SOCIAL MEDIA, NEWS MEDIA AND THE STOCK MARKET

Peiran Jiao* André Veiga† Ansgar Walther‡

This version: February 21, 2020

Abstract

We study the effect on stock volatility and turnover of coverage by traditional news media and social media. We find that coverage by traditional news media predicts *decreases* in subsequent volatility and turnover, but coverage by social media predicts *increases* in volatility and turnover. We show that these patters are consistent with a model of "echo chambers", where social networks repeat news, but some investors interpret repeated signals as genuinely new information.

JEL: D83, G02.

Keywords: Social media, News media, Asset prices.

*Department of Finance, Maastricht University; Tongersestraat 53, Maastricht 6211LM, the Netherlands (e-mail: p.jiao@maastrichtuniversity.nl)

†Imperial College Business School, South Kensington Campus, Ayrton Rd, Kensington, London SW7 2AZ (e-mail: a.veiga@imperial.ac.uk)

‡Imperial College Business School, South Kensington Campus, Ayrton Rd, Kensington, London SW7 2AZ (e-mail: a.walther@imperial.ac.uk). We thank MarketPsych for providing access to the dataset. We are grateful to Daniel Andrei, Thierry Foucault, Basile Grassi, Dan Greenwald, Robin Greenwood, Gur Huberman, Ryoko Ito, Marcin Kacperczyk, Joel Peress, Dan Quigley, Tarun Ramadorai, Selma Walther, Mungo Wilson, Peyton Young and seminar participants at the CEPR European Summer Symposium in Financial Markets for their comments. Declarations of interest: none.

1 Introduction

A survey by the Reuters Institute (2016) finds that 51% of respondents use social media to access news every week, and 12% cite it as their main source of news. Moreover, a significant share of social media content relates to stock markets, as witnessed by an emerging industry that extracts and sells market-relevant indicators from social media.[1]

An active literature studies the relationship between social media, news media and the stock market.[2] However, most existing studies focus on one source of news at a time. Therefore, an open question is whether stock markets react to social media and news media in a systematically different way. In this paper, we directly compare social and news media coverage, and thus provide new stylized facts on their relationship with stock markets. We use a unique panel dataset on media coverage from the Thomson Reuters MarketPsych Indices (TRMI) database. Our data aggregate a broad spectrum of news media sources and most popular social media into indicators of coverage ("buzz"), which facilitate a like-for-like comparison between social and news media. We merge these data with stock prices, turnover, and stock-specific characteristics.

Our main result is that coverage in social and news media are associated with markedly different patterns of subsequent return volatility and trading volume per share (or turnover). High social media buzz around a given stock predicts a statistically significant *increase* in idiosyncratic return volatility and trading activity over the following month. High news media buzz predicts a significant *decrease* in volatility and trading activity. These empirical patterns are robust to the inclusion of stock and time fixed effects, time-varying stock characteristics and measures of disagreement about asset value (the dispersion in financial analysts' opinions). Our results on volatility also apply at the market

[1] See, for example: http://www.wsj.com/articles/tweets-give-birds-eye-view-of-stocks-1436128047

[2] For example, sentiments in news and online searches predict stock returns and turnover (Tetlock, 2007), stocks with low coverage have higher returns (Fang and Peress, 2009), and press coverage reduces information asymmetries (Bushee et al., 2010). Beyond traditional news media, noise levels in trading pits predict high volatility (Coval and Shumway, 2001), activity in specialist chat rooms (e.g. *RagingBull*) predicts high volatility and turnover (Antweiler and Frank, 2004), and sentiment indicators extracted from online forums and searches can predict returns (Chen et al., 2014; Da et al., 2015).

level: high social media buzz around the stock market as a whole predicts higher return volatility, while news media buzz predicts the opposite.

We further evaluate the mechanisms at play, using a stock-level panel VAR which includes news media buzz, social media buzz, volatility, and turnover as endogenous variables. We find that an increase in news media buzz predicts an increase in subsequent social media buzz in the sense of Granger causality. The converse is not true: increases in social media buzz do not predict changes in subsequent news buzz.

The core contribution of this paper is to establish these robust stylized facts. Together, they suggest that stock markets interact with social media and news media in different ways. Moreover, it appears that news media is a leading indicator for social media. This is consistent with the view that social media contents are generated by repeating and discussing (e.g. re-tweeting) existing news. However, our findings do not have a causal interpretation (we do not observe exogenous variations in either social or news media coverage). Further work is needed to pin down the exact mechanisms at play.

As a complementary exercise, we propose a theoretical model consistent both with our evidence and with existing work. We analyze our findings through the lens of asset pricing models with imperfect information. We present a model where the processing of social media signals is subject to a "correlation neglect" or "echo chamber" effect (e.g. DeMarzo et al., 2003; Gentzkow and Shapiro, 2011; Tetlock, 2011). Social media repeat existing news media signals, and a subset of "behavioral" traders interpret these repetitions as genuinely new information. In this setting, social media and news media coverage have *opposite* effects on subsequent volatility and turnover. News media contain genuine information and therefore dampen disagreement about asset value. Periods where an asset experiences high coverage by news media are followed by *lower* return volatility and turnover. Higher coverage by social media, by contrast, increases disagreement and boosts the confidence of behavioral traders. Periods of high coverage by social media are followed by *higher* return volatility and turnover. We also show that the same patterns cannot be generated by other cognitive biases that could apply to social media such as overconfidence, conservatism, rational inattention, and confirmation bias.

Our data do not allow us to rule out all alternative mechanisms. For instance, it is possible that unobserved shocks increase both social media buzz and subsequent trading activity. However, we note that the "echo chamber" model is not only consistent with our stylized facts, but also with existing evidence showing that financial markets react to repeated signals (Huberman and Regev, 2001; Tetlock, 2011), as well as psychological evidence on repetition-induced learning (Hawkins and Hoch, 1992).

The remainder of the paper is structured as follows. In Section 2, we describe our data and present summary statistics. In Section 3, we show our main empirical results and robustness checks. Section 4 contains our theoretical framework. Section 5 concludes. Figures and Tables appear after the main text. All proofs are presented in the appendix.

2 Data

We now describe the data we use. First, we describe how we measure coverage ("buzz") and sentiment for each stock in social and news media. Second, we describe the financial data we use to measure stock prices, volatility, trading activity and stock characteristics. Then, we present summary statistics. For a summary of all variable definitions, see Table 1 in the appendix.

2.1 Measuring 'Buzz' and Sentiment

We use the Thompson Reuters MarketPsych Index (TRMI) database, which extracts measures of buzz (defined below) and sentiment from English-language news and social media content using a proprietary machine learning lexical analysis algorithm. Media contents measures in the same database have been used in several studies (e.g. Michaelides et al., 2015; Sun et al., 2016; Michaelides et al., Forthcoming). For consistency, we focus on the period between January 2009 and December 2014, since several waves of major news source additions occurred before 2009.

During this time, the main sources of traditional news media content are (i) Reuters News, (ii) a host of mainstream news sources collected by MarketPsych Data, and (iii) online content collected by Moreover Technologies from about 50,000 internet news sites that include top international and business news sources, top regional new sources, and leading industry sources. The online news content includes many finance-specific sites such as Forbes and SeekingAlpha.

The main sources of social media content are (i) content collected by MarketPsych Data from internet forums and finance-specific tweets, and (ii) a social media feed constructed by Moreover Technologies, which captures the top 30% of social media content, as ranked by popularity using incoming links, collected from around 4 million social media sources such as chat rooms (including stock-market specific chats), public Facebook posts, blogs, micro-blogs and tweets.

From these sources, the TRMI algorithm extracts high-frequency measures of media coverage ("buzz"), sentiment, and events surrounding each of about 3000 US stocks. The TRMI indicators update every five minutes. We rely on a dataset which reports them at the daily frequency.

The *total buzz* of a stock on a given day counts the number of words and phrases referring to the stock in the above sources. This number is obtained by first identifying references about a specific stock in news articles and social media posts, and then counting the total number of phrases and words referring to sentiments (for instance, fear, joy or trust) and/or events (for instance, litigation, mergers or layoffs) related to this stock. Therefore, total buzz captures not only which stocks are being mentioned, but also the intensity of discussion of a particular stock, as captured by the quantity of phrases and words. This measure is more informative than the length of the article, since meaningless words ('the', 'are', 'in', etc.) are not included. For our main analysis, we use a monthly measure of total buzz, which is obtained by summing across days. Figure 1 shows total buzz at the market level for social and news media.

The *relative buzz* of a stock is defined as the total buzz of a stock in a given month, divided by the total buzz of all stocks mentioned in that month. This calculation is done separately for social and news media content, yielding our key measures of coverage: Social media relative buzz (BuzzS*)* and news media relative buzz (BuzzN), both of which are continuous variables between zero and one.

Total buzz appears to contain significant time effects and potential structural breaks (see Figure 1). We therefore focus our analysis on *relative buzz*.[3] This amounts to using a non-linear control for total buzz in the market, and gives a stationary measure of each stock's individual coverage. Using such a measure in our main analysis assumes that investors allocate attention horizontally across stocks at each point in time.[4]

We will also include, as control variables, measures of "sentiment" from the TRMI database. In calculating the TRMI sentiment indices, the sentiment scores of words are calculated by first splitting the articles and sentences into phrases and words, then letting human annotators evaluate their sentiments with the consensus value taken, and using these labels to train a machine learning classification algorithm. Sentiment on news and social media (SentN and SentS) is the difference between the number of "positive" and "negative" references to a stock, divided by the stock's total buzz, so that their values range from -1 to 1. To account for the potential asymmetric effects of positive and negative sentiment, we also include the number of negative references in isolation (SentN(-) and SentS(-)). SentN(-) and SentS(-), are equal to 0 when sentiment is posi-

[3]However, as total buzz, rather than relative buzz, corresponds to our model specification more closely. We also conducted robustness tests of our results with total buzz. This will be explained in Section 3.3.

[4]This corresponds well to our model in Section 4. We also tried an alternative specification that assumes investors allocate attention vertically for a particular company over time. Our findings are still robust. These results are reported in Table 18 of Appendix C.

tive, and equal to the original sentiment value when they are negative. Our sentiment controls serve as a proxy for potential unobserved stock-specific events.

2.2 Measuring Volatility

We merge our measures of buzz with monthly financial data from the Center for Research in Securities Prices (CRSP) and the Compustat database. The main variables of interest are trading activity and the realized idiosyncratic return volatility of each stock.

In our main analyses, we construct the parametric measure of realized idiosyncratic volatility ("iVolp") in two steps. First, for every month m in the sample, we estimate a three-factor model of daily returns on each stock by fitting the following regression equation:

$$(R_{it} - Rf_t) = \beta_0^{(m)} + \beta_1^{(m)}(Rm_t - Rf_t) + \beta_2^{(m)} SMB_t + \beta_3^{(m)} HML_t + \epsilon_t^{(m)}.$$

R_{it} is the return to stock i on day t; Rf_t is the one-month treasury bill rate; Rm_t is the return to the value-weighted market portfolio; SMB_t is the average return on the three Fama and French (1993) small-cap portfolios minus the average return on the three big-cap portfolios; and HML_t is the average return on the two value stock portfolios minus the average return on the two growth stock portfolios. Second, we define the idiosyncratic volatility of stock i in month m as the sum of squared errors over all days in month m from this monthly regression.

We check the robustness of our results by considering an alternative, non-parametric, measure of idiosyncratic volatility ("iVoln"), which is obtained by taking the variance of daily returns of each stock within a month at the monthly frequency (see Table 10).

2.3 Other Financial Data

Our measure of trading activity is turnover ("Turn"), which is the share trading volume of a stock divided by the number of shares outstanding. We also include in our analysis a set of financial variables which have been shown to have predictive power for volatility and trading activity. We take from CRSP the size of a firm ("Size") measured by its market capitalization, monthly stock price returns ("Return") and absolute value of return ("AbsReturn"), and we calculate the standard deviation of the last 60 monthly returns ("TotalSD"). Using Compustat data, we calculate each firm's leverage ("Leverage"), and its degree of focus as measured by the Herfindahl-Hirschman index of segment revenue ("HHI"). We include the fraction of institutional ownership ("InstOwn") from the Thomson Reuters Stock Ownership Summary.[5] We further obtain the dispersion of analyst opinions ("AnalystDisp") from the I/B/E/S summary files. Notice that all variables (HHI, InstOwn, etc) are allowed to vary over time for each stock at the monthly level.

2.4 Sample Selection and Summary Statistics

We focus on stocks which are traded on NYSE, AMEX and NASDAQ. We follow the literature in excluding regulated utilities (SIC codes 4910-4949), depository institutions (SIC 6000-6099) and holding and investment companies (SIC 6700-6799). The panel is unbalanced due to the entry and exit of stocks, so we restrict attention to a balanced panel in our main analysis. Our balanced panel includes 1848 stocks observed for 72 months (January 2009 to December 2014).

Table 2 shows sample means and standard deviations for (relative) buzz, volatility and turnover, for all stocks and disaggregated by industry.[6] Buzz is measured in per-

[5] The literature on stock market volatility demonstrates the predictive power of size (Cheung and Ng, 1992), returns (Duffee, 1995), institutional ownership (Dennis and Strickland, 2002) and trading volume (Schwert, 1989). Trading activity is commonly associated with the absolute value of returns (Karpoff, 1987; Schwert, 1989), institutional ownership (Tkac, 1999) and size (Tkac, 1999; Lo and Wang, 2000).

[6] Industries are classified according to Thomson Reuters Business Classfication (TRBC) 10 economic sectors.

centage points. We winsorize all financial data and the buzz measures at 1% to ensure that our estimates are not driven by outliers.

The average of relative buzz is about 0.03% per stock. Values of buzz range from zero, for stocks which are not covered (not mentioned in any news source) in a given month, to 0.53% for the stocks with most coverage in news media and 0.65% for the stocks with most coverage in social media.[7] The standard deviation of buzz is slightly higher for social media than for news media. Figure 2 illustrates the heterogeneity across industries. For instance, buzz is relatively high in the trade, services, manufacturing and finance industries, and relatively low in agriculture and mining/construction. Moreover, news buzz is relatively more prominent than social media buzz in industrials, but the opposite is true in healthcare. In our main analysis, we will control for firm fixed effects which absorb time-invariant variation across industries. We will also conduct robustness tests which ensure that our results are not driven by any particular industry.

Table 3 reports sample averages and standard deviations for our financial control variables. Table 5 decomposes variation in our panel data into between and within stock variation. For both news and social media, the majority of variation occurs between stocks.

Table 4 shows contemporaneous correlations, with p-values in parentheses. Social and news media buzz are strongly correlated for a given stock within the same period. Buzz also correlates with size, and this correlation is stronger for news than for social media. There is also a strong correlation between buzz and turnover, especially for social media. The contemporaneous correlation between volatility and buzz is positive for social media, and negative (but close to zero) for news media.

Figure 1 shows 30-day moving averages of the *total* news and social media buzz of all stocks in our sample. At the market level, both types of buzz go through noticeable swings, and they are positively but not perfectly correlated. Figure 2 shows a moving av-

[7] Before we winsorize, we observe stocks which have relative buzz close to 10% on individual days.

erage of relative buzz for two industries (technology and utilities), while Figure 3 shows the same timelines for two individual firms (Walmart and Apple).

3 Empirical Results

3.1 Volatility

We run a panel regression of each stock i's volatility next month $(t+1)$ on this month's (t) buzz in social and news media, $BuzzS_{i,t}$ and $BuzzN_{i,t}$, stock-level control variables $X_{i,t}$ (described above in Sub-section 2.3), and this month's volatility $iVolp_{i,t}$. We control for stock and time fixed effects α_i and μ_t:

$$iVolp_{i,t+1} = \alpha_i + \mu_t + \beta_S BuzzS_{i,t} + \beta_N BuzzN_{i,t} + \gamma X_{i,t} + \delta \times iVolp_{i,t} + \epsilon_{i,t+1}. \quad (1)$$

Table 6 reports the results from estimating Equation (1). The most complete specification is presented in column (4).

High news media buzz predicts lower subsequent volatility and this is statistically significant at the 1% level. The effect is robust to the inclusion of time and stock fixed effects, as well as our other controls. Social media buzz predicts higher future volatility, and this relationship becomes statistically significant once we control for stock characteristics. These results are robust to the inclusion of sentiment controls and the dispersion of analyst opinions (AnalystDisp), a common measure of disagreement among investors.[8] The variables are normalized for the ease of interpretation: an increase in news media buzz in month t by 1 standard deviation decreases volatility in month $t+1$ by about 0.05 of a standard deviation; an increase in social media buzz in month t by 1 standard deviation increases volatility in month $t+1$ by about 0.04 of a standard deviation.

[8] The number of observations drops as we include our controls due to missing observations.

To further interpret the economic significance of buzz on volatility, suppose that a stock goes from having no buzz in social media to being one of the most talked-about stocks, with a relative buzz of 0.5%. Then, according to our preferred specification, our measure of the stock's subsequent idiosyncratic volatility rises on average by 0.31 of a standard deviation or 54% of average volatility. For an equivalent change in news media buzz, the stock's subsequent volatility falls by 0.39 of a standard deviation or 68% of average volatility.

A potential concern is that our estimated coefficients, with opposite signs on news and social media, are driven by the strong positive contemporaneous correlation between news and social media buzz.[9] To check whether this is the case, we introduce the two measures separately in the last two columns. This does not significantly affect the sign, magnitude or significance of the coefficients, which suggests that contemporaneous correlation is not driving our results.

3.2 Turnover

For trading activity, as measured by turnover, we estimate the analogous panel regression:

$$Turn_{i,t+1} = \alpha_i + \mu_t + \beta_S BuzzS_{i,t} + \beta_N BuzzN_{i,t} + \gamma X_{i,t} + \delta \cdot Turn_{i,t} + \epsilon_{i,t+1}. \qquad (2)$$

Table 7 reports our estimates.

High social media buzz predicts high subsequent turnover, and high news media buzz predicts low subsequent turnover. This effect is statistically significant at the 1% level, and does not change significantly when we include stock, sentiment and disagreement controls. Introducing social and news media buzz separately in the last two columns affects our estimates only marginally, suggesting that they are not driven by

[9] Intuitively, if Corr $[BuzzN_{i,t}, BuzzS_{i,t}] \approx 1$, then estimating Equation 1 may yield $\beta_S = -\beta_N > 0$, even if the true coefficients were $\beta_S = \beta_N = 0$.

the contemporaneous correlation between social and news media. Our preferred specification (column (4)) suggests that a 1 standard deviation increase in news media buzz decreases turnover by 0.065 standard deviations, while a 1 standard deviation increase in social media buzz increases turnover by 0.04 standard deviations.

To interpret the economic effects of buzz on turnover, suppose again that relative social media buzz around a stock rises from zero to the level of the most talked-about stocks at 0.5%. Then according to our preferred specification in column (4), the stock's subsequent turnover increases on average by 0.32 of a standard deviation or 30% of average turnover. For an equivalent rise in news media buzz, turnover decreases on average by 0.51 of a standard deviation or 47% of average turnover.

3.3 Robustness Checks

Tables 8 and 9 show the results of estimating Equations (1) and (2) on an unbalanced panel of data (i.e., including stocks which are not present during our balanced-panel sample). The estimated effects of buzz are quantitatively similar to the baseline model, and significant at the 1% level. Table 10 shows the results using the non-parametric measure of volatility ("iVoln", described in Section 2.2). Again, the estimates are quantitatively similar to the baseline model, and significant at the 1% level. Tables 11 and 12 repeat the estimation of our preferred specification in sub-samples where each industry (healthcare, technology, etc.) is removed in turn. Tables 13 and 14 perform the same exercise, but this time excluding stocks in each quartile of market capitalization. In both exercises, the estimated coefficients remain significant and stable, showing that the results are not being driven by any particular industry or by stocks with particularly large or small market capitalization.

We also conducted additional robustness checks by including variables that control for market conditions, including the S&P500 return and VIX index. These are reported in Table 20 of Appendix C.

Since our model was derived using the number of signals, or the absolute amount of information, instead of relative, we also conducted additional regressions using absolute buzz, and find that our results still hold. These are reported in Table 21 of Appendix C.

3.4 Vector Autoregression

To further evaluate the dynamic interaction between buzz and market outcomes, we estimate a panel vector autoregression (VAR) which includes four endogenous variables: news and social media buzz, return volatility and turnover. This will allow us to test for Granger causality between these endogenous variables. We use a balanced panel, include two lags of the endogenous variables, a set of exogenous control variables, and month and stock fixed effects.[10] The model also includes Size, InstOwn, Return and SentN as exogenous variables and stock-specific controls. Other controls were also included but were insignificant and did not change our main results. Standard deviations are clustered at the stock level.

Table 15 shows the estimated panel VAR coefficients. The results are easier to visualize via the impulse response functions in Figure 4. The estimated effects are similar to our baseline panel regressions: An increase in social media buzz (shown in the right two panels of the Figure) is associated with a significant increase in subsequent volatility and turnover which declines over time. An increase in news media buzz (in the left panels) is associated with a significant decrease in volatility. The effect of news media buzz on turnover has a negative point estimate but is imprecisely estimated.

[10]To control for fixed effects in a computationally feasible manner, we time-demean the endogenous variables to account for time fixed effects, and apply a Helmert transformation to create forward mean differenced forms which remove stock fixed effects. For a vector of endogenous variables $\hat{z}_{i,t}$, in a panel of time periods $t = 1, ..., T$, the Helmert-transformed endogenous variables are

$$\mathbf{z}_{i,t} = \sqrt{\frac{T-t}{T-t+1}}(\hat{\mathbf{z}}_{it} - \frac{1}{T-t}\sum_{n=t+1}^{T}\hat{\mathbf{z}}_{in}).$$

Based on our estimates, we test for Granger causality among the endogenous variables. The results are presented in Table 16. The null hypothesis is that for two endogenous variables i and j, a contemporaneous increase in i does not predict a significant subsequent change in j. We reject this hypothesis for all pairs (i, j) at the 1% level, with two exceptions: a shock to social media buzz does not predict a significant change in news media buzz (p-value 0.054), and the impact of turnover on social media buzz is significant only at the 5% level (p-value 0.039).

3.5 Market-level Effects

Our analysis so far has focused on stock-level news and social media buzz, stock-level trading activity and idiosyncratic volatility. We now examine whether the same effects are present at the market level, i.e. whether the *total buzz* surrounding all stocks, in either social or news media, has predictive power for aggregate volatility and trading activity.

To study the effect of media buzz on market return volatility, we obtain a daily time series of market return volatility from a generalized autoregressive conditional heteroskedasticity (ARCH) models. We use the GJR-GARCH model to capture the potential leverage effect, i.e. the asymmetry in the effect of positive and negative returns on volatility (see, for example, Duffee, 1995).

We use two alternative series for market return: the value-weighted return from CRSP (VWRet), and the return on the S&P 500 (SPRet). We construct market-level measures of buzz in news media (MktBuzzN) and social media (MktBuzzS) by summing up the total buzz for all individual stocks on each day. For analyzing SPRet, we generate measures of buzz (SPBuzzN and SPBuzzS) which aggregate total buzz only for S&P 500 stocks.

We report the results in Table 17. The negative effect of news buzz and the positive effect of social buzz on volatility can also be found at the market level, using both all stocks or only S&P 500 stocks. Our analysis of turnover, by contrast, did not yield significant results at the market level.

4 Asset Pricing Model

We now present a theoretical model of asset pricing with imperfect information, which rationalizes the empirical patterns we find. As a baseline, one might consider a model where all traders are symmetric and rational, and news media and social media both provide informative, public signals to investors. It is immediate that such a model would be inconsistent with our results. Indeed, we have shown that social and news media buzz predict opposite developments in stock turnover and return volatility.

In this section, we propose an alternative model where social media acts as an "echo chamber". In section 4.3, we discuss alternative explanations, including models which view social media as private information, which feature other behavioral biases, and in which buzz is endogenously determined.

Consider an economy with three dates $t \in \{0, 1, 2\}$ and two generations of traders, which are born at dates 0 and 1 respectively, and who we call generation 0 and 1.[11] There is a unit mass of traders indexed by i. Trader i of generation t lives for one period and maximizes the expected utility of future wealth, $E_{it}[u(W_{t+1})]$, where utility $u(.)$ is CARA with risk tolerance 1. The risk-free interest rate is 1 and the future is not discounted. There is one risky asset, in zero net supply, which yields a final payoff θ at date 2. At date 0 all traders have a common prior $\theta \sim \mathcal{N}(m_0, \frac{1}{\rho_0})$.[12]

At date 1 there are N public signals from traditional news media, of the form $s_n =$

[11] In principle, the arguments below extend to an economy with T trading dates, which could be solved using recursive methods as in He and Wang (1995). Since the key effects arise as long as there is an intermediate date at which information can arrive, we focus on a three-period setting.

[12] There are no noise traders who trade for purely exogenous reasons. While noise traders affect the overall volume of trade and return volatility, they do not affect the *changes* in these measures when the number of social and news media signals changes. Since these derivatives are the subject of our empirical analysis, we omit noise traders from the baseline model.

$\theta + \epsilon_n$, $n = 1, ..., N$, where $\epsilon_n \sim \mathcal{N}(0, \frac{1}{\rho_e})$. There are also K signals from social media: $e_k = \theta + \epsilon_k$, with $k = 1, ..., K$, where $\epsilon_k \sim \mathcal{N}(0, \frac{1}{\rho_e})$. Our measures of news and social media buzz are N and K. Let the average values of news and social media signals be $\bar{s} = \frac{1}{N} \sum_n s_n$ and $\bar{e} = \frac{1}{K} \sum_k e_k$ respectively. For ease of exposition, we restrict attention to a single stock and consider the total number of social and news media signals (N and K) about that stock.[13]

To analyze "echo chamber" effects, we make two assumptions. First, the K social media signals are repetitions of the N true signals, so they do not contain new information. Because they are repetitions, we will assume that $\bar{e} = \bar{s}$.[14] Second, traders in generation 1 are heterogeneous in how they process information: There is a mass λ of behavioral traders and $1 - \lambda$ rational traders. Rational traders ignore social media signals, but behavioral traders treat them as genuinely new information with precision ρ_e. In other words, behavioral traders misinterpret the data-generating process, and act as if the social media errors ϵ_k were independent of the news media errors ϵ_n.

Under these assumptions, rational traders' posterior mean at date 1 is

$$m_1 = m_0 + w\left(\bar{s} - m_0\right),$$

where $w = \frac{N\rho_e}{\rho_0 + N\rho_e}$ is the rational updating weight on news media. Thus, rational traders update based only on the average news media signal $\bar{s}$. By contrast, behavioral traders further update based on the average social media signal $\bar{e}$, for the posterior mean

$$\hat{m}_1 = m_1 + \hat{w}\left(\bar{e} - m_1\right),$$

where $\hat{w} = \frac{K\rho_\epsilon}{\rho_1 + K\rho_\epsilon}$ is the behavioral updating weight on social media. Due to their interpretation of social media as new information, behavioral traders also have tighter

[13]Controlling for total market-level buzz (or for time fixed effects), N and K have a one-for-one mapping into the *relative* buzz of a given stock i, which we included in our regressions.

[14]For instance, we suppose that each signal n is repeated ψ times, for a total of $K = \psi N$ echoes.

posteriors than rational ones. In particular, the posterior precision is $\rho_1 = \rho_0 + N\rho_e$ for rational traders and $\hat{\rho}_1 = \rho_1 + K\rho_e$ for behavioral traders.

4.1 Equilibrium Prices

The price at date 0 is $p_0 = E_0[p_1] = m_0$, and the price at date 2 equals the final payoff $p_2 = \theta$. At date 1, the demands of each of the rational and behavioral traders respectively are $x_1 = \rho_1(m_1 - p_1)$ and $\hat{x}_1 = \hat{\rho}_1(\hat{m}_1 - p_1)$. The equilibrium price at date 1 must clear the market, so it solves $(1-\lambda)x_1 + \lambda\hat{x}_1 = 0$, or

$$p_1 = m_1 + \eta(\bar{e} - m_1),$$

$$\eta = \frac{\lambda K\rho_e}{\rho_1 + \lambda K\rho_e}.$$

The term $\eta(\bar{e} - m_1)$ is the distortion introduced by behavioral traders. Without social media echoes, $K = \eta = 0$. This disagreement occurs to the extent that the average echo $(\bar{e})$ disagrees with the rational posterior (m_1). Intuitively, when $\bar{e} > m_1$, or equivalently when the average signal $\bar{s}$ exceeds the prior mean m_0, behavioral traders generate excess demand due to their response to social media, and the price rises above fundamental values. The distortion in price is smaller when the rational posterior is precise (large ρ_1), since this implies that behavioral traders update relatively little beyond what rational traders do. The distortion is naturally increasing in the share of behavioral traders λ and social media buzz K.

4.2 Empirical Predictions

Our empirical analysis focuses on the behavior of average stock turnover and the volatility of stock returns following the arrival of information ("buzz") from social and news media. In the model, information in the model arrives at date 1. The subsequent stock turnover[15] is

$$T_2 = (1-\lambda)|x_1|.$$

The subsequent return is

$$R_2 = p_2 - p_1,$$

and return volatility is measured by its variance.[16] We show that the comparative statics of the model match our empirical findings:

Proposition 1. *Expected turnover between dates* 1 *and* 2

$$\mathbb{E}\left[T_2\right] \quad = \quad (1-\lambda)\eta\rho_1\mathbb{E}\left[|\bar{e}-m_1|\right] \tag{3}$$

is increasing in K *and decreasing in* N*. The subsequent volatility of returns*

$$\mathbb{V}\left[R_2\right] \quad = \mathbb{V}[p_2 - p_1] = \quad \mathbb{V}\left[\theta - \bar{s}w - \bar{e}\eta\left(1-w\right)\right] \tag{4}$$

is increasing in K *and decreasing in* N*.*

These results have intuitive interpretations. For turnover, note that T_2 is driven by confidence and disagreement. In the absence of social media echoes ($\eta = 0$), all traders would have the same posterior beliefs so there would be no trade. More generally, posterior means will differ because the average echo ($\bar{e}$) differs from the posterior mean of rational individuals (m_1), hence the term $\mathbb{E}\left[|\bar{e}-m_1|\right]$. Similarly, turnover vanishes when $\lambda = 0$ or $\lambda = 1$ since disagreement vanishes at these extreme points (notice that $\eta = \frac{\lambda K \rho_e}{\rho_1 + \lambda K \rho_e}$). Finally, turnover increases with confidence: if posterior beliefs are tighter (high ρ_1), trade will be more aggressive. Regarding the volatility of subsequent returns $\mathbb{V}\left[R_2\right] = \mathbb{V}[p_2 - p_1]$, note that $p_2 - p_1$ is volatile whenever the time 1 price p_1 is not aligned

[15] Since we may normalize the number of outstanding shares without loss of generality, T_2 also captures the stock's "turnover" which we will measure in the data.

[16] As is standard in the asset pricing literature based on CARA-Gaussian setups, we consider the per-unit-of-stock return $p_{t+1} - p_t$ instead of the per-dollar return p_{t+1}/p_t (e.g. He and Wang (1995)). Our results on volatility remain valid up to a first-order approximation when the per-dollar return is considered.

with the true fundamental value $p_2 = \theta$ which is revealed at date 2. More social media buzz K drives a wedge between p_1 and θ because behavioral traders grow in confidence and are willing to buy and sell at prices that are misaligned with true fundamentals. Thus, the variance of subsequent returns increases in K. By contrast, more news media signals N offset this effect and therefore lead to a reduction in the variance of subsequent returns.[17]

These predictions are robust to our assumptions about behavioral traders' attention to news media. So far, we have assumed that behavioral traders pay attention to news as well as social media. Similar conclusions arise in a model where behavioral traders consider only social media signals, which we analyze in Appendix B.5.

Since our model has only one trading period, we cannot make predictions about the persistence of these effects. The intuition for a multi-period model would be clear: on one hand, if the underlying data generating process is stationary, traders' beliefs will tend to converge again in the long run, so that there is reversal. On the other hand, if social media generates disagreement in the short run, this disagreement will persist for a few periods. Our vector autoregression analysis suggests that the data is consistent with both of these effects. Indeed, the impulse responses shown in Figure 4 appear to revert towards zero over time. For example, the strongest effect of social media buzz is felt after about one period for volatility, and after five periods for turnover.

[17] The model's predictions regarding the contemporaneous return volatility $\mathbb{V}[R_1] = \mathbb{V}[p_1 - p_0]$ are less clear-cut. More social media buzz K increase disagreement and the confidence of behavioral traders, both of which increase the volatility of prices p_1 and, hence, the volatility of contemporaneous returns. More news media signals N create two competing effects. First, since all traders agree about the distribution of news signals, a higher ratio N/K reduces the relative disagreement in the economy and dampens price volatility. Second, as in standard models, more genuine information raises the variance of all traders' posterior beliefs m_1 because prices respond to information, which increases price volatility. The first effect runs into diminishing returns when N/K is large, in which case the second effect dominates. A formal analysis is available on request.

4.3 Alternative Mechanisms

We have argued in this section that a model of social media as an "echo chamber" is consistent with our empirical findings. In this section we discuss possible alternative explanations. We present a formal analysis in the Appendix B.

First, it is possible that social media conveys genuine information (unlike in the "echo chamber" model), but that its interpretation by behavioral traders is subject to other types of behavioral biases. Our linear-Gaussian setup can be adapted to allow for some common biases in the behavioral finance literature. The cases we consider in the appendix are as follows:

1. *Overconfidence:* Behavioral investors overstate the precision of their signals as in Scheinkman and Xiong (2003). In this case, the precision of social media signals is ρ_ϵ but behavioral investors perceive the precision to be $(1+a)\,\rho_e$, where $a > 0$ measures overconfidence.

2. *Conservatism:* Behavioral investors overstate the precision of their prior beliefs as in Barberis et al. (1998). In this case, behavioral investors attach precision $(1+b)\,\rho_0$ to their prior, with $b > 0$ measuring conservatism.

3. *Rational inattention:* Behavioral investors observe signals with cognitive noise, which must satisfy a constraint on entropy reduction as in Peng and Xiong (2006) and Kacperczyk et al. (2016).

4. *Confirmation bias*: Behavioral investors ignore signals that do not conform with their prior sentiment as in Rabin and Schrag (1999). Specifically, behavioral traders with an optimistic predisposition process negative signals $s_i < m_0$ as if they were equal to the prior mean m_0.

We show in the appendix that, for all of these biases, an increase in the number of social media signals tends to reduce volatility, because social media buzz is truly informative in this class of models. Turnover, however, is increased in most cases because social media signals increases the confidence of investors who bet on perceived disagreements. This is not consistent with our empirical results on volatility. Hence, although we are unable to rule out all possible alternative behavioral theories, our empirical results seem to favor our model over a subclass of behavioral models where social media is informative but wrongly interpreted.

Second, heightened social or news media activity could arise endogenously in response to increases in prior uncertainty (decreases in ρ_0) or increases in prior disagreement (increased dispersion in traders' prior mean μ_0) about a stock's payoff. A rigorous treatment of endogenous buzz is beyond the scope of this paper, but these effects are possibly consistent with our results. For example, more disagreement is associated in standard models with higher volatility and turnover. Hence, one can imagine our results being generated by a mechanism where disagreement shocks cause both social media (but not news media) activity in period t, and then subsequent high volatility and turnover in period $t+1$. To test the "echo chamber" model against this alternative, once would require quasi-experimental variation in social and news media coverage, which our dataset does not permit.

While some of the models described above are falsified by our empirical findings, the variation in our data do not allow us to pinpoint the causal mechanism behind the patterns we find. It is therefore possible that the patterns we find are the outcomes of a different mechanism. We now briefly explore other potential mechanisms, although a deeper treatment of these is deferred to future research.

First, it is possible that all individuals are rational, but those who pay attention to social media have different preferences from others. A sufficiently rich form of heterogeneity, and its correlation with attention to different news sources, could in principle generate the empirical patterns we observe. While our empirical results are robust to controlling for a number of observable stock characteristics, we do not observe any trader characteristics, and therefore cannot analyze which agents pay attention to one or the other news source. For instance, individuals who pay attention to social media might be risk seeking while those who pay attention to news media might be risk averse. Individuals could also differ in liquidity constraints or trading horizons as in Kondor (2012).

Second, the empirical patterns we observe could be due to heterogeneity in prior beliefs. Heterogeneity in the mean and variance of beliefs in a simple CARA-Gaussian framework is unlikely to rationalize the data, but we cannot eliminate this possibility in richer models where higher moments of beliefs matter for agent decisions.

Third, it is possible that news and social media broadcast different types of signals. For instance, social media might conveys only extreme signals (realizations greatly above or below the mean θ_0), whereas news media conveys only moderate signals. Indeed, Figures 5 and 6 in Appendix D shows that there are some systematic differences between the two media sources. For instance, news media is more likely to contain facts, while social media is more likely to contain emotions.

Finally, the formalism of our model could be given a slightly different economic interpretation: One could think of a model where social media signals are repetitions of news media signals, but behavioral traders have a bias that leads them to overreact to this particular source of information. This would lead to similar qualitative predictions to our model, where behavioral traders also effectively overreact to social media.

5 Conclusions

We have analyzed a large dataset collecting coverage of stocks in traditional news media and social media. We find the following robust stylized facts: High social media coverage at the stock level predicts high subsequent return volatility and trading activity, while high news media coverage predicts the opposite. It further appears that news media activity around a stock is a leading indicator of social media coverage. This paper is among the first to directly compare news and social media. The main contribution is to demonstrate that social and news media have different relationships with stock prices. This insight motivates the development of new theories of social and news media coverage that can explain our findings.

We have briefly discussed one such theory. We augment a standard model of trading to view social media is an “echo chamber”, where boundedly rational agents fail to account correctly for the repetition of information. This model is consistent with our empirical findings. Moreover, it is consistent with a growing literature in finance which demonstrates that investors react to “stale news”. We have also argued that alternative behavioral biases cannot easily generate the same price patterns. Due to a lack of quasi-experimental variation in media coverage in our data, we cannot conclude with certainty that this mechanism is at play. Rather, we believe that this model provides a reasonable starting point for future empirical work.

References

ANTWEILER, W. AND M. Z. FRANK (2004): "Is all that talk just noise? The information content of internet stock message boards," *Journal of Finance*, 59, 1259–1294.

BARBERIS, N., A. SHLEIFER, AND R. VISHNY (1998): "A model of investor sentiment," *Journal of Financial Economics*, 49, 307–343.

BUSHEE, B. J., J. E. CORE, W. GUAY, AND S. J. HAMM (2010): "The role of the business press as an information intermediary," *Journal of Accounting Research*, 48, 1–19.

CHEN, H., P. DE, Y. J. HU, AND B.-H. HWANG (2014): "Wisdom of crowds: The value of stock opinions transmitted through social media," *Review of Financial Studies*, 27, 1367–1403.

CHEUNG, Y. AND L. K. NG (1992): "Stock price dynamics and firm size: An empirical investigation," *Journal of Finance*, 47, 1985–1997.

COVAL, J. D. AND T. SHUMWAY (2001): "Is sound just noise?" *Journal of Finance*, 56, 1887–1910.

DA, Z., J. ENGELBERG, AND P. GAO (2015): "The sum of all FEARS: Investor sentiment and asset prices," *Review of Financial Studies*, 28, 1–32.

DEMARZO, P. M., D. VAYANOS, AND J. ZWIEBEL (2003): "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, 118, 909–968.

DENNIS, P. J. AND D. STRICKLAND (2002): "Who blinks in volatile markets, individuals or institutions?" *The Journal of Finance*, 57, 1923–1949.

DUFFEE, G. R. (1995): "Stock returns and volatility: A firm-level analysis," *Journal of Financial Economics*, 37, 399–420.

FAMA, E. F. AND K. R. FRENCH (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

FANG, L. AND J. PERESS (2009): "Media coverage and the cross-section of stock returns," *Journal of Finance*, 64, 2023–2052.

GENTZKOW, M. AND J. M. SHAPIRO (2011): "Ideological Segregation Online and Offline," *Quarterly Journal of Economics*, 126.

HAWKINS, S. AND S. HOCH (1992): "Low-involvement learning: Memory without evaluation," *Journal of Consumer Research*, 19, 212–225.

HE, H. AND J. WANG (1995): "Differential information and dynamic behavior of stock trading volume," *The Review of Financial Studies*, 8, 919–972.

HUBERMAN, G. AND T. REGEV (2001): "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar," *Journal of Finance*, 56, 387–396.

KACPERCZYK, M., S. VAN NIEUWERBURGH, AND L. VELDKAMP (2016): "A rational theory of mutual funds' attention allocation," *Econometrica*, 84, 571–626.

KARPOFF, J. M. (1987): "The relation between price changes and trading volume: A survey," *Journal of Financial and quantitative Analysis*, 22, 109–126.

KONDOR, P. (2012): "The more we know about the fundamental, the less we agree on the price," *Review of Economic Studies*, 79, 1175–1207.

LO, A. W. AND J. WANG (2000): "Trading volume: Definitions, data analysis, and implications of portfolio theory," *Review of Financial Studies*, 13, 257–300.

MICHAELIDES, A., A. MILIDONIS, AND G. P. NISHIOTIS (Forthcoming): "Private information in currency markets," *Journal of Financial Economics*.

MICHAELIDES, A., A. MILIDONIS, G. P. NISHIOTIS, AND P. PAPAKYRIAKOU (2015): "The adverse effects of systematic leakage ahead of official sovereign debt rating announcements," *Journal of Financial Economics*, 116, 526–547.

MILGROM, P. AND N. STOKEY (1982): "Information, trade and common knowledge," *Journal of Economic Theory*, 26, 17–27.

MUTHEN, B. (1990): "Moments of the censored and truncated bivariate normal distribution," *British Journal of Mathematical and Statistical Psychology*, 43, 131–143.

PENG, L. AND W. XIONG (2006): "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, 80, 563–602.

RABIN, M. AND J. L. SCHRAG (1999): "First impressions matter: A model of confirmatory bias," *Quarterly Journal of Economics*, 114, 37–82.

REUTERS INSTITUTE FOR THE STUDY OF JOURNALISM (2016): "Digital News Report," Tech. rep., .

SCHEINKMAN, J. A. AND W. XIONG (2003): "Overconfidence and speculative bubbles," *Journal of Political Economy*, 111, 1183–1220.

SCHWERT, G. W. (1989): "Why does stock market volatility change over time?" *Journal of Finance*, 44, 1115–1153.

SUN, L., M. NAJAND, AND J. SHEN (2016): "Stock return predictability and investor sentiment: A high-frequency perspective," *Journal of Banking and Finance*, 73, 147–164.

TETLOCK, P. C. (2007): "Giving content to investor sentiment: The role of media in the stock market," *Journal of Finance*, 62, 1139–1168.

——— (2011): "All the news that's fit to reprint: Do investors react to stale information?" *Review of Financial Studies*, 24, 1481–1512.

TKAC, P. A. (1999): “A trading volume benchmark: Theory and evidence,” *Journal of Financial and Quantitative Analysis*, 34, 89–114.

Figures

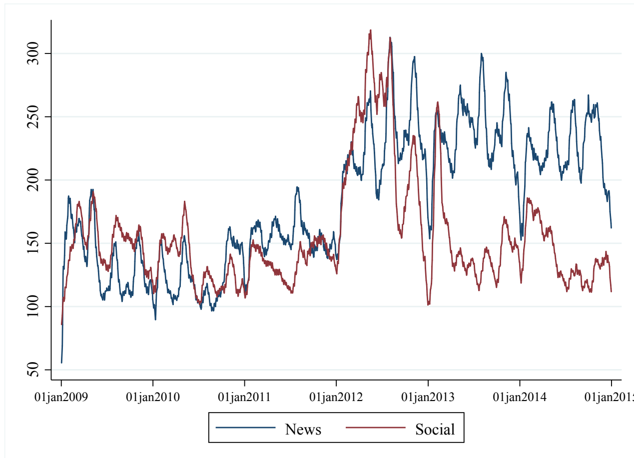


Figure 1: Moving average of total news and social media buzz at the market level, measured in thousands.

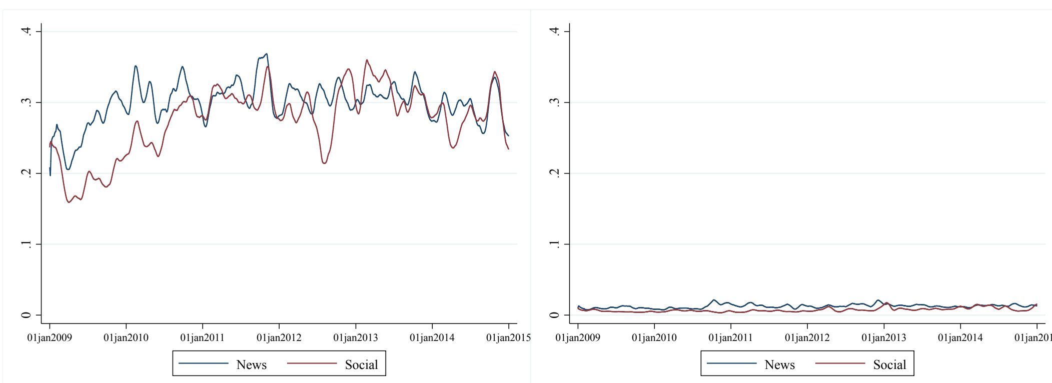


Figure 2: Moving average of relative buzz by industry. On the left, technology. On the right, utilities.

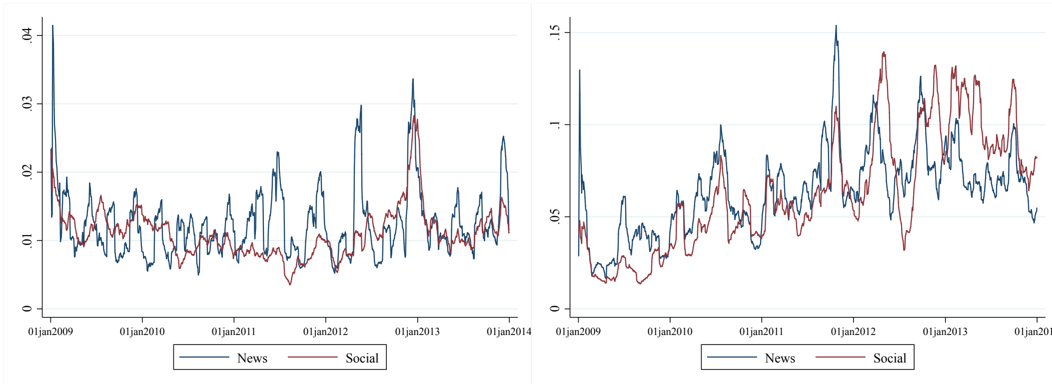


Figure 3: Moving average of relative buzz for two individual stocks. On the left, Walmart. On the right, Apple.

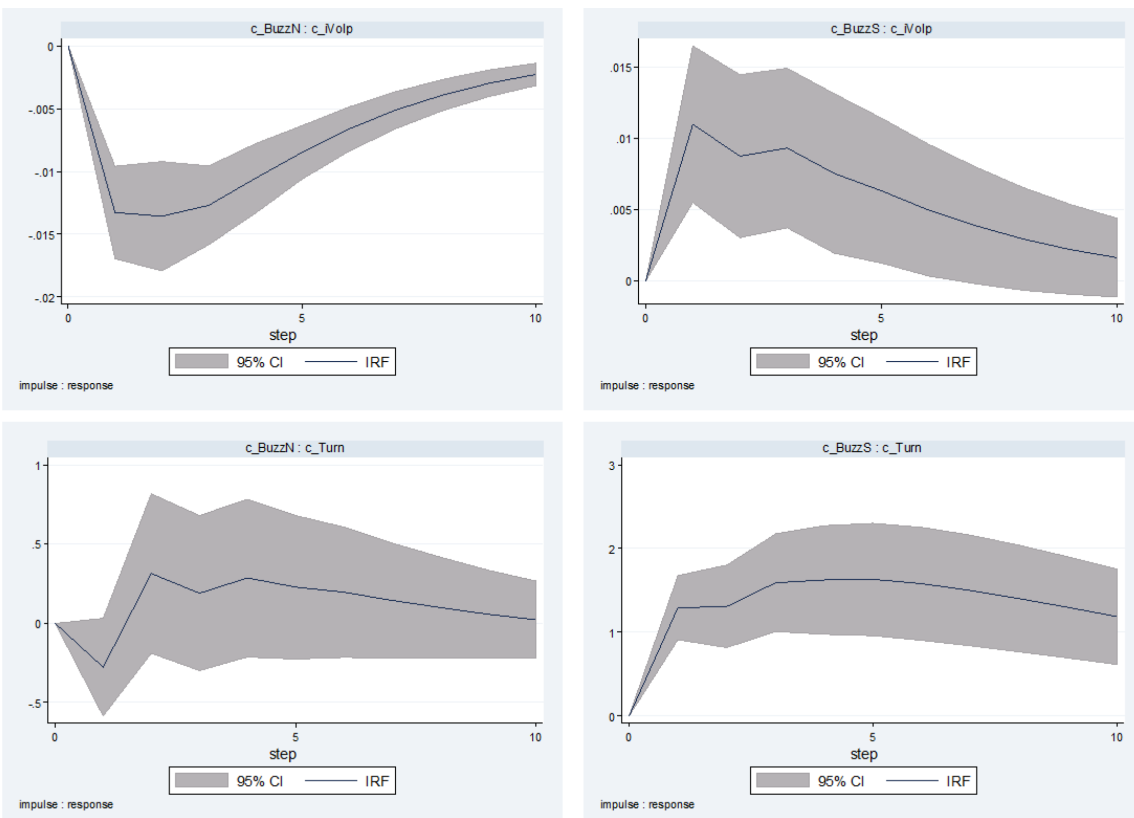


Figure 4: Impulse response functions. Top left: effect of BuzzN on volatility; Top right: effect of BuzzS on volatility Bottom left: effect of BuzzN on turnover; Bottom right: effect of BuzzS on turnover

Tables

Table 1: Variable definitions.

Variable	Definition	Calculation
iVolp	idiosyncratic volatility (parametric measure)	sum of squared residuals in Fama-French regression using daily data within a month
iVoln	idiosyncratic volatility (nonparametric measure)	variance of daily returns within a month
Turn	turnover	share trading volume divided by number of outstanding shares
Size	market value of equity	absolute value of share price multiplied by number of outstanding shares
InstOwn	institutional ownership	the fraction of shares held by 13F institutional investors
HHI	Herfindahl-Hirschman index	the Herfindahl-Hirschman index of sales at the Fama-French 48 industry level at an annual frequency
Leverage	leverage	the ratio of total long-term debt and debt in current liabilities over the sum of the numerator and shareholders' equity
Return	return	the change in the total value of an investment in a common stock over a month per dollar of initial investment, calculated using the closing price of each month
AbsReturn	absolute return	absolute value of Return
TotalSD	total standard deviation	standard deviation of monthly return in the last 60 months
AnalystDisp	analyst forecast dispersion	the natural log of one plus the standard deviation of analyst earnings-per-share forecasts, normalized by the absolute value of mean forecast in a given month
SentN	news media sentiment	relative positive sentiment net of negative sentiment in news media
SentN(-)	news media sentiment (negative)	relative positive sentiment net of negative sentiment in news media (only negative values)
SentS	social media sentiment	relative positive sentiment net of negative sentiment in social media
SentS(-)	social media sentiment (negative)	relative positive sentiment net of negative sentiment in social media (only negative values)

Table 2: Summary statistics for buzz, volatility and turnover by industry.

Industry	BuzzN	BuzzS	iVolp	Turn	Industry	BuzzN	BuzzS	iVolp	Turn	
CYC	0.04	0.03	0.01	2.33	MAT	0.02	0.02	0.01	2.20	Mean
	0.08	0.08	0.02	1.97		0.06	0.06	0.01	2.06	SD
	0.53	0.65	0.13	11.01		0.53	0.65	0.13	11.01	Max
	0.00	0.00	0.00	0.11		0.00	0.00	0.00	0.11	Min
	24747	24747	24747	24747		8309	8309	8309	8309	N
ENE	0.02	0.03	0.01	3.01	NCY	0.03	0.02	0.01	1.83	Mean
	0.05	0.07	0.02	2.36		0.05	0.06	0.01	1.65	SD
	0.53	0.65	0.13	11.01		0.53	0.65	0.13	11.01	Max
	0.00	0.00	0.00	0.11		0.00	0.00	0.00	0.11	Min
	8725	8725	8725	8725		8512	8512	8512	8512	N
FIN	0.03	0.02	0.01	1.51	TEC	0.04	0.05	0.01	2.18	Mean
	0.05	0.07	0.01	1.49		0.09	0.12	0.02	1.87	SD
	0.53	0.65	0.13	11.01		0.53	0.65	0.13	11.01	Max
	0.00	0.00	0.00	0.11		0.00	0.00	0.00	0.11	Min
	10219	10219	10219	10219		21478	21478	21478	21478	N
HLC	0.02	0.05	0.01	2.05	TEL	0.06	0.05	0.01	1.49	Mean
	0.04	0.11	0.02	1.82		0.15	0.10	0.02	1.20	SD
	0.53	0.65	0.13	11.01		0.53	0.65	0.13	11.01	Max
	0.00	0.00	0.00	0.11		0.00	0.00	0.00	0.13	Min
	19275	19275	19275	19275		1590	1590	1590	1590	N
IND	0.03	0.02	0.01	1.60	UTL	0.00	0.00	0.02	0.95	Mean
	0.07	0.05	0.01	1.26		0.00	0.01	0.03	1.34	SD
	0.53	0.65	0.13	11.01		0.01	0.03	0.13	11.01	Max
	0.00	0.00	0.00	0.11		0.00	0.00	0.00	0.21	Min
	24297	24297	24297	24297		76	76	76	76	N
Total	0.03	0.03	0.01	2.05	Mean					
	0.07	0.09	0.02	1.83	SD					
	0.53	0.65	0.13	11.01	Max					
	0.00	0.00	0.00	0.11	Min					
	133056	133056	133056	133056	N					

Notes: This table reports summary statistics of BuzzN, BuzzS, iVolp and Turn for each industry and for the whole sample. Industries are classified according to the Thomson Reuters Business Classification (TRBC) 10 economic sectors.

Table 3: Summary statistics for financial control variables.

	Size	InstOwn	Ret	HHI	Leverage	TotalSD	AnalystDisp
Mean	13.8058	0.6972	0.0196	0.7389	0.0258	0.1430	0.2932
SD	1.7366	0.2438	0.1274	0.2923	0.0546	0.0632	0.1287
Max	18.3953	1.0000	0.4822	1.0000	0.3417	0.3791	0.5336
Min	10.1016	0.0597	-0.3136	0.0000	0.0000	0.0489	0.0000
N	159358	162268	161526	141532	151185	162220	142005

Table 4: Contemporaneous correlations.

Variables	BuzzN	BuzzS	iVolp	Turn	Size	InstOwn	Ret	HHI	Leverage	TotalSD
BuzzN	1.000									
BuzzS	0.554	1.000								
	(0.000)									
iVolp	-0.032	0.100	1.000							
	(0.000)	(0.000)								
Turn	0.133	0.307	0.258	1.000						
	(0.000)	(0.000)	(0.000)							
Size	0.409	0.230	-0.386	0.176	1.000					
	(0.000)	(0.000)	(0.000)	(0.000)						
InstOwn	0.012	-0.085	-0.174	0.235	0.339	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
Ret	-0.001	-0.010	0.120	0.023	-0.042	-0.010	1.000			
	(0.691)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
HHI	-0.076	-0.008	0.076	0.049	-0.173	-0.052	0.003	1.000		
	(0.000)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)	(0.242)			
Leverage	0.059	0.051	0.065	-0.008	-0.027	-0.044	0.007	-0.026	1.000	
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.006)	(0.000)		
TotalSD	-0.132	0.077	0.331	0.185	-0.470	-0.166	0.020	0.109	0.077	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Table 5: Summary statistics between and within stocks.

Variable		Mean	Std. Dev.	Min	Max	Observations
BuzzN	overall	0.030	0.069	0	0.532	N = 133056
	between		0.063	0	0.532	n = 1848
	within		0.028	-0.258	0.517	T = 72
BuzzS	overall	0.030	0.087	0	0.650	N = 133056
	between		0.075	0	0.650	n = 1848
	within		0.043	-0.452	0.667	T = 72
iVolp	overall	0.010	0.018	0.000	0.131	N = 133056
	between		0.009	0.001	0.066	n = 1848
	within		0.015	-0.052	0.138	T = 72
Turn	overall	2.035	1.832	0.106	11.006	N = 133056
	between		1.397	0.113	10.616	n = 1848
	within		1.186	-4.206	12.273	T = 72

Table 6: Volatility with a balanced panel.

	(1)	(2)	(3)	(4)	(5)	(6)
	iVolp(+1)	iVolp(+1)	iVolp(+1)	iVolp(+1)	iVolp(+1)	iVolp(+1)
iVolp	0.256***	0.212***	0.213***	0.181***	0.182***	0.179***
	(26.53)	(21.36)	(21.40)	(15.96)	(16.16)	(15.79)
BuzzN	-0.0679***	-0.0612***	-0.0578***	-0.0517***	-0.0445***	
	(-7.55)	(-6.75)	(-6.49)	(-6.34)	(-5.42)	
BuzzS	0.00607	0.0503***	0.0514***	0.0374***		0.0296***
	(0.58)	(4.34)	(4.43)	(3.53)		(2.85)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock Controls	No	Yes	Yes	Yes	Yes	Yes
Sentiments	No	No	Yes	Yes	Yes	Yes
AnalystDisp	No	No	No	Yes	Yes	Yes
N	131208	112515	112515	99558	99558	99558
R^2	0.184	0.197	0.198	0.176	0.176	0.176

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVolp(+1) as the dependent variable, using a balanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalised: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 7: Turnover with a balanced panel.

	(1)	(2)	(3)	(4)	(5)	(6)
	Turn(+1)	Turn(+1)	Turn(+1)	Turn(+1)	Turn(+1)	Turn(+1)
Turn	0.614***	0.619***	0.620***	0.610***	0.620***	0.606***
	(83.46)	(73.07)	(73.17)	(68.50)	(65.40)	(67.41)
BuzzN	-0.0739***	-0.0663***	-0.0649***	-0.0649***	-0.0567***	
	(-8.45)	(-6.71)	(-6.60)	(-7.00)	(-6.08)	
BuzzS	0.0339***	0.0403***	0.0407***	0.0424***		0.0326***
	(4.28)	(4.36)	(4.41)	(4.31)		(3.39)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock Controls	No	Yes	Yes	Yes	Yes	Yes
Sentiments	No	No	Yes	Yes	Yes	Yes
AnalystDisp	No	No	No	Yes	Yes	Yes
N	131208	112515	112515	99558	99558	99558
R^2	0.416	0.415	0.415	0.417	0.416	0.415

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with Turn(+1) as the dependent variable, using a balanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalised: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 8: Main results for volatility (Unbalanced panel)

	(1) iVolp(+1)	(2) iVolp(+1)
iVolp	0.225***	0.155***
	(27.52)	(15.66)
BuzzN	-0.0668***	-0.0505***
	(-8.19)	(-6.35)
BuzzS	0.00249	0.0411***
	(0.24)	(4.29)
Stock FE	Yes	Yes
Month FE	Yes	Yes
Stock Controls	No	Yes
Sentiments	No	Yes
AnalystDisp	No	Yes
N	159549	116491
R^2	0.159	0.157

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVolp(+1) as the dependent variable, using an unbalanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalized: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 9: Main results for turnover (Unbalanced panel)

	(1)	(2)
	Turn(+1)	Turn(+1)
Turn	0.590***	0.583***
	(88.03)	(69.20)
BuzzN	-0.0722***	-0.0617***
	(-7.71)	(-6.29)
BuzzS	0.0388***	0.0481***
	(5.21)	(5.13)
Stock FE	Yes	Yes
Month FE	Yes	Yes
Stock Controls	No	Yes
Sentiments	No	Yes
AnalystDisp	No	Yes
N	159613	116504
R^2	0.384	0.378

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with Turn(+1) as the dependent variable, using an unbalanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalized: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 10: Non-parametric volatility.

	(1)	(2)
	iVoln(+1)	iVoln(+1)
iVoln	0.314***	0.261***
	(34.43)	(25.90)
BuzzN	-0.0617***	-0.0525***
	(-7.67)	(-6.73)
BuzzS	0.00741	0.0296***
	(0.83)	(3.00)
Stock FE	Yes	Yes
Month FE	Yes	Yes
Stock Controls	No	Yes
Sentiments	No	Yes
AnalystDisp	No	Yes
N	131208	99558
R^2	0.348	0.358

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVoln(+1) as the dependent variable, using a balanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalized: iVoln, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 11: Volatility: exclusion by industry. Balanced Panel.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Excluded	CYC	ENE	FIN	HLC	IND	MAT	NCY	TEC	TEL	UTL
iVolp	0.158***	0.180***	0.174***	0.194***	0.180***	0.182***	0.177***	0.197***	0.181***	0.181***
	(13.17)	(15.59)	(14.92)	(14.84)	(14.85)	(15.78)	(15.25)	(15.49)	(15.80)	(15.96)
BuzzN	-0.0528***	-0.0513***	-0.0537***	-0.0427***	-0.0566***	-0.0506***	-0.0513***	-0.0555***	-0.0516***	-0.0517***
	(-5.81)	(-6.22)	(-6.38)	(-5.48)	(-5.72)	(-5.97)	(-6.03)	(-5.98)	(-6.30)	(-6.34)
BuzzS	0.0285**	0.0399***	0.0371***	0.0405***	0.0391***	0.0388***	0.0374***	0.0357***	0.0361***	0.0374***
	(2.36)	(3.50)	(3.44)	(4.41)	(3.46)	(3.62)	(3.35)	(2.83)	(3.37)	(3.53)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sentiments	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AnalystDisp	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	81724	93185	92560	83918	80219	93552	93644	81239	98228	99558
R^2	0.144	0.174	0.168	0.212	0.176	0.174	0.177	0.194	0.176	0.176

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVolp(+1) as the dependent variable, using a balanced panel. Each regression excludes one industry. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalized: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 12: Turnover: exclusion by industry. Balanced Panel.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Excluded	CYC	ENE	FIN	HLC	IND	MAT	NCY	TEC	TEL	UTL
Turn	0.624***	0.596***	0.608***	0.612***	0.614***	0.601***	0.612***	0.610***	0.608***	0.610***
	(64.16)	(62.67)	(66.86)	(59.00)	(64.11)	(69.01)	(66.98)	(62.75)	(67.74)	(68.50)
BuzzN	-0.0607***	-0.0629***	-0.0693***	-0.0651***	-0.0735***	-0.0607***	-0.0630***	-0.0638***	-0.0651***	-0.0649***
	(-6.27)	(-6.72)	(-7.37)	(-6.52)	(-6.46)	(-6.57)	(-6.59)	(-5.89)	(-6.98)	(-7.00)
BuzzS	0.0408***	0.0424***	0.0391***	0.0517***	0.0460***	0.0447***	0.0405***	0.0343***	0.0428***	0.0424***
	(3.99)	(4.02)	(4.04)	(4.14)	(4.49)	(4.51)	(4.00)	(3.17)	(4.24)	(4.31)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sentiments	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AnalystDisp	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	81724	93185	92560	83918	80219	93552	93644	81239	98228	99558
R^2	0.429	0.407	0.413	0.433	0.416	0.407	0.420	0.412	0.417	0.417

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with Turn(+1) as the dependent variable, using a balanced panel. Each regression excludes one industry. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalized: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 13: Volatility: exclusion by size quartile. Balanced Panel.

	(1)	(2)	(3)	(4)
Excluded	Q1	Q2	Q3	Q4
iVolp	0.0533***	0.196***	0.183***	0.164***
	(4.06)	(14.89)	(15.42)	(14.29)
BuzzN	-0.0288***	-0.0405***	-0.0553***	-0.113***
	(-4.25)	(-5.04)	(-6.93)	(-5.32)
BuzzS	0.0284***	0.0506***	0.0334***	0.0280
	(3.17)	(4.83)	(2.68)	(1.59)
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Stock Controls	Yes	Yes	Yes	Yes
Sentiments	Yes	Yes	Yes	Yes
AnalystDisp	Yes	Yes	Yes	Yes
N	80179	73394	72669	72432
R^2	0.082	0.198	0.193	0.187

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVolp(+1) as the dependent variable, using a balanced panel. Each regression excludes one Size quartile. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalized: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 14: Turnover: exclusion by size quartile. Balanced Panel.

	(1)	(2)	(3)	(4)
Excluded	Q1	Q2	Q3	Q4
Turn	0.603***	0.586***	0.603***	0.602***
	(58.38)	(63.09)	(58.35)	(57.44)
BuzzN	-0.0577***	-0.0612***	-0.0643***	-0.0851***
	(-6.16)	(-6.25)	(-7.80)	(-4.19)
BuzzS	0.0419***	0.0545***	0.0332***	0.0389***
	(3.47)	(4.69)	(3.37)	(3.41)
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Stock Controls	Yes	Yes	Yes	Yes
Sentiments	Yes	Yes	Yes	Yes
AnalystDisp	Yes	Yes	Yes	Yes
N	80179	73394	72669	72432
R^2	0.415	0.402	0.400	0.398

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with Turn(+1) as the dependent variable, using a balanced panel. Each regression excludes one Size quartile. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. The following variables are normalized: iVolp, Turn, BuzzN, BuzzS. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp.

Table 15: Panel vector autoregression model.

	(1) iVolp(t)	(2) Turn(t)	(3) BuzzN(t)	(4) BuzzS(t)
iVolp(t-1)	0.256***	3.103***	0.136***	0.0338*
	(23.81)	(5.47)	(9.84)	(1.84)
iVolp(t-2)	0.0914***	-11.269***	-0.125***	-0.123***
	(11.82)	(-22.88)	(-11.76)	(-9.35)
Turn(t-1)	-0.00109***	0.459***	-0.00162***	-0.00155***
	(-9.07)	(51.09)	(-6.78)	(-4.77)
Turn(t-2)	0.000343***	0.247***	0.00117***	0.000543**
	(3.40)	(31.60)	(6.05)	(2.18)
BuzzN(t-1)	-0.0137***	-1.333***	0.177***	-0.0800***
	(-6.14)	(-5.70)	(4.76)	(-5.81)
BuzzN(t-2)	-0.00463**	-0.173	0.0968***	-0.0260**
	(-2.43)	(-0.96)	(3.32)	(-2.53)
BuzzS(t-1)	0.00402	0.546**	-0.0109	0.596***
	(1.26)	(2.21)	(-0.95)	(31.36)
BuzzS(t-2)	0.00118	-0.420*	-0.00510	0.217***
	(0.43)	(-1.92)	(-0.61)	(13.82)
Controls	Yes			
N	123659			

t statistics in parentheses. Standard errors clustered by stocks.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel vector autoregression results with iVolp, Turn, BuzzN and BuzzS as endogenous variable, and 2 lags of the endogenous variables. Standard errors clustered by stocks. Exogenous control variables include: Size, InstOwn, Return and SentN. We also try to include other control variables, and find their coefficients insignificant and their influence on the main coefficients of interest limited.

Table 16: Panel VAR Granger causality tests.

Equation: iVolp			Chi^2	DF	P-Value
Turn	does not Granger cause	iVolp	94.10	2	< 0.001
BuzzN	does not Granger cause	iVolp	20.34	2	< 0.001
BuzzS	does not Granger cause	iVolp	11.38	2	0.003
Equation: Turn					
iVolp	does not Granger cause	Turn	178.55	2	< 0.001
BuzzN	does not Granger cause	Turn	19.29	2	< 0.001
BuzzS	does not Granger cause	Turn	32.51	2	< 0.001
Equation: BuzzN					
iVolp	does not Granger cause	BuzzN	45.41	2	< 0.001
Turn	does not Granger cause	BuzzN	22.75	2	< 0.001
BuzzS	does not Granger cause	BuzzN	5.85	2	0.054
Equation: BuzzS					
iVolp	does not Granger cause	BuzzS	26.52	2	< 0.001
Turn	does not Granger cause	BuzzS	6.47	2	0.039
BuzzN	does not Granger cause	BuzzS	16.09	2	< 0.001

Table 17: Market-Level GJR-GARCH models.

	(1)	(2)	(3)	(4)	(5)
	VWRet	VWRet	SPRet	SPRet	SPRet
Main					
Constant	0.000687***	0.000885***	0.000573**	0.000780***	0.000728**
	(2.84)	(3.00)	(2.41)	(2.71)	(2.50)
ARCH					
ARCH(-1)	0.206***	0.271***	0.211***	0.309***	0.310***
	(6.03)	(5.01)	(5.80)	(4.95)	(5.06)
TARCH(-1)	-0.0611**	-0.146***	-0.0756**	-0.179***	-0.207***
	(-1.98)	(-2.59)	(-2.32)	(-2.82)	(-3.36)
GARCH(-1)	1.163***	0.540***	1.165***	0.544***	0.572***
	(12.49)	(8.52)	(11.90)	(8.48)	(11.33)
Constant	-0.0000462***		-0.0000435***		
	(-4.77)		(-4.41)		
HET					
MktBuzzN(-1)		-0.0000185***		-0.0000179***	
		(-4.31)		(-4.16)	
MktBuzzS(-1)		0.00000921***		0.00000876***	
		(3.23)		(3.08)	
SPBuzzN(-1)					-0.0000305***
					(-5.20)
SPBuzzS(-1)					0.0000270***
					(4.61)
Constant		-8.596***		-8.711***	-9.056***
		(-34.40)		(-35.35)	(-34.61)
N	1510	1183	1510	1183	1183
ll	4665.3	3704.0	4693.1	3721.2	3727.4

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A Proof of Proposition 1

First, we establish that $\bar{s}$ is Gaussian with mean and variance

$$\mathbb{E}\left[\bar{s}\right] = \frac{1}{N}\mathbb{E}\left[\sum_n \theta + \epsilon_n\right] = m_0$$

$$\mathbb{V}\left[\bar{s}\right] = \frac{1}{N^2}\mathbb{V}\left[N\theta + \sum_n \epsilon_n\right] = \frac{1}{\rho_0} + \frac{1}{N\rho_\epsilon}.$$

Then, since the ϵ_n are white noise,

$$Cov\left[\theta, \bar{s}\right] = Cov\left[\theta, \theta + \frac{1}{N}\sum_n \epsilon_n\right] = \mathbb{V}\left[\theta\right] = \frac{1}{\rho_0},$$

Also, $\bar{e} = \bar{s}$, then

$$\bar{e} \sim \mathcal{N}\left(m_0, \frac{1}{\rho_0} + \frac{1}{N\rho_\epsilon}\right).$$

This also implies that $\mathrm{Cov}\left[\bar{e}, \bar{s}\right] = \mathbb{V}\left[\bar{s}\right]$ and $Cov\left[\theta, \bar{e}\right] = \frac{1}{\rho_0}$.

Finally, notice that

$$\frac{\partial \eta}{\partial K} = \frac{\lambda \rho_e \rho_1}{\left(\rho_1 + \lambda K \rho_e\right)^2} > 0$$

$$\frac{\partial w}{\partial N} = \frac{\rho_e \rho_0}{\left(\rho_0 + N\rho_e\right)^2} > 0$$

A.1 Turnover

We are interested in the expected turnover

$$\mathbb{E}\left[T_2\right] = \mathbb{E}\left[|(1-\lambda)x_1|\right] = (1-\lambda)\eta\rho_1\mathbb{E}\left[|\bar{e} - m_1|\right]$$

Notice that, if there are no echo chambers ($\eta = 0$), then $T_2 = 0$ because there is no disagreement and therefore no trade.[18]

Since K affects only η, then

$$\frac{\partial \mathbb{E}\left[T_2\right]}{\partial K} = (1-\lambda)\rho_1 \mathbb{E}\left[\left|\bar{e} - m_1\right|\right] \frac{\partial \eta}{\partial K} > 0.$$

To determine $\frac{\partial \mathbb{E}[T_2]}{\partial N}$, we first describe the distribution of $\bar{e} - m_1$:

$$\mathbb{E}\left[\bar{e} - m_1\right] = m_0 - m_0 = 0.$$

$$\mathbb{V}\left[m_1\right] = \mathbb{V}\left[m_0 + w\left(\bar{s} - m_0\right)\right] = w^2 \mathbb{V}\left[\bar{s}\right]$$

$$\text{Cov}\left[\bar{e}, m_1\right] = \text{Cov}\left[\bar{e}, w\bar{s}\right] = w\mathbb{V}\left[\bar{s}\right]$$

$$\begin{aligned}
\mathbb{V}\left[\bar{e} - m_1\right] &= \mathbb{V}\left[\bar{e}\right] + \mathbb{V}\left[m_1\right] - 2Cov\left[\bar{e}, m_1\right] \\
&= \mathbb{V}\left[\bar{s}\right] + w^2 \mathbb{V}\left[\bar{s}\right] - 2w\mathbb{V}\left[\bar{s}\right] \\
&= (1-w)^2 \mathbb{V}\left[\bar{s}\right]
\end{aligned}$$

Therefore $|\bar{e} - m_1|$ is a folded normal with mean

$$\mathbb{E}\left[\left|\bar{e} - m_1\right|\right] = \sqrt{\frac{2}{\pi}(1-w)^2 \mathbb{V}\left[\bar{s}\right]} \propto (1-w)\sqrt{\mathbb{V}\left[\bar{s}\right]}$$

Since $\frac{\partial w}{\partial N} > 0$, and $\mathbb{V}\left[\bar{s}\right]$ is decreasing in N, then

[18] In what follows, we emphasize that we consider N and K to be independent. If K are interpreted to be "echoes" of the original signals, we implicitly assume that the number of repetitions adjusts when K changes. This is because we observe only the empirical analogues of N and K.

$$\frac{\partial \mathbb{E}\left[T_2\right]}{\partial N} < 0$$

A.2 Subsequent Volatility

We begin by computing.

$$\begin{aligned}
\mathbb{V}\left[R_2\right] &= \mathbb{V}\left[\theta - m_1 - \eta\left(\bar{e} - m_1\right)\right] \\
&= \mathbb{V}\left[\theta - \bar{s}\left(w + \eta\left(1-w\right)\right)\right] \\
&= \frac{1}{\rho_0} + \left(w + \eta\left(1-w\right)\right)^2 \mathbb{V}\left[\bar{s}\right] - 2\left(w + \eta\left(1-w\right)\right) \operatorname{Cov}\left[\theta, \bar{s}\right] \\
&= \frac{1}{\rho_0} + \left(w + \eta\left(1-w\right)\right)^2 \mathbb{V}\left[\bar{s}\right] - 2\left(w + \eta\left(1-w\right)\right) \frac{1}{\rho_0}
\end{aligned}$$

Then, the effect of K is characterized by

$$\frac{\partial \mathbb{V}\left[R_2\right]}{\partial K} = 2\left(w + \eta\left(1-w\right)\right)\left(1-w\right) \mathbb{V}\left[\bar{s}\right] \frac{\partial \eta}{\partial K} - 2\left(1-w\right) \frac{1}{\rho_0} \frac{\partial \eta}{\partial K}.$$

By factoring $2\left(1-w\right) \frac{\partial \eta}{\partial K} > 0$, we see that the effect has the sign of

$$\begin{aligned}
\left(w + \eta\left(1-w\right)\right) \mathbb{V}[\bar{s}] - \frac{1}{\rho_0} &= \left(w + \eta\left(1-w\right)\right)\left(\frac{1}{\rho_0} + \frac{1}{N\rho_e}\right) - \frac{1}{\rho_0} \\
&= w\frac{1}{N\rho_e} - (1-w)\frac{1}{\rho_0} + \eta(1-w)\left(\frac{1}{\rho_0} + \frac{1}{N\rho_e}\right) \\
&= \frac{N\rho_e}{\rho_0 + N\rho_e}\frac{1}{N\rho_e} - \frac{\rho_0}{\rho_0 + N\rho_e}\frac{1}{\rho_0} + \eta\frac{\rho_0}{\rho_0 + N\rho_e}\frac{\rho_0 + N\rho_e}{\rho_0 N\rho_e} \\
&= \frac{\eta}{N\rho_e} > 0
\end{aligned}$$

where the third line uses $w = \frac{N\rho_e}{\rho_0 + N\rho_e}$. Hence

$$\frac{\partial \mathbb{V}\left[R_2\right]}{\partial K} > 0.$$

We now turn to $\frac{\partial \mathbb{V}[R_2]}{\partial N}$ and compute it explicitly. The numerator is positive and the denominator is

$$-\rho_e \left(\rho_e K^3 \lambda^3 + 3\rho_e K^2 \lambda^2 N + \rho_0 K^2 \lambda^2 + 3\rho_e K L N^2 + \rho_e N^3 + \rho_0 N^2\right)$$

which is always negative. Therefore,

$$\frac{\partial \mathbb{V}\left[R_2\right]}{\partial N} < 0.$$

B Alternative Behavioral Biases

Throughout this Appendix, we focus on predictions on trading volume T_2 and returns R_2 after the arrival of information at date 1. For simplicity, we omit time subscripts and write $T_2 = T$, $R_2 = R$. We present models in which social media is informative, but the interpretation of its content is subject to behavioral biases. Our goal is to capture some of the most common biases studied in behavioral finance. We restrict ourselves to models where "behavioral investors" have Gaussian posteriors, so their demand is linear as in the standard CARA-Gaussian model. We summarize the testable predictions of these models in the text. Unless otherwise specified, we continue to use the notation of Section 4.

Investors have a common prior belief that $\theta \sim \mathcal{N}(\theta_0, \rho_0^{-1})$. There are K social media signals $\boldsymbol{s} = (s_1, ..., s_K)$, each with precision ρ_e. We assume that news media signals, if present, are processed in line with Bayes' law by all investors, and are therefore implicit in the common prior. Rational investors form the Bayesian posterior $\theta \sim \mathcal{N}\left(\theta_R, \rho_R^{-1}\right)$, where

$$\theta_R = \sum_{i=1}^{K} w_i s_i + \left(1 - \sum_{i=1}^{K} w_i\right) \theta_0, \tag{5}$$

$$\rho_R = \rho_0 + K \rho_\epsilon. \tag{6}$$

The Bayesian updating weights are $w_i = \rho_e/\rho_B$. Generically, behavioral investors will have posterior beliefs $\theta|\mathrm{s} \sim \mathcal{N}(\theta_B, \rho_R^{-1})$, where

$$\theta_B = \sum_i \hat{w}_i(s_i + \eta_i) + \left(1 - \sum_i \hat{w}_i\right)\theta_0, \tag{7}$$

$$\rho_B = \hat{\rho}_0 + \sum_i \hat{\rho}_i. \tag{8}$$

The updating rules of behavioral investors shown above can exhibit three deviations from Bayes' rule. First, the weight attributed to signals by behavioral investors ($\hat{w}_i$) may differ from the rational weights w_i. Second, the precision attributed to priors ($\hat{\rho}_0$) and each signal i ($\hat{\rho}_i$) when deriving the posterior precision can differ from the true ρ_0 and ρ_e. Finally, the perception of the levels of the signals can differ from the truth by a (potentially stochastic) term η_i.

Note that Equations (4) and (3) in the main text remain valid for equilibrium prices, the (approximate) volatility of returns $\mathbb{V}[R]$, and turnover $\mathbb{E}[T]$. We now characterize these quantities as a function of social media buzz K under various behavioral biases. If behavioral biases are absent ($\lambda = 0$), trading activity is zero by the "no trade theorem" (Milgrom and Stokey, 1982), and volatility is simply the posterior variance of rational traders ρ_R^{-1}, which is decreasing in buzz K.

B.1 Alternative Biases: Overconfidence

We model overconfidence by assuming that individuals perceive the correct signals ($x_i = 0$) and prior variance ($\hat{\rho}_0 = \rho_0$), but believe that social media signals have precision $\hat{\rho}_i = (1 + a)\rho_\epsilon$, where $a > 0$ measures overconfidence. Thus they use the overconfident updating weights $\hat{w}_i = \hat{\rho}_i/\rho_B > w_i$.

The disagreement between behavioral and rational investors is $\theta_B - \theta_R = K(\hat{w}_i - w_i)(\theta_0 - \bar{s})$, where $\bar{s} = N^{-1}\sum_i s_i$ is the average signal. The disagreement is normally distributed with mean zero and variance $\mathbb{V}[\theta_B - \theta_R] = K^2(\hat{w}_i - w_i)^2\mathbb{V}[\bar{s}]$. The absolute value of disagreement $\|\theta_B - \theta_R\|$ has a folded normal distribution with mean $\mathbb{E}[\|\theta_B - \theta_R\|] = (2\mathbb{V}[\theta_B - \theta_R]/\pi)^{1/2}$. Substituting into (4) and (3) and differentiating yields

$$\frac{\partial \mathbb{V}[R]}{\partial K} \overset{\text{sign}}{=} a\lambda - (1+a\lambda)^2 \frac{K\rho_e}{\rho_0} - 1$$
$$\frac{\partial \mathbb{E}[T]}{\partial K} \overset{\text{sign}}{=} (1-a\lambda)\frac{K\rho_e}{\rho_0} + 1.$$

Therefore, K decreases volatility and increases turnover.

B.2 Alternative Biases: Conservatism

We capture conservatism by assuming that behavioral investors correctly perceive signals ($x_i = 0$) and their precision ($\hat{\rho}_i = \rho_e$), but believe the precision of their prior to be $\hat{\rho}_0 = (1+b)\rho_0$ for $b > 0$. Now, behavioral investors use the conservative weights $\hat{w}_i = \rho_e/\rho_B < w_i$. The analysis is analogous to the case of overconfidence, and we find that buzz unambiguously decreases volatility but increases turnover:

$$\frac{\partial \mathbb{V}[R]}{\partial K} < 0,$$
$$\frac{\partial \mathbb{E}[T]}{\partial K} > 0.$$

B.3 Alternative Biases: Rational Inattention

We model rationally inattentive traders who observe social media signal s_i with cognitive noise $\eta_i \sim \mathcal{N}\left(0, \rho_{\eta,i}^{-1}\right)$, but optimally choose the $\rho_{\eta,i}$ subject to an upper bound on entropy reduction. Letting $c_i = \rho_e - \left(\rho_e + \rho_{\eta,i}^{-1}\right)^{-1}$ be the decline in the precision of signal i due to inattention, the entropy reduction achieved by behavioral traders is determined by and increasing in the signal-to-noise ratio $\left(K\rho_e - \sum_{i=1}^{K} c_i\right)/\rho_0$, so the attention constraint is $\sum_i c_i \geq \underline{c}$ for an appropriate $\underline{c}$. To ensure that the attention constraint is meaningful, we assume that behavioral investors cannot infer information from prices for free.[19]

By Theorem 1 in Peng and Xiong (2006), behavioral traders wish to maximize the posterior precision $K\rho_e - \sum_{i=1}^{K} c_i$, and are therefore indifferent between all choices which satisfy the binding attention constraint $\sum_i c_i = \underline{c}$. For the simplest possible exposition, we assume here that behavioral traders observe the first $k < K$ signals perfectly, but do not pay attention to the remaining $K - k$ signals ($\rho_{\eta,i} = +\infty$ for $i \leq k$, and $\rho_{\eta,i} = 0$ for $i > k$). This is exactly optimal when $\bar{c}/\rho_e$ is an integer, and a convenient approximation otherwise. Rational investors process all K signals.

Let $\bar{s}_1 = k^{-1} \sum_{i \leq k} s_i$ denote the average signal observed by behavioral investors, and $\bar{s}_2 = (K-k)^{-1} \sum_{i>k} s_i$ the average of the remaining signals. The weights placed on each signal by behavioral and rational investors are, respectively, $\hat{w}_i = \rho_e/(\rho_0 + k\rho_e)$ and $w_i = \rho_e/(\rho_0 + K\rho_e)$. Using these weights in (5) and (7) we obtain posterior means:

$$
\begin{aligned}
\theta_R &= w_i \left(k\bar{s}_1 + (K-k)\,\bar{s}_2\right) + (1 - Kw_i)\theta_0 \\
\theta_B &= \hat{w}_i k\bar{s}_1 + (1 - \hat{w}_i k)\theta_0.
\end{aligned}
$$

The disagreement $\theta_B - \theta_R$ has mean zero and variance $\mathbb{V}\left[\theta_B - \theta_R\right] = \rho_B^{-1} - \rho_R^{-1}$, implying $\mathbb{E}\left[\|\theta_B - \theta_R\|\right] = \left(2\mathbb{V}\left[\theta_B - \theta_R\right]/\pi\right)^{1/2}$. Substituting into (4) and (3) and differentiating, we find that buzz decreases the variance of returns, but increases turnover:

[19] Generally, the reduction in entropy from observing a Gaussian signal with precision ρ is $I = \frac{1}{2}\log_2\left(1 + \frac{\rho}{\rho_0}\right)$. Here behavioral traders observe signals $(s_i + \eta_i)_{i=1}^{N}$, with respective precision $\rho_e - c_i$. By the linearity of posteriors, this is equivalent to observing one signal with precision $\rho = \sum_i (\rho_e - c_i)$, and so entropy reduction I is determined by the proposed ratio:

$$
\frac{\rho}{\rho_0} = \frac{K\rho_e - \sum_{i=1}^{K} c_i}{\rho_0}
$$

The attention constraint is $I \leq \bar{I}$, which holding ρ_0 and ρ_e fixed, can be written equivalently as $\sum c_i \geq \underline{c}$.

$$\frac{\partial \mathbb{V}[R]}{\partial K} < 0,$$
$$\frac{\partial \mathbb{E}[T]}{\partial K} > 0.$$

B.4 Alternative Biases: Confirmation Bias

We model confirmation bias by assuming that behavioral traders use the rational precisions ($\hat{\rho}_i = \rho_e$ and $\hat{\rho}_0 = \rho_0$) and rational weights $w_i = \rho_e/\rho_R$ for updating, but have an optimistic predisposition (the pessimistic case is analogous). Thus they interpret positive signals $s_i > \theta_0$ correctly, but take negative signals $s_i < \theta_0$ to be equal to their prior θ_0. The perceived signal is therefore $s_i + \eta_i$ where $\eta_i = \max\{0, \theta_0 - s_i\} \geq 0$ is the misperception due to confirmation bias. The misperception has a censored Gaussian distribution, and it is possible to show, extending the argument of Muthen (1990), that the joint moments of any two misperceptions (η_i, η_j) satisfy

$$\begin{aligned}
\mathbb{E}[\eta_i] &= \sqrt{\frac{1}{2\pi}\mathbb{V}[\theta_0 - s_i]} = \sqrt{\frac{1}{2\pi}\left(\frac{1}{\rho_0} + \frac{1}{\rho_e}\right)}, \\
\mathbb{V}[\eta_i] &= \left(\frac{1}{\rho_0} + \frac{1}{\rho_e}\right)\left(\frac{1}{2} - \frac{1}{2\pi}\right), \\
Cov[\eta_i, \eta_j] &= \left(\frac{1}{\rho_0} + \frac{1}{\rho_e}\right)\left[\zeta(r)r - \frac{1}{2\pi}\left(1 - \sqrt{1 - r^2}\right)\right],
\end{aligned}$$

where $r = \rho_e/(\rho_0 + \rho_e)$ is the correlation between two signals s_i and s_j, and

$$\zeta(r) = \frac{1}{4} + \frac{1}{2\pi}\text{ArcSin}(r)$$

denotes the probability that two signals both lie below the prior. The disagreement between behavioral and rational traders is $\theta_B - \theta_R = N w_i \bar{\eta}$, where $\bar{\eta} = K^{-1}\sum_i \eta_i$ is the average misperception, and has moments

$$\mathbb{E}\left[\theta_B-\theta_R\right]=Kw_i\mathbb{E}\left[\eta_i\right]=Kw_i\sqrt{\frac{1}{2\pi}\left(\frac{1}{\rho_0}+\frac{1}{\rho_e}\right)},$$
$$\mathbb{V}\left[\theta_B-\theta_R\right]=K^2w_i^2\left\{\frac{1}{K}\mathbb{V}\left[\eta_i\right]+\left(1-\frac{1}{K}\right)Cov\left[\eta_i,\eta_j\right]\right\}.$$

Note further that $\mathbb{E}\left[\theta_B-\theta_R\right]=\mathbb{E}\left[\|\theta_B-\theta_R\|\right]$ since behavioral traders are weakly more optimistic than rational ones. Substituting into (4) and (3) and differentiating, we can sign the effect of buzz on volatility turnover in general, and the effect of buzz on volatility in the limiting case with a large number of signals:

$$\lim_{K\to\infty}\frac{\partial\mathbb{V}[R]}{\partial K}\overset{\text{sign}}{=}\frac{2-r}{r}\left[\zeta(r)r-\frac{1}{2\pi}\left(1-\sqrt{1-r^2}\right)\right]-\left(\frac{1}{2}-\frac{1}{2\pi}\right)-(1-r). \quad (9)$$
$$\frac{\partial\mathbb{E}[T]}{\partial K}>0.$$

To check the sign in (9), we note that the right-hand side is negative for all $r\in(0,1)$ as long as

$$1-r\geq\frac{2-r}{r}\left[\zeta(r)r-\frac{1}{2\pi}\left(1-\sqrt{1-r^2}\right)\right]-\left(\frac{1}{2}-\frac{1}{2\pi}\right)\equiv\tau\left(r\right).$$

We have $\tau\left(1\right)=0$ and $\lim_{r\to0}\tau\left(r\right)=1/2\pi<1$, so the above holds at both boundaries of the set $(0,1)$. It is sufficient to show that $\tau'\left(r\right)\geq-1$, which rules out any crossings with $\tau\left(r\right)=1-r$ on the interior of the set. We have

$$\begin{aligned}\tau'\left(r\right)&=\frac{1-\sqrt{1-r^2}}{\pi r^2}-\frac{1}{4\pi}\left(\pi+2\text{ArcSin}\left(r\right)\right)\\&\geq\frac{1}{2\pi}-\frac{1}{2}>-1,\end{aligned}$$

where the second line uses the facts that the term $\frac{1-\sqrt{1-r^2}}{\pi r^2}$ is strictly increasing in r and that $\text{ArcSin}\left(r\right)\leq\pi/2$. Thus, for large enough K, social media buzz decreases volatility and increases turnover.

B.5 Behavioral traders ignore news media

In our baseline model, behavioral traders pay attention to N news media signals, as well as to K repetitions of these signals in social media. Therefore, they effectively respond to $N + K$ signals.

An alternative assumption would be one where behavioral traders pay attention only to social media. However, the initial signal updating equation is the same for a trader who responds to N true news media signals, as for a trader who (wrongly) responds to the first K repetitions of those signals. Effectively, the alternative assumption therefore reduces the number of repetitions that behavioral traders respond to by N.

From this logic, it follows easily that the equilibrium in the alternative setup is the same as the equilibrium in our baseline model where the number of social media signals is reduced to $K - N$. Formally, $T_2^{base}(K, N)$ and $R_2^{base}(K, N)$ be turnover and returns in the baseline model. Then in the alternative model we have

$$T_2 = T_2^{base}(K - N, N)$$
$$R_2 = R_2^{base}(K - N, N)$$

It follows that the effect of social media on turnover and volatility are:

$$\frac{d\mathbb{E}[T_2]}{dK} = \frac{\partial \mathbb{E}\left[T_2^{base}\right]}{\partial K} > 0$$
$$\frac{d\mathbb{E}[R_2]}{dK} = \frac{\partial \mathbb{E}\left[R_2^{base}\right]}{\partial K} > 0;$$

and effects of news media are:

$$\frac{d\mathbb{E}[T_2]}{dN} = -\frac{\partial \mathbb{E}\left[T_2^{base}\right]}{\partial K} + \frac{\partial \mathbb{E}\left[T_2^{base}\right]}{\partial N} < 0$$
$$\frac{d\mathbb{E}[R_2]}{dK} = -\frac{\partial \mathbb{E}\left[R_2^{base}\right]}{\partial K} + \frac{\partial \mathbb{E}\left[R_2^{base}\right]}{\partial N} < 0,$$

where the inequalities follow from Proposition 1.

C Alternative Specifications

This section reports regression results from alternative specifications. Table 18 uses abnormal buzz (BuzzN.ab and BuzzS.ab) calculated as the relative buzz of a month minus the mean buzz of previous 6 months divided by the standard deviation. Table 19 tests whether there is sample selection bias in Tables 6 and 7.

Table 18: Regressions using abnormal buzz.

	(1) iVolp(+1)	(2) iVolp(+1)	(3) Turn(+1)	(4) Turn(+1)
iVolp	0.0903***	0.0862***	-7.440***	-7.917***
	(7.38)	(7.05)	(-10.00)	(-10.74)
Turn	0.000581***	0.000524***	0.600***	0.591***
	(7.13)	(6.30)	(56.46)	(58.12)
BuzzN.ab	-0.000171***		-0.0208***	
	(-5.59)		(-5.15)	
BuzzS.ab	2.59e-10***		7.30e-08***	
	(4.51)		(15.88)	
BuzzN		-0.0111***		-1.738***
		(-5.78)		(-6.59)
BuzzS		0.00565**		0.951***
		(2.58)		(4.47)
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Stock Controls	Yes	Yes	Yes	Yes
Sentiments	Yes	Yes	Yes	Yes
AnalystDisp	Yes	Yes	Yes	Yes
N	81248	81248	81248	81248
R^2	0.053	0.053	0.382	0.383

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVolp(+1) or Turn(+1) as the dependent variable, using a balanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp. Regressions 1 and 3 use abnormal buzz (BuzzN.ab and BuzzS.ab) calculated as the relative buzz of a month minus the mean buzz of previous 6 months divided by the standard deviation. Regressions 2 and 4 use our main regressors BuzzN and BuzzS, but on a restricted sample consistent with regressions 1 and 3, because the calculation of abnormal buzz cannot use the first 6 months of each stock.

Table 19: Regressions testing for sample selection bias.

	(1) iVolp(+1)	(2) iVolp(+1)	(3) iVolp(+1)	(4) Turn(+1)	(5) Turn(+1)	(6) Turn(+1)
iVolp	0.212***	0.180***	0.181***		-6.395***	-6.311***
	(17.10)	(15.90)	(15.98)		(-11.51)	(-11.35)
Turn		-0.000113	-0.0000986	0.600***	0.608***	0.610***
		(-1.24)	(-1.09)	(70.48)	(68.24)	(68.53)
BuzzN	-0.0153***	-0.0142***	-0.0134***	-1.935***	-1.767***	-1.718***
	(-6.65)	(-6.62)	(-6.38)	(-7.56)	(-7.14)	(-6.99)
BuzzS	-0.000829	0.00760***	0.00780***	0.751***	0.888***	0.896***
	(-0.37)	(3.47)	(3.55)	(3.61)	(4.26)	(4.31)
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sentiments	Yes	Yes	Yes	Yes	Yes	Yes
AnalystDisp	Yes	Yes	Yes	Yes	Yes	Yes
N	99558	99558	99558	99558	99558	99558
R^2	0.146	0.176	0.176	0.409	0.416	0.417

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVolp(+1) or Turn(+1) as the dependent variable, using a balanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp. These are regressions 1-3 in Tables 6 and 7 but using the restricted sample consistent with regression 4 in those tables.

Table 20: Volatility and Volume Regressions with Market Controls.

	(1)	(2)
	iVolp(+1)	Turn(+1)
iVolp	0.181***	-6.310***
	(15.96)	(-11.36)
BuzzN	-0.0134***	-1.718***
	(-6.34)	(-7.00)
BuzzS	0.00772***	0.896***
	(3.53)	(4.31)
Stock FE	Yes	Yes
Month FE	Yes	Yes
Stock Controls	Yes	Yes
Market Controls	Yes	Yes
Sentiments	Yes	Yes
AnalystDisp	Yes	Yes
N	99558	99558
R^2	0.176	0.417

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports panel regression results with iVolp(+1) or Turn(+1) as the dependent variable, using a balanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Market Controls include S&P500 return and VIX index. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp. These regressions are based on the specification in Regression (4) of Table 6, adding Market Controls.

Table 21: Volatility and Volume Regressions Using Absolute Buzz.

	(1) iVolp(+1)	(2) iVolp(+1)	(3) Turn(+1)	(4) Turn(+1)
iVolp	0.161***	0.161***	-6.661***	-6.661***
	(13.17)	(13.17)	(-10.83)	(-10.83)
ABuzzN	-0.000000133***	-0.000000133***	-0.0000367***	-0.0000367***
	(-3.93)	(-3.93)	(-6.79)	(-6.79)
ABuzzS	0.000000158***	0.000000158***	0.0000162***	0.0000162***
	(3.46)	(3.46)	(3.88)	(3.88)
Stock FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Stock Controls	Yes	Yes	Yes	Yes
Market Controls	No	Yes	No	Yes
Sentiments	Yes	Yes	Yes	Yes
AnalystDisp	Yes	Yes	Yes	Yes
N	87364	87364	87364	87364
R^2	0.155	0.155	0.413	0.413

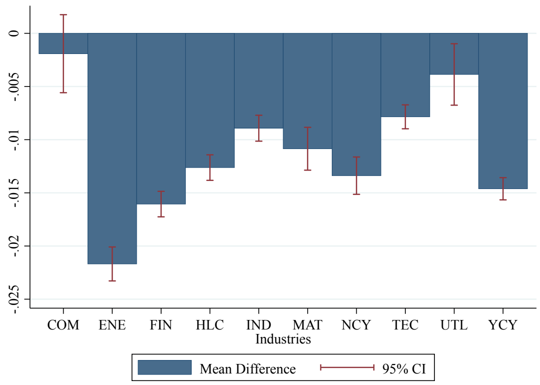
t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

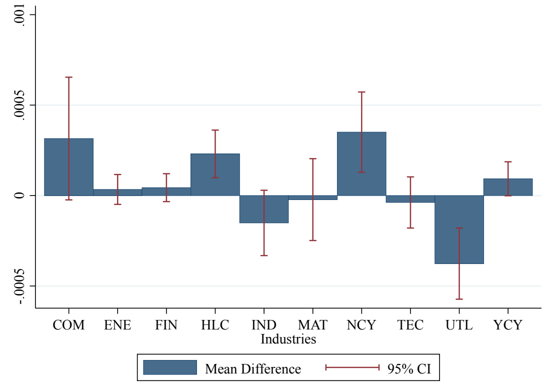
Notes: This table reports panel regression results with iVolp(+1) or Turn(+1) as the dependent variable, using a balanced panel. All regressions include month fixed effects, stock fixed effects and standard errors clustered by stocks. Stock controls include iVolp, Turn, Size, InstOwn, HHI, Leverage, Return, TotalSD. Market Controls include S&P500 return and VIX index. Sentiment controls include SentN, SentN(-), SentS and SentS(-). Dispersion in the analyst opinions is AnalystDisp. The main explanatory variables are ABuzzN and ABuzzS, which represent the absolute amount of buzz in news and social media respectively.

D Differences between the Two Media Sources

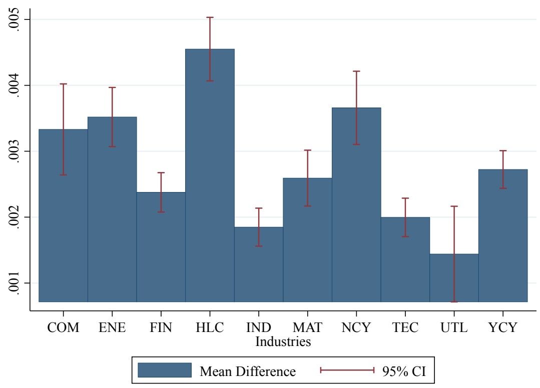
In this section, we visually show that there are some systematic differences between the two media sources in terms of their contents. For instance, news media is more likely to contain facts, while social media is more likely to contain emotions.



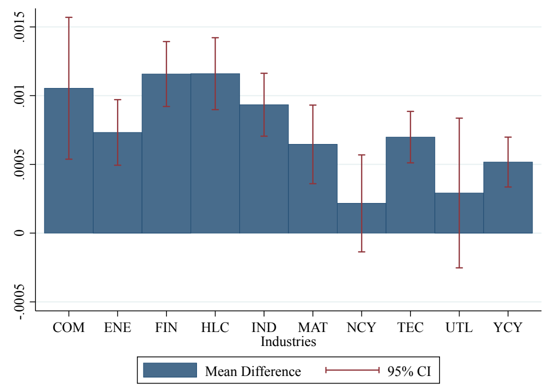
(a) Emotion vs Fact



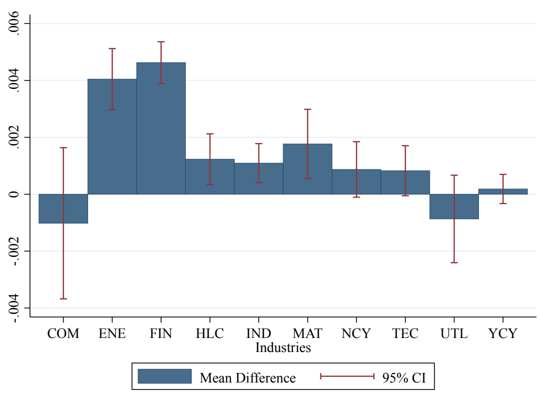
(b) Mentioning of Layoffs



(c) Mentioning of Litigation



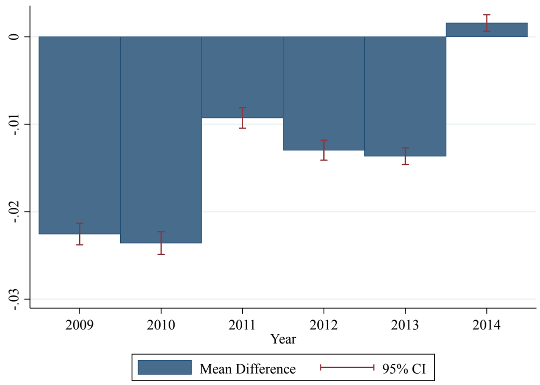
(d) Mentioning of Management Change



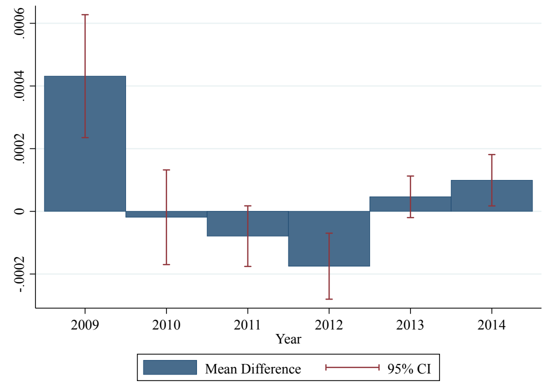
(e) Mentioning of Mergers

Figure 5: DIFFERENCES BETWEEN THE TWO MEDIA SOURCES BY INDUSTRY

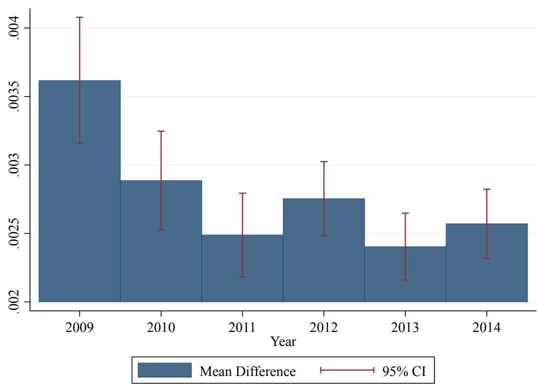
Note: This figure demonstrates the differences between news and social media coverages across industries. In each panel, we subtract social media value from news media value. Panels (a) to (e) are repsectively media coverage of emotions net of fact, layoffs, litigation, management change, and mergers.



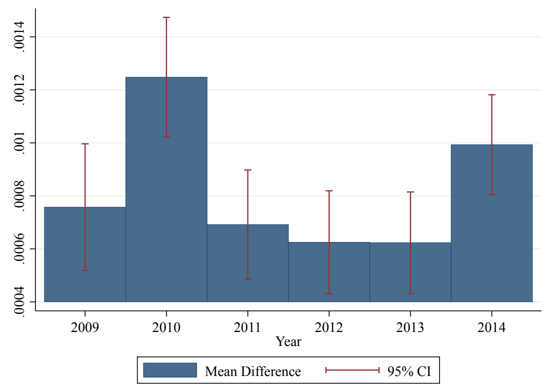
(a) Emotion vs Fact



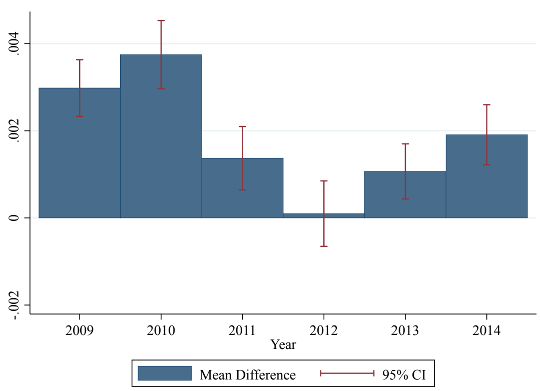
(b) Mentioning of Layoffs



(c) Mentioning of Litigation



(d) Mentioning of Management Change



(e) Mentioning of Mergers

Figure 6: DIFFERENCES BETWEEN THE TWO MEDIA SOURCES BY YEAR

Note: This figure demonstrates the differences between news and social media coverages over the years. In each panel, we subtract social media value from news media value. Panels (a) to (e) are repsectively media coverage of emotions net of fact, layoffs, litigation, management change, and mergers.

Dear Referee,

Thank you so much for your additional suggestions. We have revised our paper accordingly and added some additional analyses.

1. We agree with you that absolute buzz is more closely connected to our model than relative buzz, and that S&P500 return and VIX index would provide good controls for fluctuations in the market. Using absolute buzz is an important issue, so we added footnote 3 in Section 2.1 to prepare the reader for this. Then we added a discussion in Section 3.3, showing the regression results in Table 21 of Appendix C. On the other hand, we also tested the effect of buzz after controlling for market fluctuations in both Table 20 and Table 21 of Appendix C. All of these tests reveal that our main conclusions are robust.

2. We have corrected the errors in referencing tables. Thank you for your careful reading.

We are really grateful for your thoughtful comments and suggestions. Thank you so much!