Do Local Investors Know More?
A Direct Examination of Individual Investors' Information Set

ROBERT GIANNINI

PAUL IRVINE

TAO SHU*

March 2015

* Giannini is at the BlueCrest Capital Management. Irvine is the C. R. Williams professor of finance at Texas Christian University. Shu is from the Terry College of Business at the University of Georgia and is visiting Hong Kong University of Science and Technology (HKUST). We thank Stocktwits.com for providing the stock tweets data. We thank Kenneth Ahern, Jonathan Berk, Fangjian Fu, Mark Grinblatt, Bige Kahraman, Harold Mulherin, Jeffry Netter, Mark Seasholes, Sheridan Titman, Eric Yeung, and seminar participants at the 2014 European Finance Association Meeting, 2014 Utah Winter Finance Conference, the 2013 FIRS Conference, the 2013 Asian Finance Association Conference, the 2012 China International Conference of Finance, the 3rd TCFA Best Paper Symposium, and University of Georgia for helpful comments. Shu appreciates the financial support from the Terry Sanford Award at the University of Georgia. Correspondence may be sent to Dr. Shu at: taoshu@terry.uga.edu.

Do Local Investors Know More?
A Direct Examination of Individual Investors' Information Set

March 2015

Abstract

We examine 176,375 Twitter posts from nearly two thousand individuals covering 1,015 U.S. companies from 2009 to 2011. While these individuals on average exhibit a negative stock picking ability, they are significantly more informed about local companies than about nonlocal companies, with the differential return predictability being 18 basis points per week. Local advantage is much larger in firms without public news coverage and firms with greater information asymmetry. Compared to local investors, nonlocal investors exhibit significantly greater overreaction to analyst opinions. These results indicate that local advantage is attributable to individual investors' private information, which reduces investors' behavioral biases.

1. Introduction

Do investors have more value-relevant information about local firms? This question has important implications for financial market efficiency, the information diffusion process, and investment practices. In addition, local advantage can shed light on the puzzling "home bias" in which investors exhibit a strong preference for locally headquartered stocks (e.g., Coval and Moskowitz 1999, 2001; Ivkovic and Weisbenner 2005). The pioneering work of Coval and Moskowitz (2001) provides evidence of local advantage by showing that mutual fund managers earn abnormal returns on their local investments.[1] Researchers have also documented local advantage for various financial market participants including analysts (Malloy 2005; Bae, Stulz, and Tan 2008), commercial and investment banks (Butler 2008; Agarwal and Hauswald 2010), institutional shareholders as monitors (Gaspar and Massa 2007; Ayers, Ramalingegowda, and Yeung 2011), and acquirers (Almazan, Motta, Titman, and Uysal 2010).

Despite the extensive studies on local advantage, the empirical evidence is mixed for individual investors. Ivkovic and Weisbenner (2005) observe local advantage in a sample of 34,517 households from 1991 to 1996. They find that individual investors earn an abnormal return of 3.2% per annum on their local holdings relative to their nonlocal holdings. However, Seasholes and Zhu (2010) use a calendar-time portfolio approach and show that individual investors earn only zero alphas on their local holdings relative to their nonlocal holdings, and their purchases of local stocks significantly *underperform* their sales of local stocks. Seasholes and Zhu conclude that "individuals do not seem to have value-relevant information about the local stocks they trade."

In this paper we take a novel approach to investigate whether individuals have value-relevant information about local stocks. While previous studies focus on the performance of local investments

[1] Baik, Kang, and Kim (2010) and Bernile, Kumar, and Sulaeman (2013) further find that local ownerships of general institutions positively predict stock returns.

by individual investors, we use a large sample of Twitter posts to directly examine investors' information about local and nonlocal firms. Twitter is an electronic social network where users post short thoughts of no more than 140 characters, called tweets. Twitter has been documented to have fast growing impact on the financial markets. Since quick and broad information dissemination is available through Twitter, in April 2013, the U.S. Securities and Exchange Commission approved using Twitter to communicate company announcements.

We use a unique dataset of Twitter posts on publicly traded U.S. companies from the website Stocktwits.com, a popular platform for Twitter users to tweet about stocks. These Twitter posts come from the users of Stocktwits.com who provide location information at the city (county) level and live in the continental U.S. Our final sample contains 176,375 Twitter posts from 1,903 Twitter users, covering 1,015 publicly traded U.S. companies from July 11, 2009 to June 10, 2011.

We begin by examining the overall stock-picking ability of Twitter messages. This test can also provide some insight on why some institutional investors take an active interest in an open source stream of 140 character bits of information. We extract the sentiment in the Twitter posts using a maximum likelihood classification (ME) approach and examine whether these evaluations predict stock returns. The ME classification technique does more than just count the numbers of positive and negative words but allows both individual words and phrases and their strength to indicate sentiment. Further tests on abnormal returns suggests that the ME approach can better identify sentiment than the naïve Bayesian method currently standard in the literature. Our examination of investors' information sets rather than their trading offers a novel test on the informedness of individual investors because the trading decisions of investors, particularly individual investors, can differ from their information due to various factors such as diversification, liquidity provision, and trading constraints.[2] Thus, our novel data can provide new evidence on investor informedness and local advantage.

We find that the sample individual investors on average have a *negative* return predictability. For example, a one standard deviation increase in the two-week evaluation measure is associated with a 10 basis point *decrease* (t-stat -5.27) in abnormal stock returns in the subsequent week. This result is robust to the alternative return windows from two days to one month, alternative measurements of evaluations, and skipping a week before return measurements to control for microstructure effects.

Twitter users' local advantage completely ameliorates this underperformance. We classify Twitter posts into local and nonlocal posts according to whether the distance between a user's location and corporate headquarters is within 100 miles.[3] Nonlocal posts still exhibit strong negative return predictability, but local posts have no significant return predictability disadvantage. For example, a one standard deviation increase in nonlocal evaluations predicts a decrease of 20 basis points (t-stat -6.15) in weekly stock returns, but a one standard deviation increase in local evaluations predicts a decrease of 2 basis point (t-stat -0.89) in weekly stock returns. Local advantage, measured as the differential predictability between local and nonlocal evaluations, is a statistically and economically significant 18 basis points (t-stat 4.17) per week.

The local advantage persists for return measurement windows from two days to one month, and for alternative constructions of the evaluation measures. The local advantage is also robust to controls for priced factors, return autocorrelations, microstructure effects, and geographical factors. To address the concern of potential noise in the Twitter posts, we repeat the tests using a sub-sample of users in Stocktwits.com's list of "recommended" contributors. These users generally have a large

[2] For example, Kaniel, Saar, Titman (2008) find evidence that individuals make trading profits by providing liquidity.
[3] We conduct robustness tests by classifying local and nonlocal posts using the 50-mile criterion (Ivkovic and Weisbenner 2007) instead of the 100-mile criterion, or whether the Twitter user and the company are located in the same state. The results of local advantage hold under both alternative approaches.

following, a long track record, post meaningful or interesting comments, and are notable within the social network. The result of local advantage is robust when we restrict the sample to recommended users. Additionally, our evidence indicates that the differential predictability is not likely to be caused by Twitter followers' stronger responses to local posts.

Next, we examine whether local advantage is due to individual investors acquiring value-relevant private information about local firms ("information hypothesis"). In the model of Van Nieuwerburgh and Veldkamp (2009), investors concentrate on acquiring private information about local firms to amplify their initial information advantage in these stocks. To explore this hypothesis, we examine local advantage across firms with and without public news coverage during the period of Twitter posts.[4] If local advantage is due to local investors' access to private information, then we expect the local advantage to be stronger among firms with no public news coverage. Our results show that local advantage is much larger in the firms without public news coverage (56 basis points per week) than for firms with public news coverage (13 basis points per week). This finding lends strong support to the information hypothesis and also supports Tetlock's (2010) conclusion that public news releases can level the playing field for uninformed investors. The results using the proxies for information asymmetry proposed by the existing literature also show a significantly positive association between local advantage and information asymmetry. These results indicate that local advantage is caused by individual investors' access to private information about local firms.

To further understand the local information advantage, we investigate if local investors better predict earnings surprise than nonlocal investors. For an earnings announcement, we measure earnings surprise using standardized unexpected earnings (*SUE*) constructed as the difference between actual earnings and consensus forecast, scaled by stock price. We find that evaluation of local investors prior

[4] We collect news articles from the three major news wires including Reuters News, Dow Jones News Wire, and PR News Wire.

to earnings announcements better predicts *SUE* than that of nonlocal investors, which suggests that superior information about earnings is a source of local information advantage.

Despite the average negative return predictability for Twitter posts, the Stocktwits.com stream is actively sold to hedge funds and other market participants. To demonstrate that Twitter posts do contain useful investing information, we calculate returns to a rolling long-short strategy that goes long the portfolio containing firms with more favorable local evaluations than non-local evaluations, and goes short the portfolio containing firms with less favorable local evaluations than nonlocal evaluations. The average daily abnormal profits of this zero-investment strategy are significantly positive, ranging from 5.9 basis points to 8.5 basis points per day. The abnormal profits increase significantly when we exclude firms with less information asymmetry (large firms, high analyst coverage firms, or low idiosyncratic volatility firms).

Our results are consistent with a large literature finding that individuals lose significantly from their trading and that individuals are subject to various behavioral biases (e.g., Barber and Odean 2000, 2008; Barber, Odean, and Zhu 2009; Barber, Lee, Liu, and Odean 2009). We attempt to understand the exact mechanisms of the underperformance of individual investors in our setting and how local information advantage helps alleviate this underperformance.

First, it is possible that local advantage is associated with return reversal (or price impact). For example, nonlocal investors might blindly follow contemporaneous stock returns and therefore suffer a loss from subsequent short-term return reversal. It is also possible that nonlocal investors represent the sentiment of general investors that causes price to move and then reverse subsequently. Return reversal, however, is unlikely to explain our findings because our regressions control for lagged returns contemporaneous to investor evaluation. To further evaluate the effect of return reversal on local advantage, we repeat the regressions without controlling for lagged returns (short-term reversal) or controlling for lagged returns in the previous one-month window instead of the two-week window.

The results on local advantage remain similar in both settings, suggesting that return reversal is largely irrelevant to the observed local advantage.

The second mechanism we examine is based on the previous studies which find that individual investors fail to correct for the complexity of analysts' incentives. For example, Malmendier and Shanthikumar (2007, 2009) and Mikhail, Walther, and Willis (2007) show that small traders tend to overreact to analysts' recommendations, especially buy recommendations, and therefore experience significant underperformance. We hypothesize that the information advantage of local investors can alleviate their overreaction to analyst opinion. Our empirical results show that nonlocal investors respond more aggressively to overoptimistic analyst forecasts or buy recommendations than local investors. The nonlocal overreaction to analyst opinions is not warranted by subsequent stock performance and explains a significant fraction of their underperformance. This evidence suggests that local investors appear to rely more on their own private information rather than the analyst opinions, which allows them to overcome the overreaction and the associated underperformance. This is the first test that we are aware of that investigates the mechanism of local investors' relative advantage.

Our findings shed light on whether or not individual investors have a local advantage (e.g., Ivkovic and Weisbenner 2005; Seasholes and Zhu 2010). We are the first to directly examine investors' *information* about local and nonlocal companies, and our evidence suggests that individual investors can have a statistically and economically significant informational advantage about local firms. Our findings also demonstrate that local information advantage can overcome the significant negative performance of uninformed investors, yielding important insight into market efficiency and individual investors' performance. The recent work by Seasholes and Zhu (2010) suggests that individuals do not have a local advantage because they earn zero or negative abnormal returns on local investments.

We differ from their approach in that we document local investors' information advantage that may or may not turn into trading profits due to individual investors' constraints or behavioral biases.[5]

We also contribute to a rapidly growing literature on the informedness of individual investors. Previous studies provide evidence that individual investors lose significantly from their trading. However, several recent papers suggest that individual investors can be informed prior to earnings announcements (Kaniel, Saar, Titman, Liu 2012), takeovers (Griffin, Shu, and Topaloglu 2012), or when they submit market orders (Kelley and Tetlock 2013). Our paper performs a unique examination of the informedness of individual investors since we directly study the information possessed by individual investors and this test design is free of trading complications such as liquidity or price impact. We find that, on the one hand, the evidence from Twitter posts shows that individual investors exhibit *negative* stock return predictability, which underlines the lack of stock-picking ability for individual investors. On the other hand, individual investors are able to acquire private information when they live close to a firm's headquarters.

Our study also extends the literature on internet communication about the financial markets. Motivated by the rapid growth in internet communication in the past two decades, financial researchers have started to investigate whether stock specific internet messages contain value-relevant information or just noise. Several studies find that messages posted on internet stock message boards (e.g., Yahoo! Finance) have little to no predictive power for stock returns, suggesting that these messages may simply contain noise (Tumarkin and Whitelaw 2001; Antweiler and Frank 2004; Das, Martinez-Jerez, and Tufano 2005). Consistent with these studies, we also document a negative stock return predictability for the Twitter posts in our sample. However, we extend the previous studies by

[5] For example, short sales constraints (borrowing constraints) can prevent investors from capitalizing on negative (positive) information. Investor irrationalities may also prevent individual investors from utilizing information. For instance, the disposition effect (Odean 1998; Grinblatt and Han 2005) can make investors hold on to a losing position despite negative information.

documenting that the Twitter posts may contain value-relevant information about local companies. Therefore, our results therefore provide interesting evidence that instead of all noise, internet messages also contain value-relevant information about financial markets.[6]

The outline of our paper is as follows. Section 2 describes the Twitter data and sample construction. Section 3 presents the evidence of local advantage. Section 4 analyzes the source of local advantage, and Section 5 concludes.

2. Data and Sample Selection

2.1. Twitter and Financial Markets

Twitter is a micro blogging application where users are able to post short thoughts of no more than 140 characters, called tweets. While Twitter was started as a social network, its worldwide popularity and broad user base have garnered it a fast growing impact on many aspects of people's lives. One prominent example is that during the 2011 Egyptian Revolution, Egyptian bloggers and journalists widely used Twitter to report on the strike, organize legal protection, and draw attention to their efforts.

Twitter has also been related to the financial markets. Paul Hawtin, founder of Twitter hedge fund Derwent Capital, claims "Today, social media creates a vast amount of information and it has been proven that the sentiment derived from it can predict stock market movements." In April 2013, the U.S. Securities and Exchange Commission approved using Twitter to communicate company announcements. On April 24, 2013, the Dow Jones industrial average immediately plunged by more than 140 points after a hacker sent out a false tweet from Associated Press's account.

We collect Twitter posts from Stocktwits.com, an open micro-blogging site which is powered by Twitter with a focus on financial markets. Stocktwits.com was founded in 2008 and has since then

[6] Chen, De, Hu, and Hwang (2014) provide supporting evidence on the value relevance of particular internet sites.

become a popular website for Twitter users to exchange investment information. Since its inception, Stocktwits.com has been covered by major news media such as The New York Times and CNNMoney.com. In 2010, Stocktwits was named Time.com's top 50 best websites as well as Fast Company's top 10 innovative companies in finance.

2.2. Why do People Share Financial Information in Twitter Posts?

A natural question is why people would share value-relevant information in their Twitter posts. More broadly, why do people post internet messages about the stock markets? Despite the rapid growth in internet communication about financial markets, there is not much theoretical literature that rationalizes this type of communication. DeMarzo, Vayanos, and Zwiebel (2003) propose a model in which investors fail to account for the repetition of opinions ("persuasion bias"). In equilibrium, well-connected agents in a social network can have significant influence on the actions of other members, and therefore, impact the market. This model can potentially explain why Twitter users may have incentives to gain popularity and followers by sharing value-relevant information in their Twitter posts. Consistent with this view, the models in Cao, Coval, and Hirshleifer (2002), Colla and Mele (2010), and Hong, Hong, and Ungureanu (2011) show that information sharing among investors can cause trading and affect the outcomes of the financial markets.[7]

Additionally, popularity gained in the Twitter world may also bring direct financial benefits. For example, Stocktwits.com now offers a full "marketplace" of premium blogs to users of the site. These premium blogs are based on the tweets and trading ideas of successful investors from the Stocktwits.com community. The themes of the premium streams range from value investing to swing

[7] Empirical evidence also suggests that people listen to ideas from friends to make financial decisions (Duflo and Saez 2002). Additionally, Hong, Kubik, and Stein (2004, 2005) show that institutional and individual investors' investment decisions are affected by other institutions in the same area or neighbors.

trading, and annul subscriptions can cost in excess of $800. Therefore the potential financial benefits can also motivate Twitter users to gain followers by sharing value-relevant information.

A more malevolent alternative is that some Stocktwits users are attempting to manipulate the price. Since Leinweber and Madhavan (2001) find a "pump and dump" strategy is by far the most common form of price manipulation, we undertake several examinations of the data to look for opinion and return patterns consistent with this activity. However, even a focus on stocks priced below our $2 threshold failed to uncover any evidence consistent with this strategy.

2.3. Collection of Twitter Posts and Construction of Sample

Figure 1 provides an example of the stream of Twitter posts that comprise our sample. Twitter users comment about a specific company by referring to the company's ticker preceded by a "$" hashtag. An example of this would be "$MSFT and $AAPL are a buy!". Hashtagging allows us to extract specific company references with a high level of accuracy by looking for the "$" hashtag followed by one to four capital letters that constitute the ticker symbol. In the case of multiple company references in one post, like the example above, each reference is counted as a unique post.[8]

Our initial sample contains all the twitter posts about publicly traded companies from Stocktwits.com from July 11, 2009 to June 10, 2011. For each post in the data, we have the content of the post, the associated ticker symbol(s), the date and the time of the post, and the blogger's account ID and the number of followers. This initial sample contains 1,048,575 posts covering 7,757 security symbols. Since some of the symbols represent non-stock assets such as gold, foreign currencies, or indices, we further identify stock tweets by matching to stock tickers in CRSP. This procedure yields in total 782,904 stock tweets covering 5,927 stock tickers, with each post associated to a unique ticker and author. We further match stock tickers to PERMNOs, which is the unique firm identifier in our

[8] Among the original posts, 88% cover only one symbol, 7% covering two symbols, 5% covering more than two symbols.

analysis, and the matched sample contains 778,764 posts covering 5,806 unique firms (PERMNOs). Section A1 of the Appendix describes the details of the matching procedures.

We collect a Twitter user's location on the user's profile page on Stocktwits.com by searching the account ID. Out of the 9,723 users in the initial sample, 3,052 users provide some kind of location information. We then require the users to live in the continental U.S. and provide location information at the city (county) level because we require both the state and the city information to calculate the distance to corporate headquarters. Section A.2 of the Appendix provides details about the identification of user locations.

We require the sample firms to have available CRSP data, have headquarters located in the continental U.S., and have at least ten Twitter posts and one local Twitter post over our sample period.[9] To control for microstructure effects, we drop penny stocks that are priced below two dollars at the end of the previous year.[10] Our final sample of Twitter posts contains 176,375 posts from 1,903 users covering 1,015 publicly traded companies from July 11, 2009 to June 10, 2011. Figure 2 plots the locations of the sample Twitter users. The users live in all of the states in the continental U.S. except North Dakota. The highest percentages of users are in the states of New York (18%), California (17%), Illinois (9%), Texas (7%), and Florida (7%). The remaining users reside in 42 other states and Washington D.C., with no other state accounting for over 5 percent of sample users. The geographical distribution of Twitter users is consistent with the distributions of U.S. population and economic activity. This dispersed distribution avoids the issue of geographical clustering.[11] For robustness, we also use the approach in Seasholes and Zhu (2010) to construct state-adjusted returns and find similar results on local advantage.

---

[9] Our results are robust when we require the firm to have at least one Twitter posts over the sample period, or when we do not require the firm to have at least one local Twitter posts over the sample period.
[10] Our results are similar when we include penny stocks into the sample.
[11] Geographic clustering is a common problem in U.S. financial research as many financial intermediaries are clustered in the New York City area (see Anand et al. 2011).

We obtain accounting data for our sample firms from Compustat, and data on analyst coverage, analyst forecasts, and analyst recommendations from the IBES summary file. We also obtain the daily returns of the three Fama-French factors and the momentum factor (UMD) from Kenneth French's data library for the construction of abnormal returns.[12] Some of our tests use news articles collected from Factiva, and we will describe the details of the construction of news data when we discuss the corresponding tests.

2.4. Classifying Local and Nonlocal Posts

We calculate the straight line geographic distance between the location of each Twitter user and the headquarters of each company in our sample using longitude and latitude coordinates. We assign longitude and latitude coordinates to the user locations according to the state and city information on their profiles. We further obtain zip codes of corporate headquarters from Compustat and assign the corresponding longitude and latitude coordinates.[13] We then calculate the distance between Twitter user and company headquarters using the following equation:

$$Distance = 7921 * arcsin\,(sqrt((sin((0.017 * lat2 - 0.017 * lat1)/2))^2 + cos(0.017 * lat1) * cos(0.017 * lat2) * (sin((0.017 * long2 - 0.017 * long1)/2))^2)) \quad (1)$$

where *lat1* and *long1* are the latitude and longitude coordinates of a Twitter user and *lat2* and *long2* are the latitude and longitude coordinates of corporate headquarters.[14]

We classify a Twitter post as local (nonlocal) if the distance between the Twitter user and the corporate headquarters is within (more than) 100 miles. Previous studies use various criteria of distance to classify local stocks, from 62 to 250 miles (e.g., Coval and Moskowitz 2001; Ivkovic and

[12] We thank Professor Kenneth French for making the data available.
[13] We match longitude and latitude coordinates to zip codes using the database from http://www.getzipcodedata.com/#.
[14] This equation is provided by SAS at http://www2.sas.com/proceedings/sugi31/143-31.pdf. This approach is based on the great circle distance model which is similar to the distance equations used in the literature (e.g., Ivkovic and Weisbenner 2007) but provide greater accuracy at small distances. More details about the distance models can be found at http://en.wikipedia.org/wiki/Great-circle_distance.

Weisbenner 2005; Malloy 2005; Seasholes and Zhu 2010). We adopt the moderate 100-mile criterion and classify 20,570 posts as local and 155,805 posts as nonlocal. For robustness we also try classifying a post as local (nonlocal) using two alternative approaches: 1) If the distance between the Twitter user and the corporate headquarters is within (more than) 50 miles; 2) If the Twitter user is in the same state as the firm's (a different state from the firm's). We find similar results from our tests using the two alternative approaches (results discussed in Section 3).

2.5. Quantifying the Information in Twitter Posts

The unique features of Stocktwits posts, such as addressing replies, and the 140 character restriction produce a language that is markedly different from standard English. Conventional word counts using standard English dictionaries are unlikely to be useful in interpreting Stocktwits posts.

We use the maximum entropy (ME) approach, which endogenously creates a dictionary of terms, to classify the information in Twitter posts. The ME approach derives sentiment from the statements in posts by applying a maximum likelihood algorithm to the data. The information in Twitter posts can be subtle. For example, the statement "You would be crazy to sell $GOOG right now" contains the word "sell" which unconditionally we would assume has a negative connotation. However, the statement "crazy to sell" is obviously a positive statement. ME classification is considered the most robust technique for information classification because it controls for the conditional dependence of words (Pang, Lee, and Vaithyanathan 2002). Unlike the less sophisticated procedures which handle each word as an unconditional feature, ME classification uses the information contained in multiple word phrases such as "crazy to sell" to more accurately classify information.

In addition to controlling for the conditional dependence of words, the ME classification also avoids the misidentification issue associated with alternative approaches that simply rely on key-word frequencies. For example, Loughran and McDonald (2011) show that in the textual analysis of 10-K reports, almost three-fourths (73.8%) of the negative word counts according to the widely used Harvard Dictionary are attributable to words that are typically not negative in a financial context (e.g., tax, cost, capital, board, liability). Other words on the Harvard list (e.g., mine, cancer, crude, tire, capital) are more likely to identify a specific industry segment than reveal a negative financial event. ME classification does not suffer the noise introduced by key-word selection because the identification is based on a large training sample of Twitter posts that we hand classify.[15,16]

The general idea of ME classification is that when nothing is known about a distribution, the distribution should be uniform, i.e., have maximum entropy. Consider the example of trying to classify a document as positive, negative, or neutral, where we are only told that 50% of documents that contain the word "buy" are considered positive. Intuition tells us that if the document has the word "buy" in it then there is a 50% chance that it is a positive post, a 25% chance of being negative, and a 25% to of being neutral. If our document did not have the word "buy" in it then we would just assume an equal distribution of a 33% chance that the document falls into each category. Thus, if we knew nothing about our document, we begin with a uniform distribution with equal likelihoods for each sentiment category. This is the essence of ME classification. In practice, this process is constrained by many features, and the calculations for conditional probabilities become complex, but the logic is still the same as our simple example.

To formally describe the ME procedure, we define the following set of terms. Let $F = (f_1, \ldots f_m)$ be a set of predefined features that can appear in a post. From our previous example, the word "sell" would be a feature, and the tri-gram "crazy to sell" would also be a feature. Let $n_i(d)$ be the number of

[15] Additionally, many previous studies using the Harvard list only count negative words because they find little incremental information in the Harvard positive word list (e.g., Tetlock 2007; Engelberg 2008). In contrast, ME classification is based on both positive and negative comments in the messages.
[16] We interpret our approach as similar in spirit to Loughran and McDonald (2011) who exogenously define a dictionary which is suitable to particular types of financial information and then use that dictionary to evaluate a large data set. We exogenously classify sentiment in a training data set and allow the ME program to determine the likelihood that a particular word or phrase represents a particular sentiment.

times that the feature f_i occurs in a post d. Thus, each post is represented by a post vector that takes the form: $\bar{d} = (n_1(d), n_2(d), , n_m(d))$. Lastly, let c be a post category that takes the value of c_0 (positive, negative, or neutral). Given this set of variables, the estimate of $P(c{=}c_0 \mid d)$ is as follows:

$$P_{ME}(c = c_0 \mid d) = \frac{1}{Z(d)} (\textstyle\sum_i \lambda_{i,c} F_{i,c}(d, c)) \tag{2}$$

where $Z(d)$ is a normalization function, and $F_{i,c}$ is a feature category function for the feature i and for each category c defined as

$$F_{i,c}(d,c) = \begin{cases} 1, \text{ if } n_i(d) > 0 \text{ and } c_i = c_0 \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

For example, this feature category function only returns a value of one if the post contains the tri-gram "crazy to sell" and the post is hypothesized to be of positive sentiment. $\lambda_{i,c}$ is a weighting parameter that determines the relative strength of each of the features f_i contained in a document. If the value of $\lambda_{i,c}$ is very large then the feature f_i is considered to be very strong for a specific category c_0. Using the weighting parameter allows us to implement Jegadeesh and Wu's (2013) finding that weighting can be an important tool in content analysis.

We implement the ME classifier by hand classifying a corpus of 2,000 twitter posts. This out-of-sample set of categorized data is called training set, and is used to calculate the expected values of $F_{i,c}$. Next, we use all the Twitter posts to estimate the conditional probabilities $P_{ME}(c{=}\ c_0 \mid d)$ by calculating the maximum likelihood solution across the three different categories while satisfying the constraint that the expected values of the feature category functions $F_{i,c}$ are equal to their training data expected values. Each post in our dataset is then assigned a value of (-1, 0, 1) based on the highest conditional probability of a post being positive, negative, or neutral. We test the accuracy of this procedure by running the ME classifier on a set of 300 posts that are hand classified. The ME classifier worked well in this out of sample test, and it was able to correctly classify 67% of all posts in the test sample. This accuracy rate is similar to the accuracy level that is achieved in other sentiment classification studies, such as Pang, Lee and Vaithyanathan (2002).[17]

We also try classifying the information in Twitter posts using the Naïve Bayesian (NB) approach proposed by the existing literature (Li 2010). We conduct robustness tests using the NB approach and find similar results on local advantage. The NB approach and the corresponding results are discussed in Section 3.

3. Do Individual Investors Have Local Advantage?

3.1. Summary Statistics

Panel A of Table 1 summarizes the characteristics of our sample. A typical firm in our sample has a market capitalization of $8,579 million, a book-to-market ratio of 0.63, and is followed by 11.21 analysts. For a comparison, an average firm in the contemporaneous CRSP universe has a market cap of $2,292 million, a book-to-market ratio of 0.96, and is followed by 3.82 analysts. Since we require sample firms to have at least ten Twitter posts during the sample period, these comparisons suggest that the firms covered by Twitter users tend to have larger size, higher analyst coverage, and lower book-to-market ratios.[18] We also report average idiosyncratic volatility and daily return for sample firms. Idiosyncratic volatility for a firm-day is the standard error of residuals from the time-series regressions of a firm's excess returns on the daily market factor (MKT) in the one-year window up to the end of previous month.[19] A typical sample firm has an idiosyncratic volatility of 0.029 and an

[17] It is difficult to compare the accuracy of ME classification with previous studies in the finance literature because they generally use key-word counts directly in the empirical analyses without examining the proportions of correct and incorrect identifications of sentiments. Loughran and McDonald (2011) report that 73.8% of the negative word counts based on Harvard Dictionary are not associated with negative meanings in a financial context, but their sample is 10-K reports instead of internet messages.
[18] We drop penny stocks priced below $2, which also makes our sample firms larger than the CRSP universe.
[19] We require at least 100 daily return observations in the estimation window.

average daily return of 12 basis points, similar to the 0.030 and 11 basis points for the contemporaneous CRSP universe.

Panel B presents the summary statistics of the sample Twitter posts. The average evaluation of Twitter posts is positive, which is consistent with the recovery of stock market during the sample period. While both the local and nonlocal tweets are positive, local tweets are less so than nonlocal tweets. Regarding the Twitter coverage, a sample firm receives on average 174 posts during our sample period, including 154 nonlocal posts and 20 local posts.

To examine the why certain stocks receive interest on Twitter,, we present in Panel C the firm-level regression of Twitter coverage on firm characteristics and stock market metrics. In the first model, the dependent variable is the total number of Twitter posts for a firm during the sample period. The independent variables include market capitalization, book-to-market ratio, analyst coverage, idiosyncratic volatility, and average daily return. Because of the existing literature on "attention attracting" for individual investors (e.g., Barber and Odean 2008; Hirshleifer, Lim, and Teoh 2011), we also include into the independent variable the total number of news articles for a firm during the sample period (the collection of news article will be discussed in Section 4). The coefficient on the book-to-market ratio is significantly negative but that on analyst coverage, idiosyncratic volatility, and average daily return is significantly positive. These results suggest that Twitter coverage is higher for growth firms, and firms with more news coverage, analyst coverage, stock return volatility, and better performance. Since we standardize the dependent and independent variables, the results suggest that news coverage has a first-degree influence on Twitter coverage. We also repeat the regressions for local and nonlocal Twitter coverage respectively. While both local coverage and nonlocal coverage are strongly affected by news, local coverage is less sensitive to the other firm characteristics or stock market metrics than is nonlocal coverage.

3.2. Stock Return Predictability of Sample Twitter Users

We examine local advantage by estimating the following daily panel regression:

$$CAR\ [t,\ t+k]_i = a_1 Local\_Eval_{it} + a_2 NonLocal\_Eval_{it} + \sum \beta_j\ AR_{it\text{-}j} + \sum \gamma_i D_i + \varepsilon_{it}, \qquad (4)$$

where $CAR\ [t,\ t+k]_i$ is cumulative abnormal returns of firm i from day t to $t+k$. For our tests, we examine abnormal returns in the two- (k=1), five- (k=4), ten- (k=9) and twenty-day (k=19) windows. We follow the literature (e.g., Fama, Fisher, Jensen, and Roll 1969) to calculate daily abnormal return as residuals from the four-factor model. For each firm i on day t, we calculate abnormal returns using the factor loadings for the three Fama-French (MKT, SMB, HML) factors and the momentum factor (UMD) estimated the daily four-factor model in the [t-150, t-31] window.[20] The abnormal return AR_{it} therefore captures price response to the new information arriving on day t.

The independent variable $Local\_Eval_{it}$ is the aggregate evaluations of local Twitter users for firm i over the two-week period prior to day t. Specifically, we first assign the scores of either -1 (negative), 0 (neutral), or 1 (positive) to each local Twitter post about firm i in the two weeks prior to day t using the ME classification techniques described in Section 2.5, and then sum up the scores. We assign zero to the evaluation measure if a firm is not covered by any local Twitter post in the two-week period. Similarly, $NonLocal\_Eval_{it}$ is the aggregate evaluations of non-local Twitter users for firm i over the two-week period prior to day t. To ease the assessment of economic significance, we standardize the local and nonlocal evaluations. The coefficients α_1 and α_2 indicate the stock return predictability of local and nonlocal investors. If local investors are better at predicting returns than nonlocal investors, then we expect that $(\alpha_1 - \alpha_2) > 0$.

We further include firm fixed effects (D_i) to control for firm-specific characteristics, and ten lags of daily returns ($AR_{it\text{-}j}$) to control for short-term return reversals and microstructure effects. We

[20] We require at least 30 daily return observations in the estimation window.

calculate t-statistics using Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations.[21]

Before the examination of local advantage, we first examine the overall stock return predictability of sample Twitter users. Specifically, we estimate the following daily panel regression:

$$CAR\,[t,\, t+k]_i = a_0 Eval_{it} + \sum \beta_j AR_{it\text{-}j} + \sum \gamma_i D_i + \varepsilon_{it} \tag{5}$$

This regression is similar to equation (4) except that the independent variable is the sum of local and non-local evaluations in equation (4). The coefficients α_0 measures the overall stock return predictability of our sample Twitter users. This test offers an investigation of the informedness of Twitter posts about stock returns.

Panel A of Table 2 presents the results of the return regressions in equation (5). Notably, we observe that the coefficients on investor evaluations are significantly negative for all of the return windows. For example, in the model of five-day returns, the estimated coefficient on investor evaluation is -0.102 (t-stat -5.27), indicating that a one standard deviation increase in the evaluation measure is associated with a 10.2 basis point *decrease* in subsequent weekly returns. In Panel B, we control for microstructure effects by skipping a week before return measurements and find similar results. Panel C repeats the regressions but with evaluations in the one-week period prior to return measurement instead of two-week period, and the negative return predictability persists for all return windows.

Table 2 paints a pessimistic picture about the investing ability of individual investors. This finding is consistent with previous studies which suggest that individual investors generally lose significantly from their trading (Barber and Odean 2000; Barber, Odean, and Zhu 2009), and that when they make trading profits they do so by providing liquidity (Kaniel, Saar, and Titman 2008). This

[21] The Driscoll-Kraay standard errors are similar in spirit to the Newey-West standard errors but corrects both time-series and cross-sectional correlations in the panel regression setting.

result is in line with Antweiler and Frank (2004) who find that an early sample of messages from Yahoo! Finance message boards have little stock return predictability. This finding also illustrates that one may lose money in the stock market by simply following the opinions of Twitter posts. In the next subsection, we demonstrate that despite the unimpressive average returns, there are value-relevant differences between local and nonlocal investors' information sets.

3.3. Do Individual Investors Have Local Advantage?

In this section, we examine whether local Twitter users are more informed of future stock returns than their nonlocal peers. We estimate the regressions in equation (4) and report the results in the Panel A of Table 3. Interestingly, the significantly negative return predictability reported in Table 2 remains for nonlocal investors but disappears for local investors. For example, in the model of five-day returns, the estimated coefficient is -0.202 (t-stat -6.15) for nonlocal evaluations but -0.021 (t-stat -0.89) for local evaluations. The local advantage (α_1- α_2) is 0.180 (t-stat 4.17), suggesting that a one standard deviation increase in local evaluation relative to nonlocal evaluation predicts an 18.0 basis point increase in weekly stock returns. This local advantage is both economically and statistically significant. Additionally, local advantage is large and significant for the other return windows of two, ten, and twenty days. For robustness, we further repeat the regressions with one-week evaluation period instead of a two-week period in Panel B of Table 3 and observe similar results. The results in Table 3 provide strong evidence of local advantage among the individual investors in our sample.

3.4. Robustness Tests

We conduct various robustness tests on the local advantage results in Table 3. First, we examine whether the results on local advantage are sensitive to the alternative classification criteria of local posts. In Panel A of Table 4, we classify local and nonlocal posts according to whether the distance between Twitter users and corporate headquarters are within 50 miles. In Panel B of Table 4, we classify local and nonlocal posts according to whether the Twitter users are in the same states as the firms' headquarters. Our finding of local advantage persists with the local classification based on both alternative criteria. For example, when we use the 50-mile criterion in Panel A, in the model of five-day window the estimated local advantage (α_1- α_2) is 0.191 (t-stat 4.47), slightly larger than the 0.180 in Panel A of Table 3.

When no Twitter post covers a firm during the two-week evaluation period, we do not drop the observation but treat it as a neutral evaluation by assigning zero to the evaluation measure. For a robustness test, we repeat the regression analysis but include only the firm-days with at least one Twitter post in the evaluation period. We present the results in Panel C of Table 4, which shows that local advantage persists for all windows of abnormal returns. For example, in the model of five-day abnormal returns the estimated local advantage (α_1- α_2) is 0.206 (t-stat 3.39), slightly larger than the 0.180 in Panel A of Table 3.

Twitter posts could contain noise. For example, a user who had an unpleasant experience at the local Wal-Mart store may post a message to recommend selling Wal-Mart's stocks. If the noise is randomly distributed, then it will bias against us finding local advantage. We nevertheless address this concern by a robustness test that focuses on a sub-group of sophisticated users in Stocktwits.com's list of "recommended" contributors. Although the website does not use a quantitative rule for selecting the recommended contributors, in general, the recommended contributors have a large following, a long track record, post meaningful or interesting comments, and are influential within the social network. There are 101 recommended users in our sample. In Panel D of Table 4, we report the regression results using the posts from recommended users and find that local advantage persists. For example, in the model of five-day abnormal returns the estimated local advantage (α_1- α_2) is 0.129 percent, both economically and statistically significant (t-stat 4.14).

Seasholes and Zhu (2010) point out the importance of controlling for geographical return factors in the examination of local advantage. For example, if both sample firms and sample investors cluster in certain areas (e.g., New England or the Bay area), and if stocks of firms in these areas happen to perform well during the sample period, then one can observe a mechanically positive relation between local investment and stock performance. Although this concern is alleviated by the firm fixed effects in our regressions, we nevertheless construct state-adjusted return for a firm-day by subtracting the average daily returns of all firms located in the same state. Table 4 Panel E presents the regression analyses with state-adjusted abnormal returns, and Table 4 Panel F presents the analyses using state-adjusted raw returns. The local advantage is both statistically and economically significant in all models.

3.5. Classifying Information in Twitter Posts using the Bayesian Approach

To corroborate the results using the Maximum Entropy (ME) approach, we repeat the analysis but classify information in the Twitter posts using a Naïve Bayesian (NB) approach rather than the ME approach. The existing literature (Li 2010) shows that the NB approach is superior to dictionary and word count methods for predicting the sentiment of forward-looking statements in corporate filings. Similar to ME estimation, the NB classifier is a maximum likelihood application of Bayes' rule to a set of document features that makes the simplifying assumption that all features are independent. In the field of machine learning, NB is a popular technique to use as a baseline approach for document classification or spam filtering. We follow a similar NB approach to Li (2010) to classify information.[22]

In practice, the NB approach is similar to the ME approach in that it relies on a training set of 2,000 hand-classified posts to determine the probability that each word reflects a positive, negative or

[22]The only difference is that Li (2010) uses four categories (positive, negative, neutral, and uncertain) whereas we maintain the three categories (positive, negative, neutral) to be consistent with our main tests Li's uncertain category refers to specific words in his dictionary that referred to uncertainty, rather than being uncertain as to how to classify the document.

neutral sentiment. Each word is a feature used to classify any document in the full data set. The classification method uses the same equations: (2) and (3). Practically, NB is a constrained ME technique; the constraint being that the NB algorithm can only use single words, and not word combinations, to classify sentiment. The robustness test using the NB approach allows us to relate to the classification approach used in the existing literature and provides a strict test of the ME approach used in this paper. Specifically, if the ME approach successfully identifies the information in the sample Twitter posts, then the unconstrained ME approach should perform at least as well as the constrained NB approach.

After classifying the information in sample tweets, we repeat the regression analysis in Table 3 but with the investor evaluation based on the NB approach. In Panel A of Table 5, the local advantage is robust to the use of NB approach. For example, in the model of five-day return, the local advantage is 12.7 basis point (t-stat 3.23) per week. As it should be given the constrained nature of the NB approach, this local advantage is smaller than that using the ME approach, but remains both statistically and economically significant. Panel B of Table 5 presents the tests using one-week evaluation, and the results are similar to Panel A. Overall, Table 5 shows that our finding of local advantage is robust to the alternative approach of classifying information. Additionally, the smaller local advantage using the NB approach suggests that using the ME approach, which controls for the conditional dependence of words, is a methodological improvement relative to the literature standard.

3.6. Can Followers' Responses Explain the Observed Local Advantage?

We further examine whether the results on local advantage are due to the price impact of local posts. Specifically, followers of a Twitter user may buy (sell) after reading the users' positive (negative) evaluation, causing a positive relation between the evaluation and subsequent stock returns. If followers expect local posts to contain more reliable information than nonlocal posts and therefore respond more strongly to local posts than nonlocal posts, then one will observe that local posts predict returns better than nonlocal posts. Our findings on local advantage are not likely to be driven by price impact. While price impact is temporary, our results on local advantage persist for the return window up to one month (twenty trading days). We nevertheless conduct three robustness tests to investigate this explanation.

First, we skip one week between the measurement of evaluations and returns. Investor responses to the posts should concentrate in the week after the posts, so if the results on local advantage are caused by investor responses, then local advantage should be significantly reduced in the skip-a-week setting. Panel A of Table 6 presents the results of the skip-a-week regressions, which show that the magnitude of local advantage in all models is similar to those in Table 3. For example, the skip-a-week local advantage in terms of five-day abnormal returns is 0.127 percent (t-stat 2.75), both statistically and economically significant.

Second, we construct weighted evaluation measures that assign larger weights to the Twitter users with more followers. Specifically, we first multiply the evaluation score of each post by the number of followers of the Twitter user, and then sum up the weighted scores for local and nonlocal posts. If the result on local advantage is caused by stronger investor responses to local posts, then we should observe a stronger local advantage with the weighted evaluation measure. Panel B of Table 6 presents the results of the regressions with the weighted evaluation measures. The local advantage using the weighted measure (0.107, t-stat 3.19) is smaller than that in Table 3 (0.180). The fact that the weighted measures do not lead to a larger local advantage suggests that price impact, if at work, is not driving the local advantage.

Finally, we examine the corresponding trading volume in the return windows. If our finding on local advantage is caused by followers' stronger response to local posts than nonlocal posts, then we expect to observe a greater increase in trading volume for local posts than nonlocal posts. Panel C of Table 6 repeats the regressions in Table 3 but the dependent variable is cumulative abnormal turnover. We calculate daily turnover as daily trading volume scaled by total shares outstanding, and then follow the literature (e.g., Tkac 1999; Gebhardt, Lee, and Swaminathan 2001) to control for firm-specific and market-wide factors that affect volume. Specifically, we first calculate daily excess turnover by subtracting market turnover of the CRSP universe, and then obtain abnormal turnover for a firm-day by subtracting the firm's average daily excess turnover in the previous 180-day rolling window. Panel C shows that while the coefficient on local evaluation is insignificant, that on nonlocal evaluation is significantly positive. The volume reaction to non-local posts is significantly larger than the volume reaction to local posts. We further estimate regressions of abnormal turnover on the *absolute value* of evaluation since both positive and negative evaluation could trigger abnormal volume. The unreported results show that in this setting, the volume reaction to non-local posts is also significantly larger than the volume reaction to local posts. Overall, the results on trading volume suggest that our findings of local advantage are not driven by the stronger response of followers to local posts.

4. Is Local Advantage Private Information?

In this section, we perform a number of cross-sectional analyses to investigate if local advantage is due to individual investors acquiring private information about local firms ("information hypothesis").

4.1. The Effect of Public News Coverage on Local Advantage

If local advantage is caused by investors' access to private information about local firms, then local advantage will be stronger among firms with no public news coverage. We therefore set out to examine the effect of public news coverage on local advantage. To ensure the reliability of news sources, we collect news articles from the three major news wires including Reuters News, Dow Jones News Wire, and PR News Wire using Factiva. Section A.3 of the Appendix provides details about the collection of news stories. We collect 254,600 news articles that cover 1,000 of our 1,015 sample firms during the two-year sample period. Since our sample is comprised of relatively large firms, this coverage is consistent with Fang and Peress (2009) who examine a sample of large firms (NYSE stocks plus 500 randomly selected NASDAQ stocks) and find that annual news coverage by the four nationwide newspapers ranges between 57% and 77% during 1993 to 2002.

For each day of our sample period, we sort firms into two groups based on whether the firms have public news coverage in the previous two weeks (the measurement window for Twitter posts). We then estimate regressions of abnormal returns for the sub-samples of firms with and without news coverage. Table 7 presents the regressions of abnormal stock returns for the no-news (Panel A) and news (Panel B) sub-samples. We observe a local advantage in the no-news sample that is much larger than the local advantage found in the full sample. For example, for the five-day window of abnormal returns in Panel A, local advantage is 55.7 basis points per week (t-stat 4.60). In contrast, Panel B shows that the corresponding local advantage is 13.3 basis points (t-stat 3.51) for firms with news coverage. Panel C further presents the difference in local advantage between no-news and news samples. The spread of local advantage in the five-day return window is a large 42.4 basis points (t-stat 3.34), both economically and statistically significant. These results suggest that local advantage is significantly larger in the firms that have no public news coverage. This finding lends strong support to the hypothesis that individual investors have access to private information about locally headquartered firms. The results of this test also support Tetlock's (2010) conclusion that public news releases can level the playing field for uninformed investors.

4.2. The Effect of Information Asymmetry on Local Advantage

If local advantage is caused by local investors' access to private information, then we would expect a positive association between local advantage and information asymmetry. In this section we examine the effect of information asymmetry on local advantage using a number of commonly used proxies proposed by the previous studies.

4.2.1. The effect of firm size on local advantage

Our first proxy of information asymmetry is firm size, which is widely used in the literature (e.g., Coval and Moskowitz 1999; Hong, Lim, and Stein 2000). Previous studies suggest that small firms have greater information asymmetry than large firms because investors, facing fixed information costs, may exert more effort to learn about large firms in which they can make larger investments (Van Nieuwerburgh and Veldkamp, 2009). Therefore, if local advantage is caused by private information then we expect local advantage to be stronger in small firms.

For each day in our sample period, we classify firms into quartiles according to their market capitalizations at the end of the previous year, and calculate local advantage based on regressions of abnormal returns as in Panel A of Table 3 for small (bottom quartile) and large firms (top quartile), respectively. Panel A of Table 8 shows that local advantage for small firms is significantly larger than that of the full sample. For example, in the five-day window of abnormal returns, local advantage for small firms is 0.744 (t-stat 3.19), much larger than that of the full sample (0.180, Panel A of Table 3). In contrast, the corresponding local advantage for large firms (0.091) is only half of the full sample. The differences in local advantage between small and large firms are statistically significant in all models. Therefore, the results of the sub-sample analysis based on firm size support the information hypothesis. Since firm size also captures many other aspects of a firm, we perform more cross-sectional analyses using two other proxies for information asymmetry.

4.2.2. The effect of analyst coverage on local advantage

Analyst coverage is another commonly used proxy for information asymmetry (e.g., Brennan and Subrahmanyam 1995; Hong, Lim, and Stein 2000; Irvine 2004). Specifically, firms followed by larger numbers of analysts tend to have lower information asymmetry. Since analyst coverage and firm size are strongly correlated, we construct size-adjusted analyst coverage as the residual from cross-sectional regressions of analyst coverage on firm size. For each day in our sample period, we sort firms into quartiles according to their size-adjusted analyst coverage for the month, and examine local advantage for low coverage firms (bottom quartile of coverage) and high coverage firms (top quartile of coverage), respectively.

In Panel B of Table 8, we observe that local advantage for low coverage firms is significant in all return windows and much larger than that of the full sample (Panel A of Table 3). In contrast, local advantage is much smaller for the high coverage firms. The spread in local advantage between low and high coverage firms are also quite large and statistically significant. Thus, the analyst coverage results are consistent with the information hypothesis.

4.2.3. The effect of idiosyncratic volatility on local advantage

We also use idiosyncratic stock return volatility as a proxy for information asymmetry. A number of studies suggest that higher idiosyncratic volatility indicates a larger amount of firm-specific information not shared by the market, and therefore, greater information asymmetry (e.g., Bhagat, Marr, and Thomson 1985; Blackwell, Marr, and Spivey 1990; Krishnaswami and Subramaniam 1999; Zhang 2006). For each day of our sample period, we sort firms into quartiles based on idiosyncratic volatility, and examine local advantage among high volatility firms (top quartile of volatility) and low volatility firms (bottom quartile of volatility), respectively.

Panel C of Table 8 presents the results, which show that local advantage among high volatility firms is strong for all return windows examined. For example, the local advantage is 25.2 basis points (t-stat 2.76) in the five-day return window, both economically and statistically significant. On the contrary, local advantage among low volatility firms is only 0.6 basis points (t-stat 0.23). The differences in local advantage between high and low volatility groups are large and statistically significant. Therefore, the evidence from idiosyncratic volatility is also consistent with the information hypothesis.

To summarize, our results using the proxies for information asymmetry generally present a positive association between information asymmetry and local advantage. These results suggest that individual investors' local advantage is due to their access to value-relevant private information about local firms.

4.3. Local Information Advantage and Earnings Predictability

Our results thus far suggest that individual investors' local advantage is associated with information acquisition. We therefore examine whether local investors have better information about firm fundamentals. Specifically, we examine if local evaluation prior to earnings announcement can better predict earnings surprise than nonlocal evaluation. To measure earnings surprise, we construct the measure of standardized unexpected earnings (*SUE*) as below:

$$SUE = \frac{Actual - Expected}{P} \tag{7}$$

where *Actual* is actual earnings, *Expected* is the median analyst forecasts prior to earnings announcements, and *P* is the stock price at the end of the fiscal quarter. To control for outliers, we winsorize *SUE* at the 1 percent and 99 percent cutoffs.

Table 9 presents the cross-sectional regression of *SUE* on local and nonlocal evaluations for the 6,889 earnings announcements of our sample firms with available data. The local and nonlocal evaluations are standardized to facilitate the comparison of economic significance. The coefficient on local evaluation is positive and marginally insignificant, while that on nonlocal evaluation is significantly negative. The difference between local and nonlocal is statistically significant (t-stat 3.55). This result is consistent with our findings that individual investors are generally misinformed but local information advantage helps offset their disadvantage.

4.4. Performances of Long-Short Trading Strategies based on Local Advantage

In this subsection, we examine the profitability of zero-investment trading strategies based on local advantage. This examination is not only of interest to practitioners, but also helps verify the validity of our local advantage findings. On each day of our sample period, we form two portfolios based on the contrasts between non-local evaluations and local evaluations. The first portfolio, "locally favorable portfolio", contains firms for which the difference between local and nonlocal evaluations in the past two weeks is greater than or equal to zero. The second portfolio, "locally unfavorable portfolio", contains firms for which the difference between local and nonlocal evaluations in the past two weeks is less than zero. We then hold the two portfolios for J days, where J=2, 5, 10, or 20. This strategy is similar to the rolling momentum strategy proposed by Jegadeesh and Titman (1993) except that we form portfolios based on contrasting evaluations rather than momentum.

Table 10 reports the average daily abnormal profits of the zero-investment strategies that go long the locally favorable portfolio and go short the stocks with locally unfavorable portfolio. Specifically, we calculate for each day the difference in average abnormal returns between the two portfolios ("locally favorable portfolio" – "locally unfavorable portfolio") and then report the time-series means. Daily abnormal returns are constructed based on the four-factor model as defined in Section 3. To control for time-series correlations, we report t-statistics using Newey-West robust standard errors with 10 lags.

We observe in Table 10 that the daily abnormal profits range from 5.9 basis points to 8.5 basis points, and are statistically significant for all windows. These results provide strong evidence for the existence of local advantage. Since Section 4.2 shows that local advantage is associated with firm characteristics including firm size, analyst coverage, and idiosyncratic volatility, in an attempt to improve the performance of the local-advantage-based strategy, we further exclude large firms, high analyst coverage firms, or low idiosyncratic volatility firms (their classifications are defined in Section 4.2). Table 10 shows that the daily abnormal profits increase in these sub-samples. For example, excluding the large firms (bottom tercile of firm size), the daily abnormal profits range from 8.8 basis points to 12.5 basis points across different holding windows. These results are also consistent with our previous finding that local advantage is increasing in information asymmetry, suggesting that local advantage is likely caused by investors' access to private information about locally-headquartered companies.

4.5. Potential Mechanisms of Individual Underperformance and Local Advantage

Our results show that individual investors on average have a negative stock return predictability, which is consistent with a large literature documenting that individuals suffer various behavioral biases (Hirshleifer 2001) and lose significantly from trading. Additionally, our results suggest that local advantage helps correct the mistakes by individuals. In this subsection, we examine two potential mechanisms that can contribute to the underperformance of individual investors in our setting and how local information advantage helps alleviate such underperformance.

First, we examine the possibility that local advantage is associated with return reversal. It is possible, for example, that nonlocal investors chase contemporaneous stock return and then earns a negative return due to short-term return reversal. It is also possible that nonlocal investors represent the sentiment of general investors which may temporarily move stock prices up or down followed by a reversal. Furthermore, it is possible that nonlocal investors use the "pump and dump" strategy to manipulate stock price, and the subsequent return reversal makes their evaluation negatively predict future returns. It is worth noting that we control for lagged returns contemporaneous to investor evaluation in all the regressions, so return reversal is unlikely to explain our findings. We nevertheless evaluate the effect of return reversal on local advantage by repeating the regressions without controlling for lagged returns (short-term reversal) in Panel A of Table 11. If the observed local advantage is accounted for by return reversal, then we will observe much stronger local advantage without controlling for return reversal. Panel A shows that, inconsistent with the return reversal explanation, the local advantage remains almost the same without controlling for return reversal. In Panel B of Table 11, we control for lagged returns in the previous one-month window instead of that in the previous two-week window, and the local advantage remains almost the same, suggesting that our results on local advantage is largely irrelevant to the control of return reversal. [23]

Next, we explore a second potential mechanism based on the existing literature. Previous studies find evidence that retail investors fail to correct for the complexity of analysts' incentives and tend to slavishly trade in the direction of the recommendation. For example, Malmendier and Shanthikumar (2007, 2009) and Mikhail, Walther, and Willis (2007) sort investors by trade size and find that small traders tend to overreact to analysts' recommendations, especially buy recommendations. These authors find that, as a result, small investors experience significant underperformance compared to large traders. We therefore hypothesize that investors' overreaction to analyst opinion is, at least in part, responsible for nonlocal investor underperformance. As a corollary to this hypothesis, we predict that the private information possessed by local investors helps them overcome this bias.

To test our hypothesis, we examine the responses of local and nonlocal investors to overoptimistic analyst forecasts and analyst recommendations. We follow the literature (e.g., Bradshaw, Richardson, and Sloan 2006) and construct monthly measure of analyst optimism as mean earnings forecast minus the corresponding actual earnings, scaled by stock price at the summary date. Both mean forecasts and actual earnings are obtained from the IBES monthly summary file.[24] We then examine local and nonlocal investors' responses to analyst optimism in a regression setting.

---

[23] To further evaluate the possibility that the observed local advantage is caused by nonlocals using "pump and dump" strategy, we examine penny stocks priced below $2 because "pump and dump" can be more profitable for penny stocks. Inconsistent with the "pump and dump" explanation, we find no evidence that local advantage is greater in penny stocks than our sample of non-penny-stocks.

[24] We further adjust analyst optimism by controlling for the average of other firms in the same two-digit SIC industry to control for any industry effects.

Panel A of Table 12 reports panel regressions of monthly local or nonlocal investors' evaluations on the analyst optimism measure in the previous month for 988 firms with available data in the sample period. The independent variable is the sum of evaluations of all Twitter posts from local or nonlocal investors for a firm-month. We standardize both the dependent and the independent variables to facilitate the comparison of economic significance. We observe that the coefficient on the local evaluation is an insignificant -0.009 (t-stat -1.27) but the coefficient on the nonlocal evaluation is a significantly positive 0.027 (t-stat 2.22). The difference in the two coefficients is statistically significant at the 0.05 level. These results suggest that nonlocal investors respond much more aggressively to overoptimistic analyst forecasts than do local investors.

We also examine whether nonlocal investors react more strongly to analyst recommendations than local investors do. We first obtain monthly median analyst recommendation from the IBES summary file, where an individual analyst recommendation takes the values of 1 (strong buy), 2 (buy), 3 (hold), 4 (sell), or 5 (strong sell). We then construct a binary variable "sell recommendation" ("buy recommendation") that equals 1 if the consensus recommendation is higher (lower) than 3, and 0 otherwise.

Panel B of Table 12 presents panel regressions of monthly local or nonlocal evaluations on lagged monthly buy and sell recommendations. We also standardize the evaluation variables to facilitate the comparison of coefficients. The coefficients on sell recommendations are significantly negative for both local evaluations (-0.116, t-stat -3.37) and nonlocal evaluations (-0.104, t-stat -6.67), indicating that both local and nonlocal investors tweet negatively about firms with sell recommendations. Although the coefficient for nonlocal investors is slightly larger, the difference is not statistically significant. For buy recommendations, the coefficient is insignificantly positive for local evaluation (0.021, t-stat 1.13) but significantly positive for nonlocal evaluation (0.071, t-stat 3.04).

The significant difference in the coefficients (0.050, t-stat 1.68) indicates that nonlocal investors respond more positively to buy recommendations than local investors.

We also estimate panel regressions of monthly cumulative abnormal returns (CAR) in the month t+1 or t+2 on the analyst optimism measure in month t and analyst recommendations in month t, respectively. The unreported results show that analyst optimism and buy recommendations both negatively predict subsequent abnormal returns, which indicates that individual investors' overreaction to analyst optimism and buy recommendation cause underperformance and that local advantage helps alleviate such underperformance. The evidence from this experiment therefore is supportive of the view that locally-obtained information advantage helps investors overcome their behavioral biases.

5. Conclusion

This paper investigates the local advantage of individual investors using a unique dataset of Twitter posts that cover publicly traded U.S. companies. While previous studies on individual investors' local advantage focus on the abnormal returns on investors' local investments, we directly examine individual investors' information about local and nonlocal companies.

We first examine the overall stock-picking ability of the sample Twitter users. Our examination focuses on investors' information rather than their trading, which offers a novel test on the informedness of individual investors because the trading decisions of investors, particularly individual investors, can be affected by various factors such as behavioral biases, trading constraints, diversification, and liquidity provision. We find that these individual investors exhibit significantly negative stock return predictability. We then contrast the stock return predictability between local and nonlocal investors and observe a large and significant local advantage. For example, when we examine weekly returns subsequent to investor evaluations, local advantage is 18 basis points per week, both economically and statistically significant. Further analyses show that local advantage is much larger in firms without public news coverage, and firms with severe information asymmetry. These results indicate that local advantage is due to individual investors' access to private information about local firms.

To examine a potential source of local advantage, we hypothesize that locally-obtained information helps reduce the behavioral biases of individual investors that harms their investment performance. Our results show that nonlocal investors exhibit stronger overreaction to analysts' forecasts and recommendations than do local investors. This overreaction can explain a significant fraction of the nonlocal underperformance we document. This experiment is the first into the source of local outperformance that we are aware of and we conclude that indeed, local advantage can improve performance by reducing investors' behavioral biases.

We contribute to the debate on whether local advantage exists for individual investors. While the recent work by Seasholes and Zhu (2010) finds little evidence that individual investors earn abnormal returns on their local investments, we directly examine investors' information set and document a significant local advantage. Together with Seasholes and Zhu, our results suggest the possibility that individual investors may fail to convert their value-relevant information about local firms into trading profits.

Our findings also have interesting implications for the rapidly growing internet communication about financial markets. Many people perceive that internet messages on the stock markets simply contain noise or reflect investor sentiment that is unrelated to firm fundamentals. We find that, indeed, the Twitter posts on average have a large negative return predictability. However, we also observe that local posts significantly outperform nonlocal posts and such advantage seems to result from contributors possessing private information about local firms. This finding suggests that internet communication about the financial markets contains value-relevant information as opposed to just noise.

References

Agarwal, Sumit, and Robert Hauswald, 2010, Distance and private information in lending, *Review of Financial Studies* 23, 2757-2788.

Almazan, Andres, Adolfo De Motta, Sheridan Titman, and Vahap Uysal, 2010, Financial structure, acquisition opportunities, and firm locations, *Journal of Finance* 65, 529-563.

Anand, Amber, Vladimir Gatchev, Leonardo Madureira, Chrito Pirinsky, and Shane Underwood, 2011, Geographic proximity and price discovery: Evidence from the Nasdaq, *Journal of Financial Markets,* 14, 193-226.

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.

Ayers, Benjamin, Santhosh Ramalingegowda, and Eric Yeung, 2011, Hometown advantage: The effects of monitoring institution location on financial reporting discretion, *Journal of Accounting and Economics* 52, 41-61.

Bae, Kee-Hong, Rene M. Stulz, and Hongping Tan, 2008, Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts, *Journal of Financial Economics* 88, 581-606.

Baik, Kang, 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, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773-806.

Barber, Brad M., Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean, 2009, Just how much do investors lose from trade? *Review of Financial Studies* 22, 609-632.

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.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets? *Review of Financial Studies* 22, 151-186.

Bernile, Gennaro, Alok Kumar, and Johan Sulaeman, 2013, Home away from home: Economic relevance and local investors, Working Paper, University of Miami.

Bhagat, Sanjai, Wayne M. Marr, and Rodney Thomson, 1985, The Rule 415 experiment: Equity markets, *Journal of Finance* 85, 1385-1401.

Blackwell, David W., Wayne M. Marr, and Michael F. Spivey, 1990, Shelf registration and the reduced due diligence argument: Implications of the underwriter certification and the implicit insurance hypotheses, *Journal of Financial & Quantitative Analysis* 25, 245-259.

Bradshaw, Mark T., Scott A. Richardson, and Richard G. Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics* 42, 53–85.

Brennan, Michael J., and Avanidhar Subrahmanyam, 1995, Investment analysis and price formation in securities markets, *Journal of Financial Economics* 38, 361-381.

Butler, Alexander W., 2008, Distance still matters: Evidence from municipal bond underwriting, *Review of Financial Studies* 21, 763-784.

Cao, Henry H., Joshua D. Coval, and David Hirshleifer, 2002, Sidelined investors, trade-generated news, and security returns, *Review of Financial Studies* 15, 615-648.

Chen, Hailiang, Prabuddha De, Yu 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.

Colla, Paolo, and Antonio Mele, 2010, Information linkages and correlated trading, *Review of Financial Studies* 23, 203-246.

Coval, Joshua D., and Tobias J. Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance* 54, 2045-2073.

Coval, Joshua D., and Tobias J. Moskowitz, 2001, The geography of investment: informed trading and asset prices, *Journal of Political Economy* 109, 811-41.

Das, Sanjiv, Asis Martinez-Jerez, and Peter Tufano, 2005, eInformation: A clinical study of investor discussion and sentiment, *Financial Management* 34, 103-137.

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

Driscoll, John, and Aart Kraay, 1998, Consistent covariance matrix estimation with spatially dependent panel data, *Review of Economics and Statistics* 80, 549-560

Duflo, Esther, and Emmanuel Saez, 2002, Participation and investment decisions in a retirement plan: The influence of colleagues' choices, *Journal of Public Economics*, 85,121-148.

Engelberg, Joseph, 2008, Costly information processing: Evidence from earnings announcements, Working paper, University of North Carolina.

Gaspar, Jose-Miguel, and Massimo Massa, 2007, Local ownership as private information: Evidence on the monitoring-liquidity trade-off, *Journal of Financial Economics* 83, 751-792.

Fama, Eugene F., Lawrence Fisher, Michael C. Jensen, and Richard Roll, 1969, The adjustment of stock prices to new information. *International Economic Review* 10, 1-21.

Fang, Lily H., and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64, 2023-2052.

Gebhardt, William R., Charles M. C. Lee, and Bhaskaran Swaminathan, 2001, Toward an implied cost of capital, *Journal of Accounting Research* 39, 135-176.

Griffin, John M., and Michael Lemmon, 2002, Book-to-market equity, distress risk, and stock returns, *Journal of Finance* 57, 2317-2336.

Griffin, John M., Tao Shu and Selim Topaloglu, 2012, Examining the dark side of financial markets: Do institutions trade on information from investment banks connections? *Review of Financial Studies* 25, 2155-2188.

Grinblatt, Mark, and Bing Han, 2005, Prospect theory, mental accounting, and momentum, *Journal of Financial Economics* 78, 311-339.

Hirshleifer, David, Sonya Lim, and Siew Hong Teoh, 2011, Limited investor attention and stock market misreactions to accounting information, *Review of Asset Pricing Studies* 1, 35-73.

Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533-1597.

Hong, Dong, Harrison Hong, and Andrei Ungureanu, 2011, An epidemiological approach to opinion and asset price-volume dynamics, Working Paper, Princeton University.

Hong, Harrison, Jeffery Kubik, and Jeremy Stein, 2004, Social interactions and stock market participation, *Journal of Finance* 59, 137 – 163.

Hong, Harrison, Jeffery Kubik, and Jeremy Stein, 2005, The neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *Journal of Finance* 60, 2801 – 2824.

Hong, Harrison, Terence Lim, and Jeremy C. Stein, 2000, Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies, *Journal of Finance* 55, 265-295

Irvine, Paul J., 2004, Analysts' forecasts and brokerage-firm trading, *Accounting Review* 79, 125-149.

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.

Ivković, Zoran, and Scott Weisbenner, 2007, Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices, *Review of Financial Studies* 20, 1327-1357.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-91.

Jegadeesh, Narasimhan and Di Wu, 2013. Word power: A new approach for content analysis, *Journal of Financial Economics* 110, 712-729.

Kaniel, Ron, Gideon Saar, and Sheridan Titman, 2008, Individual investor trading and stock returns, *Journal of Finance* 63, 273-310.

Kaniel, Ron, Gideon Saar, Sheridan Titman, and Shuming Liu, 2012, Individual investor trading and return patterns around earnings announcements, *Journal of Finance* 67, 639-680.

Kelley, Eric, and Paul Tetlock, 2013, How wise are crowds? Insights from retail orders and stock returns, *Journal of Finance* 68, 1229-1265.

Krishnaswami, Sudha, and Venkat Subramaniam, 1999, Information asymmetry, valuation, and the corporate spin-off decision, *Journal of Financial Economics* 53, 73-112.

Leinweber, David J., and Ananth N. Madhavan, 2001, Three hundred years of stock price manipulations, *Journal of Investing* 10, 1-10.

Li, Feng, 2010, The information content of forward-looking statements in corporate filings - a Naïve Bayesian machine learning approach, *Journal of Accounting Research* 48, 1049-1102.

Loughran, Tim, and Bill McDonald, 2011, When is liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal Finance* 66, 35-65.

Malloy, Christopher J., 2005, The geography of equity analysis, *Journal of Finance* 60, 719–755.

Malmendier, Ulrike, and Devin Shanthikumar, 2007, Are small investors naïve about incentives, *Journal of Financial Economics* 85, 457-489.

Malmendier, Ulrike, and Devin Shanthikumar, 2009, Do security analysts speak in two tongues? Working paper, UC Berkeley.

Mikhail, Michael, Beverly Walther, and Richard Willis, 2007, When security analysts talk, who listens? *The Accounting Review* 82, 1227-1253.

Odean, Terrance, 1998, Are investors reluctant to realize their losses? *Journal of Finance* 53, 1775-1798.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan, 2002, Thumbs up? Sentiment classification using machine learning techniques, Proceedings, *ACL-02 Conference on Empirical methods in natural language.*

Seasholes, Mark S., and Ning Zhu, 2010, Individual investors and local bias, *Journal of Finance* 65, 1987-2010.

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

Tetlock, Paul C., 2010, Does public financial news resolve asymmetric information? *Review of Financial Studies* 23, 3520-3557.

Tkac, Paula A., 1999, A trading volume benchmark: Theory and evidence, *Journal of Financial and Quantitative Analysis* 34, 89-114.

Tumarkin, Robert, and Robert F. Whitelaw, 2001, News or noise? Internet postings and stock prices, *Financial Analysts Journal* 57, 41-51.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2009, Information immobility and the home bias puzzle, *Journal of Finance* 64, 1187-1215.

Zhang, Frank, 2006, Information uncertainty and stock returns, *Journal of Finance* 61, 105-137.

Appendix

A1. Matching Tickers to PERMNOs

We use PERMNOs to identify sample firms to assist the merging between data sets. Since both the Stocktwits messages and the news articles are based on stock tickers, we create a linking file that assigns PERMNO to a TICKER-date during 2009 to 2011. We first download CRSP daily stock file from January 2009 to December 2011, and identify the first and last dates of each PERMNO-ticker pair. Then, for each calendar day from January 2009 to December 2011, we assign the corresponding PERMNO to a ticker as long as the day is between the first and the last days of the PERMNO-ticker pair. We then examine the resulting matches and find that while most of the PERMNO-ticker pairs are one-one matches for a given day, there are a very small number multiple matches between PERMNO and ticker on a day. We address these multiple matches as follows:

1) One PERMNO matched to two tickers: Two PERMNOs 90469 and 91501 are each matched to two tickers on some days. This is due to the change in tickers during an interim period. For example, PERMNO 90469's ticker is ARBX for most of the time during our sample period, but for the one-month interim period from June 14, 2010 to July 12, 2010, its ticker changes to ARBXD. Therefore, our procedure of using the start and end dates assigns both the tickers ARBX and ARBXD to the PERMNO 90469 for this one-month period. We address this issue by keeping only the valid tickers (ARBXD in the case of PERMNO 90469) for these two tickers in the sub-periods.
2) One ticker matched to two PERMNOs: During 2009 to 2011, there are 52 tickers each matched to two PERMNOs for either the whole period or a sub-period. We find that these cases are due to a firm issuing shares of two classes which correspond to two different PERMNOs (e.g., shares with voting power vs. shares without voting power). To address this issue, for each of these 52 tickers, we calculate the total share volume for two PERMNOs respectively during 2009 to 2011, and keep the PERMNO with the larger share volume. In most cases, the share volume of one PERMNO is much larger than the other.

A2. Collect Twitter Users' Location Information

We identify a poster's location using the following approach:

1) For each user, we first search the user's account ID on Stocktwits.com to pull up the profile page and record the user location(s).
2) If a user's Stocktwit profile page does not contain location information, we then search the account ID on Twitter.com to pull up the user profile. A small number of these users provide location information on their Twitter.com profile. Since the same account ID can correspond to different users on Stocktwits.com and Twitter.com (e.g., the account "Tony" on Twitter could be a different user than "Tony" on Stocktwits.com), we use a poster's location from Twitter.com only when we have enough evidence that the account belongs to same user as on Stocktwit.com - most often times it is the profile pictures that are the same in the Stocktwits profile and Twitter profile. Only a small number of user locations are collected using this approach.

In a rare situation, a user provides more than one locations. In this case, we include the user in our sample as long as one of the locations is in the continental U.S. Additionally, when we identify local posts in this case, a post is considered local as long as one of the user locations is local to the company discussed in the posts (within 100 miles of the corporate headquarters).

The user locations for our sample contain the state and city (county) information. We convert user locations into coordinates using *http://itouchmap.com/latlong.html*, which provides coordinates for the center of a city (county). We then use the coordinates to calculate distance between a user and a corporate headquarter according to equation (1) in the paper.

A3. Collection of News Stories

The news search is based on the tickers and firm names. For each stock tickers covered by Stocktwits, we collect the corresponding firm name (names) from the CRSP monthly stock file during the sample period. We then search the news stories from Dow Jones Newswire, Reuters News, and PR Newswire from July 10, 2009 and June 10, 2011. When we search a firm, we first enter the ticker, and then pick a name from Factiva's suggested list of firm names that matches the firm's name in CRSP. We also eliminate the duplicates of news stories for a given firm. We then matched the articles to PERMNOs using the approach described in Section A1. Overall, 96.1% of the articles are matched to PERMNOs. The unmatched articles are outside the date ranges of CRSP for the corresponding tickers. This happens because even when a firm is not traded in the exchange, it can still have news coverage. For example, General Motors (PERMNO 12079) stopped trading on June 1, 2009 and resumed trading on November 18, 2010 with a new PERMNO of 12369. GM's news articles during this interim period are therefore unmatched to a PERMNO.

Table 1: Summary Statistics

Panel A reports summary statistics for the 1,015 firms in our sample from July 11, 2009 to June 10, 2011. For a firm-day, market capitalization is measured at the end of previous year. Book-to-market ratio is book equity divided by market capitalization measured at the end of fiscal year. A firm's book-to-market ratio of fiscal year ending in calendar year *t* is matched to firm-days from July of *t*+1 to June of *t*+2. Book-to-market ratios are winsorized at 1 percent and 99 percent cutoff points. For a firm-day, analyst coverage is the number of analysts covering the firm in the previous month; idiosyncratic volatility the standard error of residuals from time-series regressions of the firm's excess returns on the market excess returns (MKT) in the one-year window ending in previous month; daily return is the daily raw return. Idiosyncratic volatilities are winsorized at the 99 percent cutoff points. We first calculate the average of firm characteristics for a firm across the firm-days, and then report the distribution of average characteristics across firms. Panel B presents the summary statistics of the sample Twitter posts, including the evaluation of a post identified using the Maximum Entropy (ME) approach, and the number of Twitter posts for a sample firm during the sample period. A Twitter user is local (nonlocal) to a firm if the user's location is less than (more than) 100 miles from the firm's headquarters. Panel C presents firm-level cross-sectional regressions of Twitter coverage on firm characteristics. The independent variable is the number of Twitter posts, local Twitter posts, or nonlocal Twitter posts for a firm during the sample period. The dependent variables include firm characteristics measured as in Panel A. *Ln(ME)* is natural log of market capitalization. *Ln(Coverage)* is natural log of 1 plus analyst coverage. # *News Articles* is the total number of news articles covering a firm during the sample period. We standardize both the dependent variables and independent variables. We also repeat the regression using alternative dependent variables including the number of local posts and the number of nonlocal posts. The regressions include a constant term which is not reported for brevity. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

Panel A: Characteristics of Sample Firms

	Mean	**STD**	**P10**	**P25**	**P50**	**P75**	**P90**
Market Capitalization ($M)	8,579	24,682	205	519	1,607	5,567	17,984
Book/Market Ratio	0.63	0.66	0.12	0.27	0.50	0.84	1.30
Analyst Coverage	11.21	7.79	1.98	5.01	9.96	16.33	21.81
Idiosyncratic Volatility	0.029	0.014	0.015	0.019	0.026	0.035	0.047
Daily Stock Return (%)	0.12	0.15	-0.03	0.04	0.11	0.18	0.28

Panel B: Summary Statistics of Sample Twitter Posts

	Mean	**STD**	**P10**	**P25**	**P50**	**P75**	**P90**
Evaluation of Posts	0.34	0.69	-1.0	0.0	0.0	1.0	1.0
Evaluation of Local Posts	0.26	0.69	-1.0	0.0	0.0	1.0	1.0
Evaluation of Nonlocal Posts	0.35	0.69	-1.0	0.0	0.0	1.0	1.0
# Posts per Firm	173.77	695.77	15.0	26.0	58.0	123.0	285.0
# Local Posts per Firm	20.27	123.40	1.0	2.0	5.0	13.0	32.0
# Nonlocal Posts per Firm	153.50	582.29	11.0	21.0	48.0	109.0	266.0

Panel C: Regressions of Twitter Coverage on Firm Characteristics

	Dependent Variable		
	#Twitter Posts	#Local Posts	#Nonlocal Posts
Ln(ME)	-0.017	-0.143***	0.010
	(-0.34)	(2.83)	(0.19)
Book-to-Market Ratio	-0.053*	-0.031	-0.056**
	(-1.88)	(-1.13)	(-1.99)

	Dependent Variable		
	#Twitter Posts	#Local Posts	#Nonlocal Posts
Ln(Analyst Coverage)	0.106***	0.068*	0.112***
	(2.65)	(1.74)	(2.78)
# News Articles	0.514***	0.595***	0.488***
	(16.16)	(18.98)	(15.25)
Idiosyncratic Volatility	0.089**	-0.000	0.106***
	(2.29)	(-0.01)	(2.72)
Ave. Daily Ret.	0.084***	0.044	0.091***
	(2.94)	(1.55)	(3.17)
Adjusted R^2	0.280	0.300	0.269
Number of Obs	1,015	1,015	1,015

Table 2: Panel Regressions of Abnormal Stock Returns on Investor Evaluations

Panel A presents panel regressions of abnormal stock returns on prior investor evaluations. The dependent variable is two-, five-, ten-, or twenty-day cumulative abnormal returns (measured in percent), respectively. To calculate daily abnormal return for a firm-day, we first estimate a Fama-French 4 Factor regression for the firm in the previous 150-day rolling window, and then use the estimated factor loadings to calculate abnormal returns for the firm-day. The independent variables include investor evaluations in the two-week windows prior to return measurements. Investor evaluation is the sum of evaluations from local and nonlocal posts in the two weeks prior to return measurement. All regressions include firm fixed effects with lagged returns in the previous ten trading days as controls. Panel B is similar to Panel A but skips one week before the return measurement. Panel C is similar to Panel A but with one-week evaluations rather than two-week evaluations. T-statistics (reported in parentheses) are calculated using Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variable			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Panel A: Regressions on Two-Week Evaluations				
Two-Week Evaluation	-0.044***	-0.102***	-0.193***	-0.391***
	(-5.49)	(-5.27)	(-5.21)	(-6.11)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel B: Regressions on Two-Week Evaluations: Skip-a-Week Returns				
Two-Week Evaluation	-0.038***	-0.091***	-0.189***	-0.377***
	(-4.59)	(-4.67)	(-5.32)	(-6.60)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,255	435,093	434,817	434,226
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel C: Regressions on One-Week Evaluations				
One-Week Evaluation	-0.043***	-0.097***	-0.172***	-0.340***
	(-5.94)	(-5.98)	(-5.89)	(-6.58)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015

Table 3: Panel Regressions of Stock Returns on Local and Nonlocal Evaluations

Panel A presents panel regressions of stock returns on prior local and nonlocal evaluations. The dependent variable is cumulative two-, five-, ten-, or twenty-day abnormal returns (measured in percent), respectively. To calculate daily abnormal return for a firm-day, we first estimate a Fama-French 4 Factor regression for the firm in the previous 150-day rolling window, and then use the estimated factor loadings to calculate abnormal returns for the firm-day. The independent variables include local evaluation and nonlocal evaluation in the two-week window prior to return measurement. To calculate local and nonlocal evaluations, we first classify Twitter posts into local and nonlocal posts according to whether the Twitter users' locations are within 100 miles of the headquarters of the firms mentioned in the posts. We use maximum entropy classification to measure the evaluation of each post, and then sum the evaluation measures of the local and nonlocal posts, respectively, in the two weeks prior to return measurement. We standardize the independent variables for each regression. For each regression, we further report the difference between the coefficients on local evaluation and nonlocal evaluation. All regressions include firm fixed effects with lagged returns in the previous ten trading days as controls. T-statistics (reported in parentheses) are calculated using Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. Panel B is similar to Panel A except that the local and nonlocal evaluations are measured in the one-week window prior to return measurements. ***, **, and * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Panel A: Regressions of Abnormal Returns				
	Dependent Variable			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Local Evaluation	-0.009	-0.021	-0.057	-0.112
	(-0.75)	(-0.89)	(-1.33)	(-1.45)
Nonlocal Evaluation	-0.088***	-0.202***	-0.362***	-0.736***
	(-5.88)	(-6.15)	(-5.62)	(-6.86)
Local – Nonlocal	0.079***	0.180***	0.305***	0.624***
	(3.65)	(4.17)	(3.76)	(4.52)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel B: Regressions of Abnormal Returns: One-Week Evaluations				
Local Evaluation	-0.016	-0.023	-0.043	-0.109*
	(-1.58)	(-1.14)	(-1.30)	(-1.83)
Nonlocal Evaluation	-0.075***	-0.187***	-0.331***	-0.624***
	(-5.57)	(-6.64)	(-6.52)	(-6.93)
Local – Nonlocal	0.059***	0.163***	0.288***	0.516***
	(3.08)	(4.36)	(4.65)	(4.62)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015

Table 4: Panel Regressions of Stock Returns: Alternative Construction of the Measure or Alternative Sample Selection

Panel A presents the regressions of abnormal returns on local and nonlocal evaluations. The definition of abnormal returns, local and nonlocal evaluations, and regression settings are similar to the Panel A of Table 3 except that we classify Twitter posts into local and nonlocal posts according to whether the Twitter users' locations are within *50 miles* of the headquarters of the firms mentioned in the posts. Panel B is similar to the Panel A of Table 3 except that we classify Twitter posts into local and nonlocal according whether the users and the company headquarters locate in the same state. Panel C presents the regressions of abnormal returns similar to the Panel A of Table 3 except that we only include firm-days that have at least one Twitter post in the two-week period of evaluation measurement. Panel D is similar to Panel A of Table 3 except that we include only the users recommended by Stocktwits.com. Panel E is similar to Panel A of Table 3 except that the dependent variables are two-, five, ten- or twenty-day cumulative state-adjusted abnormal returns (measured in percent). We calculate a firm's daily state-adjusted abnormal return as the firm's daily abnormal return minus the average daily abnormal returns of all firms in the same state as the firm. Panel F is similar to Panel E but with state-adjusted raw returns as dependent variable. Daily state- adjusted raw return for a firm is calculated as the daily raw return of the firm minus the average daily raw return of all firms in the same state as the firm. All regressions include firm fixed effects and lagged returns of the previous ten trading days as controls. T-statistics (reported in parentheses) are calculated with the Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variable			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Panel A: Local Posts Identified Using the 50-Mile Criterion				
Local Evaluation	-0.008	-0.014	-0.030	-0.061
	(-0.73)	(-0.67)	(-0.79)	(-0.88)
Nonlocal Evaluation	-0.089***	-0.206***	-0.376***	-0.764***
	(-5.87)	(-6.19)	(-5.82)	(-7.15)
Local – Nonlocal	0.081***	0.191***	0.345***	0.703***
	(3.77)	(4.47)	(4.35)	(5.35)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel B: Local Posts Identified Using the State Criterion				
Local Evaluation	-0.021*	-0.050**	-0.104***	-0.197***
	(-1.89)	(-2.06)	(-2.69)	(-2.74)
Nonlocal Evaluation	-0.083***	-0.186***	-0.327***	-0.686***
	(-5.90)	(-6.38)	(-6.24)	(-8.22)
Local – Nonlocal	0.062***	0.136***	0.222***	0.489***
	(3.12)	(3.50)	(3.69)	(5.12)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	500,631	500,424	500,079	499,360
Number of PERMNOs	1,174	1,174	1,174	1,174

	Dependent Variable			

	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Panel C: Require At Least One Post in the Evaluation Window				
Local Evaluation	-0.011	-0.023	-0.069	-0.154
	(-0.63)	(-0.68)	(-1.16)	(-1.51)
Nonlocal Evaluation	-0.106***	-0.230***	-0.384***	-0.776***
	(-4.78)	(-5.12)	(-4.45)	(-5.66)
Local – Nonlocal	0.095***	0.206***	0.315***	0.621***
	(2.84)	(3.39)	(2.93)	(3.57)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	196,311	196,237	196,131	195,939
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel D: Include Only Recommended Users				
Local Evaluation	-0.014	-0.035*	-0.074*	-0.181**
	(-1.60)	(-1.77)	(-1.93)	(-2.45)
Nonlocal Evaluation	-0.071***	-0.165***	-0.306***	-0.585***
	(-5.73)	(-6.35)	(-6.33)	(-7.48)
Local – Nonlocal	0.057***	0.129***	0.232***	0.403***
	(3.87)	(4.14)	(3.84)	(4.02)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	430,259	430,115	429,871	429,354
Number of PERMNOs	996	996	996	996
Panel E: Regressions of State-Adjusted Abnormal Returns				
Local Evaluation	-0.012	-0.028	-0.065	-0.129
	(-1.00)	(-1.11)	(-1.48)	(-1.59)
Nonlocal Evaluation	-0.101***	-0.235***	-0.427***	-0.838***
	(-6.82)	(-7.07)	(-6.80)	(-8.19)
Local – Nonlocal	0.089***	0.207***	0.361***	0.709***
	(3.94)	(4.57)	(4.69)	(5.59)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	434,606	434,444	434,170	433,593
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel F: Regressions of State-Adjusted Raw Returns				
	Dependent Variables			
	2-Day Ret.	5-Day Ret.	10-Day Ret.	20-Day Ret.
Local Evaluation	-0.008	-0.019	-0.051	-0.099
	(-0.69)	(-0.83)	(-1.29)	(-1.48)
Nonlocal Evaluation	-0.072***	-0.160***	-0.272***	-0.528***
	(-4.63)	(-4.70)	(-4.40)	(-5.82)
Local – Nonlocal	0.064***	0.141***	0.221***	0.429***
	(3.00)	(3.21)	(2.99)	(4.18)
	Dependent Variable			

	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	434,606	434,444	434,170	433,593
Number of PERMNOs	1,015	1,015	1,015	1,015

Table 5: Panel Regressions of Stock Returns on Local and Nonlocal Evaluations: Evaluations of the Twitter Posts Measured Using Bayesian Approach

Panel A presents panel regressions of stock returns on prior local and nonlocal evaluations. The dependent variable is cumulative two-, five-, ten-, or twenty-day abnormal returns (measured in percent), respectively. To calculate daily abnormal return for a firm-day, we first estimate a Fama-French 4 Factor regression for the firm in the previous 150-day rolling window, and then use the estimated factor loadings to calculate abnormal returns for the firm-day. The independent variables include local evaluation and nonlocal evaluation in the two-week window prior to return measurement. To calculate local and nonlocal evaluations, we first classify Twitter posts into local and nonlocal posts according to whether the Twitter users' locations are within 100 miles of the headquarters of the firms mentioned in the posts. We use the *Bayesian* classification approach to measure the evaluation of each post, and then sum the evaluation measures of the local and nonlocal posts, respectively, in the two weeks prior to return measurement. We standardize the independent variables for each regression. For each regression, we further report the difference between the coefficients on local evaluation and nonlocal evaluation. Panel B is similar to Panel A except that the local and nonlocal evaluations are measured in the one-week window prior to return measurements. All regressions include firm fixed effects with lagged returns in the previous ten trading days as controls. T-statistics (reported in parentheses) are calculated using Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. ***, **, and * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Panel A: Regressions of Abnormal Returns				
Local Evaluation	0.003	0.004	0.010	0.028
	(0.39)	(0.24)	(0.27)	(0.45)
Nonlocal Evaluation	-0.055***	-0.123***	-0.219***	-0.478***
	(-4.83)	(-4.43)	(-3.99)	(-5.28)
Local – Nonlocal	0.057***	0.127***	0.229***	0.507***
	(3.70)	(3.23)	(2.85)	(4.10)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel B: Regressions of Abnormal Returns: One-Week Evaluations				
Local Evaluation	-0.006	-0.005	-0.010	-0.012
	(-0.97)	(-0.33)	(-0.37)	(-0.26)
Nonlocal Evaluation	-0.045***	-0.109***	-0.188***	-0.369***
	(-5.29)	(-5.74)	(-4.89)	(-5.55)
Local – Nonlocal	0.039***	0.104***	0.178***	0.356***
	(3.65)	(4.17)	(3.35)	(4.08)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015

Table 6: Panel Regressions of Stock Returns: Examine the Price Impact Explanation

Panel A presents regressions of abnormal returns on local and nonlocal evaluations in the two-week period ending one week before the return measurements. The constructions of abnormal returns, investor evaluations, and regression settings are defined in the heading of Table 3. Panel B presents regressions of abnormal returns on the weighted evaluation measures in the two-week period before return measurement. Specifically, during the two-week period before return measurement, we multiply the evaluation of each Twitter post by the number of followers of the Twitter user, and then sum up the weighted evaluations. Panel C presents regressions of turnovers on local and nonlocal evaluations in the two-week period prior to turnover measurements, where the independent variables are two-, five-, ten-, or twenty-day cumulative abnormal turnovers. Daily turnover is a firm's daily trading volume scaled by total shares outstanding. We obtain daily excess turnover by subtracting cross-sectional average turnover of the CRSP universe, and then calculate abnormal turnover for a firm-day by subtracting average daily excess turnover of the firm in the previous 180-day rolling window. All regressions include firm fixed effects with lagged returns of the previous ten trading days as controls. T-statistics (reported in parentheses) are calculated with the Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Panel A: Regressions of Abnormal Returns: Skip One Week				
Local Evaluation	-0.011	-0.034	-0.071	-0.084
	(-1.16)	(-1.49)	(-1.59)	(-1.24)
Nonlocal Evaluation	-0.070***	-0.161***	-0.335***	-0.737***
	(-4.80)	(-4.50)	(-5.13)	(-7.75)
Local – Nonlocal	0.059***	0.127***	0.264***	0.652***
	(3.18)	(2.75)	(3.06)	(5.32)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,255	435,093	434,817	434,226
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel B: Return Regressions on Weighted Evaluation Measures				
Local Weighted Evaluation	-0.014	-0.033	-0.078*	-0.184**
	(-1.22)	(-1.26)	(-1.89)	(-2.12)
Nonlocal Weighted Evaluation	-0.058***	-0.140***	-0.270***	-0.525***
	(-4.18)	(-4.82)	(-5.12)	(-6.08)
Local – Nonlocal	0.044**	0.107***	0.191***	0.341***
	(2.44)	(3.19)	(3.54)	(4.05)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,501	435,342	435,073	434,506
Number of PERMNOs	1,015	1,015	1,015	1,015

Panel C: Regressions of Trading Volume on Local and Nonlocal evaluations				
	Dependent Variables			
	2-Day CAV	5-Day CAV	10-Day CAV	20-Day CAV
Local Evaluation	0.026	0.043	0.038	0.080
	(1.39)	(1.06)	(0.54)	(0.79)
Nonlocal Evaluation	0.226***	0.434***	0.657***	0.808***
	(5.82)	(4.88)	(3.90)	(2.91)
Local – Nonlocal	-0.200***	-0.392***	-0.618***	-0.728**
	(-4.12)	(-3.48)	(-2.91)	(-2.14)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	435,512	435,395	435,196	434,769
Number of PERMNOs	1,015	1,015	1,015	1,015

Table 7: Panel Regressions of Stock Returns: Stocks without Public News vs. Stocks with Public News

This table reports regressions of two-, five-, ten-, or twenty-day abnormal returns on local and nonlocal evaluations in the two-week period prior to return measurement for stocks with and without public news coverage, respectively. We collect news articles from PR News Wire, Dow Jones News Wire, and Reuters News and classify stocks into two groups based whether they have news coverage in the two-week period of evaluation measurement. We then estimate regressions for the non-news firms in Panel A and for the news firms in Panel B. We further report the difference in local advantage between no-news and news samples in Panel C. The regression settings and the independent variables are as defined in the heading of Table 3. All regressions include firm fixed effects with lagged returns of the previous ten trading days as controls. T-statistics (reported in parentheses) are calculated with Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Panel A: Stocks without Public News Coverage in the Period of Twitter Posts				
Local Evaluation	-0.017	-0.016	-0.176	-0.540***
	(-0.51)	(-0.21)	(-1.51)	(-3.24)
Nonlocal Evaluation	-0.243***	-0.573***	-1.015***	-1.782***
	(-8.01)	(-6.37)	(-4.96)	(-5.22)
Local - Nonlocal (1)	0.261***	0.557***	0.839***	1.242***
	(5.82)	(4.60)	(3.39)	(3.12)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	135,696	135,648	135,566	135,398
Number of PERMNOs	904	904	904	904
Panel B: Stocks with Public News Coverage in the Period of Twitter Posts				
Local Evaluation	-0.011	-0.022	-0.046	-0.073
	(-1.15)	(-1.18)	(-1.34)	(-1.12)
Nonlocal Evaluation	-0.068***	-0.155***	-0.276***	-0.589***
	(-4.43)	(-4.89)	(-4.50)	(-5.90)
Local - Nonlocal (2)	0.057**	0.133***	0.230***	0.516***
	(2.79)	(3.51)	(3.16)	(4.25)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	299,805	299,694	299,507	299,108
Number of PERMNOs	1,000	1,000	1,000	1,000
Panel C: Difference in Local Advantage: Non-News versus News Stocks				
(1) - (2)	0.204***	0.424***	0.609**	0.726*
	(4.15)	(3.34)	(2.36)	(1.74)

Table 8: Local Advantage across Proxies of Information Asymmetry

Panel A reports local advantage for small and large firms. On each day of our sample period, we sort stocks into four groups based on their market capitalizations. We then estimate regressions of abnormal returns as in the Panel A of Table 3 for small firms (lowest quartile of market capitalization) and large firms (highest quartile of market capitalization), respectively. We then report local advantage ('Local – Nonlocal' in the Panel A of Table 3) for small firms, large firms, and their differences. For Panel B, on each day of our sample period, we sort stocks into four groups based on size-adjusted analyst coverage, where size-adjusted analyst coverage is residual from cross-sectional regression of analyst coverage on size. We then report local advantage for low coverage firms (lowest quartile of coverage), high coverage firms (highest quartile of coverage), and their differences. For Panel C, on each day of our sample period, we sort stocks into four groups based on idiosyncratic volatility. Idiosyncratic volatility for a firm-day is standard deviation of the residuals from the time-series regression of daily stock returns on the market factor (MKT) in the one-year window up to the end of previous month. We then report local advantage for high volatility firms (highest quartile of volatility), low volatility firms (lowest quartile of volatility), and their differences. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Panel A: Local Advantage for Small versus Large Firms				
Small Firms	0.368***	0.744***	1.150***	1.786**
	(3.45)	(3.19)	(2.66)	(2.26)
Large Firms	0.043***	0.091**	0.137**	0.365***
	(2.36)	(2.46)	(2.03)	(2.64)
Small – Large	0.325***	0.653***	1.012**	1.421*
	(3.00)	(2.76)	(2.31)	(1.77)
Panel B: Local Advantage for Low versus High Analyst Coverage Firms				
Low Coverage Firms	0.225***	0.515***	0.839***	1.666***
	(2.96)	(3.41)	(3.26)	(4.13)
High Coverage Firms	0.050***	0.122***	0.187*	0.442**
	(2.14)	(2.37)	(1.94)	(2.48)
Low – High	0.175**	0.394**	0.652**	1.224***
	(2.20)	(2.47)	(2.37)	(2.78)
Panel C: Local Advantage for High versus Low Idiosyncratic Volatility Firms				
High Idio. Volatility Firms	0.131***	0.252***	0.395**	0.662*
	(2.97)	(2.76)	(2.19)	(1.95)
Low Idio. Volatility Firms	0.006	0.006	-0.013	0.039
	(0.51)	(0.23)	(-0.27)	(0.51)
High – Low	0.125***	0.246***	0.408**	0.623*
	(2.72)	(2.59)	(2.19)	(1.79)

Table 9: Regression of SUE on Investor Evaluations

This table presents cross-sectional regression of unexpected earnings on local and nonlocal evaluations. The sample includes 6,889 quarterly earnings announcements of sample firms that have available data to estimate unexpected earnings. The dependent variable is standardized unexpected earnings (*SUE*), calculated as the difference between the actual earnings and expected earnings (based on IBES median estimates) divided by stock price. To control for outliers, we winsorize *SUE* at the 1 and 99 percent cutoff points. The independent variables include local and nonlocal evaluations measured in the two weeks prior to the earnings announcement date. To facilitate the comparison of economic significance, I standardize the local and nonlocal evaluations. The regression also include firm fixed effects. T-statistics are reported in parentheses. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variable: SUE
Local Evaluation	0.038
	(1.52)
Nonlocal Evaluation	-0.057**
	(-2.01)
Local - Nonlocal	0.094***
	(3.55)
Firm Fixed Effects	Yes
Number of Obs	6,889
Number of PERMNOs	982
Adj. R^2	0.324

Table 10: Daily Abnormal Profits (%) of Rolling Long-Short Strategies Based on Local Advantage

This table presents daily abnormal profits (%) of rolling long-short strategies based on (local – nonlocal) evaluations. For each firm-day, we contrast the local evaluation in the previous two weeks versus nonlocal evaluation in the previous two weeks. Then on each day, we form a portfolio containing firms for which the non-local evaluation measures are lower than the local evaluations ("locally unfavorable portfolio") and a portfolio containing firms for which the non-local evaluation measures are greater than or equal to the local evaluation measures ("locally favorable portfolio"). We then hold these portfolios for J days, where J=2, 5, 10, or 20. This strategy is similar to the rolling momentum strategy proposed by Jegadeesh and Titman (1993) except that we form portfolios based on differential evaluations rather than momentum. We then calculate the daily abnormal profits of a strategy that long the "locally favorable portfolio" and short the "locally unfavorable portfolio". Specifically, we first calculate for each day the difference in average abnormal returns between the two portfolios, and then report time-series means of the daily abnormal profits. Daily abnormal return is constructed based on Fama and French 4 Factor model and is defined in the heading of Table 2. To control for time-series correlations, we calculate t-statistics (in parentheses) using Newey-West robust standard errors with 10 lags. To control for microstructure effects we follow the literature and skip one week before return measurement. We report daily abnormal profits of the long-short strategy for firms in our full sample as well as and sub-samples that exclude large firms, high analyst coverage firms, and low idiosyncratic volatility firms. The classifications of sub-samples are defined in the heading of Table 9. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Hold 2 Days	**Hold 5 Days**	**Hold 10 Days**	**Hold 20 Days**
Full Sample	0.085***	0.076***	0.065***	0.059***
	(5.62)	(6.07)	(6.86)	(6.22)
Exclude Large Firms	0.125***	0.116***	0.097***	0.088***
	(6.27)	(6.32)	(7.08)	(7.00)
Exclude High Coverage Firms	0.105***	0.141***	0.092***	0.094***
	(7.08)	(6.73)	(7.39)	(8.86)
Exclude Low Volatility Firms	0.111***	0.097***	0.085***	0.086***
	(5.22)	(5.50)	(6.58)	(7.81)

Table 11: Panel Regressions of Abnormal Return: Analyze the Explanations Based on Short-Term Return Reversal

Panel A presents the regressions of abnormal returns on local and nonlocal evaluations. The definition of abnormal returns, local and nonlocal evaluations, and regression settings are similar to the Panel A of Table 3 except that we do not control for lagged returns. Panel B is similar to the Panel A of Table 3 except that we control for lagged returns in the previous one-month window instead of the two-week window. All regressions include firm fixed effects and lagged returns of the previous ten trading days as controls. T-statistics (reported in parentheses) are calculated with the Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

Panel A: Regressions Without Controlling for Lagged Returns				
	Dependent Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Local Evaluation	-0.009	-0.020	-0.053	-0.101
	(-0.80)	(-0.85)	(-1.26)	(-1.35)
Nonlocal Evaluation	-0.080***	-0.182***	-0.322***	-0.655***
	(-5.17)	(-5.43)	(-4.99)	(-6.11)
Local – Nonlocal	0.071***	0.162***	0.269***	0.555***
	(3.18)	(3.67)	(3.28)	(4.01)
Controls of Lagged Returns	No	No	No	No
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	444,381	444,222	443,953	443,386
Number of PERMNOs	1,015	1,015	1,015	1,015
Panel B: Regressions Controlling for Monthly Abnormal Returns				
Local Evaluation	-0.010	-0.023	-0.060	-0.117
	(-0.82)	(-0.95)	(-1.42)	(-1.51)
Nonlocal Evaluation	-0.093***	-0.211***	-0.379***	-0.742***
	(-5.95)	(-6.13)	(-5.59)	(-6.73)
Local – Nonlocal	0.083***	0.188***	0.318***	0.626***
	(3.78)	(4.25)	(3.83)	(4.50)
Controls of Lagged Returns in the Previous Month	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	424,833	424,674	424,405	423,838
Number of PERMNOs	1,015	1,015	1,015	1,015

Table 12: Panel Regressions of Monthly Local or Nonlocal Evaluations on Analyst Optimism and Consensus Analyst Recommendation

Panel A reports panel regressions of monthly local or nonlocal evaluations on lagged analyst optimism measure. The dependent variable is local or nonlocal evaluation for each firm-month in the sample period. The independent variable is the analyst optimism measure in the month prior to the month of return, where the analyst optimism measure is calculated as mean analyst forecast (obtained from the IBES summary file) minus the corresponding actual earnings, scaled by stock price of the summary date. We further adjust analyst optimism of a firm by the average of other firms in the same two-digit SIC industry. We standardize the independent and dependent variables to facilitate the comparison of economic significances, and include firm fixed effects in both regressions. T-statistics (reported in parentheses) are calculated with Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations. We also report the difference in the coefficients on analyst optimism and the associated t-statistics. Panel B is similar to Panel A except that the independent variable is monthly consensus analyst recommendations prior to the return month. We first obtain monthly consensus analyst recommendation as median recommendation from the IBES summary file, where recommendation takes the values of 1 (strong buy), 2 (buy), 3(hold), 4(sell), or 5 (strong sell). "Sell recommendation" ("buy recommendation") is a binary variable that equals 1 if the consensus recommendation is higher (lower) than 3, and 0 otherwise. We standardize the dependent variables to facilitate the comparison of economic significances. ***, **, and * represent statistical significances at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependent Variables		
	Local Evaluation	Non-Local Evaluation	Non-Local - Local
Panel A: Regressions on Analyst Optimism			
Constant	0.000	0.002	
	(0.01)	(0.02)	
Analyst Optimism	-0.009	0.027**	0.037**
	(-1.27)	(2.22)	(2.56)
Firm Fixed Effects	Yes	Yes	
Number of Obs	21,344	21,344	
Number of PERMNOs	988	988	
Panel B: Regressions on Buy and Sell Recommendations			
Constant	-0.011	-0.039	
	(-0.22)	(-0.44)	
Sell Recommendation	-0.116***	-0.104***	-0.011
	(-3.47)	(-6.67)	(0.30)
Buy Recommendation	0.021	0.071***	0.050*
	(1.13)	(3.04)	(1.68)
Firm Fixed Effects	Yes	Yes	
Number of Obs	21,806	21,806	
Number of PERMNOs	994	994	

Figure 1
Example of Twitter Stream

This figure shows the interface that a Stocktwits.com user will see. Company tickers can be seen in blue after the $ hashtags.

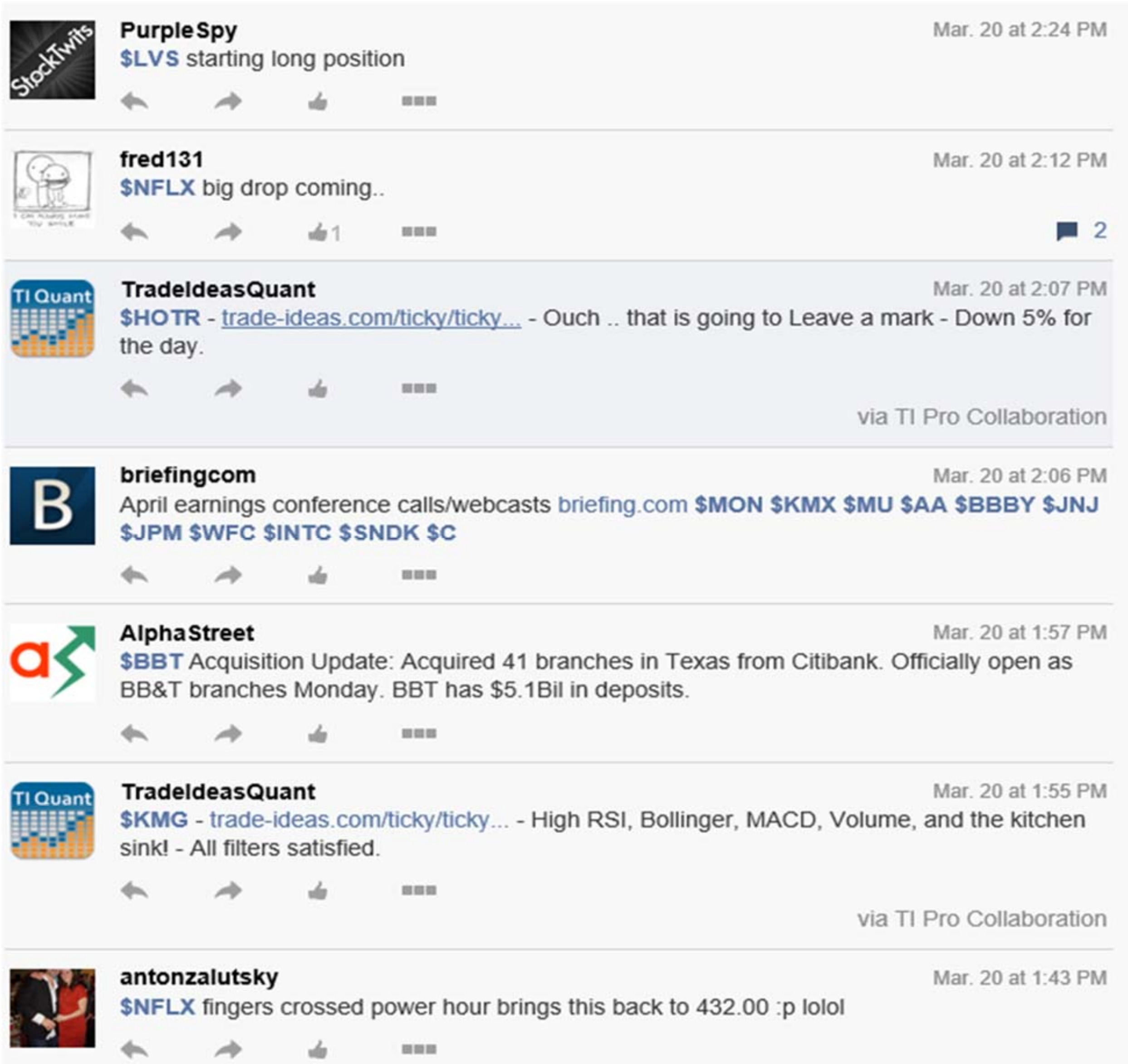

Figure 2
Geographical Distribution of Sample Twitter Users

This figure plots the geographical distribution of the Twitter users in our sample. We divide the states with six groups depending on the percentages of sample users: 0 percent; between 0 and 0.5 percent; between 0.5 and 1.5 percent; between 1.5 and 2 percent; between 2 and 3 percent; greater than 3 percent. The states with higher percentages are marked with darker colors.