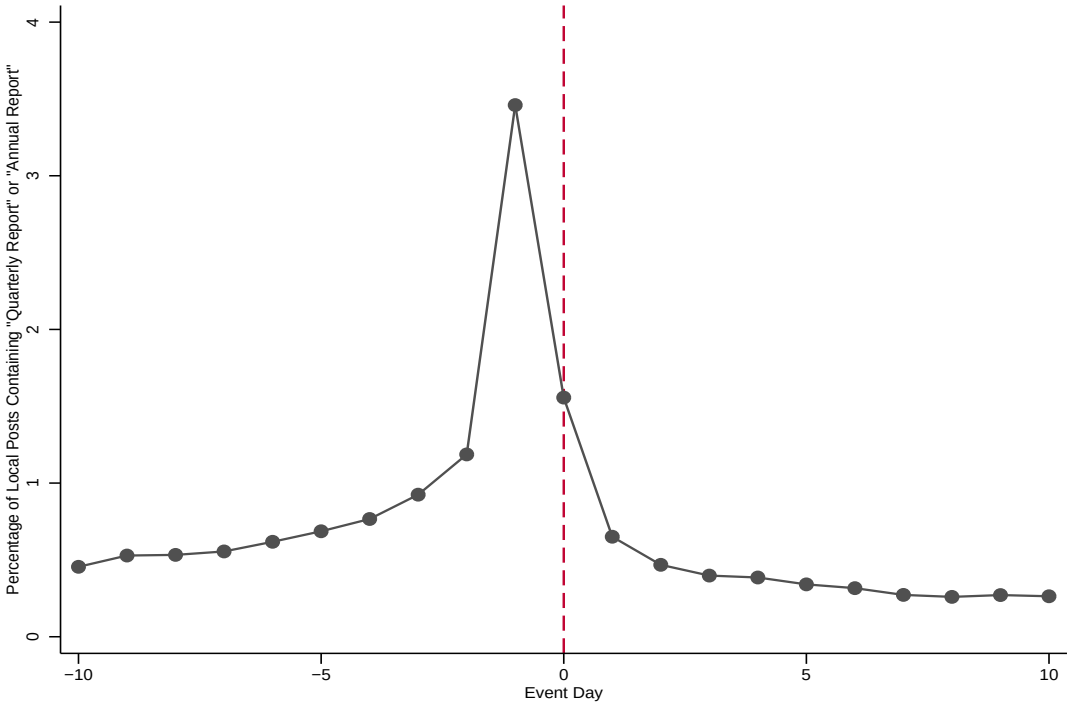


Figure IA.5
Percentage of Local Posts Containing "Quarterly Report" or "Annual Report" Around Earnings Announcements

This figure plots the percentage of local posts that contain the terms "quarterly report" or "annual report" around firms' earnings announcements. The x-axis is the number of trading days relative to the quarterly earnings announcement date. The sample period is from June 2007 to May 2013.