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

ROBERT GIANNINI

PAUL IRVINE

TAO SHU*

November 2013

---

* Giannini is at the BlueCrest Capital Management. Irvine is the C. R. Williams professor of finance at Texas Christian University and Shu is from the Terry College of Business at the University of Georgia. We thank Stocktwits.om for providing the stock twits data. We thank Fangjian Fu, Bige Kahraman, Harold Mulherin, Jeffry Netter, Mark Seasholes, Sheridan Titman, Eric Yeung, and seminar participants at the 2012 China International Conference of Finance, the 3rd TCFA Best Paper Symposium, the 8th Financial Intermediation Research Society conference, the 2013 Asian Finance Association annual meeting, 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

November 2013

Abstract

We examine 216,266 Twitter posts from over two thousand individuals covering 1,819 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 19 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, recently 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 by individual investors, we use a large sample of Twitter posts to directly examine

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

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 216,266 Twitter posts from 2,008 Twitter users, covering 1,819 publicly traded U.S. companies from July 11, 2009 to June 10, 2011.

Our sample Twitter users are geographically dispersed, living in all of the states in the continental U.S. except North Dakota. The highest percentages of users come from the states of New York (18%), California (17%), Illinois (9%), Texas (7%), and Florida (7%), and none of the remaining states accounts for over 5 percent of users. The geographical distribution of users is consistent with the distributions of population and economic activity, and the dispersed distribution avoids the issue of geographical clustering.

We start by examining the overall stock-picking ability of our sample Twitter users. We extract the evaluations in the Twitter posts using a maximum likelihood classification approach and examine whether these evaluations predict stock returns. Our examination of investors' information set rather than their trading 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 not related to informedness, such as diversification, liquidity provision, and trading constraints.[2] Thus, absent gaming behavior (which we find no evidence of), our unique data can better represent

[2] For example, Kaniel, Saar, Titman (2008) find evidence that individuals make trading profits by providing liquidity.

investors' true opinions than traded portfolios. Our setting also provides new evidence on the informedness of internet messages about the stock markets. 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 9 basis point *decrease* (t-stat -5.66) 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 -7.19) in weekly stock returns, but a one standard deviation increase in local evaluations predicts a decrease of 1 basis point (t-stat -0.36) in weekly stock returns. Local advantage, measured as the differential predictability between local and nonlocal evaluations, is a statistically and economically significant 19 basis points (t-stat 5.61) 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 of price 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 following, a long track record, post meaningful or interesting comments, and are notable

[3]Ivkovic and Weisbenner (2007) show that local information diffusion among individual investors dissipates dramatically beyond 50 miles. For robustness, we also classify local and nonlocal posts by whether the Twitter user and the company are located in the same state, and find similar results.

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.

The superior stock picking ability of local investors is consistent with individual investors possessing value-relevant private information about local firms ("information hypothesis"). However, an alternative explanation is that local investors are better at analyzing public information regarding local firms ("sophistication hypothesis"). To disentangle the two hypotheses, 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 (51 basis points per week) than for firms with public news coverage (14 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.

For a more thorough investigation of our competing explanations, we also conduct cross-sectional analyses of local advantage across levels of information asymmetry. If local advantage is due to investors' private information, then we expect the local advantage to be more pronounced in firms with more severe information asymmetry. Previous studies provide evidence that information asymmetry is severe in small firms (e.g., Coval and Moskowitz 1999), non-S&P 500 firms (e.g., Ivkovic and Weisbenner 2005), low analyst coverage firms (e.g., Brennan and Subrahmanyam 1995), and firms with high idiosyncratic volatility (e.g., Krishnaswami and Subramaniam 1999). Our results based on these proxies consistently show a significantly positive association between local advantage

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

and information asymmetry, indicating that local advantage is caused by individual investors' access to private information about local firms.

We further examine local advantage using a rolling long-short strategy based on local advantage that is similar to the rolling momentum strategy proposed by Jegadeesh and Titman (1993). For each day of our sample period, we construct the "locally favorable portfolio" that contains firms with negative nonlocal evaluations but non-negative local evaluations, and the "locally unfavorable portfolio" that contains firms with positive nonlocal evaluations but non-positive local evaluations. We then design a rolling strategy that goes long the "locally favorable portfolio" and goes short "locally unfavorable portfolio". The average daily abnormal profits of this zero-investment strategy are significantly positive, ranging from 2.6 basis points to 6.2 basis points per day. The abnormal profits increase significantly when we exclude firms with less information asymmetry (large firms, S&P 500 firms, high analyst coverage firms, or low idiosyncratic volatility firms). For example, when we exclude firms in top tercile of firm size, we observe daily abnormal profits from 5.5 basis points to 11.6 basis points per day.

Last, we examine a potential mechanism that can contribute to the negative performance of individual investors and how local information advantage helps alleviate such underperformance. Previous studies 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 (Hirshleifer 2001).[5]

We also contribute to a rapidly growing literature on the informedness of individual investors. Some studies provide evidence that individual investors lose significantly from their trading (Barber and Odean 2000, Barber, Odean, and Zhu 2009) or that individuals make trading profits but only through liquidity provision (Kaniel, Saar, and Titman 2008). 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 use particular order types (Kelley and Tetlock 2013). Our paper performs a unique examination of the

[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.

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 predicting 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 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.

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 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 Zweibel (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.[6]

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 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.

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.[7]

[6] 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.

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

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 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 over our sample period.[8] To control for microstructure effects, we drop penny stocks that are priced below two dollars at the end of the previous year. Our final sample of Twitter posts contains 216,266 posts from 2,008 users covering 1,819 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%),

[8] Our results are robust when we require the firm to have at least one Twitter posts over the sample period.

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.[9] For robustness, we also use the approach in Seasholes and Zhu (2010) to construct state-adjusted returns and find similar results on local advantage.

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.[10] 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.[11] We then calculate the distance between Twitter user and company headquarters using the following equation:

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

[9] 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).
[10] We thank Professor Kenneth French for making the data available.
[11] We match longitude and latitude coordinates to zip codes using the database from http://www.getzipcodedata.com/#.

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.[12]

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 Weisbenner 2005; Malloy 2005; Seasholes and Zhu 2010). We adopt the moderate 100-mile criterion because Ivkovic and Weisbenner (2007) show that local information diffusion among individual investors dissipates beyond 50 miles. Our approach classifies 20,570 posts as local and 195,696 posts as nonlocal. For robustness we also try classifying a post as local (nonlocal) if the Twitter user is in the same state as the firm's (a different state from the firm's) and find similar results from our tests.

2.5. Quantifying the Information in Twitter Posts

We use the maximum entropy (ME) approach to classify the information in Twitter posts. The ME approach derives meaning from natural language in the posts by applying a maximum likelihood algorithm to qualitative data. Since the information in a Twitter post can be subtle, using key word frequencies alone can cause misclassification. 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

[12] 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.

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, or 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.[13]

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.

[13] 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.

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 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 $P(c=c_0 | d)$ is as follows:

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

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

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

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. 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 $P_{ME}(c= c_0 | d)$ by maximizing the entropy 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 100 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).[14]

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 $6,341 million, a book-to-market ratio of 0.68, and is followed by 10.51 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.[15] 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.[16] A typical sample firm has an idiosyncratic volatility of 0.028 and an average daily return of 12 basis points, similar to the 0.030 and 11 basis points for the contemporaneous CRSP universe.

Panel B further presents firm characteristics sorted on the total number of Twitter posts, local posts, and nonlocal posts, respectively. The results show that Twitter users, whether local or nonlocal, tend to cover larger firms, growth firms, and high analyst coverage firms. Non-local users

---

[14] 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.
[15] We drop penny stocks priced below $2, which also makes our sample firms larger than the CRSP universe.
[16] We require at least 100 daily return observations in the estimation window.

seem to have a slight preference for covering winner stocks, which is not the case for local users. In addition, the number of Twitter posts does not vary with idiosyncratic volatility.

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.[17] 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 maximum entropy 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 (α_1- α_2)>0.

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

We further include firm fixed effects (D_i) to control for firm-specific characteristics, and ten lags of daily returns (AR_{it-j}) to control for short-term return reversals and microstructure effects. We calculate t-statistics using Driscoll and Kraay (1998) robust standard errors that control for both cross-sectional and time-series correlations.[18]

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:

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

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 messages about stock returns.

Panel A of Table 2 presents the results of the return regressions in equation (5). Interestingly, 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 -0.091 (t-stat -5.66), indicating that a one standard deviation increase in the evaluation measure is associated with a 9.1 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

[18] 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.

when they make trading profits they do so by providing liquidity (Kaniel, Saar, and Titman 2008). This 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 messages. In the next subsection, we explore whether there is a difference in the stock return predictability between local and nonlocal posts.

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 -0.196 (t-stat -7.19) for nonlocal evaluations but -0.006 (t-stat -0.36) for local evaluations. The local advantage (α_1- α_2) is 0.189 (t-stat 5.61), suggesting that a one standard deviation increase in local evaluation relative to nonlocal evaluation predicts an 18.9 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 raw returns instead of abnormal returns 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 in Table 3 are specific to the two-week sentiment measurement window of evaluations. Panel A of Table 4 repeats the regressions in Table 3 but with the local and nonlocal evaluations measured in the previous one-week window instead of the two-week window. These results consistently show statistically and economically significant coefficients on (α_1- α_2) in all models, suggesting that local advantage persists with the alternative measurement window.

We also examine whether our results on local advantage are sensitive to an alternative classification of local posts. 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 the state criterion. For example, in the model of five-day window the estimated local advantage (α_1- α_2) is 0.154 (t-stat 4.76), only slightly smaller than the 0.189 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 0.226 (t-stat 4.74), slightly larger than the 0.189 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 0.123 percent, both economically and statistically significant (t-stat 4.76).

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 5 presents the regression analyses with state-adjusted abnormal returns (Panel A) and state-adjusted raw returns (Panel B). The local advantage is both statistically and economically significant in all models.

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 0.152 percent (t-stat 4.23), similar to the 0.189 percent in Table 3.

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 on the regressions with the weighted evaluation measures. The local advantage using the weighted measure (0.108, t-stat 4.25) is smaller than that in Table 3 (0.187). 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 A of Table 7 repeats the regressions in Table 2 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 A of Table 7 shows that while the coefficients on local evaluations are insignificant, those on nonlocal evaluations are 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 values of evaluations since both positive and negative evaluations could trigger abnormal volume. Panel B of Table 7 confirms that the volume reaction to non-local posts is significantly larger than the volume reaction to local posts. Overall, the results in Table 7 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 or Investor Sophistication?

Our finding of a local advantage is consistent with individual investors possessing private information about local firms ("information hypothesis"). However, an alternative explanation is that local investors are better at analyzing public information than nonlocal investors ("sophistication hypothesis"). In this section, we perform a number of cross-sectional analyses to investigate whether local advantage is accounted for by access to private information or investor sophistication.

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 484,201 news articles that cover 1,789 of our

1,819 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 8 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 50.5 basis points per week (t-stat 5.92). In contrast, Panel B shows that the corresponding local advantage is 14.2 basis points (t-stat 4.69) 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 36.4 basis points (t-stat 4.02), 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. 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 three groups 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 tercile) and large firms (top tercile), respectively. Panel A of Table 9 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 0.526 (t-stat 3.91), much larger than that of the full sample (0.189, 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 a broad set of proxies for information asymmetry.

4.2.2. S&P 500 composite index as a proxy for information asymmetry

Previous studies on local advantage use the S&P 500 composite index to classify high and low information asymmetry firms. For example, Ivkovic and Weisbenner (2005) document a much stronger local advantage for firms that are not a component of the S&P 500 index. Seasholes and Zhu (2010) also use the S&P 500 index as a measure of information asymmetry, but find no evidence of local advantage in either index or non-index firms. For each day of our sample period, we sort sample firms into two groups according to whether they are in the S&P 500 composite index at the beginning of the month and then report, in Panel B of Table 9, the local advantage for index and non-index firms, respectively.

Panel B of Table 9 demonstrates a sharp contrast between non-index firms and index firms. Specifically, local advantage among non-index firms is significantly larger than that among index firms. For example, for the model of five-day window, the difference in local advantage between non-index and index firms is 15.3 basis points per week (t-stat 2.30), both economically and statistically significant. To summarize, these results present some evidence that, consistent with the information hypothesis, local advantage is stronger in non-index firms than index firms.

4.2.3. 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 terciles according to their size-adjusted analyst coverage for the month, and examine local advantage for low coverage firms (bottom tercile of coverage) and high coverage firms (top tercile of coverage), respectively.

In Panel C of Table 9, we observe that local advantage for low coverage firms is significant in all return windows and about twice as large as 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.

4.2.4. 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 three groups based on idiosyncratic volatility, and examine local advantage among high volatility firms (top tercile of volatility) and low volatility firms (bottom tercile of volatility), respectively.

Panel D of Table 9 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 29.8 basis points (t-stat 4.28) in the five-day return window, both economically and statistically significant. On the contrary, local advantage among low volatility firms is only 5.7 basis points (t-stat 2.08). 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, all our results using proxies for information asymmetry consistently 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. 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 non-local evaluations in the past two weeks are negative but the corresponding local evaluations are non-negative. The second portfolio, "locally unfavorable portfolio", contains firms for which the non-local evaluations in the past two weeks are positive but the corresponding local evaluations are non-positive. 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 2.6 basis points to 6.2 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, index identity, analyst coverage, and idiosyncratic volatility, in an attempt to improve the performance of the local-advantage-based strategy, we further exclude large firms, S&P 500 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 5.5 basis points to 11.6 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.

4.4. Nonlocal underperformance

A striking finding in our study is the persistent underperformance of nonlocal investors. Theoretically, in an efficient market, significant underperformance should be as difficult to identify as is persistent outperformance. This subsection explores one potential mechanism that contributes to the underperformance of nonlocal investors and whether local investors' private information advantage mitigates this underperformance. 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.[19] We then examine local and nonlocal investors' responses to analyst optimism in a regression setting.

Panel A of Table 11 reports panel regressions of monthly local or nonlocal investors' evaluations on the analyst optimism measure in the previous month for 1,751 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.[20] We observe that the coefficient on the local evaluation is an insignificant 0.002 (t-stat 0.27) but the coefficient on the nonlocal evaluation is a significantly positive 0.030 (t-stat 2.04). The difference in the two coefficients is statistically significant at the 0.10 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 11 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 (-0.104, t-stat -2.51) and nonlocal evaluations (-0.155, t-stat -6.57), indicating that both local and nonlocal investors tweet negatively about firms with sell

[19] 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.
[20] We winsorize both the local and nonlocal evaluations at the 1 and 99 percent cutoffs points to control for outliers.

recommendations. Although the coefficient for nonlocal investors is larger, the difference is not statistically significant. For buy recommendations, local and nonlocal investor sentiment is positive and significant for both local evaluations (0.041, t-stat 2.19) and nonlocal evaluations (0.115, t-stat 4.35). For buy recommendations, the significant difference in the coefficients (0.074, t-stat 2.27) indicates that nonlocal investors respond more positively to buy recommendations than local investors.

Our results indicate that the reactions of nonlocal investors to analysts' overoptimistic earnings forecasts and buy recommendations are significantly stronger than those of local investors. We further examine if this overreaction hurts their investment performance. Panel A of Table 12 presents panel regressions of monthly cumulative abnormal returns (CAR) in the month t+1 or t+2 on the analyst optimism measure in month t. The results show negative coefficients on analyst optimism in both regressions. One standard deviation increase in analyst optimism in month t is associated with a significantly negative CAR in month t+1 of -0.501 percent. Multiplying this result by the 0.029 difference in response between nonlocal and local investors (Panel A of Table 11) shows that a one standard deviation increase in analyst optimism is associated with approximately 1.5 basis point of underperformance for nonlocal investors relative to local investors.

Nonlocal underperformance attributable to overoptimism is considerably larger for analysts' recommendations. Panel B of Table 12 present regressions of monthly CARs on analyst recommendations, which indicate that buy recommendations in month t lead to negative CARs of -1.51 percent in month t+1. Multiplying this number by the 0.074 difference in response for buy recommendations (Panel B of Table 11) shows that overreaction to buy recommendation can cause -11.2 basis points of nonlocal investor monthly underperformance.[21]

[21] Although the differences in sentiment exposure to sell recommendations is not statistically significant, at the means the Table 11 coefficients indicate that 23 basis points of monthly underperformance can be attributed to overreaction to sell

From this experiment, we conclude that the source of the relative advantage of local investors is that their locally-obtained information advantage helps investors overcome their behavioral biases. In this specific test, we show that local investors have smaller sentiment responses to analysts' opinions. This attenuation of the overreaction bias improves local investors' investment performance.

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 19 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.

recommendations. Sell recommendations in our sample significantly outperform, producing CARs of 4.7% in the following month (Table 12, Panel B).

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 can hurts 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., 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.

Chan, Wesley S., 2003, Stock price reaction to news and no-news: Drift and reversal after headlines, *Journal of Financial Economics* 70, 223-260.

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, 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.

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.

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.

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 1,819 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 reports average firm characteristics for the subsets of firms sorted into two groups according to the numbers of all Twitter posts, local posts, and nonlocal posts, respectively. 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 A: Characteristics of Sample Firms

	Mean	**STD**	**P10**	**P25**	**P50**	**P75**	**P90**
Market Capitalization ($M)	6,341	20,098	160	396	1,192	4,153	13,303
Book/Market Ratio	0.68	0.66	0.12	0.29	0.54	0.90	1.43
Analyst Coverage	10.51	7.33	1.60	4.24	8.65	14.62	20.04
Idiosyncratic Volatility	0.029	0.014	0.015	0.019	0.026	0.036	0.047
Daily Stock Return (%)	0.11	0.16	-0.04	0.04	0.11	0.19	0.28

Panel B: Average Firm Characteristics Sorted on the Numbers of Posts

	All Posts		Local Posts		Nonlocal Posts	
	Low	High	Low	High	Low	High
Market Capitalization ($M)	2,065	10,566	3,516	8,579	2,254	10,361
Book/Market Ratio	0.76	0.61	0.74	0.64	0.75	0.61
Analyst Coverage	7.23	12.71	8.44	11.21	7.30	12.63
Idiosyncratic Volatility	0.030	0.029	0.029	0.029	0.029	0.029
Daily Return	0.10	0.12	0.11	0.11	0.10	0.13

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 variables are two-, five-, ten-, and 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.

	Dependant Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Panel A: Regressions on Two-Week Evaluations				
Two-Week Evaluation	-0.039***	-0.091***	-0.176***	-0.352***
	(-5.79)	(-5.66)	(-5.67)	(-6.57)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,694	768,305	767,647	766,279
Number of PERMNOs	1,819	1,819	1,819	1,819
Panel B: Regressions on Two-Week Evaluations: Skip-a-Week Returns				
Two-Week Evaluation	-0.034***	-0.084***	-0.172***	-0.338***
	(-5.03)	(-5.18)	(-5.74)	(-7.05)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,073	767,677	767,008	765,596
Number of PERMNOs	1,819	1,819	1,819	1,819
Panel C: Regressions on One-Week Evaluations				
One-Week Evaluation	-0.037***	-0.085***	-0.155***	-0.304***
	(-6.30)	(-6.42)	(-6.52)	(-7.17)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,694	768,305	767,647	766,279
Number of PERMNOs	1,819	1,819	1,819	1,819

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 variables are cumulative two-, five-, ten- and 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 dependent variables are raw returns instead of abnormal returns. ***, **, and * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Panel A: Regressions of Abnormal Returns

	Dependant Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Local Evaluation	-0.003	-0.006	-0.023	-0.023
	(-0.33)	(-0.36)	(-0.72)	(-0.83)
Nonlocal Evaluation	-0.084***	-0.196***	-0.327***	-0.722***
	(-6.82)	(-7.19)	(-6.87)	(-8.07)
Local – Nonlocal	0.081***	0.189***	0.307***	0.674***
	(4.89)	(5.61)	(5.41)	(6.12)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,694	768,305	767,647	766,279
Number of PERMNOs	1,819	1,819	1,819	1,819

Panel B: Regressions of Raw Returns

	Dependant Variables			
	2-Day Ret.	5-Day Ret.	10-Day Ret.	20-Day Ret.
Local Evaluation	-0.009	-0.030	-0.070*	-0.120*
	(-1.11)	(-1.56)	(-1.91)	(-1.96)
Nonlocal Evaluation	-0.071***	-0.159**	-0.290**	-0.558***
	(-2.90)	(-2.76)	(-2.69)	(-3.10)
Local – Nonlocal	0.062**	0.129**	0.220*	0.438**
	(2.32)	(2.09)	(1.90)	(2.27)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,694	768,305	767,647	766,279
Number of PERMNOs	1,819	1,819	1,819	1,819

Table 4: Panel Regressions of Stock Returns: Alternative Construction/ Sample Selection

Panel A presents the regressions of abnormal returns on local and nonlocal evaluations in the one-week window prior to return measurements. The definition of abnormal returns, local and nonlocal evaluations, and regression settings are similar to the Panel A of Table 3. Panel B presents the regressions of abnormal returns on local and nonlocal evaluations in the two-week window prior to return measurements. These regressions are 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 firms that have at least one Twitter post in the two-week period of evaluation measurement. 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.

	Dependant Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Panel A: Return Regressions: Evaluations Measured in the Past One Week				
Local Evaluation	-0.011	-0.013	-0.020	-0.059
	(-1.38)	(-0.81)	(-0.80)	(-1.31)
Nonlocal Evaluation	-0.069***	-0.172***	-0.318***	-0.599***
	(-6.40)	(-7.55)	(-7.78)	(-8.14)
Local – Nonlocal	0.058***	0.159***	0.299***	0.539***
	(4.07)	(5.54)	(6.12)	(6.12)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,694	768,305	767,647	766,279
Number of PERMNOs	1,819	1,819	1,819	1,819
Panel B: Local Posts Identified Using the State Criterion				
Local Evaluation	-0.014	-0.031	-0.063*	-0.119**
	(-1.64)	(-1.60)	(-1.95)	(-1.99)
Nonlocal Evaluation	-0.079***	-0.185***	-0.343***	-0.700***
	(-6.55)	(-7.21)	(-7.29)	(-9.56)
Local – Nonlocal	0.064***	0.154***	0.280***	0.581***
	(4.05)	(4.76)	(5.18)	(7.26)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,694	768,305	767,647	766,279
Number of PERMNOs	1,819	1,819	1,819	1,819
Panel C: Return Regressions: At Least One Post in the Evaluation Period				
Local Evaluation	-0.004	-0.008	-0.035	-0.088
	(-0.31)	(-0.30)	(-0.72)	(-1.04)
Nonlocal Evaluation	-0.107***	-0.234***	-0.405***	-0.803***
	(-5.93)	(-6.41)	(-5.75)	(-7.26)
Local – Nonlocal	0.102***	0.226***	0.370***	0.715***
	(3.93)	(4.74)	(4.32)	(5.11)

	Dependant Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR.
Panel C: Return Regressions: At Least One Post in the Evaluation Period				
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	297,583	297,470	297,288	296,903
Number of PERMNOs	1,819	1,819	1,819	1,819
Panel D: Include Only Recommended Users				
Local Evaluation	-0.010	-0.025*	-0.053*	-0.131**
	(-1.55)	(-1.70)	(-1.86)	(-2.37)
Nonlocal Evaluation	-0.061***	-0.149***	-0.290***	-0.556***
	(-5.94)	(-6.71)	(-7.07)	(-8.40)
Local – Nonlocal	0.051***	0.123***	0.237***	0.425***
	(4.35)	(4.76)	(4.75)	(5.28)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	750,760	750,410	749,817	748,579
Number of PERMNOs	1,758	1,758	1,758	1,758

Table 5: Panel Regressions of Stock Returns: State Adjusted Returns

Panel A presents the regressions of state-adjusted abnormal returns on local and nonlocal evaluations in the two-week period prior to return measurements. 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. The independent variables and regression settings are defined in the heading of Table 3. Panel B repeats regressions in Panel A but with state-adjusted raw returns. Daily state adjusted return for a firm is calculated as the daily return of the firm minus the average daily return of all firms in the same state as the firm. All regressions include firm fixed effects with lagged returns of the previous ten trading days of 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.

Panel A: Regressions of State-Adjusted Abnormal Returns

	Dependant Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Local Evaluation	-0.006	-0.012	-0.029	-0.060
	(-0.61)	(-0.62)	(-0.87)	(-0.98)
Nonlocal Evaluation	-0.094***	-0.223***	-0.417***	-0.810***
	(-8.22)	(-8.67)	(-8.56)	(-9.72)
Local – Nonlocal	0.089***	0.212***	0.388***	0.750***
	(5.34)	(6.26)	(6.58)	(7.47)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	766,537	766,145	765,482	764,104
Number of PERMNOs	1,819	1,819	1,819	1,819

Panel B: Regressions of State-Adjusted Raw Returns

	Dependant Variables			
	2-Day Ret.	5-Day Ret.	10-Day Ret.	20-Day Ret.
Local Evaluation	-0.004	-0.010	-0.030	-0.063
	(-0.50)	(-0.61)	(-1.01)	(-1.24)
Nonlocal Evaluation	-0.063***	-0.142***	-0.250***	-0.471***
	(-5.34)	(-5.62)	(-5.47)	(-7.17)
Local – Nonlocal	0.059***	0.132***	0.220***	0.409***
	(3.81)	(4.21)	(4.10)	(5.54)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	766,537	766,145	765,482	764,104
Number of PERMNOs	1,819	1,819	1,819	1,819

Table 6: Panel Regressions of Stock Returns: Alternative Measurements of Evaluations or Returns

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. 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.

	Dependant Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Panel A: Regressions of Abnormal Returns: Skip One Week				
Local Evaluation	-0.004	-0.016	-0.036	-0.029
	(-0.62)	(-0.91)	(-1.05)	(-0.55)
Nonlocal Evaluation	-0.071***	-0.167***	-0.338***	-0.715***
	(-5.87)	(-5.81)	(-6.36)	(-9.10)
Local – Nonlocal	0.066***	0.152***	0.303***	0.687***
	(4.46)	(4.23)	(4.54)	(5.16)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,703	767,677	767,008	765,596
Number of PERMNOs	1,819	1,819	1,819	1,819
Panel B: Return Regressions on Weighted Evaluation Measures				
Local Weighted Evaluation	-0.009	-0.020	-0.051*	-0.125*
	(-1.03)	(-1.07)	(-1.68)	(-1.93)
Nonlocal Weighted Evaluation	-0.052***	-0.128***	-0.252***	-0.484***
	(-4.75)	(-5.47)	(-5.88)	(-7.11)
Local – Nonlocal	0.043***	0.108***	0.200***	0.361***
	(3.19)	(4.25)	(4.84)	(5.86)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,694	768,305	767,647	766,279
Number of PERMNOs	1,819	1,819	1,819	1,819

Table 7: Panel Regressions of Trading Volumes

Panel A presents regressions of turnovers on local and nonlocal evaluations in the two-week period prior to turnover measurements. 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. Investor evaluations are defined in the heading of Table 3. We standardize the independent variables for each regression. Panel B is similar to Panel A but the independent variables are absolute values of the local and nonlocal evaluation measures. 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.

	Dependant Variables			
	2-Day Vol.	5-Day Vol.	10-Day Vol.	20-Day Vol.
Panel A: Regressions of Trading Volumes on Local and Nonlocal evaluations				
Local Evaluation	0.014	0.020	0.008	0.027
	(0.97)	(0.65)	(0.15)	(0.35)
Nonlocal Evaluation	0.202***	0.387***	0.590***	0.750***
	(6.68)	(5.57)	(4.49)	(3.50)
Local – Nonlocal	-0.188***	-0.367***	-0.582***	-0.723***
	(-4.97)	(-4.16)	(-3.52)	(-2.76)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,736	768,479	768,041	767,107
Number of PERMNOs	1,819	1,819	1,819	1,819
Panel B: Regressions of Trading Volumes on Un-Signed Local and Nonlocal Evaluations				
|Local Evaluation|	-0.073***	-0.135**	-0.196*	-0.197
	(-2.71)	(-2.28)	(-1.83)	(-1.15)
|Nonlocal Evaluation|	0.298***	0.547***	0.795***	0.965***
	(7.26)	(5.86)	(4.61)	(3.62)
|Local| – |Nonlocal|	-0.371***	-0.681***	-0.991***	-1.161***
	(-5.95)	(-4.78)	(-3.75)	(-2.81)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	768,736	768,479	768,041	767,107
Number of PERMNOs	1,819	1,819	1,819	1,819

Table 8: 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	0.014	-0.000	-0.094	-0.329**
	(0.52)	(-0.00)	(-1.08)	(-2.62)
Nonlocal Evaluation	-0.201***	-0.505***	-0.985***	-1.772***
	(-9.07)	(-7.83)	(-6.75)	(-6.80)
Local - Nonlocal (1)	0.215***	0.505***	0.891***	1.443***
	(6.37)	(5.92)	(5.13)	(4.90)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	272,678	272,532	272,267	271,686
Number of PERMNOs	1,687	1,687	1,687	1,687
Panel B: Stocks with Public News Coverage in the Period of Twitter Posts				
Local Evaluation	-0.005	-0.008	-0.020	-0.030
	(-0.67)	(-0.58)	(-0.77)	(-0.61)
Nonlocal Evaluation	-0.065***	-0.150***	-0.269***	-0.557***
	(-5.20)	(-5.79)	(-5.48)	(-6.94)
Local - Nonlocal (2)	0.061***	0.142***	0.249***	0.526***
	(3.82)	(4.69)	(4.27)	(5.51)
Controls of Lagged Returns	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Number of Obs	496,016	495,773	495,380	494,593
Number of PERMNOs	1,788	1,788	1,788	1,788
Panel C: Difference in Local Advantage: Non-News versus News Stocks				
(1) - (2)	0.155***	0.364***	0.643***	0.917***
	(4.15)	(4.02)	(3.51)	(2.96)

Table 9: 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 three 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 tercile of market capitalization) and large firms (highest tercile 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 two groups based whether they are in the S&P 500 Composite Index at the beginning of the month. We then report local advantage for non-S&P 500 firms, S&P 500 firms, and their differences. For Panel C, on each day of our sample period, we sort stocks into three 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 tercile of coverage), high coverage firms (highest tercile of coverage), and their differences. For Panel D, on each day of our sample period, we sort stocks into three 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 tercile of volatility), low volatility firms (lowest tercile 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	0.231***	0.526***	1.027***	1.767***
	(3.75)	(3.91)	(4.17)	(4.26)
Large Firms	0.040***	0.091***	0.146***	0.333***
	(2.76)	(3.00)	(2.65)	(3.21)
Small – Large	0.191***	0.436***	0.881***	1.434***
	(3.02)	(3.16)	(3.49)	(3.35)
Panel B: Local Advantage for Non-S&P versus S&P Firms				
Non-S&P Firms	0.105***	0.244***	0.434***	0.811***
	(3.70)	(4.16)	(3.63)	(3.69)
S&P Firms	0.040***	0.092***	0.163***	0.388***
	(3.06)	(3.57)	(3.42)	(4.64)
Non-S&P – S&P	0.065**	0.153**	0.270**	0.422*
	(2.09)	(2.30)	(2.10)	(1.79)
Panel C: Local Advantage for Low versus High Analyst Coverage Firms				
Low Coverage Firms	0.161***	0.384***	0.732***	1.402***
	(3.54)	(4.12)	(4.41)	(4.97)
High Coverage Firms	0.051***	0.120***	0.197***	0.403***
	(2.95)	(3.32)	(3.02)	(3.58)
Low – High	0.111**	0.265***	0.535***	0.999***
	(2.28)	(2.65)	(3.00)	(3.29)
Panel D: Local Advantage for High versus Low Idiosyncratic Volatility Firms				
High Idio. Volatility Firms	0.132***	0.298***	0.551***	1.008***
	(4.17)	(4.28)	(4.09)	(4.32)

	Dependent Variables			
	2-Day CAR	5-Day CAR	10-Day CAR	20-Day CAR
Panel D: Local Advantage for High versus Low Idiosyncratic Volatility Firms				
Low Idio. Volatility Firms	0.021*	0.057**	0.100**	0.239***
	(1.91)	(2.08)	(2.08)	(3.07)
High – Low	0.0112***	0.242***	0.450***	0.768***
	(3.29)	(3.23)	(3.26)	(3.12)

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 positive but the local evaluations are not ("locally unfavorable portfolio") and a portfolio containing firms for which the non-local evaluation measures are negative but the local evaluation measures are not ("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, S&P 500 index 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	0.062***	0.049**	0.037**	0.026*
	(2.62)	(2.35)	(2.17)	(1.87)
Exclude Large Firms	0.116**	0.077***	0.060**	0.055**
	(2.45)	(2.62)	(2.19)	(2.12)
Exclude S&P Firms	0.103**	0.068***	0.054**	0.0051**
	(2.38)	(2.69)	(2.36)	(2.57)
Exclude High Coverage Firms	0.077**	0.050*	0.055**	0.047**
	(2.15)	(1.78)	(2.27)	(2.14)
Exclude Low Volatility Firms	0.088**	0.071***	0.052**	0.033***
	(2.33)	(2.38)	(2.06)	(1.55)

Table 11 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.

	Dependant Variables		
	Local Evaluation	Non-Local Evaluation	Non-Local - Local
Panel A: Regressions on Analyst Optimism			
Constant	0.001	0.001	
	(0.02)	(0.01)	
Analyst Optimism	0.002	0.030**	0.029*
	(0.27)	(2.04)	(1.79)
Firm Fixed Effects	Yes	Yes	
Number of Obs	37,366	37,366	
Number of PERMNOs	1,751	1,751	
Panel B: Regressions on Buy and Sell Recommendations			
Constant	-0.023	-0.065	
	(-0.39)	(-0.56)	
Sell Recommendation	-0.104**	-0.155***	-0.050
	(-2.51)	(-6.57)	(-1.05)
Buy Recommendation	0.041**	0.115***	0.074**
	(2.19)	(4.35)	(2.27)
Firm Fixed Effects	Yes	Yes	
Number of Obs	38,466	38,466	
Number of PERMNOs	1,769	1,769	

Table 12 Panel Regressions of Monthly Abnormal Stock Returns on Analyst Optimism and Recommendations

Panel A presents panel regressions of monthly abnormal stock returns on previous monthly analyst optimism measures. The dependent variables are monthly cumulative abnormal returns (measured in percent). 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 are the analyst optimism measures one month or two months before the return month. 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. We further report the difference between the coefficients on analyst optimism of the two regressions. Both regressions include firm fixed effects. 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 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 significance at the 0.01, 0.05, and 0.10 levels, respectively.

	Dependant Variables	
	CAR of Month t+1	CAR of Month t+2
Panel A: Regressions on Analyst Optimism		
Constant	-0.278***	-0.267***
	(-2.59)	(-2.90)
Analyst Optimism (Month t)	-0.501***	-0.246***
	(-8.85)	(-3.50)
Firm Fixed Effects	Yes	Yes
Number of Obs	37,354	37,330
Number of PERMNOs	1,751	1,751
Panel B: Regressions on Buy and Sell Recommendations		
Constant	0.595***	0.451***
	(4.57)	(3.19)
Buy Recommendation (Month t)	-1.513***	-1.218***
	(-23.05)	(-16.66)
Sell Recommendation (Month t)	4.694***	3.801***
	(6.57)	(-9.80)
Firm Fixed Effects	Yes	Yes
Number of Obs	38,442	38,341
Number of PERMNOs	1,769	1,769

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.

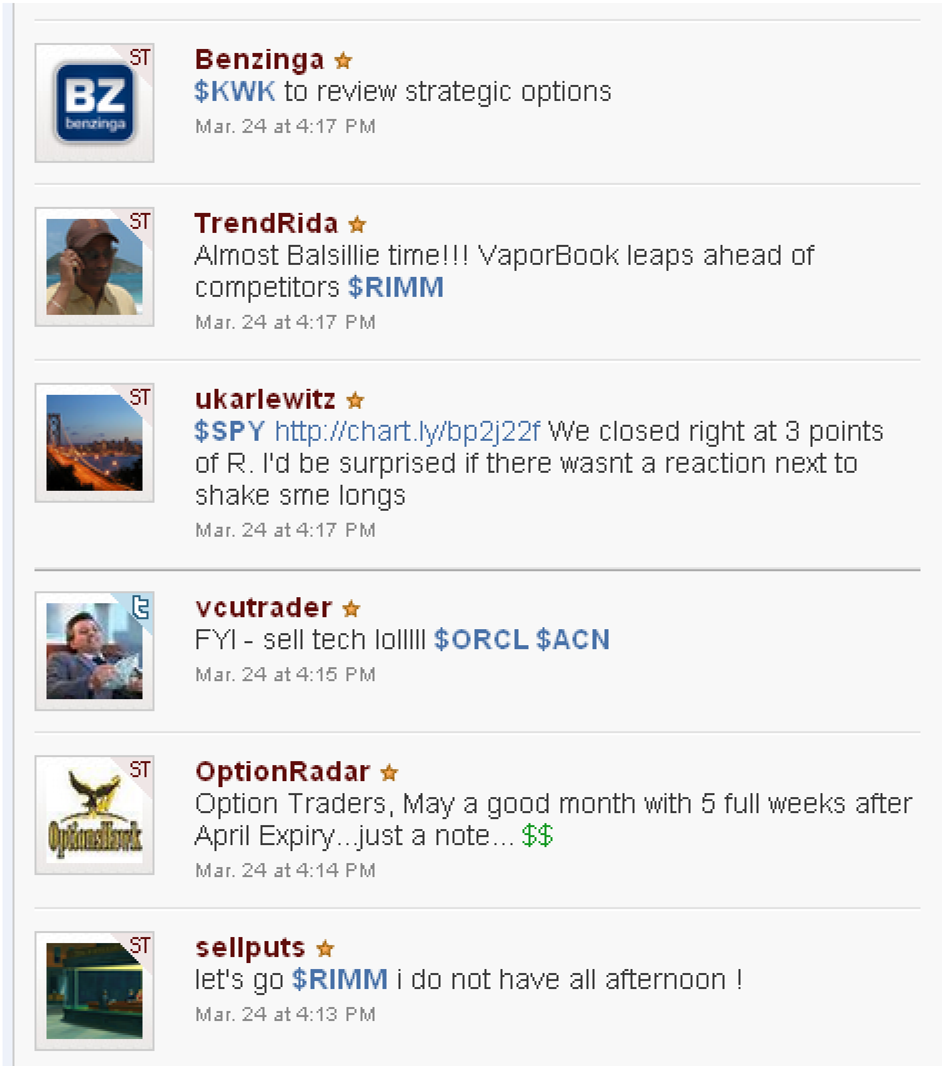

Figure 2
Geographical Distribution of Sample Twitter Users

This figure plots the geographical distribution of the 2,008 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.