Biased News and Irrational Investors: Evidence from Biased Beliefs about Uncertainty and Information Acquisition*

Jiatao Liu†

December 25, 2021

Abstract

Investors who use biased information from news media subsequently tend to make irrational decisions about acquiring firm-specific information compared to rational expectations. This model of information acquisition yields testable predictions that are verified by using a novel dataset of news stories. First, when sentiment in news articles, as a proxy for biased public information, is more optimistic, investors tend to acquire less earnings-relevant information before the earnings announcement and vice versa. Second, the return predictability from firm-specific news sentiment confirms that it contributes to variations in asset information risk due, in a biased belief equilibrium, to the proportion of informed investors deviating from rational expectations. Overall, these findings suggest that biased public information inherent in news sentiment serves to irrationalize investors' acquisition of firm-specific information through a biased perception of uncertainties in the risky asset payoff.

JEL Classification: G11; G12; G14; G41

Keywords: Biased Beliefs; Information Acquisition; News Sentiment; Information Risk; Risk Premium

*I thank Marco Pagano at University of Naples Federico II, Thierry Foucault at HEC Paris, Giovanni Cespa at Bayes Business School, Alessandro Rebucci at Johns Hopkins University, Jack Bao at University of Delaware, Zipeng Wang at Westlake University and participants at Cass PhD Research Day for helpful comments. I am also grateful to Jefferson Duarte from Rice University to provide the GPIN data and Richard Peterson as the managing director from Thomson Reuters MarketPsych to offer the novel dataset of news stories to support my study. All errors are my own.

†Bayes Business School, City, University of London, 106 Bunhill Row, London UK, EC1Y 8TZ; Email: jiatao.liu@bayes.city.ac.uk

1 Introduction

A theory of what drives investors' decision-making on acquiring information is explored in models of how rational investors perceive the uncertainty surrounding risky assets (Grossman and Stiglitz, 1980; Veldkamp, 2006; Andrei et al., 2019; Benamar et al., 2019). By contrast, in studies adopting the behavioral perspective, researchers customarily assume that investors suffer from psychological bias such as overconfidence, which causes the equilibria achieved in the information acquisition model to diverge from rational expectations equilibria (Odean, 1998; García et al., 2007; Ko and Huang, 2007). One argument in the study by Tirole (2002) is that rationalists have legitimate concerns about the shortcomings of the *Homo economicus* paradigm, and that the field of neoclassical economics study can be enriched by contributions from behavioral studies without losing the rigor of quantitative economics analysis.

Adopting insights from behavioral studies, an interesting question arises in situations in which the perception of uncertainty in the risky asset's payoff is not rationally formed. A question of similar interest relates to the cause of the irrationality that drives investors' biased decision-making on information acquisition. Therefore, in line with the inspiration of Tirole's (2002) study, I seek to answer the question of how an irrational decision concerning the acquisition of further information can be made by investors by drawing on behavioral views to investigate the drivers of irrationality.

The traditional view of investors' irrationality originating from psychological bias fails to adequately address how biased information transmission contributes to irrational decision-making by investors. Specifically, linguistic or rhetorical tone measured by sentiment, as a partial order on reporting strategy in publicly available news stories through newswires or online media, may bias investors toward irrational decisions concerning whether or not to acquire private information in investment. This paper addresses this gap by examining how, by using biased public information about the market or companies as measured by sentiment from news stories, investors' acquisition of private, firm-specific information deviates from the rational expectations equilibrium.

Building on the model by Grossman and Stiglitz (1980), I develop a three-period model by extension from the seminal study by Andrei et al. (2019), who argue that investors' rational perception of economic uncertainty affects their attention to firm-specific information. I introduce an additional medium to relax the assumption of rationality in the model, namely, the consideration of biased public information from news to which investors are exposed exogenously before they begin to trade. Although rational agents are found to be subject to biased information in the media for decision-making (Baron, 2006; Kamenica and Gentzkow, 2011), to simplify the analysis, I adhere to Hirshleifer's (2020) study and add a parsimonious friction-naïve assumption in the model. As stated in Hirshleifer (2020), information receivers' naïveté about bias in the messaging is due to people's general tendency to take the information at face value, rather than adjusting for the features of the data-generating process. Therefore, investors' are naïve about bias in the news when considering their investment choices; as a result, their acquisition of firm-specific information will deviate from the equilibrium in rational expectations.

The key difference in the model I develop in this paper compared to existing studies on biased information acquisition is that irrationality arises from the bias in the news information, rather than from investors' behavioral irrationality as the sole cause. The investors' biased acquisition decision about firm-specific information is made through the channel of their beliefs about the uncertainties in the risky asset payoff, which are biased by the public information from news articles that tend to be either optimistically or pessimistically reported.[1] When there is a positive (upward biased) tone in the news that investors read, they feel more optimistic or less uncertain about economic conditions or a firm's individual performance surrounding the risky investment. Accordingly, investors are biased towards an under-perception of the systematic uncertainty or idiosyncratic uncertainty in the payoff of a risky asset, which causes investors either to overstate the informativeness of price or understate the value of firm-specific information respectively. In a biased belief equilibrium, investors eventually acquire less firm-specific information than they would if the decision were made under rational expectations. By the same token, when the news is marked by a negative tone (downward biased), it leads investors to acquire more firm-specific information, due to them feeling more uncertain about the economy or the firm itself. This more uncertain perception leads investors to understate the informativeness of price or overstate the value of firm-specific information.

The model yields three testable predictions. First, since investors' perception of uncertainties in risky assets is inversely related to the tone in the news media, news sentiment, as a proxy for biased public information in the model, negatively predicts a acquisition of firm-specific information. Second, the deviation of firm-specific information acquisition, especially from firm-specific news sentiment, indicates a different degree of price informativeness and hence a deviation of risky assets' information asymmetry risk from the rational expectations equilibrium. As proposed by O'Hara (2003), investors require a risk premium to hold the risky assets which are subject to high information risk; thus, the compensation of the information risk in this model varies with the biased decision to acquire firm-specific information. This bias is caused by sentiment in firm-specific news. Third, firm-specific news sentiment predicts positive cross-sectional variation of stock returns in the form of variation in information risk, led by a shift from the rational expectations equilibrium of the proportion of informed investors.

To test these predictions, I use a novel dataset from Thomson Reuters MarketPsych (TRMI). To collect this dataset, Thomson Reuters develops an algorithm to conduct textual analysis of

[1] In section 4.1, I first verify this channel of irrationality as motivation from empirical evidence, arguing that the tone in the news biases the variance of distribution rather than the mean in the risky asset payoff components.

worldwide news and online media sources to provide a sentiment index. This takes the form of linguistic tone measured by counting the usage of positive and negative words in the news stories about the aggregate market or individual firms. Therefore, I use TRMI news sentiment indices as measures of biased tone in the news to test its impact on information acquisition behavior. I find strong evidence of an inverse relationship between news sentiment and uncertainties. On the one hand, it is clear that stock market news sentiment is significantly and negatively correlated with customary measures of systematic uncertainties, such as the stock market expectation of volatility on S&P500 index options (VIX) or the Economic Policy Uncertainty (EPU) indices (Baker et al., 2016). On the other hand, by using a bundle of proxies for firm-specific uncertainties – such as the variance of regression residuals from an AR(1) process of firm earnings per share (Griffin, 1977), the absolute value of unexpected earnings (Hirshleifer et al., 2008) and idiosyncratic volatility shock (Bali et al., 2018) – firm-specific news sentiment is found to be consistent in negatively predicting all proxies of firm-specific uncertainties.

Next, I examine how the news sentiment indices affect investors' decision to acquire firm-specific information. First, a proxy for firm-specific information acquisition, in line with the study by Weller (2018), is measured by earnings-related information incorporated into price before announcements. Second, I show the empirical evidence to confirm the model's theoretical implication of an inverse relationship between news sentiment and investors' acquisition of firm-specific information. In fact, when a more optimistic tone is found in the news about either the stock market or a particular firm, investors tend to acquire less earnings-relevant information before it is released and vice versa. These results hold after controlling for fixed effects, firms' fundamental variables and benchmark uncertainty measures, namely, the VIX and EPU. Overall, these findings confirm my theoretical results that the biased public information contained in news shifts investors' acquisition of firm-specific information away from the rational expectations equilibrium. I also show the predictability of the effect of sentiment in firm-specific news on cross-sectional variation of stock returns by proposing an argument that information risk in risky assets varies with firm-specific news sentiment. Specifically, I conduct daily cross-sectional Fama-Macbeth (1973) regressions to show that firm-specific news sentiment positively predicts future stock returns without reversal. These empirical results hold after including firm-fundamental control variables, volume–return predictors, and other influential effects from news variables such as value-relevant information (Tetlock et al., 2008; Chen et al., 2014) and reduction of information asymmetry (Tetlock, 2010). These findings are consistent with theoretical results. Sentiment in the firm-specific news drives a biased belief equilibrium in investors' firm-specific information acquisition which deviates from rational expectations; the information risk in the risky assets eventually becomes relatively higher or lower to traders through a price discovery process.

As an additional test of the risk premium argument, I conduct a factor pricing test by constructing a zero-cost portfolio sorted by daily cross-sectional firm-specific news sentiment. On average, the news sentiment factor earns around a 6.6-basis point return per day, which is equal to annualized return of about 16.63%. In addition, controlling for classical asset pricing factors such as the Fama–French five factors (Fama and French, 2015), the momentum factor (Carhart, 1997), the Pastor and Stambaugh liquidity factor (Pástor and Stambaugh, 2003), and short- and long-term reversal factors does not accommodate for abnormal returns as fully as the news-sentiment portfolio does. In sum, the factor pricing results support this study's theoretical proposition that sentiment, particularly from firm-specific news, affects information risk in risky assets, in that the proportion of informed investors, in a biased belief equilibrium, departs from rational expectations.

My study makes a unique contribution to the literature on information acquisition by investors. Through both neoclassical and behavioral economics perspectives, prior studies have addressed how investors' perceptions of uncertainty or the value of signals create demand for information about assets' fundamental payoff (Grossman and Stiglitz, 1980; Veldkamp, 2006; Odean, 1998; Garća et al., 2007; Ko and Huang, 2007; Andrei et al., 2019; Benamar et al., 2019). In line with the behavioral school's tendency to relax strict rationality in economic studies, this research is enriched by the introduction of a new biased channel that is motivated by Hirshleifer's (2020) seminal study, which argues that biased information or signals stemming from information transmission significantly affect investors' decision-making and may cause asset mispricing. Therefore, in contrast to the majority of extant behavioral studies in finance and economics that examine the behaviors of irrational agents, this paper focuses on biased information percolation as argued for by Hirshleifer (2020) and proposes that investors should not necessary be presumed to be irrational agents. Investors can, in fact, be `forced' into behaving sub-optimally when they receive and apply biased public information from news in their decision-making on acquisition of firm-specific information.

My study also contributes to the growing body of research that makes use of textual data in finance and economics. This literature includes studies by Tetlock (2007), Akhtar et al. (2011) and Garcia (2013) on negative news sentiment regarding aggregate markets predicting market returns; studies by Tetlock et al. (2008), Tetlock (2010), Chen et al. (2014) and Ke et al. (2019) on firm-specific news or online media sentiment containing valuable information for predicting positive future returns; and studies on the effect of media on stock markets by Bhattacharya et al. (2009), Engelberg and Parsons (2011), Peress (2014), Hillert et al. (2014), and Bonsall IV et al. (2020). However, news sentiment plays a key role in my study in demonstrating investors' biased decision-making on firm-specific information acquisition, which has not been addressed in the literature. Additionally, contrary to the argument that value-relevant information may be found in the news, the empirical result that sentiment in firm-specific news predicts positive future stock returns supports the theoretical prediction that information risk varies with firm-specific news sentiment.

Furthermore, my study sheds light on other studies that address how information purveyors such as journalists or media companies reflect different tones in news or media which bias or slant audiences' economic or political opinions (Mullainathan and Shleifer, 2005; Baron, 2006; Gentzkow et al., 2015). More importantly, media bias can be persistent as information in the news is suppressed or withheld by news organizations, in that the bias cannot be undone by rational or sophisticated agents since they do not know how much information the news supplier has and when information is being withheld (Bernhardt et al., 2008; Anderson and McLaren, 2012). In financial markets, preference for or disagreement with a journalist's report or media channels' views can affect stock market behaviors and financial valuation (Dougal et al., 2012; Gurun and Butler, 2012; Hillert et al., 2018). In line with these studies on media bias, I provide additional evidence that tone in the news, measured by sentiment, leads investors to form a biased perception of uncertainties in risky assets, and thus make a biased decision to acquire firm-specific information in equilibrium. To the best of my knowledge, this paper is the first study to bridge this gap on the effect of biased public information in the news on investors' acquisition of firm-specific information.

Finally, the theoretical result regarding investors' biased information acquisition decision in this paper is also in line with studies on information rigidity (Sims, 2003; Coibion and Gorodnichenko, 2012, 2015; Bouchaud et al., 2019) and extrapolation (Alti and Tetlock, 2014; Greenwood and Shleifer, 2014; Hirshleifer et al., 2015; Choi and Mertens, 2019). On the one hand, an investors' reluctance to take on board new information, as expounded in information rigidity studies, is similar to the implications of the model developed in this paper. Sticky information acquisition, whereby investors are less willing to acquire firm-specific information in a biased belief equilibrium, is caused by positive sentiment in the news. On the other hand, the overweighted amount of recent information used by investors in information extrapolation research is similar to the present study's understanding of negative sentiment in the news. Investors acquire too much firm-specific information compared to what they would acquire in a rational expectations scenario. Although the biased incorporation of information for the purposes of making an investment decision in the model presented in this paper shares similar psychological behaviors to those described in the information rigidity and extrapolation studies, the channel for bias in this study's model is different, as bias mainly originates from the news media itself, rather than from investors.

The paper is organized as follows. Section 2 introduces a theoretical model of biased information acquisition and develops testable predictions. Section 3 describes the dataset used for the empirical studies and provides data summary statistics. Section 4 details the empirical results of the tests, which show that with news sentiment held as a proxy for biased public information, investors' biased perception of uncertainties gives rise to biased information acquisition. Section 5 entails a test conducted on the pricing power of firm-specific news sentiment on stock returns. Section 6 offers the study's conclusions. Robustness tests are in the Online Appendix.

2 Information Acquisition Model with Biased Beliefs

This study reports the development of a model for how investors become informed as a way of reducing the uncertainty of risky asset investments. I assume that the acquisition of firm-specific information is costly. This cost can be understood as, among other things, hiring financial advisers, analyzing financial reports, gathering information about consumers' preferences, buying financial data or outsourcing financial data analysis. Therefore, only a fraction of investors will choose to pay for such costly information. This paper demonstrates how the tone of exogenous costless public information from news media may give investors a biased rather than rational perception of the uncertainty surrounding risky assets. As a consequence, firm-specific information acquisition deviates from the rational expectations equilibrium.

2.1 Model Setup

The principles of this static model for information acquisition are based on Grossman and Stiglitz (1980), and those of biased public information are based on the proposition of biased information transmission by Hirshleifer (2020). The economy of the current model is similar to that of Kacperczyk et al. (2016) and Andrei et al. (2019). The biased belief draws on work by Odean (1998), García et al. (2007), Ko and Huang (2007) and Heller and Winter (2020) in allowing irrationality in the economy. However, the key argument of biased belief in this model is the result of biased public information such as news sentiment and not investors' psychological bias, which has been broadly addressed in the behavioral literature.

In a hypothetical economy populated by a continuum of investors indexed by $i \in [0,1]$, there are three periods $t \in \{0,1,2\}$. At $t = 0$ investors read costless news about the market or particular firms they are considering an investment in and make a decision on whether or not to acquire more private information about firm-specific conditions to inform their investment decision. Investors trade competitively at $t = 1$ in the financial market. At $t = 2$ the payoff of financial assets will be realized and investors will consume their terminal wealth.

Investors trade a risk-free asset and a risky asset. The risk-free asset pays a gross interest rate of r_f and the supply is infinitely elastic. The risky asset (stock) has an equilibrium price P_1 at $t = 1$ and pays a risky dividend at $t = 2$:

$$D_2 = \overline{D} + m_2 + e_1 \tag{1}$$

The risky dividend payoff has three components: a mean payoff $\overline{D} > 0$, a market component $m_2 \sim N(0, \sigma_m^2)$ and a firm-specific component $e_1 \sim N(0, \sigma_e^2)$. The firm-specific component will be available at $t = 1$ to investors who choose to become informed. Therefore, informed investors will perfectly observe e_1. Additionally, m_2 and e_1 are independent.

The mean payoff $\bar{D}$ is common knowledge for all investors at $t = 0$. Investors with rational expectations know the variance (uncertainty) of the market component, σ_m^2, and the variance (uncertainty) of the firm-specific component, σ_ϵ^2 at $t = 0$. However, investors' knowledge about σ_m^2 and σ_ϵ^2 are biased by reading news with non-neutral tones about the market or a firm at $t = 0$.[2]

This understanding of biased information in the news sheds light on one of the major propositions stated by Hirshleifer (2020), namely, that information transmission bias results from misreporting, in that a signal received by investors is subject to an upward or downward bias in the signal itself. In addition, information receivers interpret the biased information from news naively and without adjusting for the bias in the news. In fact, investors' unawareness or naivety about the bias in the news can be easily relaxed, because Bernhardt et al. (2008) and Anderson and McLaren (2012) developed models to confirm that rational agents cannot undo this bias caused by the suppression or withholding of information by suppliers.

The assumption of rational or sophisticated investors may make the model in the current study even more parsimonious or generalized, but without including a verification of the biased effect from public information in the news, I retain the customary assumption of naivety in the model proposed by Hirshleifer (2020). Hence, following the Hirshleifer (2020) study, this paper defines the tone from news – which is measured by sentiment in the way news providers describe the stock market or particular firms – as tending to be either more optimistic or pessimistic. This is the bias (b) in costless information reporting to investors. Investors' prior beliefs of both market or firm-specific components' uncertainty is subject to bias through the tone of the market- or firm-specific news respectively, which they receive at $t = 0$. Furthermore, all investors are homogeneously biased by the tone of news.[3]

For simplicity, I assume that the biased effect of the news sentiment about the whole market (S_m) is independent of the firm-specific news sentiment (S_ϵ).[4] Therefore, the uncertainty of the market component, σ_m^2 is only biased by the market news sentiment, and σ_ϵ^2 is only biased by the sentiment in firm-specific news. Finally, as investors are naive about the validity of news tone, they make trading or investment decisions based on their unconscious, biased beliefs.

As argued by Odean (1998), Ko and Huang (2007) and Heller and Winter (2020), I assume that all investors' subjective beliefs follow a bias function $b(S_j; \sigma_j^2)$, where σ_j^2 is a constant of correct beliefs, and $j \in (m, \epsilon)$. This posits that the biased prior belief of both market and firm-

[2] I outline a simple model to describe why news or media always has bias $E[b] \neq 0$ in the Online Appendix.

[3] Since news is costless and available to all investors at $t = 0$, I assume all investors have the same biased beliefs about the uncertainties for tractability.

[4] Even though I make this assumption of independence in the theoretical model, I control the market news sentiment in all the empirical testing for robustness.

specific components' uncertainty is parameterized by the bias function:

$$b(S_j;\sigma_j^2)_{j\in(m;e)} = \sigma_{b;j}^2 \begin{cases} S_j \uparrow & \sigma_{b;m}^2 < \sigma_m^2,\ \sigma_{b;e}^2 < \sigma_e^2 \\ S_j = 0 & \sigma_{b;m}^2 = \sigma_m^2,\ \sigma_{b;e}^2 = \sigma_e^2 \\ S_j \downarrow & \sigma_{b;m}^2 > \sigma_m^2,\ \sigma_{b;e}^2 > \sigma_e^2 \end{cases} \tag{2}$$

where b denotes the investors' subjective biased belief throughout the paper. Notably, bias in the news is not intended to advance a false perception or convince investors to alter their own perceptions. In fact, the effect of bias can be understood as presented in the study of Gentzkow et al. (2015), who defined the bias as a partial order on reporting strategies that shift agents' beliefs about a firm strategy to either the right or the left. In my study, the bias shifts investors' beliefs towards either more optimistic or more pessimistic perceptions of the uncertainties. Therefore, the biased information from news media slants investors' perception, causing them either to overestimate or underestimate σ_m^2 and σ_e^2, and does not mislead investors into changing the mean of the distribution about m_2 and e_1.

The rationale for biased beliefs in the model is as follows: as the tone in news about the market or a particular firm grows more positive or optimistic, investors' certainty regarding the market or the firm's future performance will also grow, and vice versa.[5] If the tone in the news is neutralized ($S = 0$), meaning that the information from news is genuinely objective and devoid of bias, investors have a rational prior belief about the uncertainties. Since all investors are naive about the validity of biased information from news, they are behaving optimally by believing that their biased understanding of those uncertainties is indeed correct, even though, in fact, it is not.

At $t = 1$, there is a public signal about the market in the economy and the signal is available for all investors:

$$M_1 = m_2 + h_1 \tag{3}$$

where $h_1 \sim N(0;\sigma_h^2)$ and is independent from m_2 and e_1.

Following the Grossman and Stiglitz (1980) information acquisition model, at $t = 0$, all investors decide if they want to acquire the private information about e_1, which will be perfectly observed at $t = 1$. I denote the decision of investor i with variable L_0^i, where $L_0^i = 1$ denotes when investor i chooses to become informed and $L_0^i = 0$ indicates that she wishes to stay uninformed.

I assume that investors have CARA utility function with zero initial wealth[6] and maximize

[5] In other words, the biased uncertainty is a monotonically decreasing function of news sentiment. I do not assume a particular form of the function between biased uncertainty and news sentiment. However, without loss of generality, one can simply assume a linear form $\sigma_{b;j}^2 = (1 - S_j)\sigma_j^2$.

[6] Without loss of generality, I suppress $W_0 = 0$ because the CARA utility maximization problem is independent of initial wealth.

their expected utility with biased beliefs :

$$U_b^i = E_{b;0}^i\left[-e^{-a(W_2^i - cL_0^i)}\right] \tag{4}$$

where a is the risk aversion coefficient and c is a positive information cost for those who choose to become informed about ϵ_1 at $t = 1$. W_2^i is investor i's terminal wealth at $t = 2$.

Investors choosing to become informed by perfectly observing ϵ_1 at $t = 1$ are denoted by I. Investors who choose to remain uninformed are denoted by U. Following the noisy rational expectations model proposed by Grossman and Stiglitz (1980), the uninformed investors are still able to learn ϵ_1 partially through the perceived equilibrium price. This is described below.

At $t = 1$ investors choose their optimal portfolios:

$$q_1^i = \frac{E_{b;1}^i[D_2] - r_f P_1}{a\,Var_{b;1}^i[D_2]}; \quad \text{for } i \in \{I;U\} \tag{5}$$

where $E_{b;1}^i$ and $Var_{b;1}^i$ are subject to investor i's biased beliefs. Following O'Hara (2003), I assume that the risky asset random supply x_1 is independent of μ_2, ϵ_1, η_1, and that x_1 is normally distributed with mean $\bar{x}$ and variance σ_x^2, or $N(\bar{x};\sigma_x^2)$. With the exception of the case in which the random supply prevents a perfect revelation of ϵ_1 through the price, the positive expected supply of the risky asset implies a risk premium in the model as traders demand compensation to hold the risky assets in equilibrium. Finally, with λ_1 denoting the proportion of informed investors, the equilibrium price of the risky asset is determined by the market clearing condition:

$$\lambda_1 q_1^I + (1-\lambda_1) q_1^U = x_1 \tag{6}$$

Because investors are naive about the validity of the news tone, investors with biased perceptions of uncertainties believe they are acting optimally and the equilibrium is determined by investors' biased beliefs. Similar rationales can be found in Heller and Winter (2020). In two-player games, the authors argue that players are blind to their biased beliefs regarding the opponent's strategy and choose the best response strategy to their biased beliefs. The equilibrium yielded by the model of Heller and Winter's (2020) study is subject to the players' biased belief. Therefore, the equilibrium achieved in the model I discuss in this paper falls within the ambit of the *biased belief equilibrium* proposed by Heller and Winter (2020).

2.2 Equilibrium

By virtue of investors' naivety about biased tones in the news information they consume, the biased belief equilibrium (BBE) in my study is obtained in the same manner as in a noisy rational expectations equilibrium model (REE). I posit that the investors' perceived pricing function is:

$$P_1 = A\bar{D} + BM_1 + Ge_1 - Kx_1 + H\bar{x} \tag{7}$$

As uninformed investors are able to partially learn the e_1 for free from the price, the informative signal from price revealing is defined as:

$$\hat{p}_1 = \frac{P_1 - A\bar{D} - BM_1 + (K - H)\bar{x}}{G} = e_1 - \frac{K}{G}(x_1 - \bar{x}) \tag{8}$$

The information set for informed and uninformed investors is different. For informed investors, the information set is $F_I = \{\bar{D}; M_1; e_1; \hat{p}_1\}$. For uninformed investors, the information set is $F_U = \{\bar{D}; M_1; \hat{p}_1\}$. Therefore, the following equations define optimal portfolio choice from (5) for informed and uninformed investors (see Appendix A.2 for the derivation):

$$q_1^I = \frac{\bar{D} + \frac{s_{b;m}^2}{s_{b;m}^2 + s_h^2} M_1 + e_1 - r_f P_1}{a \mathrm{Var}_{b;1}^I [D_2]} \tag{9}$$

$$q_1^U = \frac{\bar{D} + \frac{s_{b;m}^2}{s_{b;m}^2 + s_h^2} M_1 + \frac{s_{b;e}^2}{s_{b;e}^2 + \frac{K^2}{G^2} s_x^2} \hat{p}_1 - r_f P_1}{a \mathrm{Var}_{b;1}^U [D_2]} \tag{10}$$

The optimal portfolio from equations (9) and (10) clearly indicates that, on average, informed investors hold more of the risky asset ($q_1^I > q_1^U$) when the expected return is positive. This is because they perfectly observe e_1 at $t = 1$, thus reflecting a lower risk ($\mathrm{Var}_{b;1}^I[D_2] < \mathrm{Var}_{b;1}^U[D_2]$) which is bestowed on them by their superior information (O'Hara, 2003).

As noted above, investors are naive about their biased beliefs and use $s_{b;m}^2$ and $s_{b;e}^2$ instead of rational perceptions ($s_m^2; s_e^2$) to make their optimal investment decision. Therefore, the model is solved by the standard procedure introduced by Grossman and Stiglitz (1980) which uses the market clearing condition (6) to find the equilibrium price with investors' biased beliefs. The proof is provided in the Appendix A.2.

Proposition 1. In equilibrium, the coefficients on the fundamental, public signal, private signal and supply noise in the investors' perceived pricing function are given by:

$$A = \frac{1}{r_f}; \quad B = \frac{s_{b;m}^2}{(s_{b;m}^2 + s_h^2) r_f}; \quad G = \frac{l_1 g f_I + (1 - l_1) g f_U F}{r_f Z}; \quad K = \frac{(1 - l_1) g f_U F \frac{K}{G} + 1}{r_f Z};$$

$$H = \frac{(1 - l_1) g f_U F \frac{K}{G}}{r_f Z}; \quad F = \frac{s_{b;e}^2}{s_{b;e}^2 + \frac{K^2}{G^2} s_x^2}; \quad \frac{K}{G} = \frac{a \mathrm{Var}_{b;1}^I [D_2]}{l_1}; \quad f_I = \frac{1}{\mathrm{Var}_{b;1}^I [D_2]}; \quad f_U = \frac{1}{\mathrm{Var}_{b;1}^U [D_2]}; \tag{11}$$

$Z = (\lambda_1 \gamma \phi_I + (1 - \lambda_1)\gamma \phi_U)\rho_f; \ \gamma = \frac{1}{\alpha}$

2.3 Information Acquisition in Investors' Biased Belief Equilibrium

As stated in Grossman and Stiglitz (1980), in equilibrium, investors must be indifferent when choosing whether to become informed or uninformed. The indifference condition yields the following equation (see the proof in Appendix A.3):

$$\frac{U_b^I}{U_b^U} = e^{\alpha c} \sqrt{\frac{Var^I_{b;1}[D_2]}{Var^U_{b;1}[D_2]}} = 1 \tag{12}$$

Proposition 2. In investors' biased belief equilibrium, the proportion of investors who become informed λ_1 can be solved by the benefit and cost function:

$$\Pi(\lambda) = \frac{\lambda_1^2 \sigma^2_{b;\epsilon}\delta + \alpha^2 Var^I_{b;1}[D_2]^2 \sigma_x^2 \delta - \alpha^2 Var^I_{b;1}[D_2]\sigma_x^2 \sigma^2_{b;\epsilon}}{\alpha^2 Var^I_{b;1}[D_2]\sigma_x^2 \sigma^2_{b;\epsilon}\delta} = 0; \ \text{where } \delta = e^{2\alpha c} - 1 \tag{13}$$

The implicit function (13) is jointly determined by λ_1 and the uncertainties $(Var^I_{b;1}[D_2], \sigma^2_{b;\epsilon})$. The model yields investors' biased belief equilibrium λ_1 which depends on how investors perceive the uncertainties of market and firm-specific components. Therefore, the proportion of investors who are willing to observe ϵ_1 in this model deviates from the rational expectations equilibrium which is customarily implied by Grossman and Stiglitz (1980).

On the one hand, if investors hold correct beliefs about σ_m^2 and σ_ϵ^2, in which $\Sigma_j = 0$, the model yields the same results as would be found under rational expectations. This is mainly addressed by Andrei et al. (2019), who argue that investors' information demand depends on systematic (market) uncertainty. In fact, their study rests on the assumption that investors do not suffer information transmission bias, which is represented as $\Sigma_j = 0$ in the current study.

On the other hand, this paper will relax the assumption of investors being devoid of biased beliefs. The model developed in this study comprehensively analyzes comparative statics concerning how investors' information acquisition about ϵ_1 deviates from rational expectations. This is explained by information transmission bias derived from news sentiment. Correspondingly, the positive expected supply of the risky asset ($E[x_1] = \bar{x}$) in this study's model contributes an additional implication for how firm-specific news sentiment has return predictability as an information risk premium on the risky asset.

Corollary 1. In equilibrium, from equation (13) under a necessary condition $\Pi'(\lambda_1) > 0$, since $\frac{\partial \lambda_1}{\partial \sigma_{b;j}} > 0$ and from equation (2), $\sigma^2_{b;j}$ monotonically decreases with Σ_j, the model predicts $\frac{\partial \lambda_1}{\partial \Sigma_j} < 0$, where $j \in \{m, \epsilon\}$. (The Proof is available in Appendix A.4)

2.4 Information Acquisition with Biased Beliefs of Systematic Uncertainty

On the basis of Proposition 1, the price informativeness is defined as (see Appendix A.2 for the proof):

$$n_\beta = \frac{\rho^2}{1-\rho^2} = \frac{\lambda_1^2 \sigma_{\beta;\varepsilon}^2}{\alpha^2 \mathrm{Var}^l_{\beta;1}[D_2]^2 \sigma_x^2} \tag{14}$$

where ρ is the correlation between ε_1 and $\hat{p}_1$. Holding $\sigma_{\beta;\varepsilon}^2$ constant, price informativeness increases as more investors become informed ($\lambda_1 \uparrow$), are less risk-averse ($\alpha \downarrow$), have less systematic uncertainty ($\mathrm{Var}^l_{\beta;1}[D_2] \downarrow$) or the random supply is less volatile ($\sigma_x^2 \downarrow$).

Figure 1 depicts the relationship between λ_1 which is the investors' information demand and biased belief of systematic uncertainty $\mathrm{Var}^l_{\beta;1}[D_2]$. It should be noted that, since investors' belief about the uncertainty of m_2 is biased by sentiment from the consumption of news on the condition of the market, as a consequence, the $\mathrm{Var}^l_{\beta;1}[D_2]$ is directly biased by linear projection of $\sigma_{\beta;m}^2$ and σ_h^2 (see Appendix A.1 for the proof).

[Insert Figure 1 here]

First, if we keep σ_ε^2 unbiased, the blue line in Figure 1 shows zero bias ($S_m = 0 \to \beta = 0$) in the news consumed by investors about the market. Therefore, the model is reconciled with the rational expectations as argued by Andrei et al. (2019). The theoretical maximum of information demand is reached when the systematic uncertainty is at:

$$\mathrm{Var}^l_{\beta=0;1}[D_2] = \mathrm{Var}^l_1[D_2] = \frac{\sigma_\varepsilon^2}{2(e^{2\alpha c}-1)} \tag{15}$$

and the informed investors' information quality under rational expectations is defined as:

$$v = \frac{\sigma_{\beta=0;\varepsilon}^2}{\mathrm{Var}^l_{\beta=0;1}[D_2]}. \tag{16}$$

The hump shape is due to the trade-off between price informativeness (n) and informed investors' quality of information (v). Before the systematic uncertainty reached at $\mathrm{Var}^l_1[D_2]$, as the market becomes more uncertain, higher systematic uncertainty, which reduces price informativeness, motivates investors' desire to acquire private information about ε_1. Nevertheless, if the market becomes too uncertain (above the level $\mathrm{Var}^l_1[D_2]$), it is worthless for investors to acquire information about ε_1, because the significantly decreased quality of informed investors' information makes them reluctant to pay anything at all to observe ε_1. This link between investors' information demand and economic uncertainty is mainly addressed by Andrei et al. (2019).

The novel study of Dougal et al. (2012) finds evidence that journalists are significant predictors of the positive–negative words balance of writing in the “Abreast of the Market” column in

The Wall Street Journal Their persistent bullishness and bearishness has a significant impact on the financial market. As a consequence, investors consume news about the market or economic conditions before they make investment or trading decisions, and as long as the sentiment from market news is not neutral ($S_m \neq 0$), their beliefs are biased by the market news sentiment, either overstating ($\sigma^2_{b;m} > \sigma^2_m$ as $S_m \downarrow$) or understating ($\sigma^2_{b;m} < \sigma^2_m$ as $S_m \uparrow$).

Tesser and Rosen (1975) state that people's reluctance to report bad news is a means of shielding discomforting feelings from public display. This drives more positive reporting by information disseminators, as acknowledged by Hirshleifer (2020). The green line in Figure 1 shows that as market news sentiment S_m grows to be more optimistic, the fraction of investors who want to become informed about ϵ_1 in the biased belief equilibrium is always less than that seen in the rational expectations equilibrium at every level of rationally perceived systematic uncertainty ($Var^I_{b=0;1}[D_2]$) before it reaches $Var^I_1[D_2]$. This is because, at each level of $Var^I_1[D_2]$, investors' belief about $\sigma^2_{b;m} \downarrow$ is negatively biased. Similarly, $Var^I_{b;1}[D_2] \downarrow$, from the rational perception σ^2_m is due to investors consuming news containing an optimistic tone or sentiment about the market. Investors irrationally place more aggressive orders with the negatively biased systematic uncertainty; thus, investors with this biased belief of $Var^I_{b;1}[D_2] \downarrow$ perceive the price as more informative than the price informativeness in rational expectations. Because of the systematic uncertainty's inverse relationship with price informativeness and its dominant effect on investors' information demand to observe ϵ_1, investors are less willing to acquire information about ϵ_1 in the biased belief equilibrium due to the positively biased price informativeness ($\nu_b \uparrow$) differing from the negatively biased $Var^I_{b;1}[D_2] \downarrow$.

Negativity bias has been broadly addressed in the psychological literature. Rozin and Royzman (2001) and Baumeister et al. (2001) argue that people have a tendency to pay more attention to negative information and to interpret information negatively. Hence, journalists use negative tones in their work to attract investors' attention to consume news and improve the profit of selling news (Arango-Kure et al., 2014). The red line in Figure 1 shows that as the market news sentiment becomes more pessimistic, the proportion of informed investors in the biased belief equilibrium is greater than the proportion of investors who want to become informed in the equilibrium under rational expectations at every level of $Var^I_{b=0;1}[D_2]$ before it reaches $Var^I_1[D_2]$. When investors consume market news with a negative tone, this engenders greater perception of uncertainty about economic conditions and investors tend to perceive a higher $\sigma^2_{b;m} \uparrow$, then $Var^I_{b;1}[D_2] \uparrow$. Thus, the positive biased $Var^I_{b;1}[D_2] \uparrow$ drives investors irrationally to trade less aggressively. As a consequence, investors with the biased belief of $Var^I_{b;1}[D_2] \uparrow$ perceive that price is not as informative ($\nu_b \downarrow$) as in the rational expectations model. This negatively biased price informativeness ($\nu_b \downarrow$) motivates investors to pay costs for observing ϵ_1. In equilibrium, the positively biased perception of market uncertainty from negative news sentiment leads to more information acquisition regarding ε_1 than observed in the rational expectations scenario.

2.5 Information Acquisition with Biased Beliefs of Firm-Specific Uncertainty

To study the comparative statics of the impact of the biased perception of $\sigma^2_{\beta,\varepsilon}$ on investors' information acquisition, I first reconcile the model with rational expectations ($S_e = 0 \to \beta = 0$) regarding the relationship between λ_1 and $\sigma^2_{\beta=0,\varepsilon}$. Equation (13) implies that λ_1 is a non-decreasing function of $\sigma^2_{\beta=0,\varepsilon}$ in the range of $P'(\lambda_1) \geq 0$ and it yields $\frac{\partial \lambda_1}{\partial \sigma_{\beta=0,\varepsilon}} > 0$ (see Appendix A.4 for the proof). Increasing $\sigma^2_{\beta=0,\varepsilon}$ for a given λ_1 and $Var^I_{\beta=0,1}[D_2]$ indicates that the variance of ε_1 ($Var^U_{\beta=0,1}[\varepsilon_1|\hat{p}_1]$) perceived by the uninformed investors must be increased and that the indifference condition function shifts downward from the equilibrium level. As a result, and to maintain the indifference condition at the equilibrium level, there must be more investors becoming informed, and thus a higher λ_1 in equilibrium (Grossman and Stiglitz, 1980). This intuition is also consistent with the findings presented in Veldkamp (2006), whereby the uncertainty of the given price of asset payoff is largely relative to the uncertainty of given information (here, ε_1) on the payoff. Therefore, when $\sigma^2_{\beta=0,\varepsilon}$ is high, information that reveals ε_1 is more valuable because the degree of reduction of $Var^U_{\beta=0,1}[D_2|\hat{p}_1]$ is considerable. Thus, risk-averse investors are more willing to become informed to remove the firm-specific uncertainty $\sigma^2_{\beta=0,\varepsilon}$ when it is higher, more specifically, at every level of market uncertainty.

If we assume that the market news is not biased by any tone ($S_m = 0$), the blue line in Figure 2 is the λ_1^e denoted as the equilibrium level under the rational expectations ($S_e = 0$) as λ_1 increases with σ^2_ε. Despite investors' optimal behavior in the market, their perception of $\sigma^2_{\beta,\varepsilon}$ may be biased by the tone (sentiment) in the firm-specific news. Investors are unconscious of their being biased by the news sentiment; consequently, the λ_1^e deviates to $\lambda_1^{b,e}$ and the b denotes biased belief equilibrium.

[Insert Figure 2 here]

As discussed by Berger and Milkman (2012) and Berger (2014), people are more likely to share and discuss positive content in the news or media rather than negative content, in order to maintain a reputation for providing useful information. For example, when choosing a wide range of products, advising on what to buy is more helpful than advising on what not to buy, as discussed in the marketing study of Hirshleifer (2020). Gurun and Butler (2012) find the evidence that local media tend to provide a positive slant when reporting on local firms, typically to encourage advertising expenditure from local firms. Additionally, as argued in the accounting literature, managers tend to release good news vs. bad news strategically for their own benefit - a manifestation of the agency problem in corporations (Kothari et al., 2009; Bao et al., 2019; Ahn et al., 2019).

The green line in Figure 2 shows that the curve of biased belief equilibrium $I_1^{b;e}$ is shifted downward and ends earlier in comparison to I_1^{e} in rational expectations at every level of σ_e^2. The decrease in information acquisition from investors is due to an increment in the positive tone of firm-specific news ($S_e \uparrow$) which leads to a negatively biased perception of firm-specific uncertainty $\sigma_{b;e}^2 \downarrow$. Since investors are biased to believe that a firm's future performance is more certain, ceteris paribus, the benefit derived from a reduction in the uncertainty about the payoff $Var_{b;1}^{U}[D_2]$ by knowing e_1 is underestimated by investors. Additionally, the quality of information v_b is under-perceived, because the negatively biased uncertainty about $\sigma_{b;e}^2$ makes investors feel less inclined to shed risk while keeping the systematic uncertainty unchanged. Overall, in the equilibrium with a biased belief that is more optimistic about firm-specific uncertainty $\sigma_{b;e}^2 \downarrow$, investors are less willing to pay the extra cost of acquiring the private information about e_1 and $I_1^{b;e} < I_1^{e}$ as $S_e \uparrow$.

As argued in the financial textual analysis literature, researchers find evidence that the frequency of negative words found in firm-specific news or online media dictates the overall tone of the report (Tetlock et al., 2008; Chen et al., 2014). However, the impact of negative tone in firm-specific news on investors' information acquisition decisions is unexplored. As shown in Figure 2, the red curve is investors' positively biased information demand from rational expectations. This is due to investors' positive bias about the firm-specific uncertainty $\sigma_{b;e}^2 \uparrow$ giving rise to an increment in the negativity or more pessimism in the tone of the firm-specific news. Intuitively, by reading firm-specific news with a more pessimistic tone, investors tend to predict that the firm's performance will be more uncertain in the future. As a consequence, investors over-perceive the value of information e_1 and the benefit of the reduction in $Var_{b;1}^{U}[D_2]$ by acquiring the information about e_1. Additionally, the quality of information is also overstated by a positively biased $\sigma_{b;e}^2$ while holding the systematic uncertainty constant. In sum, investors are willing to become informed as more negative sentiment ($S_e \downarrow$) exists in the firm-specific news; thus, there is an excess information acquisition in equilibrium.

In the Online Appendix, I plot another figure as a different view to show the fraction of informed investors as a function of rational perception of market uncertainty respecting biased beliefs of firm-specific uncertainty. Overall, the tone in either market news or firm-specific news raises a deviation of investors' information acquisition in equilibrium. As long as there is a non-neutral tone ($S_j \neq 0$) in the news, investors are either "sticky" or "extrapolated" to acquire private information about the firm-specific component.

2.6 Deviation of Information Risk from Rational Expectations

As argued in previous sections, news sentiment deflects investors' information acquisition about e_1 away from rational expectations due to biased beliefs about uncertainties arising from the biased tone in the news. Consequently, the monotonically decreasing relationship between the proportion of informed investors (λ_1) and news sentiment, especially for firm-specific news (S_e), results in a deviation of information risk in the risky asset and, as a consequence, in the predictability of expected returns.

Proposition 3. Expected risky asset return is $E[R_2] = \frac{a\bar{x}}{\lambda_1 \phi_I + (1-\lambda_1)\phi_U}$ and $\frac{\partial E[R_2]}{\partial \lambda_1} < 0$. From Corollary 1, $\frac{\partial \lambda_1}{\partial S_e} < 0$, therefore, sentiment in the firm-specific news has a positive predictability on the risky asset expected return $\frac{\partial E[R_2]}{\partial S_e} > 0$. (see Appendix A.5 for the proof.)

When news is not biased in its tone, the positive expected supply $\bar{x}$ implies a risk premium ($E[R_2]$) by holding the risky asset, as proposed by O'Hara (2003), due to the information risk between informed and uninformed investors in forming their investment portfolios. However, in my study, firm-specific news sentiment generates deviations in information risk because of the deviation in the firm-specific information acquisition by investors. As a consequence, there is an implied return predictability by firm-specific news sentiment. The theoretical foundation of sentiment predictability on stock returns from firm-specific news is under-explored and quite different from studies in the extant literature.[7] Therefore, this paper discusses the theoretical implications of Proposition 3 through deviations in information risk resulting from firm-specific news sentiment, to argue why firm-specific news sentiment can predict expected returns.

First, as long as price-revealing does not perfectly uncover the private information acquired by informed investors (here, e_1), this causes a non-diversified information risk to arise in the risky asset (O'Hara, 2003).[8] Additionally, as implied in a partially revealing rational expectations model, it is not possible for all investors to acquire the private information for all assets. This is because investors will value the benefit and cost in line with the indifference condition in equilibrium in order to make information acquisition decisions (Grossman and Stiglitz, 1980). Therefore, the extent to which information is private differs across assets based on the different degree of information risk in the assets. Consequently, traders demand extra compensation or expected returns to hold the assets when the information risk is large. Intuitively, the more investors choose to acquire private information ($\lambda_1 \uparrow$), the more the price will become informative in reflecting private information e_1. This will also serve to reduce the rate of privateness of the information, since the price discovery becomes more effective in revealing the private information (O'Hara, 2003). This intuition yields

[7] Additionally, there is a paucity of studies on positive firm-specific news sentiment in the existing literature. See related studies of firm-specific news sentiment predictability by Busse and Green (2002); Antweiler and Frank (2004); Tetlock et al. (2008); Chen et al. (2014); and Ke et al. (2019).

[8] As stressed by O'Hara (2003), even where investors hold portfolios with the same assets, they will have different beliefs about the expected payoff of each asset due to different information advantages between informed investors and uninformed investors. As a consequence, uninformed and informed investors hold different relative weights of risky assets in their portfolios.

$\frac{\partial E[R_2]}{\partial \lambda_1} < 0$, and the expected return in the model can be seen as a risk premium to compensate the information risk of the risky asset in a rational expectations equilibrium.

Second, as discussed above, the firm-specific news can be biased with either a more positive ($S_e \uparrow$) or more negative tone ($S_e \downarrow$). Investor's information acquisition about ε_1 deviates through the channel of a biased belief about $\sigma^2_{b,e}$ determined by S_e. In the biased belief equilibrium, the risky asset has a proportion of informed investors that is greater or lower than that which deviates from the number of investors who become informed about ε_1 under the rational expectations. This $\lambda_1^{b;e}$ deviation causes the information risk-compensating expected return of the risky asset to be higher or lower than the expected return at λ_1^{e}.

Corollary 2. If the tone (sentiment) in the firm-specific news tends to be more positive ($S_e \uparrow$), in a biased belief equilibrium, this positive tone predicts relatively higher expected returns than the rational expectations equilibrium, $E_b[R_2] > E_r[R_2]$, where b and r denote the biased belief and rational expectations equilibrium respectively. (See the proof in Appendix A.5.)

The more positive sentiment in the firm-specific news results in investors feeling less uncertain about the firm-specific component ε_1 and perceiving a negatively biased $\sigma^2_{b,e}$. In the biased belief equilibrium, there are fewer informed investors than the situation brought about by rational expectations ($\lambda_1^{b;e} < \lambda_1^{e}$). When less informed investors trade in the market, their trading incorporates little new information into the price through the price discovery process. Correspondingly, uninformed investors face more information risk, because they cannot learn much from the equilibrium price about the private information obtained by informed investors. Compared to the rational expectations equilibrium, the risky asset in this biased belief equilibrium is in fact riskier because the price discovery process is not as informative as it should be. Intuitively, traders require greater compensation to hold this risky asset since its information risk is increased by the incremental "privateness" of information. This incremental information risk comes from investors with the biased belief of $\sigma^2_{b;e} \downarrow$ as $S_e \uparrow$ being reluctant in their acquisition of private information. In sum, the more positive sentiment bias in the firm-specific news generates more information risk which is compensated for by a higher expected return of the risky asset. Finally, Corollary 2 yields an empirical prediction:

Hypothesis 1: As sentiment increases or becomes more positive or optimistic in firm-specific news, the expected return of the risky asset increases.

A more negative sentiment in the firm-specific news yields the opposite effect. In fact, the theoretical implication of negative tone in firm-level news implies less information risk in the equilibrium with biased beliefs.

Corollary 3. If the tone (sentiment) in the firm-specific news tends to be more negative ($S_e \downarrow$), in a biased belief equilibrium, this negative tone predicts relatively lower expected returns than in the rational expectations equilibrium, $E_b[R_2] < E_r[R_2]$, where b and r denote the biased belief and rational expectations equilibrium respectively. (See the proof in Appendix A.5.)

The more pessimistic or negative tone in the firm-specific news causes investors to feel more uncertain about the firm's future performance, resulting in a positively biased perception of $\sigma^2_{b,e}$. Because investors are risk-averse and may place more value on information about ϵ_1 to reduce the uncertainty, in the biased belief equilibrium, more investors are inclined to acquire the information about ϵ_1 than in the case of rational expectations ($\lambda_1^{b,e} > \lambda_1^{e}$). Since there are more informed investors trading in this biased belief equilibrium, the price discovery process sees additional new information incorporated into the price. Uninformed investors can learn more about the firm-specific component ϵ_1 from the equilibrium price through the trading process. Compared to the rational expectations equilibrium, the asset traded in the market is less risky due to an excess of investors becoming interested in being informed, causing the price discovery process to be more informative than it should be in respect of the asset. Hence, uninformed investors face relatively less information risk than they face in the rational expectations model.[9] Traders require less compensation or a lower expected return to hold the asset in equilibrium, as there is less information risk than when there is more negative sentiment in the firm-specific news. Finally, Corollary 3 yields the following empirical prediction:

Hypothesis 2: As sentiment decreases and tends to be more negative or pessimistic in the firm-specific news, the expected return of the risky asset decreases.

2.7 Discussion

The theoretical model in my study shows that the effect of biased tone or sentiment found in the news deflects investors' acquisition of firm-specific information regarding the asset's fundamental payoff, in contrast to rational expectations. Essentially, investors' eagerness or reluctance to acquire private information in this model shares similar characteristics with studies concerned with information rigidity and extrapolation.[10] Although the model discussed in this paper shares a key

[9] One could think of the extreme case as $\lambda_1^{b,e} = 1$, where, if the tone in the news about a company is strongly pessimistically biased, all risk-averse investors will panic and seek to acquire the information about ϵ_1 to reduce their positively biased uncertainty. Intuitively, the asset is no longer risky as a consideration of information asymmetry, because the effect of excess information demand minimizes the information risk in the asset.

[10] First, the information rigidity model indicates that investors tend to undervalue new information and overvalue old information. Thus, predictability comes from the slow update of new information. Second, the information extrapolation model argues that investors overweight recent information and incorporate too much of it into forecasting. As a long-run correction, there is a reversal effect. See related studies by Coibion and Gorodnichenko (2012; 2015); Bouchaud et al. (2019); Alti and Tetlock (2014); and Bordalo et al. (2019).

premise with these studies - namely, that investors' biased belief formation drives different information acquisition behaviors concerning their forecast or investment decision - the rationale for the deviation from the null to full information in equilibrium is quite different.

Most studies in the literature address investors' psychological irrationality including overcon- dence, representative bias, etc. as proposed by Tversky and Kahneman (1974). However, in the present study, the main driver of biased decisions made by investors is the consumption of biased public information in the news. It may be objected that the naivety assumption still contributes to the factor of agents' psychological bias as a trigger of irrational decision-making based on the concept of *Homo economicus* As a matter of fact, the naivety assumption can be thought of as a concession to the main argument that biased information in the news as another channel results in investors' irrational decision-making in addition to behavioral irrationality. In fact, media bias is persistent, and even rational or sophisticated consumers can not perfectly adjust for it. Information suppliers can manipulate the bias by suppressing or withholding information, motivated by either pro t-seeking or political preference (Bernhardt et al., 2008; Anderson and McLaren, 2012). By the same token, the broadly addressed issue of information withholding in nancial markets[11] contributes this particular type of bias to the process of information supply, and as a result, the Bayesian investors cannot perfectly adjust for the bias in the nancial news they receive.

Moreover, in a seminal psychological study, Le Mens and Denrell (2011) propose that even when the naivety assumption is relaxed, systematic judgment errors are still made by rational agents. This is due to the possibility that they may be subject to asymmetry of information access or their information search is interested, rather than disinterested.[12] Le Mens and Denrell stress that even when rational agents without cognitive limitations apply legitimate corrections to the bias in the sample, the corrected bias might be skewed. Thus, using skewed estimators for judgment or decision-making causes either overestimation or underestimation by the population of interest in a study.

Altogether, naivety is not necessarily a key assumption in the model in order to cause systematically biased decision-making and can be easily relaxed.[13] Therefore, agents can be rational and behave optimally as they are under rational expectations, but to some extent they are affected by the biased news. Alternatively stated, the generation of a biased belief equilibrium by biased

[11]Studies in the accounting literature have comprehensively addressed managers strategically disclosing both negative and positive news to investors (Sletten, 2012; Amir et al., 2018; Baginski et al., 2018; An et al., 2020).

[12]For example, when rational investors receive news, they may have their own preference on searching or analyzing information from the news based on their rational choice for constructing portfolios to maximize the payoff.

[13]The assumption of naivety only serves to simplify the study without solving a sub-game between investors and information suppliers such as news companies or journalists. In fact, the model can be extended to the solution of a sub-game, rst between rational investors and news suppliers as studied in Kamenica and Gentzkow (2011) and Baron (2006) who show that even rational investors are subject to bias in the news. The rest of the analysis is followed by section 2.

decision-making need not necessarily be the product of an investor's psychological irrationality.

In fact, if news sentiment can be seen as the impact of investor sentiment generating incorrect beliefs about firms' fundamentals, it should also have a short-term momentum followed by a long-run reversal correction (Tetlock, 2014). However, instead of arguing for the biased belief in the value of fundamental payoff, which is broadly addressed in the literature,[14] this paper argues that sentiment from news is the cause of investors' biased beliefs about fundamental uncertainty; and that this results in biased decisions on information acquisition. Finally, in equilibrium, the private information reflected in the price through the price discovery process is subject to these biased beliefs. Therefore, the "mispricing" in the presented theoretical model is not the result of the deviation in assets' fundamental value, but deviation in information acquisition. As a consequence, the theory suggests an empirical and testable prediction that firm-specific news sentiment has predictability on cross-sectional stock returns. Furthermore, the informativeness of the price is synchronized with investors' information acquisition in the biased belief equilibrium. Hence, firm-specific news sentiment is expected to have persistent predictability on cumulative stock returns, up to a certain length of trading periods without reversal correction.

In addition to the return predictability of firm-specific news sentiment as discussed in section 2.6, one might question whether or not the sentiment from market- or economy-wide news is comparatively predictive of stock returns. As mentioned in section 2.4, Figure 1 shows a non-monotonically increasing relationship between fractions of informed investors and biased perception of systematic uncertainty. Therefore, under normal economic conditions, the market news sentiment yields positive predictability, much like the firm-specific news sentiment. Under very uncertain economic conditions - for example, an economic bubble or recession - the market news sentiment has a reverse effect in biasing investors' information acquisition. For instance, optimistic market news sentiment makes investors under-perceive genuine market uncertainty; when the market is very uncertain above the $Var^{I}_{b;1}[D_2]$, investors acquire more private information than they should according to rational expectations. As a result, the positive market news sentiment negatively predicts stock returns under highly uncertain economic conditions and vice versa. These reversal effects of the predictability of market news sentiment are consistent with studies by Tetlock (2007) and Garcia (2013). Although the compelling non-monotonic predictability from market news sentiment, subject to different economic conditions, yields interesting theoretical and empirical predictions, a more comprehensive study on this topic is an opportunity for future research.

[14] See related studies by De Long et al. (1990); Barberis et al. (1998); Baker and Wurgler (2006); and Huang et al. (2015).

3 Data

The daily stock-level news data used in the empirical study are collected from Thomson Reuter MarketPsych (TRMI). TRMI derives newsfeeds of newly published content from approximately 40,000 internet news sites. More specifically, the news or social media content of information is assembled via TRMI crawls through hundreds of financial news sites, including, for example, The New York Times, The Wall Street Journal, The Financial Times, Seeking Alpha and many other sources that are widely read by financial professionals. In contrast to the traditional method of lexical analysis used in textual study, the technology used to create TRMI overcomes several shortcomings of the conventional approach broadly used in extant finance and economics studies (detailed information can be found in Peterson (2016)).

All daily measures from TRMI are calculated from newsfeeds before 3:30 PM ET each day. The key variables used in the present study are: U.S. stock market news sentiment S_m as a proxy for biased tone in the market news received by investors; public company news sentiment S_i as a proxy of biased tone in the firm-specific news received by investors; and the sum (absolute value) of all relevant references to an asset extracted by the algorithm called *Buzz*. *Buzz* can be thought of as a measure of the intensity of media coverage. The higher the value of *Buzz*, the more the firm is discussed in news articles or online media.

Sentiment is calculated by taking overall positive references net of negative references to an asset or market:

$$\mathrm{Sentiment}_j = \frac{\mathrm{Positive} - \mathrm{Negative}}{\mathrm{Total\ Buzz}}; \quad \text{where } j \in \{m, i\} \tag{17}$$

where $\mathrm{Positive}$ is the sum of the count of all $\mathrm{Positive}$ terms and phrases, and $\mathrm{Negative}$ is the sum of the count of all $\mathrm{Negative}$ terms and phrases; $\mathrm{Total\ Buzz}$ is the sum of total $\mathrm{Positive}$ and $\mathrm{Negative}$ counts of terms and phrases; m and i denote market and a particular firm respectively.

News sentiment varies daily, and the following empirical tests are based on firms' quarterly events and key financial variables from yearly data estimation. Therefore, this paper follows the TRMI instruction in the user guide to aggregate the news sentiment into longer frequencies such as quarterly or yearly by *Buzz*-weighted average:

$$\mathrm{Sentiment}_{j,T} = \frac{\mathrm{Buzz}_{j,t}\,\mathrm{Sentiment}_{j,t}}{\sum_t^T \mathrm{Buzz}_{j,t}}; \quad \text{where } j \in \{m, i\} \tag{18}$$

Intuitively, the higher *Buzz* at day t the more weight will be assigned to the sentiment at day t; as a result, sentiment values with high *Buzz* are more influential in contributing to the mean sentiment in a particular period. In addition, sentiment measures from news released during weekends or U.S. Federal holidays are aggregated into the next trading day.[15]

[15] In fact, the empirical study is not sensitive to how news is aggregated for non-trading days.

The TRMI contains about 4036 U.S. listed companies and the sample period is from 1998 to 2018 in this empirical study. Daily stock returns are taken from the Center for Research in Security Prices (CRSP) and financial fundamentals data are taken from the CRSP/Compustat merged database. I retain all U.S.-based common stocks with share code (SHRCD) value 10 or 11 listed on the NYSE, AMEX, and NASDAQ with exchange code (EXCHCD) 1 or 31, 2 or 32 and 3 or 33 respectively. I exclude stocks priced at less than $5 for consideration of illiquid stocks bias. Analyst forecast information is collected from the Institutional Brokers' Estimate System (IBES) and institutional ownership data are retrieved from Thomson Reuters Institutional (13f) Holdings data file. I consider two measures as benchmark systematic uncertainty proxies: VIX and the Economic Policy Uncertainty Index (EPU) introduced by Baker et al. (2016). VIX data are obtained from the WRDS CBOE index, and EPU data are assembled from the Baker et al. (2016) research lab website. Additionally, the Generalized Probability of Informed (GPIN) Trading data from NYSE stocks are gathered from the Duarte et al. (2020) website. Finally, Fama–French asset pricing factors are downloaded from the Kenneth R. French - Data Library.

Panel A in Table 1 provides summary statistics of the key daily news variables and stock financial fundamental variables in the sample. *Buzz*, market value of equity, book value to market value and Amihud's (2002) illiquidity are positively skewed and are taken as the natural log to reduce positive skewness in the subsequent regressions. Sentiment ranges from -1 (most pessimistic) to 1 (most optimistic) with a score of 0 indicating perfectly neutral sentiment. The average of sentiment in stock market news is slightly negative in tone. However, the sentiment mean in the firm-specific news is slightly positively biased.[16] Notably, the firm-specific news sentiment is much more varied than the market news sentiment. The difference between the 75th and 25th percentile in the firm-specific news sentiment is about 0.5, which is almost twice as much as the spread of sentiment in the stock market news, which is 0.28 between the percentiles. Intuitively, this is not surprising because idiosyncratic news about a variety of companies from a wide variety of news reports should understandably be divergent when compared to news about the market, which is very standardized. Hence, the variety of firm-specific news has an anticipated large spread of biased tones.

[Insert Table 1 here]

Panel B in Table 1 shows the Pearson correlation between stock market news sentiment and systematic uncertainty. First, the systematic uncertainty measures, VIX and EPU, have the expected positive significant correlation and incorporate information to represent uncertainty in economic conditions. Second, the stock market sentiment from news has a significant negative correlation with both of the systematic uncertainty proxies, and this negative relationship is consistent

[16] The average positively biased tone in the firm-specific news is consistent with Berger and Milkman (2012) and Hirshleifer (2020), who assert that $E[b] > 0$ indicating media content is more likely to be positively than negatively biased.

with extant uncertainty studies in economics.[17] The negative correlation between stock market sentiment and the VIX is even more compelling, as it is approximately -0.32. More importantly, the negative relationship between stock market news sentiment and the systematic uncertainty measures confirms the assumption in the theoretical model that more positive sentiment in the market news ($S_m \uparrow$) biases investors to understate the uncertainty of market component ($\sigma^2_{b,m} \downarrow$) and vice versa.[18]

Finally, Panel C in Table 1 shows the Pearson correlation coefficients in stock level. In general, the correlation between sentiment and other variables does not yield a significant economic relationship. However, the *Buzz* measure is positively correlated with firm size and trading turnover, but negatively correlated with illiquidity. This evidence is consistent with existing textual studies,[19] which find that larger and more liquid firms tend to be better covered in the media and thus attract more investor attention. Therefore, the *Buzz* of both the stock market and firm-specific news are important controls for the news coverage (attention) effect in the subsequent empirical tests. Finally, since there is a very high negative correlation between the size variable and the illiquidity measure after taking natural logs, to alleviate the potential multicollinearity problem in the regression analysis, only one of them is included, usually the size, as one of the control variables.[20]

4 Empirical Results

By using this novel news dataset, I first validate the proposed channel of irrationality. This particularly applies to firm-specific news sentiment as the proxy for biased public information about firm-specific condition negatively predicting firm-specific uncertainty. Next, I conduct empirical tests to verify the theoretical results including the biased effect of investors' acquisition decisions about firm-specific information resulting from either market or firm-specific news sentiment. Lastly, I verify the proposition that the deviation of information risk leads to investors' requirement for a risk premium, which is in line with the cross-sectional variation of stock returns caused by firm-specific news sentiment.

[17] Chernenko et al. (2016) study investors' over-optimism in credit markets and under-perception of the downside risk - a combination that amplifies credit booms. Baker et al. (2016) find evidence of a negative correlation (-0.742) between their uncertainty index and the Michigan Consumer Sentiment index. Da et al. (2015) construct a FEARS index as a proxy for time varying parameter uncertainty to capture investors' pessimism about market recession.

[18] In Online Appendix, I also conduct a fixed effect regressions test to verify the negative relationship between market news sentiment and systematic uncertainty.

[19] For example, Fang and Peress (2009) argue that large firms are much more likely to be covered in the media. Engelberg and Parsons (2011) study the local media impact on local trading about S&P500 index firms.

[20] In fact, all the results are unchanged, regardless of size or illiquidity.

4.1 Firm-Specific Uncertainty and Firm-Specific News Sentiment

As argued in section 2.5, investors' beliefs about the firm-specific uncertainty ($\sigma^2_{b,e}$) is biased by the sentiment in the firm-specific news. Therefore, it is important to verify this theoretical presumption before showing the evidence of biased information acquisition.

I use three measures as proxies for uncertainty in the firm-specific component. I take companies' quarterly earnings to stand for e_1 in the theory model; thus the uncertainty about quarterly earnings per share (EPS) represents the firm-specific uncertainty. With a minor abuse of notation, in the following tests, I denote $\sigma^2_{e,t}$ as the proxy for firm-specific uncertainty with investors' rational perception when $S_e = 0$ in the firm-specific news. First, I start with a simple model to estimate the uncertainty of e_1 by following the time series of firm earnings in the accounting literature. Specifically, the non-Martingale process of firm quarterly earnings has been addressed by Griffin (1977), who proposes several models to illustrate how a stationary first-order autoregressive process can be found in the data. I assume that the firm's earnings follow a simple AR(1) process; therefore, the mean squared errors (MSE) from the regression model yield firm earnings uncertainty.[21] I then conduct the AR(1) regression for company quarterly earnings as follows:

$$EPS_{i,t+1} = \gamma_0 + \gamma_1 EPS_{i,t} + \varepsilon_{i,t}$$
$$\hat{\sigma}^2_{e,t} \text{ for firm } i = \frac{\sum_{t=1}^{T} \varepsilon^2_{i,t}}{T-2} \tag{19}$$

For each firm, I conduct rolling regressions to estimate the $\hat{\sigma}^2_{b,e}$ as the first proxy of firm-specific uncertainty. I require companies to have at least 16 quarters of earnings for the estimation.

Second, the unexpected earnings (SUE) has been broadly addressed in the literature[22] and captures realized firms' fundamental performance. However, instead of using the traditional measure of SUE, I follow Hirshleifer et al. (2008) to measure the absolute value of SUE and use $Abs(SUE_{i,t})$ to identify the intensity of the seasonal random walk of unexpected earnings. Intuitively, the large SUE with a significant seasonal difference indicates a seasonal drift that is significantly different from zero between past earnings or expected earnings and future earnings. Accordingly, regardless of the seasonal difference being negative and positive, the greater the magnitude of $Abs(SUE_{i,t})$, the more difficult it is for investors to forecast either unexpectedly favorable or unfavorable company earnings using available information such as past earnings or other forecasts. Therefore, I

[21] The higher the MSE from the regression, the more uncertain the forecast earnings from the model by assuming the AR(1) process. Additionally, this AR(1) process is also in the spirit of the theoretical model setting from the study of Veldkamp (2006).

[22] Livnat and Mendenhall (2006) review related studies of SUE in accounting and corporate finance literature.

first measure the unexpected earnings, SUE, following Livnat and Mendenhall (2006) as:

$$
\begin{aligned}
\text{Compustat } SUE_{i,t} &= \frac{X_{i,t} - X_{i,t-4}}{P_{i,t}} \quad \text{(a)} \\
\text{IBES: } SUE_{i,t} &= \frac{X_{i,t} - E[X_{i,t}]}{P_{i,t}} \quad \text{(b)}
\end{aligned}
\tag{20}
$$

where $SUE_{i,t}$ (a) is calculated by using Compustat quarterly earnings data while adjusting for stock splits on $X_{i,t-4}$ and $SUE_{i,t}$ (b) is calculated by using IBES investors forecast data for robustness purposes. The $E[X_{i,t}]$ is the most recent month's median earnings forecast by analysts for the quarter t. I then take the absolute value of each measure of $SUE_{i,t}$ as the second proxy of $\sigma^2_{\epsilon,t}$.

Importantly, Bali et al. (2018) develop a new measure of idiosyncratic volatility shock, arguing that such shock is more appropriate than the level of volatility in the identification of unusual news events. Instead of arguing for the utility of measuring unusual news flow, I investigate the relationship between news sentiment and idiosyncratic volatility shock as another proxy of firm-specific uncertainty. In fact, idiosyncratic volatility shock measures the difference between future idiosyncratic risk and expected idiosyncratic risk. Intuitively, where investors use expected idiosyncratic volatility (risk) to infer future firm idiosyncratic uncertainty (risk), increased or decreased certainty in the firm-specific component will yield a smaller or higher unexpected idiosyncratic volatility respectively. As a result, the more optimistically biased tone in firm-specific news predicts a lesser volatility shock. This is because positive news sentiment induces investors to believe there will be less idiosyncratic risk in the firm-specific business condition relative to their expectation. Following Bali et al. (2018), I estimate the idiosyncratic shock as:

$$
\begin{aligned}
R^e_{i,t} &= \alpha_i + \sum_{m=1}^{M} \beta_{i,m} f_{m,t} + \epsilon_{i,t}; \\
IVOL_{i,t} &= \sqrt{var(\epsilon_{i,t}) \times \text{no. of trading days}}
\end{aligned}
\tag{21}
$$

where $f_{i,m}$ is the benchmark pricing factor. I begin by estimating the Fama–French five factor and momentum factor model for each stock. I require a firm to have had at least 60 daily returns. I then conduct daily cross-sectional regressions for each firm to estimate the idiosyncratic shock as:

$$
IVOL_{i,t} = \phi_{0,t} + \phi_{1,t}\overline{IVOL}_{i,t-1} + \sum_{j=1}^{10} \Phi_{j,t} D_{i,j} + \nu_{i,t}
\tag{22}
$$

where $IVOL_{i,t}$ from (21) and $\overline{IVOL}_{i,t-1}$ is the past average stock idiosyncratic volatility as investors' expectation about firms' idiosyncratic risk calculated by the moving average window between $t-24$ and $t-4$. $D_{i,j}$ is the 10 industry classifications dummy from Kenneth French's Data

Library. Thus, the daily unexpected shock to idiosyncratic volatility is defined as $IDIO^{shock}_{i,t} \equiv v_{i,t}$.

Finally, I use the three measures of firm-specific uncertainty to conduct the test as follows:

$$
\begin{aligned}
\hat{\sigma}^2_{e,t} &= \beta_0 + \beta_1 Sentiment_{i,[t-30,t-1]} + X\delta + \epsilon_{i,t} \quad \text{(a)} \\
IDIO^{shock}_{i,t} &= \beta_0 + \beta_1 Sentiment_{i,t-1} + X\delta + \epsilon_{i,t} \quad \text{(b)}
\end{aligned}
\tag{23}
$$

where $\hat{\sigma}^2_{e,t}$ is the proxy from (19) or (20) as representing the firm-specific uncertainty. The model (a) in (23) is based on quarterly earnings data and the $Sentiment_{i,[t-30,t-1]}$ is firm-specific news sentiment in the most recent month before quarter t calculated by the $Buzz_{i,t}$-weighted average as equation (18) from daily data. The model (b) is based on daily idiosyncratic volatility shock analysis. The X in both (a) and (b) is a vector of control variables (see Appendix B.1 for details) and δ as the coefficient vector. I use fixed effect regression for model (a) and daily Fama–Macbeth (1973) cross-sectional regressions for model (b) to test whether firm-specific news sentiment negatively predicts the proxy of firm-specific uncertainty and idiosyncratic shock respectively. In sum, the β_1 in both model (a) and (b) is expected to be both significant and negative.

Table 2 summarizes the regression results from models (a) and (b). All proxies of firm-specific uncertainty variables are winsorized at the 1% level to reduce the impact of extreme outliers. Additionally, I take the natural log of regression variance from equation (19) to reduce extreme positive skewness. Columns (1)-(3) are fixed effect regressions with standard errors clustered by firm and year-quarter. It should be noted that I use regression variance as the dependent variable in the model. Chen et al. (2018) use residuals as the dependent variable in the second step regression, and they argue that estimation of the interest explanatory variable (β_1 here) might be biased if the independent variable (sentiment) is correlated with the variables used in the first step regression. Therefore, it is necessary to include the independent variables used in the first step regression in the second step regression. I then include $EPS_{i,t-1}$ as an additional control variable.[23]

[Insert Table 2 here]

First, column (1) clearly shows that firm-specific news sentiment negatively predicts the firm earnings AR(1) regression variance $\hat{\sigma}^2_{e,t}$ estimated from equation (19). At an increase of two standard deviations of firm-specific news sentiment (2×0.2871) the firm earnings which can not be explained by the AR(1) decreased by about 0.8%. This is strong evidence for the claim that a more optimistic tone in firm-specific news may induce investors to believe that quarterly company earnings are less uncertain by applying the AR(1) model to the forecast. Second, and unsurprisingly, columns (2) and (3) show that $Abs(SUE_{i,t})$ is significantly negatively predicted by sentiment in firm-specific news. An increase in firm-specific news sentiment by two standard deviations,

[23] By the same token, I also include $\overline{VOL}_{i,t-1}$ in the model (b) from equation (23).

$Abs(SUE_{i,t})$ decreases by about 0.6% and 4.6% of its mean value, respectively, to two measures of SUE. The more optimistic tone in the news causes investors to be more confident about expected earnings or about past earnings as a reliable forecast for future earnings either up or down; thus, they feel less uncertain about the company's earnings performance, and perceive less dispersion of unexpected earnings. The reverse is also true in relation to a more pessimistic tone in the news.

There is an intriguing finding that the IBES measured $Abs(SUE_{i,t})$ has much more economic significance - about 7.6 times larger than Compustat-measured $Abs(SUE_{i,t})$. The large impact that arises from applying IBES data is consistent with studies in the accounting and corporate finance literature, which argue that the analyst earnings forecasts are more likely to be subject to bias due to irrationality from optimism.[24]

Third, column (4) shows daily cross-sectional Fama–Macbeth (1973) regressions of $IDIO^{shock}_{i,t}$ on firm-specific news sentiment; standard errors are Newey-West corrected. The regression coefficient on firm-specific news sentiment shows consistent results with columns (1)-(3). The more optimistic tone in the daily firm-specific news leads investors to believe that their understanding of firm idiosyncratic risk is less uncertain. This negatively biased firm-specific uncertainty causes investors to perceive less future idiosyncratic risk in the firm, which results in them perceiving a lower value in the unexpected idiosyncratic volatility. As a consequence, a lesser idiosyncratic volatility shock is predicted where there is more positive sentiment in the firm-specific news and vice versa.

In sum, if we assume that the econometric model uses the correct fundamental variables which are widely considered to be rational or objective, then the model should be impartial in predicting future firm-specific uncertainty (risk) of earnings. However, all three tests using either quarterly earnings data or daily idiosyncratic volatility data show strong evidence that by conditioning on biased tone in the firm-specific news, the sentiment negatively predicts every proxy of firm-specific uncertainty. Therefore, as investors read the news prior to making investment or trading decisions, their beliefs about future firm-specific uncertainty are biased either upward from negative sentiment or downward from positive sentiment in the news. This biased belief, caused by news sentiment transmits to the biased effect on investors' decision to acquire firm-specific information.

4.2 Firm-Specific Information Acquisition and News Sentiment

As the model predicts, the biased beliefs about uncertainty shift investors' acquisition of firm-specific information, component η_1, in comparison to the acquisition decision under rational expectations. However, investors' information acquisition, η_1, is not directly observed, so I conduct an event study, based on existing literature, of earnings announcements to test the inverse relation-

[24]See relevant studies by De Bondt and Thaler (1990), Abarbanell and Bernard (1992), and Easterwood and Nutt (1999), which argue that analysts are more likely to give optimistic forecasts.

ship between news sentiment and firm-specific information acquisition predicted by the theory.

I follow a novel measure of firm-specific information acquisition developed by Weller (2018) to estimate a jump ratio which is calculated within a certain window before and after companies' quarterly earnings announcements.First, I define the pre-announcement window, as starting from 21 ($a = 21$) trading days before the announcement, as the period of identification of earnings-related information entering into the price before the announcement. Second, the identification of earnings-related information incorporated into prices when the earnings information is released spans two trading days ($b = 2$) after the announcement.[25] Based on the defined study windows, I first estimate the ACAR for both pre- and post-announcement as the price drift net of predicted returns from the Fama–French five-factor model. The momentum factor is also included:

$$CAR_{i,t}^{j_1,j_2} = \sum_{t=j_1}^{j_2} \left(R^e_{it} - a_i - \sum_{m=1}^{M} b_{i,m} f_{m,t} \right) = \sum_{t=j_1}^{j_2} e_{i,t} \tag{24}$$

where $R^e_{i,t}$ is stock excess return and $f_{m,t}$ is the Fama-French and momentum factors. The a_i and $b_{i,m}$ is estimated by using 252 daily return data points and 90 days before the earnings announcement. I require stocks that have observations on at least 63 trading days to estimate the factor model.

The jump ratio is estimated by using the post-announcement ACAR scaled by the total ACAR including before and after the earnings announcement as:

$$Jump_{i,t}^{a,b} = \frac{CAR_{i,t}^{T-1,T+b}}{CAR_{i,t}^{T-a,T+b}} \tag{25}$$

where $a = 21$ and $b = 2$ as the pre-announcement and post-announcement window respectively. As indicated in Weller (2018), the denominator $CAR_{i,t}^{T-a,T+b}$ may be close to zero. Therefore, I follow the instruction proposed by Weller (2018) to set up a threshold as $|CAR_{i,t}^{T-21,T+2}| > \sqrt{24}\,\hat{s}_{i,t}$ where $\hat{s}_{i,t}$ is daily return volatility during the 24-day event window.

Intuitively, if investors decide to acquire earnings-relevant information before the announcement day, the price incorporates more information about earnings. As a consequence, informed traders drive a greater price drift ($CAR_{i,t}^{T-a,T+b}$) than that which would be expected when earnings information becomes public. Conversely, if few investors decide to acquire information about firm earnings before the information is publicly revealed, on the announcement day the price drift jumps to incorporate the newly released earnings information once it becomes available. As in the model setting, if more informed investors conduct trading before the earnings announcement, the price is more informative and reflects earnings information (θ_1 in the model) which can be partially

[25] The additional two days are to accommodate for the post-earnings announcement drift effect.

gleaned by uninformed investors as well. Hence, when the earnings information becomes available, as price has reflected the earnings information before it is revealed, the price will not jump as much as in the case in which no or few investors are informed about the earnings before the information is released. Thus, the higher the price jump ratio, the less information is incorporated in the price (less information acquisition, e_1) relative to the post-announcement information set and vice versa (Weller, 2018). Therefore, aggressive and informed traders who trade before the earnings announcement drive the price jump close to 0, while an absence of informed trading drives the price jump towards 1. To test how news sentiment biases investors' firm-specific information acquisition, I conduct a fixed effect regression as follows:

$$jump_{i,t} = b_0 + b_{j,1} Sentiment_{j,[t-21,t-1]} + Xd + e_{i,t}; \quad \text{where } j \in \{m, i\}. \tag{26}$$

where $Sentiment_{j,[t-21,t-1]}$ is the $Buzz_{i,t}$-weighted average news sentiment from 21 trading days up to 1 day before the earnings announcement. X is a vector of control variables (see detailed definitions in Appendix B.1) and d as the coefficient vector. The theoretical model in Andrei et al. (2019) indicates that economic uncertainty in the fundamental payoff matters for investors' information acquisition decision. Therefore, I add customary systematic uncertainty measures, either VIX or EPU, as an additional control variable to identify the impact of news sentiment more clearly. From the theory model predictions in Corollary 1, the $b_{j,1}$ is expected to be positively significant to indicate a more positive or optimistic tone in market or firm-specific news, predicting a higher price jump which implies less information acquisition, e_1, and vice versa.

Panel A in Table 3 shows the results from equation (26) regarding the impact of stock market news sentiment on firm-specific information acquisition. First, the specification in column 1 only controls for month- and firm-fixed effects, and indicates that sentiment from stock market news strongly predicts positive price jumps. A one-unit increase in the optimism of market news sentiment causes the price jump to increase by 0.089 (relative to the median jump ratio of 0.3365). In line with firm-specific information acquisition interpretation, an increase in stock market news sentiment by one standard deviation (0.09) is associated with a 2.38% decrease in the proportion of earnings announcement-related price impact that arises pre-announcement.

[Insert Table 3 here]

Columns (2) and (3) in Panel A include additional controls to identify the impact of stock market news sentiment on firm-specific information acquisition. In addition to fundamental controls, I also add the $Buzz_{j,t}$ variable to control for a potential asymmetric information reduction effect as stated in Tetlock (2010). As he argues, public information from news can dissipate private information held by informed investors. $Buzz_{j,t}$ is the proxy of intensity of news coverage; therefore, based on the findings from Tetlock (2010), a higher value for $Buzz_{j,t}$ indicates there is more public information available to investors and may resolve information asymmetry. Because VIX and EPU may be strongly correlated, I control for each of the measures one at a time.

Including additional controls, the second specification in column (2) – the impact of market news sentiment on the jump ratio – shows very similar results. The VIX, as expected, is negatively significant in predicting the jump ratio, which is consistent with the model under rational expectations: as systematic uncertainty increases, firm-specific information acquisition increases (Andrei et al., 2019). Column (3) shows a slightly higher magnitude of impact from stock market news sentiment on information acquisition. The EPU index has an expected negative sign as consistent with the VIX implication, but is not statistically significant.

Panel B in Table 3 shows empirical results from equation (26) with respect to the impact of firm-specific news sentiment on information acquisition about θ_1. Column (1) is the specification-only controls for month- and firm-fixed effects. The biased tone in firm-specific news shows significantly positive predictive power on the jump ratio. With a one-unit increase in firm-specific news sentiment, the price jump ratio increases by 0.057 (relative to the median value of jump ratio 0.393). With regard to firm-specific information acquisition, an increase in the optimism of firm-specific news sentiment by one standard deviation (0.31) causes investors to acquire less earnings-related information by 4.5% before the earnings announcement. Columns (2) and (3) are specifications including additional controls. I also control stock market news sentiment in specification 2 and 3. In fact, the magnitude of economic significance from firm-specific news sentiment is not compromised after adding additional control variables. Although stock market news sentiment maintains its explanatory influence on the jump ratio, it is promising that bias in the firm-specific news is inversely related to investors' firm-specific information acquisition. Moreover, there are intriguing findings between Panel A and B in Table 3. The two $Buzz_{j;t}$ controls, stock market and firm-specific news, have entirely opposite impacts on the jump ratio. In fact, more $Buzz_{f;t}$ from firm-specific news significantly decreases the jump ratio, which implies an increase in firm-specific information acquisition before the earnings announcement. However, in the case of stock market news, $Buzz_{m;t}$ has the reverse effect. This intriguing evidence is consistent with the key argument of Tetlock (2007; 2008; 2010) that market news sentiment does not contain value-relevant information, unlike firm-specific news sentiment regarding firms' fundamentals.

5 Information Risk from Firm-Specific News Sentiment

5.1 Probability of Informed Trading and Firm-Specific News Sentiment

As the model proposed that information risk is affected by variations in the proportion of informed investors as a result of firm-specific news sentiment, I investigate this proposition by testing the relationship between the probability of informed trading (PIN) developed by Easley et al. (1996) and news sentiment from particular firms. The PIN has been empirically tested as a proxy of information risk and the risk premium can be found in cross-sectional asset returns.[26]

Intuitively, as more investors choose to become informed about b_1 and trade in the market, the equilibrium price becomes more informative and is of more utility to uninformed investors. Thus, to hold the indifference condition as proposed by Grossman and Stiglitz (1980), informed investors are not willing to trade aggressively by submitting additional more-informed orders (i.e. buying when asset value is high and selling when asset value is low). In fact, when price is informative, there is a high proportion of informed traders in the market, which leads to a decline in the knowledge disparity between informed and uninformed investors. Correspondingly, submitting more informed orders does not contribute extra benefits to informed investors, since they do not want uninformed investors to gain a free ride by learning from the equilibrium price, which is itself an increment of uninformed utility. As a consequence, a reduction in aggressively informed orders submitted by informed traders decreases the order arrival rate of informed traders, m, in the PIN model. Therefore, as informed order arrival rate decreases, the PIN value decreases and there is less information risk in the asset.

As stated in Corollary 2 and 3, a more optimistic tone in firm-specific news decreases firm-specific information demand by investors and results in more information risk in the biased belief equilibrium and vice versa. Therefore, following the literature that argues that PIN can be seen as a proxy for information risk, I conduct a hypothesis test on the relationship between PIN and sentiment from firm-specific news. However, the traditional measure of PIN is subject to bias, which is that it cannot match a large amount of variation in turnover initiated by noise trade (Duarte et al., 2020).[27] Hence, in considering the limitations of the PIN model, which may result in inaccurate statistical inference, I use Generalized Probability of Informed Trade (GPIN) from Duarte et al. (2020) as an information risk proxy. It should be noted that GPIN data are only available for NYSE stocks. Consequently, the empirical results are intended to be very conservative and understate the impact of news sentiment on information risk, because companies traded on NYSE are, in general, large liquid stocks presumed to have fewer information asymmetry problems. I conduct the fixed effect regression as :

$$GPIN_{i,t} = b_0 + b_1 Sentiment_{i,t-1} + Xd + e_{i,t}; \tag{27}$$

[26] For a comprehensive study and review, see Duarte et al. (2020). See studies by Easley et al. (2002); Easley and O'hara (2004) and Easley et al. (2010) for information risk premium implied by PIN.

[27] Duarte et al. (2020) state that the implied variability of buys and sells from the PIN model, in general, is 550 times smaller than the realized variability in the data. The biased estimation from PIN derives from the failure of the model to capture large amounts of variability in noise trading.

where $GPIN_{i,t}$ is the stocks' generalized probability of informed trade, estimated with daily trade tick data, and $Sentiment_{i,t-1}$ is $Buzz_{i,t}$-weighted average firm-specific news sentiment in year $t-1$. The X includes a bundle of control variables (see Appendix B.1 for details) and α as the coefficient vector. Since the news data begins in 1998, the regression starts from 1999. The reason I use a lagged sentiment variable as the explanatory variable is due to a concern about potential inverse causality in contemporaneous periods. More specifically, since firm-specific news comes randomly throughout the year, a contemporaneous regression cannot be guaranteed to be free of endogenous issues about companies' information asymmetry, which may potentially affect sentiment in the firm-specific news. All in all, I expect a positively significant b_1 in equation (27), indicating that positive firm-specific news sentiment predicts high information risk.

Table 4 presents results from equation (27). The specification in column (1) only controls for year- and firm-fixed effects, and it confirms that biased tones in firm-specific news predict positive GPIN. With a one-unit increase in the optimism of sentiment, information risk as measured by GPIN increases by approximately 0.017 relative to its mean value (0.26). By adding more controls in columns (2) and (3), which are also variables with considerable explanatory power in respect to information asymmetry, it still maintains strong positive significance in explaining the variation of GPIN. In fact, a two-standard deviation (0.18) increase in the positivity of news sentiment concerning a particular firm increases the GPIN by about 2%. This indicates that the buy or sell orders are 2% more likely to be from informed traders who hold private information about the risky asset. Therefore, information risk in risky assets increases as tone, reported in the firm-specific news, becomes more optimistic. This high information risk caused by positive news sentiment implies a reduction of the benefit of price informativeness gained by the uninformed traders to alleviate the information asymmetry risk trading against informed traders and vice versa.

[Insert Table 4 here]

O'Hara (2003) proposed that information asymmetry existed in risky assets as the disparity in the information held by informed and uninformed investors. This information risk is perceived by traders who require compensation to hold risky assets. As shown by the results in Table 4, sentiment in the firm-specific news affects information risk as measured by GPIN in risky assets. Next, I investigate the variation in cross-sectional asset returns (the risk premium) using the deviation in information risk caused by biased tones in the firm-specific news through the biased information acquisition in equilibrium.

5.2 Firm-Specific News Sentiment Impact on Cross-Sectional Stock Returns

In the model in section 2, Corollary 2 and 3 indicate a monotonic relationship between firm-specific news sentiment and expected returns on the risky assets. This reflects the variation in information risk. Hence, to evaluate whether firm-specific news sentiment induces a deviation in information risk, which causes variation in the expected returns of assets, I examine whether sentiment from firm-level news on day t predicts positive firm excess return on $t+1$. In addition, since I argue that this predictability of cross-sectional returns stems from the investors' biased firm-specific information acquisition in equilibrium rather than from the mispricing of fundamental value, I expect that the positive predictability is not reversed and is persistent in cumulative returns for the subsequent trading days. Therefore, I conduct tests on cumulative returns up to 5 and 10 days after the firm-specific news is released.

The dependent variable in the regression model is day $t+1$ stock excess return (R^e_{t+1}) and either 5 or 10 days' worth of cumulative returns ($R^e_{t+2;t+5/10}$), where the day $t+1$ return is omitted from cumulative return as a consideration of bid-ask bounce. The control variables, all firm characteristics that affect the predictability of expected returns, include measures of company size[28] ($Size_{i;t}$), book to market ratio ($BM_{i;t}$), operating profitability ($OP_{i;t}$), investment ($INV_{i;t}$), yearly return momentum ($MOM_{i;t}$) excluding the most recent month, the last month return volatility ($RV_{i;t}$) and the last month return ($ST_{i;t}$) as short-term reversal effects. To consider return reversal predictability,[29] I add day t abnormal return $AbRet_{i;t}$ defined as the raw return minus the value-weighted market return from CRSP and cumulative abnormal returns from the past five trading days ($AbRet_{t-5;t-1}$). As demonstrated in the model of Llorente et al. (2002), if stock trading volumes are aligned with daily returns, this strongly predicts future returns.[30] Hence, additional controls regarding the trading volume effect include firms' abnormal trading volume ($AbTurn_{i;t}$), defined as log turnover on trading day t net of its average of log turnover from $t-5$ to $t-1$ and the interaction between day t abnormal return and trading volume ($AbRet_{i;t} \times AbTurn_{i;t}$).

The main test is on the firm-specific news sentiment on day t ($Sentiment_{i;t}$) to predict day $t+1$ or cumulative returns in the following days. There are two major concerns in the identification of the effect of firm-specific news sentiment on information risk premium. First, Tetlock (2010) proposes that public information from news resolves information asymmetry by testing the reduction of return reversal and volume-induced return on firm news days. Because of the definition of 'sentiment,'- which I argue is the tone of public information in the news, tending to induce a deviation in the proportion of information asymmetry in risky assets - it is necessary to control for the impact of news on resolving asymmetric information as stated by Tetlock (2010). Therefore, I use $Buzz_{i;t}$ as a proxy for the intensity of firm-specific news coverage interacting with firm-abnormal returns on day t. The rationale for controlling $Buzz_{i;t}$ is that, as a company is widely discussed

[28]Since Amihud (2002) illiquidity measure is highly negatively correlated with the size measure about -0.92, I use both of them one at a time and the results are unchanged.

[29]See related studies by Roll (1984); Jegadeesh (1990); and Lehmann (1990).

[30]See related studies of trading volume impact on return predictability by Campbell et al. (1993) and Lee and Swaminathan (2000).

in the news or there is more public information available to uninformed investors, it is easier for uninformed investors to infer superior information about the firm from the news and become less reluctant to provide liquidity to informed investors (Tetlock, 2010). As a consequence, when investors have more relevant public information about a firm on day t, the abnormal return at day t is conditional on the availability of firm-level news information, and is expected to lead return momentum as liquidity shock is dissipated gradually after the news is released (Tetlock, 2010).

Second, there is a growing number of studies[31] using textual data to assess the effect of sentiment in firm-specific news or online media platforms containing value-relevant information about companies' fundamentals. For example, Tetlock et al. (2008) and Chen et al. (2014) argue that negative sentiment in the firm-level news or media is instructive to investors regarding unfavorable earnings information from companies. However, the predictability of firm-specific news sentiment, as argued in this paper, mainly relates to the risk premium of information asymmetry, which is distinct from the argument regarding genuine information in extant studies. In fact, the predictability of the effect of news sentiment impact on cross-sectional stock returns, which is the main relationship evaluated in this study, is in addition to the predictability found in the growing literature.

Thus, a thorough consideration of the genuine information effect is necessary, as an essential control to conduct a return predictability test in the subsequent main regressions. If the genuine information effect dominates predictability from the news sentiment, the empirical results would not show a significant predictive power from the firm-specific news sentiment after controlling for the genuine information effect. Therefore, it is indispensable to disentangle the effect of genuine information contained in the firm-level news sentiment from the sentiment variable ($Sentiment_{i,t}$) for each firm. There is a valuable measure provided by TRMI: $EmotionVsFact_{i,t}$ and ranges from -1 to 1. This index measures the proportion of emotional references net of the factual reference from news articles. The emotional reference counts subjective words in the news article such as people's expressed opinions or feelings about the news stories. The factual reference counts objective words or fundamental firm information from the news stories, such as content related to operation, earnings, merging or accounting (see Appendix B.1 for details).

Intuitively, the closer to 1 in the $EmotionVsFact_{i,t}$ measure is, the more subjective opinion there is in the news story about a company. Conversely, the closer to -1 in the $EmotionVsFact_{i,t}$ measure is, the more factual, objective or fundamental material is in the news story about a company. In line with the evidence of firm-specific news sentiment containing genuine information about firms' fundamental payoff, we should expect the more factual (lower number of $EmotionVsFact_{i,t}$) reference to interact with $Sentiment_{i,t}$ to predict positive stock future returns.[32]

[31] Comprehensive survey studies can be found in Tetlock (2014) and Loughran and McDonald (2016).

[32] For example, a negative number of $EmotionVsFact_{i,t}$ and a negative sentiment indicate negative value-relevant

I therefore interact $EmotionVsFact_{i,t}$ and $Sentiment_{i,t}$ to control for the potentially genuine information contained in $Sentiment_{i,t}$. Furthermore, I add an interaction between $EmotionVsFact_{i,t}$ and $AbRet_{i,t}$ as another control for the effect of news resolving information asymmetry. For example, the greater the proportion of factual information in the company news, as investors read the news, the more likely they are to infer the private information from factual information in the news and vice versa.[33] Finally, all independent variables are standardized by each day before computing interaction terms for easy interpretation. I require at least 100 firms to have some news and non-missing independent variables each day.[34] For all firms with news sentiment, I estimate daily cross-sectional Fama–Macbeth (1973) regressions to evaluate whether the positively biased tone in the news predicts future returns either on day $t+1$ or the cumulative return in the following days. The cross-section regression specifications are:

$$DepVar_{i,t+1} = b_0 + b_1 Sentiment_{i,t} + dX + e_{i,t} \text{ for each trading day } t. \quad (28)$$

where $DepVar_{i,t+1}$ is either R^e_{t+1} or $R^e_{t+2,t+5/10}$ and the X is a vector of control variables and d as the coefficient vector. The main purpose of this test is to determine whether b_1 is significantly different from 0. More importantly, as per the theoretical predictions argued in section 2, it is expected that b_1 will have a positive value, indicating that a more optimistic tone in the news will bring about a higher return, as investors expect to be compensated for higher information risk in the risky asset and vice versa.

Column (1) in Table 5 is the results of day $t+1$ return prediction. As all independent variables are standardized, the regression coefficients interpret a change in the dependent variable as a change of one standard deviation on the predictors. Notably, the firm-specific news sentiment ($Sentiment_{i,t}$) at day t significantly predicts positive stock return on the next day, even after controlling for other important effects implied by news information. With an increase in firm-specific news sentiment by one standard deviation, the next day's return increases by about 3.1 basis points, which is equivalent to a 0.65% monthly return. Surprisingly, this increment in the next day return is economically significant; in fact R^e_{t+1} increases by approximately 110% relative to its mean (2.83 basis points) in the sample period.

[Insert Table 5 here]

More precisely, I calculate the marginal effect of sentiment by netting the effect of predictabil-

information for the firm fundamentals and vice versa.

[33] Intuitively, if the news contains more factual information about firm fundamentals, there should be a reduction on daily return reversal. In other words, one would expect a return momentum as more factual information in the news is reported in day t.

[34] Because the regression model contains about 20 predictors, the minimum observation is a consideration of sufficient degrees of freedom and the statistical power of the tests. However, the results are insensitive to this requirement.

ity from genuine information within sentiment, which is controlled by the interacted variable $EmotionVsFact_{i,t}$ * $Sentiment_{i,t}$. Its regression coefficient is consistent with the literature. For instance, sentiment extracted from subjective references in the news causes a reversal prediction. On the contrary, sentiment about factual or fundamental references in the firm-specific news positively predicts future returns. This evidence provides an important contribution to the debate in the behavioral finance literature by using textual analysis of whether sentiment is a form of bias affecting investors' valuation of an asset or contains genuine information about the firm fundamentals. Therefore, the marginal effect of sentiment predictability on $t+1$ return is about 0.42 basis point (representing a 15% increase from its mean), increasing on the next day return in line with a one standard deviation increase in sentiment and net of the effect of one standard deviation in factual reference in the firm-specific news (3.1 1 2.67 1).[35]

By disentangling the effect of news sentiment that may cause investors either to estimate firms' value in a biased way or to obtain firm value-relevant information, the sentiment maintains significant positive predictability on firm future returns. This effect is both statistically and economically significant. Hence this additional cross-sectional return predictability implies variation in information risk through firm-specific news sentiment, thus causing firm-specific information acquisition to deviate from the rational expectations equilibrium. Additionally, the control variables $Buzz_{i,t}$ $AbRet_{i,t}$, and $EmotionVsFact_{i,t}$ $AbRet_{i,t}$ are all consistent with the findings in the literature on capturing the effects of asymmetric information mitigation.[36]

Columns (2) and (3) are results regarding cumulative returns at 5 and 10 days respectively after firm-specific news is released. The firm-specific news sentiment remains positively significant on 5- and 10-day cumulative returns. With a one standard deviation increase in the optimism of news sentiment, the 5- or 10-day cumulative returns increase by about 26.83% and 13.15% respectively relative to their mean values (14.03 and 33.01 basis points). However, the control variable $EmotionVsFact_{i,t}$ $Sentiment_{i,t}$ is no longer significant. The insignificance of $EmotionVsFact_{i,t}$ $Sentiment_{i,t}$ on cumulative returns infers that the market is efficient as one-day turnaround to either incorporate valuable information about firms from news or to correct the mispricing resulting from investors' irrational response to public news containing more subjective references.

More importantly, the empirical evidence of persistent return predictability from firm-specific news sentiment is similar to that found in information diffusion studies, although the rationale is somewhat distinct. In general, information diffusion studies such as that of Hong and Stein (1999) argue that boundedly rational investors cause gradual information diffusion and their simple trading

[35] A one standard deviation increase on factual references is -1, in line with one standard deviation increase on sentiment of +1. Therefore, the predictability is contributed by genuine information from news sentiment is 2.67 (2.67 1 1). Thus, the net effect is calculated by subtracting 2.67 from 3.1.

[36] See related studies by Tetlock (2010) and further discussion in section 7.4 in the Online Appendix.

strategy causes short-run momentum and long-run overreaction on returns. However, in my model, the key issue is that public information from news is biased in the way in which it is reported; thus, the bias stems from the news supplier, not from the investors themselves, particularly their assumed irrationality.

The consequence is that, to some extent, investors are forced into being biased in their beliefs about the uncertainty of fundamental payoff, due to their being unduly influenced by the firm-specific news. Investors form a biased belief equilibrium regarding firm-specific information acquisition, e_1. In sum, as long as the tone of news information is biased (either positively or negatively), there is always a deviation in the acquisition of firm-specific information in equilibrium. This implies either higher or lower information risk in the asset compared to the rational expectations model. As proposed by O'Hara (2003), traders require compensation to hold risky assets containing more information risk, which, in my study, varies with sentiment in firm-specific news.

5.3 Firm-Specific News Sentiment Portfolio Analysis

The cross-sectional variation of stock returns predicted by firm-specific news with optimistic and pessimistic tones suggests news sentiment may result in variation in information risk across assets. Therefore, I conduct a portfolio formation analysis sorted by daily firm-specific news sentiment by following Fama and French (1992) to verify whether the risk premium of information risk triggered by firm-specific news sentiment can be captured by traditional asset pricing factors.

At the end of each trading day (3:30 PM EST), I first use monthly NYSE breakpoints of the last month median market capitalization from the Kenneth R. French Data Lab to split stocks into two portfolio sizes: small (S) and large (B). Independently, I rank stocks based on daily firm-specific news sentiment into three sentiment portfolios: pessimistic (N), neutral (M), optimistic (P). Stocks within the lowest 30^{th} percentile (N) have more negative sentiment (pessimistic tone) from their daily news articles; stocks within the highest 30^{th} percentile (P) have more positive sentiment (optimistic tone) in the daily news stories; and the stocks within the middle 40% (M) contain relatively neutral tones in the news discussion. The six interacted portfolios, value-weighted with respect to size and firm-specific news sentiment, are N/S, N/B, M/S, M/B, P/S, and P/B, sorted by portfolio size and news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios (P/S, P/B) and the average of short position in the two negative sentiment portfolios (N/S, N/B) each day and I calculate the next day ($t+1$) value-weighted portfolio returns from this zero-cost trading strategy.

The firm-specific news sentiment zero-cost portfolio generates significant positive average daily returns of about 6.6 basis points (which equates to a 16.63% annualized return) shown in Table 6. It should be noted that there is a concern that illiquidity could play a role in information risk, and that could itself explain the news sentiment pricing effect, as this study proposes that it can be seen as a trigger of information risk across assets. I calculate daily Pastor and Stambaugh liquidity factors (PSLIQ) by following Pástor and Stambaugh (2003)[37] as an additional important pricing factor to test the pricing capability of news sentiment. Panel A in Table 6 is the Pearson correlations between the firm-specific news sentiment factor and other customary pricing factors. In fact, the news sentiment factor has a weak negative correlation with the market factor (-0.106), the value factor (-0.102), and the short-term reversal factor (-0.142), and a weak positive correlation with the momentum factor (0.202). The remaining factors have correlations with the news sentiment factor roughly close to zero. Next, I investigate whether these existing factors can explain the abnormal return from the pricing factor constructed by daily firm-specific news sentiment.

Panel B in Table 6 shows the risk-adjusted daily returns from the zero-cost portfolio based on a trading strategy informed by news sentiment. I use the CAPM, Fama–French three factors (1993), and Fama–French five factors (2015) models in line with illiquidity factor to adjust the returns of the zero-cost portfolio. I also include additional momentum factor, short-run reversal and long-run reversal factors as a consideration of behavioral pricing effect in the news sentiment trading strategy portfolio. Columns (2)-(6) clearly show that none of the models fully explain the abnormal returns generated by the zero-cost portfolio that is based on firm-specific news sentiment. The average abnormal daily return ranges from 6.1 to 6.8 basis points across different pricing models. Notably, the liquidity factor does not contribute any significant effect to the value of the abnormal return from the zero-cost news sentiment portfolio.

Interestingly, the full model in column (6) shows that the news sentiment factor portfolio return is negatively significant in relation to both the value and the short-term reversal factors. Additionally, it has positive factor loadings on the momentum and investment factors. Intuitively, the significance of the momentum and short-term reversal factors captures potential behavioral effects from news sentiment affecting investors' valuation of stock performance. For example, if return on a stock is high on day t and, in the meantime, there is positive news about the firm, investors tend to under-react to this information and generate a momentum effect (Hong and Stein, 1999). Factor loading on the investment factor may be explained by news sentiment regarding companies' fundamentals, such as reporting on a firm's investment plan. For instance, if news stories report pessimistically about a company that plans to shrink its future investment, investors analyzing the company may suffer a high leverage issue in the firm to reduce investment and require a high expected return, and vice versa.

[Insert Table 6 here]

In sum, existing pricing factors have some explanatory power, either from behavioral finance

[37] See studies by Easley et al. (2010) and Kelly and Ljungqvist (2012) argue the relationship between liquidity risk and information asymmetry. I use the Fama–French five factors to estimate illiquidity beta for each stock.

perspectives, such as momentum and short-term reversal, or from fundamental interpretations such as value or investment factors. However, these baseline asset pricing factors' effects on the firm-specific news sentiment zero-cost portfolio are not economically significant and they only capture 1.5% to 7.5% of the variation in the daily zero-cost portfolio return.

This paper proposes a novel interpretation of the zero-cost news sentiment portfolio's risk-adjusted abnormal return; it can be seen as an information risk premium resulting from the biased tone in firm-specific news. The abnormal return from the news sentiment zero-cost portfolio offers empirical evidence to verify the theoretical study in section 2 for the argument of biased tone in the firm-specific news leading to a deviation in information risk. One could question whether the firm-specific news sentiment trading strategy can generate considerable profits. In fact, taking a moderate round-trip transaction cost, such as 5 basis points, the rough calculation for daily return (including the trading cost for the risk-adjusted abnormal daily return from the zero-cost portfolio) is about 1.17 basis points. Obviously, the profit will be lost by increasing the round-trip trading cost, since a daily-basis formation is too frequent in reality. Of course, the trading cost could be reduced through a weekly re-balance or tailoring of extreme sentiment stocks.[38] Nevertheless, the main purpose in this mimicking (zero-cost) portfolio factor analysis is to investigate the validity of the implications of news sentiment leading to a deviation in information risk, for which investors require high expected returns as compensation. The firm-specific news sentiment trading strategy leaves room for future study from a behavioral finance perspective.

In the Online Appendix, I conduct several robustness tests - for example: excluding data from earnings announcement days; sorting data into sub-samples based on financial characteristics; choosing an alternative asset pricing model (q-factor model by Hou et al. (2015)); and utilizing an innovative news pricing factor to control for a genuine or mis-valuation effect from firm-specific news sentiment. The empirical results are robust to all of the tests.

6 Conclusion

In this paper, I developed a theoretical model and empirically tested the predictions implied by the model to demonstrate that biased public information from news media gives rise to investors' biased acquisition of firm-specific information. First, the static information acquisition model derives several theoretical predictions, by introducing a channel via which clostless but biased public information is exogenously distributed to investors before they make investment decisions. Because investors naively do not adjust for the bias in public news information, their beliefs about

[38] As Table 5 shows, news sentiment can predict a positive cross-section stock cumulative return of up to 10 days. Alternatively, one could construct a trading strategy by tailoring for firm news sentiment, for example (as Ke et al. (2019) proposed) by adopting a strategy of buying the 50 stocks with the most positive sentiment and selling the 50 stocks with the most negative sentiment.

the systematic and firm-specific uncertainties included in the risky asset's payoff are biased. Thus, investors' acquisition of private information about the risky asset is subject to their biased beliefs.

Second, the empirical tests I conducted, where sentiment in the news is used as a proxy for biased public information in the model, yielded results that are consistent with my theoretical predictions. Investors' acquisition of firm-specific information is significantly inversely related to the tone (sentiment) in the news about the stock market or particular firms. In addition, firm-specific news sentiment in the model causes a deviation in firm-specific information acquisition from the rational expectations equilibrium. This causes the degree of information risk to deviate as well. Empirically, the Fama–Macbeth (1973) regression verifies the positive predictability of firm-specific news sentiment on expected returns, as the theoretical model predicts. Also, by constructing a daily zero-cost portfolio return factor for firm-specific news sentiment, the annualized risk premium is about 17% per year. This result is robust to the addition of additional traditional pricing factors and a novel news effect pricing factor as controls, and switching to alternative asset pricing model such as the q-factor model.

In sum, this study introduces a new understanding of the channel of irrationality in economic activity, specifically, information acquisition in investment. In most of the behavioral studies in finance and economics, researchers relax the assumption of rational economic agents and argue that psychological irrationality in humans plays an important role in economic study. This study does not oppose this classical theory. The key claim in this paper is based on the perspective of the behavioral studies, but challenges assumptions about how bias arises. It is difficult to claim that economic agents are rational all the time, as an advocate of behavioral economics would believe, but it is also difficult to accept that investors intend to make important decisions, particularly investment decisions, from an irrational or psychologically-biased standpoint. As emphasized by Tirole (2002), the enrichment derived by the incorporation of psychological factors in economics models should focus on parsimony and normative analysis rather than the impulsive framework of psychology. In this study, I keep the view aligned with behavioral finance to argue for the role of irrationality in conducting economic activities. Instead of stressing human behavioral irrationality, the trigger-biased information percolation proposed by Hirshleifer (2020) discussed in this paper conceptualizes irrationality within economic agents as social transmission bias through the distribution of news. In particular, irrationality forced by the biased information transmission through news has a significant impact on investors' decisions concerning further information acquisition. As the theoretical model demonstrates, investors' sub-optimal choices come down to thinking and decision-making that is affected by the transmission of biased information from sources upon which they may rely, such as the news media.

References

Abarbanell, J. S. and Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. The Journal of Finance, 47(3):1181–1207.

Ahn, M., Bonsall, S. B., and Van Buskirk, A. (2019). Do managers withhold bad news from credit rating agencies? Review of Accounting Studies, 24(3):972–1021.

Akhtar, S., Faff, R., Oliver, B., and Subrahmanyam, A. (2011). The power of bad: The negativity bias in australian consumer sentiment announcements on stock returns. Journal of Banking & Finance, 35(5):1239–1249.

Alti, A. and Tetlock, P. C. (2014). Biased beliefs, asset prices, and investment: A structural approach. The Journal of Finance, 69(1):325–361.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets, 5(1):31–56.

Amir, E., Levi, S., and Livne, T. (2018). Do firms underreport information on cyber-attacks? evidence from capital markets. Review of Accounting Studies, 23(3):1177–1206.

An, Z., Chen, C., Naiker, V., and Wang, J. (2020). Does media coverage deter firms from withholding bad news? evidence from stock price crash risk. Journal of Corporate Finance, page 101664.

Anderson, S. P. and McLaren, J. (2012). Media mergers and media bias with rational consumers. Journal of the European Economic Association, 10(4):831–859.

Andrei, D., Friedman, H. L., and Ozel, N. B. (2019). Economic uncertainty and investor attention. Available at SSRN 3128673.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. The Journal of finance, 59(3):1259–1294.

Arango-Kure, M., Garz, M., and Rott, A. (2014). Bad news sells: The demand for news magazines and the tone of their covers. Journal of Media Economics, 27(4):199–214.

Atiase, R. K. (1985). Predisclosure information, firm capitalization, and security price behavior around earnings announcements. Journal of accounting research, pages 21–36.

Back, K. (2010). Asset pricing and portfolio choice theory. Oxford University Press.

Baginski, S. P., Campbell, J. L., Hinson, L. A., and Koo, D. S. (2018). Do career concerns affect the delay of bad news disclosure? The Accounting Review, 93(2):61–95.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. The journal of Finance, 61(4):1645–1680.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. The quarterly journal of economics, 131(4):1593–1636.

Bali, T. G., Bodnaruk, A., Scherbina, A., and Tang, Y. (2018). Unusual news flow and the cross section of stock returns. Management Science, 64(9):4137–4155.

Banz, R. W. (1981). The relationship between return and market value of common stocks. Journal of financial economics, 9(1):3–18.

Bao, D., Kim, Y., Mian, G. M., and Su, L. (2019). Do managers disclose or withhold bad news? evidence from short interest. The Accounting Review, 94(3):1–26.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. Journal of financial economics, 49(3):307–343.

Baron, D. P. (2006). Persistent media bias. Journal of Public Economics, 90(1-2):1–36.

Barry, C. B. and Brown, S. J. (1984). Differential information and the small firm effect. Journal of financial economics, 13(2):283–294.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. Review of general psychology, 5(4):323–370.

Benamar, H., Foucault, T., and Vega, C. (2019). Demand for information, uncertainty, and the response of us treasury securities to.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. Journal of consumer psychology, 24(4):586–607.

Berger, J. and Milkman, K. L. (2012). What makes online content viral? Journal of marketing research, 49(2):192–205.

Bernhardt, D., Krasa, S., and Polborn, M. (2008). Political polarization and the electoral effects of media bias. Journal of Public Economics, 92(5-6):1092–1104.

Bhattacharya, U., Galpin, N., Ray, R., and Yu, X. (2009). The role of the media in the internet ipo bubble. Journal of Financial and Quantitative Analysis, pages 657–682.

Bonsall IV, S. B., Green, J., and Muller III, K. A. (2020). Market uncertainty and the importance of media coverage at earnings announcements. Journal of Accounting and Economics, 69(1):101264.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2019). Diagnostic expectations and stock returns. The Journal of Finance, 74(6):2839–2874.

Bouchaud, J.-P., Krueger, P., Landier, A., and Thesmar, D. (2019). Sticky expectations and the profitability anomaly. The Journal of Finance, 74(2):639–674.

Busse, J. A. and Green, T. C. (2002). Market efficiency in real time. Journal of Financial Economics, 65(3):415–437.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. The Quarterly Journal of Economics, 108(4):905–939.

Carhart, M. M. (1997). On persistence in mutual fund performance. The Journal of finance, 52(1):57–82.

Chan, L. K., Jegadeesh, N., and Lakonishok, J. (1996). Momentum strategies. The Journal of Finance, 51(5):1681–1713.

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

Chen, W., Hribar, P., and Melessa, S. (2018). Incorrect inferences when using residuals as dependent variables. Journal of Accounting Research, 56(3):751–796.

Chernenko, S., Hanson, S. G., and Sunderam, A. (2016). Who neglects risk? investor experience and the credit boom. Journal of Financial Economics, 122(2):248–269.

Choi, J. J. and Mertens, T. M. (2019). Extrapolative expectations and the equity premium. Available at SSRN 3462056.

Coibion, O. and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? Journal of Political Economy, 120(1):116–159.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. American Economic Review, 105(8):2644–78.

Da, R. and Xiu, D. (2019). When moving-average models meet high-frequency data: Uniform inference on volatility. Chicago Booth Research Paper, (17-27).

Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all fears investor sentiment and asset prices. *The Review of Financial Studies*, 28(1):1–32.

De Bondt, W. F. and Thaler, R. H. (1990). Do security analysts overreact? *The American Economic Review*, pages 52–57.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738.

Dougal, C., Engelberg, J., Garcia, D., and Parsons, C. A. (2012). Journalists and the stock market. *The Review of Financial Studies*, 25(3):639–679.

Duarte, J., Hu, E., and Young, L. (2020). A comparison of some structural models of private information arrival. *Journal of Financial Economics*, 135(3):795–815.

Dyck, A. and Zingales, L. (2003). The media and asset prices. Technical report, Working Paper, Harvard Business School.

Easley, D., Hvidkjaer, S., and O'hara, M. (2002). Is information risk a determinant of asset returns? *The journal of finance*, 57(5):2185–2221.

Easley, D., Hvidkjaer, S., and O'hara, M. (2010). Factoring information into returns. *Journal of Financial and Quantitative Analysis*, pages 293–309.

Easley, D., Kiefer, N. M., O'hara, M., and Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. *The Journal of Finance*, 51(4):1405–1436.

Easley, D. and O'hara, M. (2004). Information and the cost of capital. *The journal of finance*, 59(4):1553–1583.

Easterwood, J. C. and Nutt, S. R. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance*, 54(5):1777–1797.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1):67–97.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5):2023–2052.

Freeman, R. N. (1987). The association between accounting earnings and security returns for large and small firms. *Journal of accounting and economics*, 9(2):195–228.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300.

García, D., Sangiorgi, F., and Urošević, B. (2007). Overconfidence and market efficiency with heterogeneous agents. *Economic Theory*, 30(2):313–336.

Gentzkow, M., Shapiro, J. M., and Stone, D. F. (2015). Media bias in the marketplace: Theory. In *Handbook of media economics*, volume 1, pages 623–645. Elsevier.

Greenwood, R. and Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, 27(3):714–746.

Griffin, P. A. (1977). The time-series behavior of quarterly earnings: preliminary evidence. *Journal of Accounting Research*, pages 71–83.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408.

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

Heller, Y. and Winter, E. (2020). Biased-belief equilibrium. *American Economic Journal: Microeconomics*, 12(2):1–40.

Hendershott, T., Livdan, D., and Schürhoff, N. (2015). Are institutions informed about news? *Journal of Financial Economics*, 117(2):249–287.

Hillert, A., Jacobs, H., and Müller, S. (2014). Media makes momentum. *The Review of Financial Studies*, 27(12):3467–3501.

Hillert, A., Jacobs, H., and Müller, S. (2018). Journalist disagreement. *Journal of Financial Markets*, 41:57–76.

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*.

Hirshleifer, D., Li, J., and Yu, J. (2015). Asset pricing in production economies with extrapolative expectations. *Journal of Monetary Economics*, 76:87–106.

Hirshleifer, D. A., Myers, J. N., Myers, L. A., and Teoh, S. H. (2008). Do individual investors cause post-earnings announcement drift? direct evidence from personal trades. *The Accounting Review*, 83(6):1521–1550.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, 54(6):2143–2184.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, 28(3):791–837.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3):881–898.

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting returns with text data. Technical report, National Bureau of Economic Research.

Kelly, B. and Ljungqvist, A. (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, 25(5):1366–1413.

Ko, K. J. and Huang, Z. J. (2007). Arrogance can be a virtue: Overconfidence, information acquisition, and market efficiency. *Journal of Financial Economics*, 84(2):529–560.

Kothari, S. P., Shu, S., and Wysocki, P. D. (2009). Do managers withhold bad news? *Journal of Accounting research*, 47(1):241–276.

Le Mens, G. and Denrell, J. (2011). Rational learning and information sampling: On the “naivety” assumption in sampling explanations of judgment biases. *Psychological review*, 118(2):379.

Lee, C. M. and Swaminathan, B. (2000). Price momentum and trading volume. *The Journal of Finance*, 55(5):2017–2069.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1):1–28.

Livnat, J. and Mendenhall, R. R. (2006). Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of accounting research*, 44(1):177–205.

Llorente, G., Michaely, R., Saar, G., and Wang, J. (2002). Dynamic volume-return relation of individual stocks. *The Review of financial studies*, 15(4):1005–1047.

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.

Mullainathan, S. and Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4):1031–1053.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The journal of finance*, 53(6):1887–1934.

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance*, 58(4):1335–1354.

Pástor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3):642–685.

Peress, J. (2014). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *The Journal of Finance*, 69(5):2007–2043.

Peterson, R. L. (2016). *Trading on sentiment: The power of minds over markets*. John Wiley & Sons.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance*, 39(4):1127–1139.

Rozin, P. and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, 5(4):296–320.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Sletten, E. (2012). The effect of stock price on discretionary disclosure. *Review of accounting studies*, 17(1):96–133.

Tesser, A. and Rosen, S. (1975). The reluctance to transmit bad news. In *Advances in experimental social psychology*, volume 8, pages 193–232. Elsevier.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168.

Tetlock, P. C. (2010). Does public financial news resolve asymmetric information? *The Review of Financial Studies*, 23(9):3520–3557.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies*, 24(5):1481–1512.

Tetlock, P. C. (2014). Information transmission in finance. *Annual Review of Financial Economics*, 6(1):365–384.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467.

Tirole, J. (2002). Rational irrationality: Some economics of self-management. *European Economic Review*, 46(4-5):633–655.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Veldkamp, L. L. (2006). Media frenzies in markets for financial information. *American Economic Review*, 96(3):577–601.

Veldkamp, L. L. (2011). *Information choice in macroeconomics and finance*. Princeton University Press.

Weller, B. M. (2018). Does algorithmic trading reduce information acquisition? *The Review of Financial Studies*, 31(6):2184–2226.

Appendices

A Proof of Propositions and Corollaries

A.1 Theorems used to solve the model

Based on Bayes' rule of normal-normal updating (Back, 2010), X and Y are joint normally distributed. The expectation of X condition on Y can be projected:

$$
\begin{aligned}
E[X|Y] &= E[X] + \beta(Y - E[Y]) \\
\beta &= \frac{Cov(X;Y)}{Var(Y)} \\
Var(X|Y) &= Var(X) - \frac{[Cov(X;Y)]^2}{Var(Y)}
\end{aligned}
\tag{T1}
$$

Following Veldkamp (2011), the Wishart moment generating function of the exponential of a multi-variate quadratic form of a normal variable follows:

$$
\begin{aligned}
&z \sim N(0;\Sigma) \\
&E[e^{z'Fz+G'z+H}] = |I - 2\Sigma F|^{-1/2} \exp[\frac{1}{2}G'(I - 2\Sigma F)^{-1}\Sigma G + H]
\end{aligned}
\tag{T2}
$$

A.2 Proof of Proposition 1

Investors who pay a cost c for acquiring firm-specific information, therefore, informed investors' information set is:

$$F_I = \{\bar{D}; M_1; \varepsilon_1; \hat{p}_1\}$$

Based on T1, informed investors' expected payoff and variance of the risky assets are:

$$
\begin{aligned}
E^I_{b;1}[\mu_1] &= \frac{\sigma^2_{b;m}}{\sigma^2_{b;m} + \sigma^2_{\eta}} M_1 \\
Var^I_{b;1}[\mu_1] &= \frac{\sigma^2_{b;m}\sigma^2_{\eta}}{\sigma^2_{b;m} + \sigma^2_{\eta}} \\
E^I_{b;1}[D_2] &= \bar{D} + \frac{\sigma^2_{b;m}}{\sigma^2_{b;m} + \sigma^2_{\eta}} M_1 + \varepsilon_1 \\
Var^I_{b;1}[D_2] &= Var^I_{b;1}[\mu_1] = \frac{\sigma^2_{b;m}\sigma^2_{\eta}}{\sigma^2_{b;m} + \sigma^2_{\eta}}
\end{aligned}
\tag{A.21}
$$

Uninformed investors don't observe the firm-specific information ε_1, but they can partially learn about ε_1 from the informative signal through price revealing. Therefore uninformed investors' U information set is :

$$\mathcal{F}_U = \{\bar{D}; M_1; \hat{p}_1\}$$

Based on T1, uninformed investors learn about ε_1 based on $\hat{p}_1$ is :

$$\begin{aligned} E^U_{b,1}[\varepsilon_1 | \hat{p}_1] &= \frac{\sigma^2_{b,\varepsilon}}{\sigma^2_{b,\varepsilon} + \frac{K^2}{G^2}\sigma^2_\xi}\hat{p}_1 \\ Var^U_{b,1}[\varepsilon_1 | \hat{p}_1] &= \frac{K^2\sigma^2_{b,\varepsilon}\sigma^2_\xi}{G^2\sigma^2_{b,\varepsilon} + K^2\sigma^2_\xi} \end{aligned} \tag{A.22}$$

Therefore, for uninformed investors, the expected payoff and variance of the risky asset are:

$$\begin{aligned} E^U_{b,1}[D_2] &= \bar{D} + \frac{\sigma^2_{b,m}}{\sigma^2_{b,m} + \sigma^2_\eta}M_1 + \frac{\sigma^2_{b,\varepsilon}}{\sigma^2_{b,\varepsilon} + \frac{K^2}{G^2}\sigma^2_\xi}\hat{p}_1 \\ Var^U_{b,1}[D_2] &= Var^I_{b,1}[D_2] + Var^U_{b,1}[\varepsilon_1] \\ &= \frac{\sigma^2_{b,m}\sigma^2_\eta}{\sigma^2_{b,m} + \sigma^2_\eta} + \frac{K^2\sigma^2_{b,\varepsilon}\sigma^2_\xi}{G^2\sigma^2_{b,\varepsilon} + K^2\sigma^2_\xi} \end{aligned} \tag{A.23}$$

As defined in Grossman and Stiglitz (1980) and stated in Andrei et al. (2019), I defined price informativeness as:

$$\begin{aligned} Corr(\hat{p}_1; \varepsilon_1) &= \rho = \frac{Cov(\hat{p}_1; \varepsilon_1)}{\sigma_{b,\varepsilon}\sigma_{\hat{p}_1}} = \frac{\sigma^2_{b,\varepsilon}}{\sigma_{b,\varepsilon}\sqrt{\sigma^2_{b,\varepsilon} + \frac{K^2\sigma^2_\xi}{G^2\sigma^2_{b,\varepsilon}}}} \\ \rho^2 &= \frac{\sigma^2_{b,\varepsilon}}{\sigma^2_{b,\varepsilon} + \frac{K^2\sigma^2_\xi}{G^2\sigma^2_{b,\varepsilon}}} \\ \text{Define informativeness: } \nu &= \frac{\rho^2}{1-\rho^2} = \frac{\lambda_1\sigma^2_{b,\varepsilon}}{\alpha Var^I_{b,1}[D_2]^2\sigma^2_\xi} \\ \text{Denote } \Phi &= \frac{\nu}{1+\nu} = \rho^2 \end{aligned} \tag{A.24}$$

Following Back (2010), the customary optimal portfolios for informed and uninformed investors with CARA utility are:

$$
\begin{aligned}
q_1^I &= \frac{E^I_{b;1}[D_2] - r_f P_1}{a\,Var^I_{b;1}[D_2]} \\
q_1^U &= \frac{E^U_{b;1}[D_2] - r_f P_1}{a\,Var^U_{b;1}[D_2]}
\end{aligned}
\tag{A.25}
$$

Therefore, A.21–25 yield equations (9) and (10) for informed and uniformed investors' optimal portfolios.

To find linear conjecture equilibrium price, the market clearing condition follows equation (6). Then, using terms A.22, A.23, and A.25 to replace terms in equation (6) yields:

$$
\begin{aligned}
\lambda_1 g f_I (E^I_{b;1}[D_2 - r_f P_1]) + (1 - \lambda_1) g f_U (E^U_{b;1}[D_2 - r_f P_1])) &= x_1 \\
g = \frac{1}{a};\; f_I = \frac{1}{Var^I_{b;1}[D_2]};\; f_U &= \frac{1}{Var^U_{b;1}[D_2]}
\end{aligned}
\tag{A.26}
$$

After taking tedious algebra, the unknown coefficients A, B, G, K, and H of the linear conjectured price P_1 in equation (7) can be easily solved and showed in equation (11) of Proposition 1.

A.3 Proof of Proposition 2

To find the fraction of investors who become informed about e_1 in equilibrium, I solve equation (12), the indifference condition proposed by Grossman and Stiglitz (1980) by applying T2 as:

$$
\begin{aligned}
F &= -\frac{1}{2}Var^I[D_2]^{-1} \\
G^0 &= (E^U_{b;1}[D_2] - r_f P_1)Var^I_{b;1}[D_2]^{-1} \\
H &= -\frac{1}{2}(E^U_{b;1}[D_2] - r_f P_1)^2 Var^I_{b;1}[D_2]^{-1} \\
S &= Var^U_{b;1}[e_1 | \hat{p}_1]
\end{aligned}
\tag{A.27}
$$

Applying A.27 yields:

$$
\begin{aligned}
E_b[U^I | P_1] = &- \left| I - 2Var^U_{b;1}[e_1 | \hat{p}_1](-\frac{1}{2})Var^I_{b;1}[D_2]^{-1} \right|^{-1/2} \\
&e^{\frac{1}{2}E^U_{b;1}[D_2 - r_f P_1]^2 Var^I_{b;1}[D_2]^{-2}(I + Var^U_{b;1}[e_1 | \hat{p}_1]Var^I_{b;1}[D_2]^{-1})^{-1}Var^U_{b;1}[e_1 | \hat{p}_1] - \frac{1}{2}E^U_{b;1}[D_2 - r_f P_1]^2 Var^I_{b;1}[D_2]^{-1}}
\end{aligned}
\tag{A.28}
$$

Solving A.28 yields:

$$E_b[U^I|P_1] = \left(\frac{Var^I_{b;1}[D_2]}{Var^U_{b;1}[e_1|\hat{p}_1]+Var^I_{b;1}[D_2]}\right)^{1/2} e^{\frac{1}{2}E^U_{b;1}[D_2 \; r_f P_1]^2 Var^I_{b;1}[D_2]^{\;1}\left[\frac{Var^I_{b;1}[D_2]}{Var^U_{b;1}[e_1|\hat{p}_1]+Var^I_{b;1}[D_2]}\right]}$$

$$= \left(\frac{Var^I_{b;1}[D_2]}{Var^U_{b;1}[e_1|\hat{p}_1]+Var^I_{b;1}[D_2]}\right)^{1/2} e^{\frac{1}{2}\frac{E^U[D_2 \; r_f P_1]^2}{Var^U_{b;1}[D_2]}}$$

$$E_b[U^U|P_1] = e^{\frac{1}{2}\frac{E^U[D_2 \; r_f P_1]^2}{Var^U_{b;1}[D_2]}}$$

$$\frac{E_b[U^I]}{E_b[U^U]} = e^{\alpha\gamma}\sqrt{\frac{Var^I_{b;1}[D_2]}{Var^U_{b;1}[e_1|\hat{p}_1]+Var^I_{b;1}[D_2]}} = e^{\alpha\gamma}\sqrt{\frac{Var^I_{b;1}[D_2]}{Var^U_{b;1}[D_2]}} \qquad (A.29)$$

Therefore, applying A.22-24, it is straightforward to find the benefit and cost function $P(\cdot)$.

A.4 Proof of Corollary 1

To solve equilibrium λ_1 as a function of uncertainties ($Var^I_{b;1}[D_2]$ and $\sigma^2_{b;e}$), I set the cost and benefit function $P(\cdot) = 0$. Hence, I directly solve the numerator of $P(\cdot)$ equals to 0 as :

$$F(\cdot) = \lambda_1^2\sigma^2_{b;e}\delta + \alpha^2 Var^I_{b;1}[D_2]^2\sigma^2_x\delta \quad \alpha^2 Var^I_{b;1}[D_2]\sigma^2_x\sigma^2_{b;e} = 0; \;\; \text{where } \delta = e^{2\alpha\gamma} \quad 1 \qquad (A.30)$$

By applying the implicit theorem in a region $F'(\lambda_1) \quad 0$, the $\frac{\partial\lambda_1}{\partial Var^I_{b;1}[D_2]}$ can be found as :

$$\frac{\partial\lambda_1}{\partial Var^I_{b;1}[D_2]} = \frac{\partial F}{\partial Var^I_{b;1}[D_2]} \quad \frac{\partial\lambda_1}{\partial F}$$

$$= \frac{\alpha^2\sigma^2_x\sigma^2_{b;e} \quad 2\alpha^2\sigma^2_x\delta Var^I_{b;1}[D_2]}{2\lambda_1\sigma^2_{b;e}\delta} \qquad (A.31)$$

As long as $Var^I_{b;1}[D_2] \quad \frac{\sigma^2_{b;e}}{2\delta}$, which is the threshold of $Var^I_{b;1}[D_2]$ then $F(\lambda_1)$ increases as the λ_1 increases to reach the theoretical maximum fraction of informed investors. In that, $\frac{\partial\lambda_1}{\partial Var^I_{b;1}[D_2]} > 0$. On the one hand, as $\frac{\partial Var^I_{b;1}[D_2]}{\partial\sigma_{b;m}} > 0$ is known, by applying chain rule, it is easy to show that $\frac{\partial\lambda_1}{\sigma_{b;m}} > 0$. On the other hand, the bias function $b(S_m;\sigma^2_m)$ is inversely related to the biased perception of $\sigma^2_{b;m}$ as showed in equation (2). In other words, $\frac{\partial\sigma_{b;m}}{\partial S_m} < 0$ is monotonic decreasing. Noted that, without loss of generality , the bias function $b(\cdot)$ is not assumed for particular function forms. By applying the chain rule, as a result, $\frac{\partial\lambda_1}{\partial S_m} < 0$.

In addition, the $\frac{\partial\lambda_1}{\partial\sigma^2_{b;e}}$ can be solved in the same steps:

$$\frac{\partial\lambda_1}{\partial\sigma_{b;e}} = \frac{\partial F}{\partial\sigma_{b;e}} \quad \frac{\partial\lambda_1}{\partial F}$$

$$= \frac{\alpha^2 Var^I_{b;1}[D_2]\sigma_x^2 - \lambda_1^2\delta}{2\lambda_1\sigma^2_{b;e}\delta}$$

$$\max(\lambda_1^2) = \frac{\alpha^2 Var^I_{b;1}[D_2]\sigma_x^2}{2\delta} \text{ when } Var^I_{b;1}[D_2] = \frac{\sigma^2_{b;e}}{2\delta} \tag{A.32}$$

$$\lambda_1^2 \quad \frac{\alpha^2 Var^I_{b;1}[D_2]\sigma_x^2}{2\delta}$$

$$\frac{\partial\lambda_1}{\partial\sigma_{b;e}} = \frac{\alpha^2 Var^I_{b;1}[D_2]\sigma_x^2 - \frac{1}{2}\alpha^2 Var^I_{b;1}[D_2]\sigma_x^2}{2\lambda_1\sigma^2_{b;e}\delta}$$

$$\text{this yields } \frac{\partial\lambda_1}{\partial\sigma_{b;e}} > 0 \text{ strictly.}$$

As the bias function $b(S_e;\sigma_e^2)$ in equation (2) indicates a monotonic decreasing relationship between biased perception of rm-speci c uncertainty and rm-speci c news sentiment as the proxy of biased pubic information received by investors, therefore, $\frac{\partial\sigma_{b;e}}{\partial S_e} < 0$ is implied by equation (2), by applying the chain rule with A.32, it is straightforward to show that $\frac{\partial\lambda_1}{\partial S_e} < 0$.

A.5 Proof of Proposition 3

Following O'Hara (2003), I assume the risky asset random supply $x_1 \sim N(\bar{x};\sigma_x^2)$. Therefore, this non-zero expected random supply $\bar{x}$ implies a risky premium. Based on the market clearing condition, the expected return of the risky asset is :

$$\lambda\gamma\phi_I E^I[D_2] + (1-\lambda_1)\gamma\phi_U E^U[D_2] - x_1 = (\lambda\gamma\phi_I + (1-\lambda_1)\gamma\phi_U)P_1 r_f$$

Expected Return:

$$E[R_2] = \frac{\lambda_1\gamma\phi_I E^I[D_2] + (1-\lambda_1)\gamma\phi_U E^U[D_2]}{r_f(\lambda_1\gamma\phi_I + (1-\lambda_1)\gamma\phi_U)} - P_1 \tag{A.33}$$

$$= \frac{E[x_1]}{r_f(\lambda_1\gamma\phi_I + (1-\lambda_1)\gamma\phi_U)}$$

$$= \frac{\alpha\bar{x}}{r_f(\lambda_1\phi_I + (1-\lambda_1)\phi_U)}$$

First, the expected return is a function of λ_1 the fraction of investors who are informed about e_U. The $\frac{\partial E[R_2]}{\partial \lambda_1}$ can be found as :

$$\frac{\partial E[R_2]}{\partial \lambda_1} = \frac{a\bar{x}[r_f \phi_I - r_f \phi_U + (1-\lambda_1)\frac{\partial \phi_U}{\partial \lambda_1}]}{(r_f \lambda_1 \phi_I + r_f(1-\lambda_1)\phi_U)^2} \tag{A.34}$$

Clearly, $Var^U_{b;1}[D_2]$ the uninformed investors variance of the risky asset's payoff decreases as λ_1 increases because the price informativeness increases. Therefore, it is easy to show that $\frac{\partial \phi_U}{\partial \lambda_1} > 0$, where $\phi_U = \frac{1}{Var^U_{b;1}[D_2]}$, $\lambda_1 \leq 1$ and $\phi_I \geq \phi_U$. As a result, $\frac{\partial E[R_2]}{\partial \lambda_1} < 0$ in A.34. Because $\frac{\partial \lambda_1}{\partial S_e} < 0$ argued in Appendix A.4, applying the chain rule yields $\frac{\partial E[R_2]}{\partial S_e} > 0$.

A.6 Proof of Corollary 2 and 3

The equilibrium fraction of informed investors in rational expectations is λ^e_1 and expected return $E[R^e_2]$ reconciles to O'Hara (2003) study. As the model in this paper indicates, firm-specific news sentiment S_e deviates λ^e_1 to a biased belief equilibrium $\lambda^{b;e}_1$. On the one hand, as S_e increases and $\frac{\partial \lambda_1}{\partial S_e} < 0$, therefore $\lambda^{b;e}_1 < \lambda^e_1$. As $\frac{\partial E[R_2]}{\partial \lambda_1} < 0$ proved in A.34, $\lambda^{b;e}_1 < \lambda^e_1 \Rightarrow E_b[R_2] > E_e[R_2]$ which completes Corollary 2 proof. The Corollary 3 proof can be easily completed by the other way around.

B Detailed Information of Variables Used in Regressions

B.1 Variable Definition

Buzz This measure is the sum of all references from the news about either the stock market or particular firms that are included in one of the TRMI indexes over 24 hours.

Sentiment Overall positive references net of negative references in news about either the stock market or particular firms over 24 hours.

EmotionVsFact The sum of the absolute value of all emotions and opinions (both positive, negative, surprise and uncertainty) minus the sum of the absolute value of all facts (topics and other subjects/themes/nouns) divided by the sum of all references in the news.

VIX: Daily closing value of VIX. Source: Wharton Research Data Services-CBOE Indexes.

EPU: Daily news-based Economic Policy Uncertainty Index. Source: BBM.

SP500 Realized volatility is downloaded from Risk Lab by Da and Xiu (2019).

ME: Market value of equity in fiscal year closing price times total share of equity. Source: Compustat.

Size: Natural log of market value of equity. Source: Compustat.

BM: Book to Market Ratio as defined in Fama and French (1992). Source: Compustat.

Illiquity: Monthly illiquidity measure as per Amihud (2002). Source: CRSP.

OP: Operating Profitability, as defined in Fama and French (2015). Source: Compustat.

INV: Investment measure is defined as in Fama and French (2015) study. Source: Compustat.

RV: Return volatility is measured as standard deviation of daily return at each month. Source: CRSP.

MOM: Momentum Return Measure is defined as the cumulative return from $t-11$ to the month $t-1$ before the last month. Source: CRSP.

ST: Return from the last month to capture short-term reversal effect. Source: CRSP.

AbRet: Daily holding period return minus the value-weighted market return. Source: CRSP.

$AbRet_{t-5,t-1}$: Five days cumulative abnormal return from $t-5$ to $t-1$. Source: CRSP.

AbTurn: Natural log turnover at day t net of the average turnover in the last five days. Source: CRSP.

SUE: Unexpected earnings is calculated based on Compustat data. The calculation follows Livnat and Mendenhall (2006). Source: Compustat.

SUE^{IBES}: Unexpected earnings is calculated based on I/B/E/S data. The calculation follows Livnat and Mendenhall (2006) study. Source: Institutional Brokers Estimate System (I/B/E/S).

ForecastDispersion: The standard deviation of analysts' earnings forecasts in the most recent month before quarterly earnings announcement and scaled by the stock price. Source: Institutional Brokers Estimate System (I/B/E/S).

ForecastRevision: The median analysts' 3-month earnings forecast revision is based on Chan et al. (1996). Source: Institutional Brokers Estimate System (I/B/E/S).

Idiosyncratic Volatility (IDIOVOL): The residual standard error from Fama and French (2015) five factor plus momentum factor pricing model on a daily rolling basis. I require each company to have at least 60 observations to run the time-series regression. Sources: CRSP and Kenneth R. French Data Library.

Abs(FFCAR): Absolute value of cumulative abnormal return (CAR) is calculated from Fama and French (2015) five factors plus momentum factor pricing model. The factor betas used to calculate CAR are estimated 90 days before a quarterly earnings announcement. Sources: CRSP and Kenneth R. French Data Library.

Institutional Ownership (ITOW): This is the institutional ownership percentage from Thomson Reuters Institutional (13f) Holdings data file.

$\overline{IVOL}$: Moving average stock idiosyncratic volatility is calculated based on the window between day $t-24$ and $t-4$. Sources: CRSP and Kenneth R. French Data Library.

Price: Average daily closing price from day $t-42$ to $t-21$ before a quarterly earnings announcement. Source: CRSP.

NUMEST: Number of analyst's earnings forecasts in the most recent month before a quarterly earnings announcement. Source: Institutional Brokers Estimate System (I/B/E/S).

Turn: Turnover is total number of shares traded over a period divided by total outstanding shares. Source: CRSP.

Figure 1: Stock Market News Sentiment Impact on Information Acquisition as a function of $Var^I_{b,1}[D_2]$

This gure plots the equilibrium proportion of investors who want to acquire rm-speci c information, l_1, as a function of perceived systematic uncertainty $Var^I_{b,1}[D_2]$ by investors. l_1^e and $l_1^{b,e}$ denote the equilibrium level of the proportion of informed investors under the rational expectations and biased beliefs, respectively. I calibrate cost of information $c = 0{:}002$; the variance of supply $s_x^2 = 0{:}2$; the variance of the public signal noise s_h^2 is randomly drawn from a uniform distribution; and investors' coef cient of risk aversion $a = 3$. I set the rational perception of rm-speci c uncertainty $s_e^2 = 0{:}0623$ as the median value of the sum of squared residuals from AR(1) of firms' $EPS_{i,t}$ in the sample period. The rational perception of systematic component uncertainty s_m^2 is randomly drawn from the distribution of S&P 500 realized variance in the sample period. Without loss of generality, I assume the bias function $b(S_m, s_m^2)$ in equation (2) is linear as $s_{b,m}^2 = (1 - S_m)s_m^2$ to generate the biased perception of systematic component uncertainties. The blue curve is the equilibrium fraction of informed investors under the rational expectations without the biased impact from market news sentiment ($S_m = 0$) on s_m^2. The red curve is the equilibrium fraction of informed investors, as $s_{b,m}^2$ " is upward (positively) biased when the tone in the news about the stock market is decreased, or made more pessimistic, by one standard deviation S_m# from $S_m = 0$ under the rational expectations. The green curve is the equilibrium fraction of informed investors, as $s_{b,m}^2$# is downward (negatively) biased when the tone in the news about the stock market is increased, or made more optimistic, by one standard deviation S_m" from $S_m = 0$ under the rational expectations. The standard deviation of stock market news sentiment s_{S_m} is from the TRMI stock market news sentiment index in the U.S. and the value is 0.183.

Figure 2: Firm-Specific News Sentiment Impact on Information Acquisition as a function of $\sigma^2_{b,e}$

This figure plots the equilibrium fraction of investors who want to acquire firm-specific information, λ_1, as a function of perceived firm-specific uncertainty, $\sigma^2_{b;e}$, by investors. λ_1^e and $\lambda_1^{b;e}$ denote the equilibrium level of the proportion of informed investors under the rational expectations and biased beliefs, respectively. I calibrate cost of information about ϵ_1, $c = 0.002$; the variance of supply, $\sigma_x^2 = 0.2$; the variance of the public signal noise, σ_η^2, is randomly drawn from a uniform distribution; and investors' coefficient of risk aversion $a = 3$. The rational perception of firm-specific uncertainty, σ_e^2, is drawn randomly from the distribution of the sum of squared residuals from AR(1) of firms' $EPS_{i,t}$. I set rational perception of systematic component uncertainty, $\sigma_m^2 = 0.031$ as the mean S&P 500 realized variance in the sample period. Without loss of generality, I assume the bias function $b(S_e, \sigma_e^2)$ in equation (2) is linear as $\sigma^2_{b;e} = (1 - S_e)\sigma_e^2$ to generate the biased perception of firm-specific uncertainties. The blue curve is the equilibrium fraction of informed investors under the rational expectations without the biased impact of firm-specific news sentiment ($S_e = 0$) or (σ_e^2). The red curve is the equilibrium fraction of informed investors, as $\sigma^2_{b;e}$ " is upward (positively) biased when the tone in the news about particular firms' performance is decreased, or made more pessimistic, by one standard deviation S_e # from $S_e = 0$ under the rational expectations. The green curve is the equilibrium fraction of informed investors, as $\sigma^2_{b;e}$ # is downward (negatively) biased when the tone in the news about firm-specific information is increased, or made more optimistic, by one standard deviation S_e " from $S_e = 0$ under the rational expectations. The standard deviation of firm-specific news sentiment S_e is from the TRMI news sentiment indices of U.S.-listed firms and the value is 0.394.

Table 1: Summary Statistics and Correlations

This table presents summary statistics and correlations for sample variables. Panel A reports descriptive statistics used in empirical studies that test the impact of news sentiment on information acquisition and cross-sectional stocks returns. Panel B reports Pearson correlations (significant at the 1% level) between stock market news sentiment and proxies of economic uncertainty. Panel C reports Pearson correlations between firm-specific news sentiment and other financial fundamental variables. Detailed definitions of all variables are available in Appendix B.1.

Panel A Summary Statistics

	Mean	Std	Min	25%	50%	75%	Max	Count
$Buzz_{m,t}$	5408.554	5948.982	0.000	1230.225	3490.450	8008.125	123305.100	7670
$Sentiment_{m,t}$	-0.053	0.183	-0.870	-0.192	-0.049	0.085	0.714	7668
$Sentiment_{i,t}$	0.074	0.394	-1.000	-0.172	0.049	0.328	1.000	3458582
$Buzz_{i,t}$	223.457	1145.504	0.100	12.000	34.000	114.600	183978.300	3458582
VIX	20.208	8.500	9.140	13.885	18.540	24.035	80.860	5283
EPU	100.487	68.106	3.320	53.850	83.245	128.820	719.070	7670
$ME_{i,t}$	16701.942	43494.512	1.968	1038.649	3302.520	12375.910	867506.995	2867485
$BM_{i,t}$	0.553	2.690	0.000	0.244	0.430	0.708	359.622	2867485
$Illiquidity_{i,t}$	0.073	2.632	0.000	0.000	0.001	0.003	813.735	3451226
$OP_{i,t}$	0.457	14.556	-331.333	0.147	0.237	0.360	9423.750	2867105
$INV_{i,t}$	0.152	0.601	-0.933	-0.003	0.062	0.164	55.264	2797908
$RV_{i,t}$	0.026	0.016	0.001	0.016	0.021	0.030	1.019	3451228
$MOM_{i,t}$	0.153	0.624	-1.000	-0.109	0.100	0.318	98.571	3385220
$ST_{i,t}$	0.012	0.139	-1.000	-0.049	0.010	0.066	13.495	3451138
$AbRet_{i,t}$	0.001	0.030	-1.012	-0.009	0.000	0.009	6.979	2867485
$AbRet_{i,t-5,t-1}$	0.002	0.061	-1.077	-0.021	0.000	0.022	13.630	2867103
$AbTurn_{i,t}$	10.593	38.642	-4.304	2.905	5.707	11.337	25084.092	2867092

Panel B Systematic Variable Correlations

	Stock Market Sentiment	VIX	EPU
Stock Market Buzz	-0.154	0.012	0.079
Stock Market Sentiment		-0.319	-0.096
VIX			0.406

Panel C Idiosyncratic Variables Correlations

	$Buzz_{i,t}$	$ME_{i,t}$	$BM_{i,t}$	$Illiquidity_{i,t}$	$OP_{i,t}$	$INV_{i,t}$	$RV_{i,t}$	$MOM_{i,t}$	$ST_{i,t}$	$AbRet_{i,t}$	$AbRet_{i,t-5,t-1}$	$AbTurn_{i,t}$
$Sentiment_{i,t}$	-0.028	-0.048	-0.009	0.003	-0.002	-0.007	-0.046	0.045	0.047	0.079	0.070	-0.022
$Buzz_{i,t}$		0.427	-0.008	-0.004	0.002	0.004	-0.041	-0.002	-0.005	-0.001	-0.006	0.031
$ME_{i,t}$			-0.022	-0.007	0.011	0.001	-0.170	-0.039	-0.021	-0.006	-0.009	-0.045
$BM_{i,t}$				0.004	-0.004	-0.013	0.035	0.010	0.001	0.001	0.000	0.002
$Illiquidity_{i,t}$					0.000	-0.003	0.051	0.018	0.007	0.004	0.006	0.005
$OP_{i,t}$						-0.004	-0.019	-0.004	-0.002	-0.002	0.000	-0.004
$INV_{i,t}$							0.108	-0.005	-0.015	-0.003	-0.004	0.029
$RV_{i,t}$								0.008	0.049	0.022	0.045	0.156
$MOM_{i,t}$									0.009	0.001	0.000	0.034
$ST_{i,t}$										0.010	0.098	0.008
$AbRet_{i,t}$											-0.008	0.142
$AbRet_{i,t-5,t-1}$												0.070

Table 2: Firm-Specific News Sentiment and Firm-Specific Uncertainty

This table reports the results of regressions of proxies for firm-specific uncertainties on firm-specific news sentiment. Columns (1)–(3) are based on a quarterly data fixed effect regression model from equation (23)(a): $\hat{s}^2_{e,i,t} = b_0 + b_1 Sentiment_{i,[t-30,t-1]} + Xd + e_{i,t}$ and column (4) conducts daily cross-sectional Fama–Macbeth (1973) regressions from equation (23)(b): $IDIO^{shock}_{i,t} = b_0 + b_1 Sentiment_{i,t-1} + Xd + e_{i,t}$. For regressions in column (1)–(3), I calculate $Sentiment_{i,[t-30,t-1]}$ as daily $Buzz_{i,t}$-weighted average in the last month before quarterly earnings announcements. Control variables include: lagged one period of dependent variable, forecast revision, forecast dispersion, size, book-to-market, turnover, return volatility, idiosyncratic volatility, absolute value of last month return, absolute value of cumulative abnormal returns, and institutional ownership for quarterly data regressions in columns (1)–(3). In addition, for the regression in column (1), an additional control $EPS_{i,t-1}$ from the first step regression to estimate $\hat{s}^2_e$ is also included. Control variables in in the column (4) regression include size, book to market, turnover, firm news $Buzz_{i,t-1}$, moving average of idiosyncratic volatility $\overline{IDIO}_{i,t-1}$ from the first step regression to estimate idiosyncratic volatility shock, operating profitability, firm investment, momentum return, return volatility and short term reversal return. Detailed definitions of all variables are available in Appendix B.1. Standard errors are clustered by both firm- and time-fixed effects in columns (1)–(3). Newey-West standard errors in column (4) are robust to heteroskedasticity and twelve days of autocorrelation. ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively.

	(1)	(2)	(3)	(4)
Dependent Variable	AR(1) $\hat{s}^2_e$	Abs($SUE_{i,t}$)	Abs($SUE^{IBES}_{i,t}$)	$IDIO^{shock}_{i,t}$
$Sentiment_{i,[t-30,t-1]\to t-1}$	0.013	0.006	0.0004	0.0003
	(0.004)	(0.002)	(0.000)	(0.0001)
LagDep	0.926	0.379	0.317	0.967
	(0.006)	(0.018)	(0.012)	(0.0005)
$ForecastRevision_{i,t-1}$	0.027	0.024	0.001	
	(0.014)	(0.032)	(0.002)	
$ForecastDispersion_{i,t-1}$	0.046	0.463	0.017	
	(0.033)	(0.064)	(0.005)	
$LogME_{i,t-1}$	0.041	0.004	0.0004	0.0003
	(0.003)	(0.001)	(0.000)	(0.0002)
$LogBM_{i,t}$	0.018	0.010	0.001	0.0003
	(0.003)	(0.001)	(0.000)	(0.0004)
$LogTurn_{i,t-1}$	0.006	0.010	0.002	
	(0.003)	(0.001)	(0.000)	
$ReturnVolatility_{i,t-31}$		0.201	0.018	
		(0.019)	(0.002)	
Idiosyncratic $Volatility_{i,t-31}$	11.897	4.446	0.356	
	(2.415)	(1.667)	(0.123)	
Abs($Return_{i,t-31}$)	0.188	0.010	0.002	
	(0.041)	(0.013)	(0.001)	
Abs($FFCAR_{i,t-30,t-3}$)	0.057	0.006	0.002	
	(0.027)	(0.010)	(0.001)	
Abs($FFCAR_{i,t-2}$)	2.805	0.685	0.056	
	(2.694)	(0.534)	(0.097)	
Institutional $Ownership_{i,t-1}$	0.00021	-0.000	0.000	
	(0.000)	(0.000)	(0.000)	
$EPS_{i,t-1}$	0.0462			
	(0.0035)			
$LogBuzz_{i,t-1}$				0.00001
				(0.00002)
$\overline{IVOL}_{i,t-1}$				0.001
				(0.0001)
$OP_{i,t-1}$				0.0004
				(0.0001)
$INV_{i,t-1}$				0.0002
				(0.0001)
$MOM_{i,t-1}$				0.0004
				(0.0001)
$RV_{i,t}$				0.03
				(0.0115)
$ST_{i,t-1}$				0.0023
				(0.0003)
FE Firm	Yes	Yes	Yes	
FE Year-Quarter	Yes	Yes	Yes	
Fama-Macbeth				Yes
Constant				0.0035
				(0.0004)
Observations	61,393	89,973	89,973	2,847,177
R-squared	0.925	0.234	0.155	0.939
Number of Firms	2,589	3,042	3,042	3,592

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: News Sentiment Impact on Information Acquisition

This table presents the results of regressions of the price jump ratio as the proxy for firm-specific information acquisition on stock market news sentiment during the firm earnings announcement window. Columns (1)–(3) are based on the fixed-effect regression from equation (26): $jump_{i,t} = b_0 + b_{j,1} Sentiment_{j,[t-21,t-1]} + Xd + e_{i,t}$, where $j \in \{m, i\}$ and $jump_{i,t}$ is estimated as $CAR_{i,t}^{T-1,T+2} / CAR_{i,t}^{T-21,T+2}$ and $CAR_{i,t}^{T-a,T-b}$, the cumulative abnormal return is calculated from Fama–French five factor plus momentum factor model. The news sentiment variable: $Sentiment_{j,[t-21,t-1]}$ and $Buzz_{j,[t-21,t-1]}$ are calculated in the same way as the daily $Buzz_{j,t}$-weighted average in the study window. Control variables include $Buzz_{j,[t-21,t-1]}$ as the proxy of intensity of stock market news coverage, economic uncertainty proxies (VIX and EPU) and the numbers of analyst coverage is calculated as 21 days until one day before announcement. Size, Turn, Price, Return Volatility and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. Detailed definition of all variables are available in Appendix B.1. Standard errors are clustered by both firm- and time-fixed effect in column (1)-(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively. The different number of firms in firm-specific news sentiment regression is subject to availability of firm-level news data.

	Panel A Stock Market News Sentiment			Panel B Firm-Specific News Sentiment		
	(1)	(2)	(3)	(1)	(2)	(3)
Dependent Variable	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$	$Jump_{i,t}$
$Sentiment_{i,t-21,t-1}$				0.057	0.050	0.051
				(0.015)	(0.020)	(0.020)
$Sentiment_{m,t-21,t-1}$	0.089	0.091	0.116		0.120	0.124
	(0.010)	(0.011)	(0.011)		(0.063)	(0.063)
$Buzz_{m,t-21,t-1}$		0.027	0.029			
		(0.002)	(0.002)			
$Buzz_{i,t-21,t-1}$					-0.013	-0.012
					(0.005)	(0.005)
$VIX_{t-21,t-1}$		-0.002			-0.001	
		(0.000)			(0.001)	
$EPU_{t-21,t-1}$			-0.0002			-0.0001
			(0.000)			(0.0001)
$Size_{i,t-42,t-21}$		0.006	0.007		0.040	0.041
		(0.002)	(0.002)		(0.016)	(0.015)
$Turn_{i,t-42,t-21}$		0.002	0.003		0.018	0.019
		(0.001)	(0.001)		(0.012)	(0.012)
$Price_{i,t-42,t-21}$		-0.011	-0.010		-0.032	-0.034
		(0.003)	(0.003)		(0.016)	(0.016)
$RV_{i,t-42,t-21}$		-0.021	-0.028		-0.042	-0.047
		(0.002)	(0.002)		(0.017)	(0.015)
$NUMEST_{i,t-21,t-1}$		0.001	0.001		0.000	0.000
		(0.000)	(0.000)		(0.000)	(0.000)
$ITOW_{i,t-42,t-21}$		0.056	0.053		0.050	0.048
		(0.007)	(0.007)		(0.036)	(0.036)
FE Month	Yes	Yes	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes	Yes	Yes
Observations	93,198	91,873	91,873	3,550	3,521	3,521
R-squared		0.021	0.020		0.033	0.033
Number of Firms	10,329	10,241	10,241	1,891	1,880	1,880

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Firm-Specific News Sentiment Impact on Probability of Informed Trading

This table presents the results of regressions of Generalised Probability of Informed Trading (GPIN) as a proxy for information risk for all stocks listed on the NYSE. Columns (1)–(3) are based on fixed-effect regression from equation (27): $GPIN_{i,t} = b_0 + b_1 Sentiment_{i,t-1} + Xd + e_{i,t}$. The GPIN is estimated yearly and the regression starts from 1999 to 2018. News sentiment from either firm-specific news or stock market news is the Buzzweighted average within a year. Control Variables include $Buzz_{j,t-1}$ where $j \in \{m, i\}$ proxies coverage about firm-specific and stock market news, Size, Book to Market, Trading Volume, Idiosyncratic Volatility and Institutional Ownership. All independent variables are lagged for one year. Detailed definitions of all variables are available in Appendix B.1. Standard errors are clustered by both firm- and time-fixed effect in columns (1)–(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively.

	(1)	(2)	(3)
Dependent Variable	$GPIN_{i,t}$	$GPIN_{i,t}$	$GPIN_{i,t}$
$Sentiment_{i,t-1}$	0.017	0.014	0.014
	(0.005)	(0.005)	(0.005)
$Sentiment_{m,t-1}$			0.118
			(0.087)
$Buzz_{m,t-1}$			-0.022
			(0.014)
$Buzz_{i,t-1}$		-0.002	-0.002
		(0.001)	(0.001)
$ME_{i,t-1}$		-0.003	-0.003
		(0.002)	(0.002)
$BM_{i,t-1}$		0.001	0.001
		(0.002)	(0.002)
$Turn_{i,t-1}$		-0.0019	-0.0018
		(0.002)	(0.002)
$IDIOVOL_{i,t-1}$		0.012	0.012
		(0.009)	(0.009)
$ITOW_{i,t-1}$		0.015	0.015
		(0.006)	(0.006)
FE Year	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes
Observations	15,571	13,551	13,551
R-squared	0.150	0.148	0.148
Number of Firms	1,434	1,355	1,355

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Cross-Sectional Return Predictability from Firm-Specific News Sentiment

This table presents the results from daily cross-sectional Fama–MacBeth (1973) regressions of next-day firm-specific news sentiment $t+1$ return and cumulative returns from $t+2$ to $t+5$ or $t+10$. Variables measured by news content and all other control variables are known by day t. Columns (1)–(3) report the time-series average of the coefficients based on the model in equation (23): $DepVar_{i,t+1} = b_0 + b_1 Sentiment_{i,t} + dX + e_{i,t}$ for each trading day t, where $DepVar_{i,t+1}$ is $R^e_{i,t+1}$, $R^e_{i,t+2,t+5}$, and $R^e_{i,t+2,t+10}$, respectively. The variable $Sentiment_{i,t}$ is firm-specific news sentiment as a proxy for biased information related to the firm-specific component. The news-related interacted variables including $EmotionVsFact_{i,t} \times Sentiment_{i,t}$, $EmotionVsFact_{i,t} \times AbRet_{i,t}$, and $Buzz_{i,t} \times AbRet_{i,t}$ control for potential effects of genuine information or biased valuation regarding firm fundamentals from $Sentiment_{i,t}$. Additionally, abnormal return $AbRet_{i,t}$ at day t and its related interactions such as $AbRet_{i,t} \times Size_{i,t}$ and $AbRet_{i,t} \times AbTurn_{i,t}$ measure return reversal and volume induced predictability. Other control variables include: Size, Book to Market, Operating Profitability, Firm Investment, Momentum Return, Return Volatility, Short Term Reversal Return, Average Abnormal Return in the last five days and Abnormal Turnover. All independent variables are standardized by day before calculating interactions. Therefore, the coefficient units are basis points per standard deviation increase in the independent variables. Detailed definitions of all variables are available in Appendix B.1. Newey–West standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The t-statistics are in parentheses.

	(1)	(2)	(3)
Dependent Variable	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	3.089	3.764	4.341
	(8.188)	(5.084)	(3.799)
$EmotionVsFact_{i,t} \times Sentiment_{i,t}$	-2.673	-1.743	-1.966
	(-5.109)	(-1.743)	(-1.326)
$EmotionVsFact_{i,t} \times AbRet_{i,t}$	-3.908	1.459	-1.311
	(-3.302)	(0.799)	(-0.484)
$Buzz_{i,t} \times AbRet_{i,t}$	3.505	3.895	5.850
	(7.019)	(5.318)	(5.296)
$Buzz_{i,t} \times ME_{i,t}$	-0.034	0.755	0.700
	(-0.122)	(1.200)	(0.635)
$Buzz_{i,t}$	-0.180	-0.279	0.869
	(-0.643)	(-0.459)	(0.866)
$EmotionVsFact_{i,t}$	-0.348	-2.006	-0.715
	(-0.543)	(-1.604)	(-0.376)
$AbRet_{i,t}$	-3.383	-6.727	-6.839
	(-4.533)	(-5.006)	(-3.643)
$ME_{i,t}$	-1.550	-5.509	-12.437
	(-3.189)	(-3.692)	(-4.279)
$BM_{i,t}$	-0.541	-2.279	-2.959
	(-1.094)	(-1.545)	(-1.023)
$OP_{i,t}$	0.014	0.348	0.619
	(0.038)	(0.359)	(0.349)
$IVN_{i,t}$	0.017	-2.278	-4.974
	(0.052)	(-2.329)	(-2.638)
$RV_{i,t}$	-0.086	-0.494	-1.081
	(-0.118)	(-0.200)	(-0.217)
$MOM_{i,t}$	-0.534	1.218	3.208
	(-0.904)	(0.673)	(0.888)
$ST_{i,t}$	-0.844	-1.706	-3.518
	(-1.569)	(-1.119)	(-1.220)
$AbRet_{i,t} \times Size_{i,t}$	-2.774	-6.171	-8.232
	(-5.862)	(-7.780)	(-7.060)
$AbTurn_{i,t}$	-5.638	-1.079	-4.525
	(-4.302)	(-0.484)	(-1.251)
$AbRet_{i,t-5,t-1}$	-2.710	-4.787	-5.356
	(-4.651)	(-3.749)	(-2.525)
$AbRet_{i,t} \times AbTurn_{i,t}$	0.324	-0.708	-1.309
	(1.128)	(-1.441)	(-1.907)
Constant	3.395	15.841	33.927
	(1.939)	(2.462)	(2.561)
Daily Average Observations	540	540	539
Adjusted R-squared	0.141	0.133	0.129
Observations	2,842,780	2,840,509	2,838,805

Table 6: Firm-Specific News Sentiment Factor Risk Premium-Fama-French Factor Model Testing

This table shows daily risk-adjusted returns (a) from firm-specific news sentiment zero-cost portfolio for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalisation from the last month to split stocks into two portfolio sizes: small and big. Independently, I rank stocks based on day t news sentiment into three sentiment portfolios: pessimistic (N) 30%, neutral (M) 40%, optimistic (P) 30%. The six interacted value-weighted portfolios respecting size and news sentiment are $N=S$, $N=B$, $M=S$, $M=B$, $P=S$, $P=B$ sorting on the size and the news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios 30% ($P=S$, $P=B$) and the average of short position in the two negative sentiment portfolios 30% ($N=S$, $N=B$) each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. Panel A shows Pearson correlation between the news sentiment portfolio return and conventional factors. Panel B presents the risk-adjusted return of the news sentiment zero-cost portfolio from models of CAPM, Fama–French three or five factors with Pastor and Stambaugh liquidity factor, momentum factor and short- and long-term reversal factors. Newey–West standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The t-statistics are in parentheses.

Panel A Correlations Between Different Factors

	MKT_t	SMB_t	HML_t	RMW_t	CMA_t	UMD_t	ST_t	LT_t	$PSLIQ_t$
$Sentiment_t$	-0.106	0.029	-0.102	0.071	0.084	0.202	-0.142	0.025	0.010
MKT_t		0.070	-0.012	-0.425	-0.333	-0.257	0.355	-0.084	0.085
SMB_t			0.052	-0.298	0.055	0.029	0.014	0.283	0.042
HML_t				0.088	0.483	-0.344	-0.097	0.477	0.098
RMW_t					0.280	0.151	-0.245	-0.161	0.040
CMA_t						0.065	-0.283	0.520	0.025
UMD_t							-0.126	0.030	-0.066
ST_t								-0.138	0.061
LT_t									-0.029

Panel B Risk-Adjusted News Sentiment Zero-Cost Portfolio Return

	$Sentiment_t$	CAPM	FF3	FF5	FF5+ UMD	FF5+ Full
a	0.066	0.067	0.068	0.064	0.061	0.065
t_a	(6.397)	(6.588)	(6.756)	(6.390)	(6.143)	(6.640)
MKT_t		-0.065	-0.069	-0.031	-0.016	0.000
t_{MKT}		(-4.795)	(-5.349)	(-2.582)	(-1.382)	(0.039)
SMB_t			0.051	0.056	0.041	0.035
t_{SMB}			(1.991)	(2.304)	(1.732)	(1.430)
HML_t			-0.123	-0.202	-0.126	-0.135
t_{HML}			(-4.263)	(-7.421)	(-4.851)	(-5.091)
RMW_t				0.058	0.038	0.029
t_{RMW}				(1.670)	(1.125)	(0.784)
CMA_t				0.235	0.185	0.147
t_{CMA}				(5.430)	(4.438)	(3.196)
$PSLIQ_t$			0.021	0.018	0.020	0.024
t_{PSLIQ}			(1.151)	(1.026)	(1.166)	(1.425)
UMD_t					0.115	0.109
t_{UMD}					(6.780)	(6.806)
ST_t						-0.088
t_{ST}						(-4.669)
LT_t						0.019
t_{LT}						(0.568)
$\bar{R}^2$	0.007	0.011	0.024	0.038	0.055	0.064
Days	5241	5241	5241	5241	5241	5241

C Online Appendix

C.1 Market News Sentiment and Market Uncertainty Regression Test

Table 1 shows the negative Pearson correlation coef cients between stock market news sentiment and market uncertainty measures. In this section, I conduct a xed effect regression to further verify the assumption in equation (2) that an increase of market news sentiment biases investors to perceive a lower market uncertainty. Speci cally, I use three measures of market uncertainty: S&P 500 realized volatility (RV_{500}), EPU and VIX. For each rm earnings announcement day, I calculate the average monthly market uncertainty from the announcement day (t) to the next 21 trading days ($t+21$ for RV and VIX) or 31 calendar days ($t+31$ for EPU). The xed effect regression is as follows:

$$Dep_{i;[t;t+21=31]} = b_0 + b_1 Sentiment_{m;[t-21;t-1]} + Xd + e_{i;t} \tag{29}$$

where $Sentiment_{m;[t-21;t-1]}$ is the $Buzz$-weighted average stock market news sentiment from 21 trading days up to 1 day before the earnings announcement. X is a vector of control variables including size ($Size$), turnover ($Turn$), average price ($Price$), return volatility (RV) and institutional ownership (IOW) (See detailed de nitions in Appendix B.1) and d is the coef cient vector. Since volatility is strongly persistent, I also control the lag variable, which is the average value one month before the earnings announcement for each market uncertainty measure.

Table 7 displays the results from equation (29). I control month-, year- and rm- xed effects and the standard errors are clustered by rm- and year- xed effects. Columns (1)-(3) shows that stock market news sentiment negatively predict market uncertainty across all three measures of economic uncertainty. In sum, the results are consistent with the negative correlation shown in Table 1. More importantly, this test con rms the assumption in equation (2) to serve the biasing channel in the model that investors' perception of market uncertainty is irrationalized by reading news characterized by a non-neutral tone.

[Insert Table 7 here]

C.2 Alternative Measure of Firm-Speci c Information Acquisition

Speci cally, I calculate the average total count of search volume for the les in the most recent month before the announcement. I then take the natural logarithm of the average of total SEC les searching volumes ($LogSEC_{i;t}$). To some extent, the count of SEC EDGAR le searching volume is a more straightforward way to understand investors' acquisition of rm-speci c information. The control variables are the same as the test in Table 3.

[Insert Table 8 here]

Not surprisingly, the results in Table 8 are consistent with those of Table 3 (with price the jump measure). As news tones tend to be more positive, investors are less willing to download the company's SEC files, showing a decrease in firm-information acquisition.

C.3 Excluding Earnings Announcement Days

In this section, I re-conduct analysis to confirm the robustness of the impact of the firm-specific news sentiment on cross-sectional stock returns, by excluding earnings announcement days and sorting regression data by different financial characteristics, which may potentially affect the predictability of the deviation of information risk resulted by news sentiment.

Table 9 shows the results from running daily cross-sectional Fama–Macbeth (1973) regressions (28) with daily cross-sectional data, excluding all earnings announcement days. Because, as argued by Tetlock et al. (2008), firm-specific news is most likely to be clustered near the time of a company's earnings announcement, there is a concern is raised that the inclusion of these days may amplify the impact from firm-specific news sentiment and other effects from news related to company earnings.[39]

[Insert Table 9 here]

Columns (1)–(3) in Table 9 show that excluding news of earnings announcements slightly reduces all of the coefficients, showing the small incremental benefit from news on earnings announcement days. Nonetheless, all results remain both statistically and economically significant. Indeed, news released while an earnings announcement is being made is more likely to attract investors' attention. In addition, information from news or online media reported close to a firm's earnings announcement plays an important role in transmitting firm-fundamental information to investors and traders (Tetlock et al., 2008; Chen et al., 2014). However, as Table 9 demonstrates, excluding news reported near the time of earnings announcement does not comprise the significance of news sentiment's positive predictability on cross-sectional stock returns. In an unreported table, I also test by excluding firm-specific news on the earning announcement day, on the day that precedes it and on the day that follows it; the results are similar to Table 9.

[39] For example, Tetlock et al. (2008) did find that earnings-related news has incremental benefit to uncover firms' value-relevant information. Thus, Tetlock (2010) thoroughly considers that information asymmetry dissolution from public news may be led by earnings news. To accommodate for this, Tetlock excludes earnings-related news in the main regression analysis as a robustness concern.

C.4 Sub-sample Regression Analysis

I divide data into samples based on characteristics of firm size, illiquidity, analyst coverage, analyst forecast dispersion and institutional ownership. For each day, I divide stocks into two sub-samples, high and low, based on the daily cross-sectional median of each characteristic. Each sub-sample must have at least 50 firms to run the Fama–Macbeth (1973) regression model.

Panels A through E in Table 10 are regression results based on the sub-samples for size, analyst coverage, analyst forecast dispersion, illiquidity and institutional ownership respectively. The high and low size sub-sample regression shows similar results to Table 5. Unsurprisingly, news sentiment predictability in the small firm sub-sample has a relatively stronger effect than the big firm sub-sample. In the small size sub-sample, news sentiment is statistically significant in its prediction of all future returns for R^e_{t+1} and cumulative returns $R^e_{t+2;t+5/10}$. However, it is well addressed empirically in the existing literature that large firms generally make more information available to investors and have less information asymmetry than small firms (Banz, 1981; Barry and Brown, 1984; Atiase, 1985; Freeman, 1987), thus showing a relatively weak effect on cumulative returns $R^e_{t+2;t+5/10}$. By the same token, Panel D shows very similar results as Panel A, because small and illiquid stocks are commonly known to share similar issues, especially in respect of information asymmetry. However, results in Panel B and C (for analyst coverage and analyst forecast dispersion, respectively) do not change much compared to the results from the full sample shown in Table 5. Both the number of analysts following a company and how analysts hold different beliefs about companies' earnings performance are unable to explain the cross-sectional variation of stock returns raised by variation of information asymmetry risk implied by news sentiment. Finally, it is intriguing that variation in institutional ownership does not explain the positive predictive effect of news sentiment on cross-sectional stock returns at $t+1$, however, the significance of prediction on cumulative returns up to $t+5$ and $t+10$ are reduced in both the high and low institutional ownership sub-sample regressions. The reduction in significance is more pronounced in the low institutional ownership sub-sample. A potential reason is that institutional investors are relatively easier or cost-efficient to be informed (Hendershott et al., 2015). In other words, when it comes to making an investment into an asset, institutional investors are more likely to be biased by news sentiment in their perception of uncertainties in the risky asset. Hence, their information acquisition decision, in equilibrium, is subject to biased beliefs rather than to rational expectations. To conclude, stocks with high institutional ownership show relatively stronger empirical results from firm-specific news sentiment than stocks with low institutional ownership.

[Insert Table 10 here]

In sum, even though the above robustness test shows that news sentiment has a stronger impact of return predictability on small and illiquid stocks, on average, the cross-sectional variation of stock returns resulting in information risk implied by firm-specific news sentiment is robust in all sub-samples, which may imply potential problems in stocks such as information asymmetry (size and illiquidity), investors' alternative beliefs (analyst forecast coverage and dispersion) and better-informed investors (institutional ownership).

C.5 q-factor model testing

Hou et al. (2015) develop an empirical asset pricing model known as the *q*-factor. The *q*-factor model indicates that expected excess returns can be explained by the sensitivities of the market factor, a size factor, an investment factor and a return on equity factor. More importantly, Hou, Xue and Zhang conduct comprehensive empirical testing on existing anomalies in cross-sectional stock returns and demonstrate the strong explanatory power of the *q*-factor model. The authors argue that the *q*-factor is a very competitive alternative to the Fama–French five factors model.

Therefore, I re-conduct all the tests in section 5.3 with the *q*-factor model. Panel A in Table 11 reports Pearson correlations between the news sentiment factor and factors from the *q*-factor model. Clearly, there are almost no economically significant correlations between the firm-specific news sentiment factor and other factors. Panel B in Table 11 shows risk-adjusted alphas across different models by running time-series regressions of the zero-cost news sentiment portfolio on the base line *q*-factor model and adding additional factors as customary controls. Essentially, the results are as expected and are in line with the findings of the Fama–French factor models. The news sentiment zero-cost portfolio maintains positive significant daily abnormal returns in the range of 6.1 to 6.5 basis point. None of the factors from the base *q*-factor model have strong explanatory power on the cross-sectional variation of stock returns, where returns are the result of a deviation in information asymmetry attributed to the biased tone in firm-specific news sentiment. Overall, traditional asset pricing factors developed based on firm fundamentals, with either the Fama–French five factors or the novel *q*-factor, lack the capability to capture the pricing effect caused by firm-specific news sentiment.

[Insert Table 11 here]

C.6 News Sentiment Factors vs. Other News Factors ?

Section 4.3 demonstrates that the zero-cost portfolio formed by daily firm-specific news sentiment generates a considerable amount of daily abnormal returns (α) which cannot be fully explained by customary factors from empirical asset pricing. I argue that this abnormal return from the theoretical implication in section 1 regarding firm-specific news sentiment causes a deviation in information risk, for which investors require compensation. However, as mentioned above in section

5.2, extant studies argue that firm-specific news sentiment contains firm value-relevant information. In fact, the Fama–Macbeth (1973) regression results in Table 5 confirm this finding from the significant coefficients of the controlled variable $EmotionVsFact_{i,t} \times Sentiment_{i,t}$. Owing to the lack of explanatory power of traditional asset pricing factors, there may be a concern that the daily abnormal returns sorted by firm-specific news sentiment could be captured by other novel factors from quantified news measures.

Hence, I consider constructing an additional factor based on the empirical evidence of the news-related variables from Table 5. More specifically, $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ verifies the significance of news sentiment, including both genuine information and the biasing effect on investors' valuation of firm fundamentals. In fact, this interacted variable $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ presents an intriguing finding: news sentiment has a segmented effect between `soft' information (for example, emotional or opinion references) and `hard' information (for instance, firm-fundamental or factual references) in the news. On the one hand, the `soft' information that is more focused on emotional references is more likely to bias investors' rational decisions. On the other hand, the `hard' information – specifically, factual or fundamental information such as accounting details or earnings – is more helpful for investors to understand a company's business condition and will potentially lead investors to dissolve value-relevant information as they may be uninformed without reading the news.

Essentially, it should be noted that both of the terms from $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ range from [-1, +1]. Hence, the value of $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ has two implications depending on whether or not sentiment conditioned by the type of information (soft versus hard) generates genuine information or leads to a biased evaluation of the firm by investors.

First, taking the behavioral perspective, higher values of $EmotionVsFact_{i,t}$ aligned with firm-specific news sentiment ($Sentiment_{i,t}$) imply that sentiment is more likely to drive from emotional references, to make investors more biased about the firm valuation. In this case, for example, a positive sentiment or optimistic tone from more emotional references as the value of $EmotionVsFact_{i,t}$ increases in firm-specific news is more likely to cause investors to overprice the value of a firm. Once the value of a firm moves back to its fundamental value, firm-specific news sentiment predicting a reversal appears as the mispricing is corrected.

Second, taking the genuine information of instructing the firm fundamentals argument concerning firm-specific news, a higher value of $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ implies more negative fundamental information in the news. In this case, both $EmotionVsFact_{i,t}$ and $Sentiment_{i,t}$ decline, causing their interaction value to become higher. For instance, a value of lower $EmotionVsFact_{i,t}$ means that there is more fundamental information in the firm-specific news, necessitating a lower value of $Sentiment_{i,t}$. Therefore, the interaction indicates negative fundamental information about the company, which can be acquired through investors' reading. As a result, investors correctly lower the valuation of the firm based on news containing more negative value-relevant information about the firm. The higher value of the interaction predicts lower stock future returns.

Following the implication of $EmotionVsFact_{i,t} \times Sentiment_{i,t}$. I construct an additional news factor to capture the effect of either biased valuation or genuine information from news sentiment about particular companies. The empirical results in Table 5 demonstrate that $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ predicts negative cross-sectional stock future returns ($R^e_{i,t+1}$). Therefore, portfolio sorting is based on the standardized value of $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ at day t. As mentioned by Tetlock (2011), sorting on this standardized interacting variable produces a similar result to sorting both of the variables. Next, the construction of the mimicking portfolio is the same as the firm news sentiment portfolio in section 5.2, but based on the value of $EmotionVsFact_{i,t} \times Sentiment_{i,t}$. I construct the zero-cost portfolio to take the average of long positions, with news either containing the most emotional (E) negative sentiment (N) or the most fundamental (F) positive sentiment (P) 30% ($EN(FP) = S$; $EN(FP) = B$)[40] and the average of short positions in the stocks with either the most emotional positive sentiment or the most fundamental negative 30% ($EP(FN) = S$; $EP(FN) = B$).[41] In other words, the profit of this zero-cost portfolio drives from buying stocks with good news or which are undervalued due to bias ($EmotionVsFact_{i,t} \times Sentiment_{i,t} \downarrow$) and selling stocks with bad news or which are overvalued due to bias ($EmotionVsFact_{i,t} \times Sentiment_{i,t} \uparrow$). Lastly, I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy.

On the one hand, the zero-cost portfolio based on $EmotionVsFact_{i,t} \times Sentiment_{i,t}$ (hereafter $EFSENT_{i,t}$) earns positive significant abnormal returns by about 2.7 basis point per day (6.8% annualized abnormal return). Panel A in Table 12 presents Pearson correlations between the $EFSENT_{i,t}$ and other fundamental factors. Clearly, there is hardly any correlation between $EFSENT_{i,t}$ and extant classical factors. Panel B shows risk-adjusted alphas based on different models. In fact, there is little reduction of abnormal return earned by sorting $EFSENT_{i,t}$ when controlling for other pricing factors across column (2)–(6). The full specification in column (6) is only reduced by 0.2 basis points. Therefore, the information implied by this novel news pricing factor $EFSENT_{i,t}$ cannot be explained by the existing fundamental pricing factors.

[Insert Table 12 here]

On the other hand, I add $EFSENT_{i,t}$ as an additional novel pricing factor to adjust for doubted latent effects such as genuine information contained in the news sentiment about the firm fundamentals or mis-valuation of a company caused by news sentiment. First the Pearson correlation between the news sentiment factor and the $EFSENT_{i,t}$ factor is about -0.058. Second, as shown in Table 13, on average, the $EFSENT_{i,t}$ factor is only significantly negative around the 10% level

[40] Both these portfolios predict the positive stock future returns shown in Table 5.

[41] These two portfolios negatively predict the next day's stock returns, and the empirical result in Table 5 confirms that.

to explain the abnormal return from the news sentiment factor. Even though $EFSENT_{i,t}$ may capture some effects in the firm-specific news sentiment to decrease the abnormal return from news sentiment portfolio-for example, good fundamental information about the firm or investors' undervaluation-the total amount of explanatory power from the $EFSENT_{i,t}$ factor is not economically significant. In fact, the abnormal return owing to the news sentiment factors remains almost at the same degree seen in Table 6 without controlling for the additional news effect factor $EFSENT_{i,t}$.

In sum, the abnormal excess return generated by firm-specific news sentiment that causes the deviation for information risk in assets is robust for both traditional fundamental factors in empirical asset pricing and the novel news factor which I propose in this paper to capture either potential value-relevant information or mis-valuation effects from the firm news sentiment.

[Insert Table 13 here]

C.7 A Toy Model of News Bias

The toy model of bias in media is motivated by Dyck and Zingales (2003). The suppliers of information such as journalists or news companies supply news as a function of bias:

$$N^s = \theta^s + \beta b \tag{C.71}$$

The bias can be either an optimistic or pessimistic tone used by the information suppliers to improve readership.

However, investors demand high accuracy in news, in that, too much bias deceases demand or readership of investors. Their demand function is negatively related to the bias imposed by the information suppliers:

$$N^d = \theta^d - \gamma b \tag{C.72}$$

As mentioned by Dyck and Zingales (2003), I assume journalists or news suppliers choose to implement bias into their news, in a competitive market to equate demand and supply of news. Therefore, in equilibrium, the bias is:

$$b = \frac{\theta^d - \theta^s}{\beta + \gamma} \tag{C.73}$$

The equation (C.73) indicates information from news is always subject to some bias. Dyck and Zingales (2003)) argue that a lower degree of bias induces excess demand for news and a higher degree of bias induces excess supply of news. Noted this, the equations (C.71–3) only show the existence of bias in the news provided by the information suppliers. In other words, there is no need to specify the sign of bias.

In sum, this toy model assuredly motivates the idea that information from news contains bias, for which I argue in this paper. More speci cally, the bias is subject to news suppliers' choice of using either an optimistic (positive) or pessimistic (negative) tone (sentiment) in the news to potentially increase readership or ful l readers' demand for news.

Table 7: Market News Sentiment and Market Uncertainty

This table presents the results of regressions of market uncertainty measures on stock market news sentiment during the firm earnings announcement window. Columns (1)–(3) are based on the fixed-effect regression from equation (29): $Dep_{i,[t,t+21/31]} = b_0 + b_1 Sentiment_{m,[t-21,t-1]} + Xd + e_{i,t}$ and $Dep_{i,[t,t+21/31]}$ is the average monthly market uncertainty from the announcement day (t) to the next 21 trading days ($t+21$ for RV and VIX) or 31 calendar days ($t+31$ for EPU). The news sentiment variable $Sentiment_{m,[t-21,t-1]}$ and $Buzz_{m,[t-21,t-1]}$ are calculated in the same way as the daily $Buzz_{m,t}$-weighted average in the study window. Control variables include $Buzz_{m,[t-21,t-1]}$ as the proxy of intensity of stock market news coverage, lagged dependent variables are calculated as 21 (31) days until one day before announcement. Size, Turn, Price, Return Volatility and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. Detailed definition of all variables are available in Appendix B.1. Standard errors are clustered by both firm- and time-fixed effect in column (1)-(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively. The different number of firms in firm-specific news sentiment regression is subject to availability of firm-level news data.

	(1)	(2)	(3)
Dependent Variable	$RV_{500,[t,t+21]}$	$EPU_{t,t+31}$	$VIX_{t,t+21}$
$Sentiment_{m,[t-21,t-1]}$	0.166	121.736	15.111
	(0.002)	(1.346)	(0.243)
$RV_{500,[t-21,t-1]}$	0.254		
	(0.005)		
$EPU_{t-21,t-1}$	0.0003	0.522	0.040
	(0.000)	(0.003)	(0.001)
$Buzz_{m,[t-21,t-1]}$	0.381	58.285	38.643
	(0.005)	(2.115)	(0.479)
$VIX_{t-21,t-1}$		0.174	0.326
		(0.021)	(0.005)
Controls	Yes	Yes	Yes
FE Firms	Yes	Yes	Yes
FE Month	Yes	Yes	Yes
FE Year	Yes	Yes	Yes
Observations	91,873	91,873	91,873
R-squared	0.620	0.728	0.720
Number of Firms	10,241	10,241	10,241

Cluster standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: News Sentiment Impact on Information Acquisition Measured by Counts of SEC Files Clicks

This table presents the results of regressions of the count of SEC EDGAR file searching volume as the proxy for firm-specific information acquisition on stock market news sentiment during the firm earnings announcement window. Columns (1)–(3) are based on the fixed-effect regression from equation (26) $LogSEC_{i,t} = b_0 + b_{j,1} Sentiment_{j,t-21,t-1} + Xd + e_{i,t}$, where $j \in \{m, i\}$ and $LogSEC_{i,t}$ is the average of total SEC files searching volumes in the most recent month before the earnings announcement. The news sentiment variable $Sentiment_{j,[t-21,t-1]}$ and $Buzz_{j,[t-21,t-1]}$ are calculated in the same way as the daily $Buzz_{j,t}$-weighted average in the study window. Control variables include $Buzz_{j,[t-21,t-1]}$ as the proxy of intensity of stock market news coverage, economic uncertainty proxies (VIX and EPU) and the numbers of analyst coverage is calculated as 21 days until one day before announcement. Size, Turn, Price, Return Volatility and Institutional Ownership are calculated as 42 days up to 21 days before the announcement. Detailed definition of all variables are available in Appendix B.1. Standard errors are clustered by both firm- and time-fixed effect in column (1)-(3). ***,**,* indicate statistical significance at the two-sided 1%,5%,10% levels, respectively. The different number of firms in firm-specific news sentiment regression is subject to availability of firm-level news data.

	Panel A Stock Market News Sentiment			Panel B Firm-Specific News Sentiment		
	(1)	(2)	(3)	(1)	(2)	(3)
Dependent Variable	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$	$LogSEC_{i,t}$
$Sentiment_{i,t-21,t-1}$				-0.052	-0.042	-0.042
				(0.017)	(0.017)	(0.017)
$Sentiment_{m,t-21,t-1}$	-0.231	-0.247	-0.227		-0.42	-0.394
	(0.085)	(0.091)	(0.09)		(0.187)	(0.189)
$Buzz_{m,t-21,t-1}$		-0.016	-0.034			
		(0.025)	(0.028)			
$Buzz_{i,t-21,t-1}$					0.023	0.022
					(0.004)	(0.004)
$VIX_{t-21,t-1}$		0.004			0.001	
		(0.002)			(0.004)	
$EPU_{t-21,t-1}$			0.0004			-0.0004
			(0.000)			(0.0006)
LagDep	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
Day of Week	Yes	Yes	Yes	Yes	Yes	Yes
FE Year-Month	Yes	Yes	Yes	Yes	Yes	Yes
FE Firm	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40,412	39,971	39,971	9,183	9,121	9,121
R-squared		0.845	0.845		0.861	0.861
Number of Firms	3,660	3,641	3,641	2,586	2,568	2,568

Clustered standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9: Cross-Sectional Return Predictability from Firm-Specific News Sentiment without Earnings Announcement Days

This table presents results excluding data on firm earnings announcement days and results from daily cross-sectional Fama–MacBeth (1973) regressions of next-day firm-specific news sentiment on $t+1$ return and cumulative returns from $t+2$ to $t+5$ or $t+10$. Variables measured by news content and all other control variables are known by day t. Columns (1)–(3) report the time-series average of the coefficients based on the model in equation (28): $DepVar_{i,t+1} = b_0 + b_1 Sentiment_{i,t} + dX + e_{i,t}$ for each trading day t, where $DepVar_{i,t+1}$ is R^e_{t+1}, $R^e_{t+2,t+5}$, and $R^e_{t+2,t+10}$, respectively. The variable $Sentiment_{i,t}$ is firm-specific news sentiment as a proxy for biased information related to the firm-specific component. The news-related interacted variables including $EmotionVsFact_{i,t} \times Sentiment_{i,t}$, $EmotionVsFact_{i,t} \times AbRet_{i,t}$, and $Buzz_{i,t} \times AbRet_{i,t}$ control for potential effects of genuine information or biased valuation regarding firm fundamentals for $Sentiment_{i,t}$. Additionally, abnormal return $AbRet_{i,t}$ at day t and its related interactions such as $AbRet_{i,t} \times Size_{i,t}$ and $AbRet_{i,t} \times AbTurn_{i,t}$ measure return reversal and volume induced predictability. Other control variables include: Size, Book to Market, Operating Profitability, Firm Investment, Momentum Return, Return Volatility, Short Term Reversal Return, Average Abnormal Return in the Last Five Days and Abnormal Turnover. All independent variables are standardized by day before calculating interactions. Therefore, the coefficient units are basis points per standard deviation increase in the independent variables. Detailed definitions of all variables are available in Appendix B.1. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

	(1)	(2)	(3)
Dependent Variable	$R^e_{i,t+1}$	$R^e_{i,t+2,t+5}$	$R^e_{i,t+2,t+10}$
$Sentiment_{i,t}$	2.146	3.086	3.605
	(8.074)	(5.969)	(4.383)
$EmotionVsFact_{i,t} \times Sentiment_{i,t}$	-1.078	-0.878	-0.968
	(-4.805)	(-2.093)	(-1.562)
$EmotionVsFact_{i,t} \times AbRet_{i,t}$	-1.311	0.852	-0.172
	(-2.649)	(1.139)	(-0.150)
$Buzz_{i,t} \times AbRet_{i,t}$	3.537	4.065	6.245
	(6.753)	(5.337)	(5.377)
$Buzz_{i,t} \times ME_{i,t}$	-0.112	0.771	0.597
	(-0.377)	(1.154)	(0.507)
$Buzz_{i,t}$	-0.133	-0.288	1.034
	(-0.457)	(-0.458)	(0.985)
$EmotionVsFact_{i,t}$	-0.207	-0.780	-0.304
	(-0.781)	(-1.529)	(-0.382)
$AbRet_{i,t}$	-4.396	-5.768	-7.039
	(-6.555)	(-4.876)	(-4.462)
$ME_{i,t}$	-1.660	-5.741	-13.150
	(-3.265)	(-3.742)	(-4.397)
$BM_{i,t}$	-0.443	-2.092	-2.244
	(-0.840)	(-1.360)	(-0.742)
$OP_{i,t}$	0.079	0.433	0.929
	(0.210)	(0.429)	(0.500)
$IVN_{i,t}$	0.013	-2.317	-4.918
	(0.039)	(-2.264)	(-2.467)
$RV_{i,t}$	-0.157	-0.220	-0.966
	(-0.206)	(-0.086)	(-0.186)
$MOM_{i,t}$	-0.313	1.365	3.566
	(-0.504)	(0.718)	(0.937)
$ST_{i,t}$	-0.833	-1.792	-3.663
	(-1.478)	(-1.118)	(-1.206)
$AbRet_{i,t} \times Size_{i,t}$	-2.694	-6.420	-8.607
	(-5.458)	(-7.562)	(-6.886)
$AbTurn_{i,t}$	-5.432	-1.171	-4.822
	(-4.105)	(-0.517)	(-1.300)
$AbRet_{i,t-5,t-1}$	-2.827	-4.091	-4.665
	(-4.793)	(-3.047)	(-2.088)
$AbRet_{i,t} \times AbTurn_{i,t}$	0.326	-0.728	-1.433
	(1.080)	(-1.374)	(-1.949)
Constant	3.399	16.207	36.189
	(1.891)	(2.406)	(2.616)
Daily Average Firms	512	511	511
Adjusted R-squared	0.147	0.138	0.135
Observations	2,538,963	2,537,599	2,536,117

Table 10: Cross-Sectional Return Predictability from Firm-Specific News Sentiment Sorted by Firm Characteristics

This table presents results from daily cross-sectional Fama–MacBeth (1973) regressions of next-day firm-specific news sentiment $t+1$ return and cumulative returns from $t+2$ to $t+5$ or $t+10$ with different sub-samples sorted into two portfolios based on financial characteristics. For each day, I divide stocks into two sub-samples: high and low, based on the daily cross-sectional median of each characteristic. From panel A to E, samples are sorted based on firm size, analyst coverage, analyst forecast dispersion, illiquidity measure and institutional ownership. The low and high sub-panels report the time-series average of the coefficients from each characteristic sorted sub-samples and is based on the model in equation (28): $DepVar_{i;t+1} = b_0 + b_1 Sentiment_{i;t} + dX + e_{i;t}$ for each trading day t, where $DepVar_{i;t+1}$ is R^e_{t+1}, $R^e_{t+2;t+5}$, and $R^e_{t+2;t+10}$, respectively. The variable $Sentiment_{i;t}$ is firm-specific news sentiment as a proxy for biased information related to the firm-specific component. The news-related interacted variables including $EmotionVsFact_{i;t} * Sentiment_{i;t}$, $EmotionVsFact_{i;t} * AbRet_{i;t}$, and $Buzz_{i;t} * AbRet_{i;t}$ control for potential effects of genuine information or biased valuation regarding firm fundamentals from $Sentiment_{i;t}$. Additionally, abnormal return $AbRet_{i;t}$ at day t and its related interactions such as $AbRet_{i;t} * Size_{i;t}$ and $AbRet_{i;t} * AbTurn_{i;t}$ measure return reversal and volume induced predictability. Other control variables include: Size, Book to Market, Operating Profitability, Firm Investment, Momentum Return, Return Volatility, Short Term Reversal Return, Average Abnormal Return in the Last Five Days and Abnormal Turnover. All independent variables are standardized by day before calculating interactions. Therefore, the coefficient units are basis points per standard deviation increase in the independent variables. Detailed definitions of all variables are available in Appendix B.1. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

Panel A Size Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression

	Low			High		
	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$
$Sentiment_{i;t}$	2.702	3.386	5.098	1.222	1.142	0.060
	(6.638)	(4.383)	(4.303)	(3.963)	(1.911)	(0.063)
$EmotionVsFact_{i;t} * Sentiment_{i;t}$	-1.193	-0.549	-0.994	-0.838	-0.106	0.337
	(-3.408)	(-0.762)	(-1.003)	(-3.066)	(-0.219)	(0.441)
$EmotionVsFact_{i;t} * AbRet_{i;t}$	-2.139	1.650	-2.487	-0.971	0.422	1.756
	(-2.841)	(1.300)	(-1.306)	(-1.883)	(0.460)	(1.303)
$Buzz_{i;t} * AbRet_{i;t}$	3.539	4.108	6.853	2.374	3.161	4.010
	(4.486)	(3.270)	(3.679)	(4.786)	(3.669)	(3.069)
$BUZZ_{i;t}$	-0.163	-0.421	1.289	-0.111	-1.004	-1.108
	(-0.377)	(-0.520)	(1.016)	(-0.303)	(-1.152)	(-0.724)
$Buzz_{i;t} * Size_{i;t}$	-0.147	0.268	0.204	0.069	1.087	1.887
	(-0.335)	(0.287)	(0.136)	(0.250)	(1.496)	(1.435)
$AbRet_{i;t}$	-4.769	-2.323	-2.186	-4.760	-8.288	-11.453
	(-5.456)	(-1.558)	(-1.018)	(-7.186)	(-7.197)	(-7.179)
$ABbTurn_{i;t}$	-6.553	0.484	-2.101	-3.397	2.310	1.470
	(-3.128)	(0.147)	(-0.408)	(-2.902)	(0.963)	(0.333)
$AbRet_{i;t-5;t-1}$	-1.759	-2.554	-1.352	-4.187	-6.747	-10.216
	(-2.389)	(-1.643)	(-0.547)	(-6.638)	(-4.727)	(-4.458)
$AbRet_{i;t} * AbtTurn_{i;t}$	0.666	-1.962	-2.907	0.409	2.197	1.716
	(1.173)	(-2.182)	(-2.263)	(0.853)	(2.530)	(1.551)
$AbRet_{i;t} * Size_{i;t}$	-2.653	-4.104	-5.024	-0.769	-2.246	-2.810
	(-3.455)	(-3.217)	(-2.939)	(-1.491)	(-2.381)	(-2.059)
$ME_{i;t}$	-1.430	-4.669	-7.941	-0.709	-2.343	-5.472
	(-2.563)	(-3.095)	(-2.838)	(-1.739)	(-1.867)	(-2.264)
$BM_{i;t}$	-1.041	-3.457	-4.322	-0.706	0.188	0.869
	(-1.637)	(-1.953)	(-1.277)	(-1.222)	(0.113)	(0.275)
$OP_{i;t}$	-0.331	0.420	0.918	-0.176	0.988	1.919
	(-0.595)	(0.292)	(0.356)	(-0.383)	(0.829)	(0.856)
$IVN_{i;t}$	0.243	-4.634	-6.096	-0.794	-1.971	-4.708
	(0.459)	(-3.995)	(-3.081)	(-1.394)	(-1.366)	(-1.758)
$RV_{i;t}$	-0.489	0.029	0.271	-0.068	-2.272	-5.803
	(-0.603)	(0.011)	(0.056)	(-0.089)	(-0.919)	(-1.190)
$MOM_{i;t}$	-0.677	1.821	2.703	-0.234	2.434	5.805
	(-0.861)	(0.868)	(0.688)	(-0.360)	(1.186)	(1.420)
$ST_{i;t}$	-1.395	-1.731	-2.354	-0.761	-2.185	-6.247
	(-1.646)	(-0.985)	(-0.746)	(-1.482)	(-1.291)	(-1.982)
Constant	3.843	18.437	41.149	2.631	12.610	25.236
	(2.019)	(2.624)	(2.908)	(1.609)	(2.080)	(2.071)
Daily Average Firms	271.625	271.607	271.607	271.169	271.151	271.151
Adjusted R-squared	0.138	0.125	0.121	0.184	0.173	0.170

Panel B Analyst Coverage Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$
$Sentiment_{i;t}$	2.777	3.869	5.093	1.36	1.653	0.621
	(7.515)	(5.215)	(4.575)	(3.936)	-2.527	(0.572)
$EmotionVsFact_{i;t} * Sentiment_{i;t}$	-1.217	-0.063	-0.976	-0.904	-1.385	-0.993
	(-3.855)	(-0.099)	(-0.984)	(-3.104)	(-2.394)	(-1.159)
$EmotionVsFact_{i;t} * AbRet_{i;t}$	-1.706	1.319	1.792	-1.044	-0.129	-1.284
	(-2.401)	(1.068)	(1.027)	(-1.951)	(-0.126)	(-0.883)
$Buzz_{i;t} * AbRet_{i;t}$	3.338	4.645	7.462	2.576	2.901	4.379
	(4.128)	(3.883)	(4.352)	(4.658)	-3.132	(3.228)
$BUZZ_{i;t}$	-0.063	-0.268	1.213	0.044	-0.594	0.649
	(-0.153)	(-0.357)	(1.0)	(0.113)	(-0.605)	(0.388)
$Buzz_{i;t} * Size_{i;t}$	-0.383	-0.428	-2.229	0.313	1.904	2.89
	(-0.935)	(-0.458)	(-1.451)	(0.918)	-2.165	(1.891)
$AbRet_{i;t}$	-4.637	-2.66	-1.841	-4.513	-7.167	-9.789
	(-5.598)	(-1.858)	(-0.878)	(-6.228)	(-5.214)	(-5.202)
$ABbTurn_{i;t}$	-7.013	-0.89	-4.739	-2.686	-5.112	-8.206
	(-3.577)	(-0.273)	(-0.926)	(-2.203)	(-2.004)	(-1.826)
$AbRet_{i;t-5;t-1}$	-1.437	-3.376	-4.173	-3.988	-6.29	-8.579
	(-1.985)	(-2.266)	(-1.747)	(-5.843)	(-4.048)	(-3.447)
$AbRet_{i;t} * AbtTurn_{i;t}$	0.789	-0.43	-0.656	0.228	1.644	1.041
	(1.412)	(-0.489)	(-0.491)	(0.479)	-1.727	(0.79)
$AbRet_{i;t} * Size_{i;t}$	-3.23	-3.719	-5.496	-1.178	-2.223	-2.679
	(-4.551)	(-2.957)	(-3.017)	(-1.869)	(-2.01)	(-1.6)
$ME_{i;t}$	-2.169	-5.6	-11.767	-1.148	-4.694	-10.855
	(-3.963)	(-3.495)	(-3.862)	(-2.118)	(-2.888)	(-3.626)
$BM_{i;t}$	-0.936	-1.154	-1.084	-0.494	-2.765	-4.264
	(-1.557)	(-0.711)	(-0.342)	(-0.766)	(-1.495)	(-1.204)
$OP_{i;t}$	0.147	2.098	4.879	-0.184	-0.663	-1.322
	(0.27)	(1.576)	(2.024)	(-0.353)	(-0.527)	(-0.536)
$IVN_{i;t}$	0.088	-3.701	-5.153	-0.501	-1.263	-4.753
	(0.187)	(-3.254)	(-2.504)	(-1.043)	(-0.85)	(-1.764)
$RV_{i;t}$	0.147	-0.157	-1.094	-0.36	-0.669	-1.077
	(0.18)	(-0.063)	(-0.225)	(-0.41)	(-0.24)	(-0.194)
$MOM_{i;t}$	-1.176	0.363	0.783	-0.134	2.936	5.576
	(-1.51)	(0.175)	(0.204)	(-0.185)	-1.371	(1.368)
$ST_{i;t}$	-0.841	-2.799	-3.098	-0.812	-1.057	-4.913
	(-0.96)	(-1.606)	(-0.931)	(-1.445)	(-0.601)	(-1.507)
Constant	3.552	17.06	37.934	2.62	12.709	27.871
	(1.953)	(2.518)	(2.78)	(1.533)	-2.028	(2.212)
Daily Average Firms	277	277	277	257.409	257.396	257.396
Adjusted R-squared	0.147	0.133	0.130	0.191	0.181	0.177

Panel C Analyst Forecast Dispersion Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$
$Sentiment_{i;t}$	1.76	2.371	1.997	1.905	3.116	4.043
	(5.251)	(3.585)	(1.953)	(4.826)	(3.743)	(3.295)
$EmotionVsFact_{i;t} * Sentiment_{i;t}$	-1.269	-0.301	0.2	-0.521	-0.956	-1.714
	(-3.99)	(-0.511)	(0.25)	(-1.595)	(-1.403)	(-1.731)
$EmotionVsFact_{i;t} * AbRet_{i;t}$	-0.877	-1.648	-1.203	-1.463	1.37	-2.239
	(-1.346)	(-1.49)	(-0.677)	(-1.869)	(1.101)	(-1.221)
$Buzz_{i;t} * AbRet_{i;t}$	3.699	3.319	4.078	3.86	2.824	4.505
	(5.416)	(3.074)	(2.718)	(5.136)	(2.307)	(2.455)
$BUZZ_{i;t}$	0.368	-0.37	0.1	-0.423	-0.025	1.123
	(1.067)	(-0.505)	(0.09)	(-0.938)	(-0.027)	(0.714)
$Buzz_{i;t} * Size_{i;t}$	-0.105	1.001	1.254	0.103	0.293	-0.592
	(-0.295)	(1.241)	(0.901)	(0.243)	(0.31)	(-0.366)
$AbRet_{i;t}$	-4.8	-9.15	-11.04	-5.06	-2.029	-3.391
	(-6.44)	(-6.092)	(-5.928)	(-6.063)	(-1.283)	(-1.498)
$ABbTurn_{i;t}$	-2.435	-0.358	-5.606	-4.91	-1.188	-4.997
	(-1.817)	(-0.088)	(-0.727)	(-3.052)	(-0.41)	(-1.036)
$AbRet_{i;t-5;t-1}$	-4.345	-6.648	-9.987	-2.543	-4.293	-4.958
	(-6.426)	(-4.49)	(-4.81)	(-3.641)	(-2.589)	(-1.873)
$AbRet_{i;t} * AbtTurn_{i;t}$	-0.163	0.723	1.431	0.439	-0.208	-1.571
	(-0.339)	(0.728)	(0.856)	(0.707)	(-0.233)	(-1.159)
$AbRet_{i;t} * Size_{i;t}$	-2.868	-3.536	-5.058	-3.176	-4.184	-5.272
	(-4.278)	(-3.348)	(-2.922)	(-4.213)	(-3.343)	(-2.638)
$ME_{i;t}$	-1.896	-4.015	-10.012	-0.996	-4.411	-9.577
	(-3.288)	(-2.611)	(-3.448)	(-1.58)	(-2.575)	(-2.935)
$BM_{i;t}$	-0.415	-0.674	0.351	-0.646	-0.626	0.256
	(-0.608)	(-0.374)	(0.1)	(-0.97)	(-0.332)	(0.071)
$OP_{i;t}$	-0.379	0.108	1.435	0.362	1.486	3.445
	(-0.649)	(0.085)	(0.595)	(0.667)	(1.107)	(1.474)
$IVN_{i;t}$	0.166	-1.322	-1.578	-0.804	-4.393	-7.929
	(0.355)	(-1.106)	(-0.741)	(-1.477)	(-3.192)	(-3.177)
$RV_{i;t}$	0.658	3.278	5.933	0.166	-1.217	-3.516
	(0.869)	(1.419)	(1.301)	(0.189)	(-0.452)	(-0.677)
$MOM_{i;t}$	-0.886	0.732	1.796	-1.87	-1.5	-3.279
	(-1.229)	(0.369)	(0.471)	(-2.515)	(-0.7)	(-0.797)
$ST_{i;t}$	-0.892	-3.584	-6.866	-0.983	-1.435	-2.614
	(-1.455)	(-2.198)	(-2.377)	(-1.372)	(-0.658)	(-0.697)
Constant	4.449	17.67	37.831	1.841	11.891	25.376
	(2.951)	(3.147)	(3.37)	(0.882)	(1.566)	(1.656)
Daily Average Firms	260	260	260	259	259	259
Adjusted R-squared	0.161	0.142	0.141	0.169	0.154	0.148

Panel D Illiquidity Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$
$Sentiment_{i;t}$	1.142	1.501	0.088	2.86	3.088	4.998
	(3.638)	(2.365)	(0.087)	(6.971)	(4.025)	(4.205)
$EmotionVsFact_{i;t} * Sentiment_{i;t}$	-0.883	-0.368	0.428	-1.252	-0.23	-0.925
	(-3.16)	(-0.733)	(0.588)	(-3.613)	(-0.34)	(-0.929)
$EmotionVsFact_{i;t} * AbRet_{i;t}$	-0.533	0.635	1.103	-1.934	1.67	-1.73
	(-1.025)	(0.673)	(0.795)	(-2.674)	(1.259)	(-0.927)
$Buzz_{i;t} * AbRet_{i;t}$	2.339	3.522	4.117	4.453	4.742	8.101
	(4.464)	(3.941)	(2.948)	(5.388)	(3.604)	(4.243)
$BUZZ_{i;t}$	0.112	-0.653	0.926	-0.132	-0.396	1.011
	(0.308)	(-0.719)	(0.584)	(-0.308)	(-0.488)	(0.783)
$Buzz_{i;t} * Size_{i;t}$	0.19	1.156	1.182	-0.206	0.538	-1.462
	(0.604)	(1.429)	(0.825)	(-0.49)	(0.618)	(-1.043)
$AbRet_{i;t}$	-4.759	-8.53	-11.667	-5.213	-1.545	-1.018
	(-6.807)	(-6.772)	(-6.784)	(-6.079)	(-1.029)	(-0.478)
$ABbTurn_{i;t}$	-5.283	1.509	0.798	-6.076	0.316	-1.511
	(-3.171)	(0.589)	(0.171)	(-3.05)	(0.088)	(-0.276)
$AbRet_{i;t-5;t-1}$	-3.895	-7.317	-10.087	-1.543	-1.96	-1.696
	(-5.843)	(-5.048)	(-4.183)	(-2.033)	(-1.21)	(-0.704)
$AbRet_{i;t} * AbtTurn_{i;t}$	0.838	2.026	1.407	0.749	-2.025	-2.42
	(1.503)	(2.166)	(1.165)	(1.273)	(-2.075)	(-1.772)
$AbRet_{i;t} * Size_{i;t}$	0.318	-1.192	-1.979	-2.679	-3.745	-4.574
	(0.561)	(-1.133)	(-1.271)	(-3.668)	(-2.899)	(-2.596)
$ME_{i;t}$	-1.114	-2.809	-7.411	-1.73	-4.933	-9.84
	(-2.333)	(-2.01)	(-2.738)	(-3.1)	(-3.166)	(-3.419)
$BM_{i;t}$	-0.723	-1.413	-2.445	-0.785	-2.286	-1.666
	(-1.254)	(-0.836)	(-0.751)	(-1.269)	(-1.302)	(-0.514)
$OP_{i;t}$	0.175	0.089	0.282	-0.386	1.065	3.374
	(0.385)	(0.077)	(0.127)	(-0.689)	(0.706)	(1.329)
$IVN_{i;t}$	-0.514	-1.273	-4.151	0.276	-4.791	-6.807
	(-0.98)	(-0.92)	(-1.581)	(0.535)	(-4.155)	(-3.519)
$RV_{i;t}$	0.754	-2.117	-4.153	-0.784	0.932	1.288
	(0.893)	(-0.791)	(-0.792)	(-0.987)	(0.384)	(0.272)
$MOM_{i;t}$	-0.236	2.36	6.442	-1.302	0.631	0.369
	(-0.34)	(1.136)	(1.585)	(-1.698)	(0.305)	(0.095)
$ST_{i;t}$	-0.096	-0.127	-1.676	-1.715	-4.569	-6.553
	(-0.18)	(-0.072)	(-0.506)	(-1.865)	(-2.604)	(-1.972)
Constant	2.211	12.088	25.489	3.84	18.393	40.803
	(1.328)	(1.962)	(2.055)	(2.029)	(2.647)	(2.914)
Daily Average Firms	272	272	272	271	271	271
Adjusted R-squared	0.190	0.180	0.177	0.141	0.125	0.122

Panel E Institutional Ownership Sub-Sample Daily Cross-Sectional Fama-Macbeth Regression						
	Low			High		
	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$	$R^e_{i;t+1}$	$R^e_{i;t+2;t+5}$	$R^e_{i;t+2;t+10}$
$Sentiment_{i;t}$	2.778	1.942	2.29	2.406	3.074	3.799
	(3.365)	(1.149)	(0.905)	(3.15)	(1.899)	(1.594)
$EmotionVsFact_{i;t} * Sentiment_{i;t}$	-1.005	-0.62	-1.005	-1.106	-1.406	-3.344
	(-1.571)	(-0.436)	(-0.518)	(-1.721)	(-1.045)	(-1.858)
$EmotionVsFact_{i;t} * AbRet_{i;t}$	-2.264	-3.645	-7.842	-0.899	3.594	2.011
	(-1.794)	(-1.185)	(-1.888)	(-0.701)	(1.858)	(0.718)
$Buzz_{i;t} * AbRet_{i;t}$	-0.084	4.79	5.854	2.754	4.163	3.573
	(-0.051)	(1.769)	(1.634)	(2.057)	(1.664)	(1.013)
$BUZZ_{i;t}$	-1.33	-1.757	-0.713	-0.044	2.773	1.975
	(-1.824)	(-1.105)	(-0.25)	(-0.053)	(1.553)	(0.81)
$Buzz_{i;t} * Size_{i;t}$	0.784	0.445	-2.177	-0.847	-3.855	-7.218
	(0.841)	(0.225)	(-0.695)	(-1.096)	(-2.451)	(-2.907)
$AbRet_{i;t}$	-3.938	-6.846	0.934	-5.32	-4.632	-5.251
	(-2.399)	(-2.136)	(0.212)	(-3.195)	(-1.706)	(-1.463)
$ABbTurn_{i;t}$	-5.884	-6.452	-16.193	-0.212	-2.444	-2.599
	(-1.76)	(-0.967)	(-1.45)	(-0.076)	(-0.516)	(-0.407)
$AbRet_{i;t-5;t-1}$	-1	0.107	1.084	-3.141	-3.813	-3.996
	(-0.687)	(0.03)	(0.183)	(-2.298)	(-1.433)	(-0.918)
$AbRet_{i;t} * AbtTurn_{i;t}$	1.076	0.392	-1.48	0.005	-0.764	0.343
	(0.933)	(0.189)	(-0.419)	(0.005)	(-0.532)	(0.181)
$AbRet_{i;t} * Size_{i;t}$	-2.633	-8.505	-11.968	-1.936	-4.178	-3.111
	(-1.77)	(-2.835)	(-2.807)	(-1.488)	(-2.047)	(-1.026)
$ME_{i;t}$	0.232	-3.913	-14.418	-0.797	-3.789	-6.178
	(0.195)	(-1.079)	(-2.169)	(-0.691)	(-1.257)	(-1.138)
$BM_{i;t}$	2.019	2.113	3.055	-0.145	0.663	8.899
	(1.514)	(0.516)	(0.403)	(-0.11)	(0.185)	(1.238)
$OP_{i;t}$	1.296	1.004	4.84	1.377	4.925	13.532
	(1.162)	(0.359)	(1.023)	(1.112)	(1.579)	(2.551)
$IVN_{i;t}$	0.349	-2.145	1.307	0.102	-2.791	-7.858
	(0.308)	(-0.775)	(0.248)	(0.104)	(-1.103)	(-1.616)
$RV_{i;t}$	1.144	2.346	3.261	-0.74	-1.33	-3.605
	(0.613)	(0.415)	(0.308)	(-0.543)	(-0.297)	(-0.413)
$MOM_{i;t}$	-1.171	1.842	2.742	-0.843	4.457	12.514
	(-0.847)	(0.502)	(0.37)	(-0.55)	(0.984)	(1.538)
$ST_{i;t}$	-0.354	-5.154	-11.613	-1.496	-3.999	-8.967
	(-0.26)	(-1.458)	(-1.645)	(-1.163)	(-1.143)	(-1.45)
Constant	1.538	15.635	36.591	5.221	21.748	51.6
	(0.456)	(1.274)	(1.512)	(1.478)	(1.682)	(2.031)
Daily Average Firms	262	262	262	262	262	262
Adjusted R-squared	0.188	0.171	0.172	0.148	0.134	0.136

Table 11: Firm-Specific News Sentiment Factor Risk Premium-q-factor Model Testing

This table shows daily risk-adjusted returns (a) from a firm-specific news sentiment zero-cost portfolio for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalisation from the last month to split stocks into two portfolio sizes: small and big. Independently, I rank stocks based on day t news sentiment into three sentiment portfolios: pessimistic (N) 30%, neutral (M) 40%, optimistic (P) 30%. The six interacted value-weighted portfolios respecting size and news sentiment: N=S;N=B;M=S;M=B;P=S;P=B sorting on the size and the news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios 30% (P=S;P=B) and the average of short position in the two negative sentiment portfolios 30% (N=S;N=B). Each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. Panel A shows the Pearson correlation between the news sentiment portfolio return and pricing factors from the q-factor model. Panel B presents the risk-adjusted return of the news sentiment zero-cost portfolio from models of q-factor with Pastor and Stambaugh liquidity factor, momentum factor and short- and long- term reversal factors. Newey–West standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

Panel A Correlations Between Different Factors

	$R_{MKT;t}$	$R_{ME;t}$	$R_{IA;t}$	$R_{ROE;t}$	$R_{EG;t}$
$Sentiment_t$	-0.108	0.019	0.069	0.095	0.092
$R_{MKT;t}$		0.148	-0.308	-0.385	-0.36
$R_{ME;t}$			0.019	-0.172	-0.259
$R_{IA;t}$				0.187	0.14
$R_{ROE;t}$					0.577

Panel B Risk-Adjusted Firm-Specific News Sentiment Zero-Cost Portfolio Returns by q-factor Model

	$Sentiment_t$	q-factor	q-factor + PLS	q-factor+UMD	q-factor+Full
a	0.066	0.062	0.062	0.060	0.065
t_a	(6.397)	(6.009)	(6.005)	(5.921)	(6.453)
$R_{MKT;t}$		-0.045	-0.046	-0.030	-0.014
$t_{R_{MKT}}$		(-3.115)	(-3.198)	(-2.274)	(-0.942)
$R_{ME;t}$		0.059	0.059	0.027	0.036
$t_{R_{ME}}$		(2.131)	(2.104)	(1.076)	(1.318)
$R_{IA;t}$		0.053	0.050	0.083	0.077
$t_{R_{IA}}$		(1.437)	(1.385)	(2.412)	(1.974)
$R_{ROE;t}$		0.059	0.058	-0.073	-0.067
$t_{R_{ROE}}$		(1.544)	(1.538)	(-1.895)	(-1.712)
$R_{EG;t}$		0.090	0.093	0.081	0.045
$t_{R_{EG}}$		(2.173)	(2.264)	(2.055)	(1.109)
$PSLIQ_t$			0.014	0.019	0.019
t_{PSLIQ}			(0.849)	(1.132)	(1.187)
UMD_t				0.155	0.154
t_{UMD}				(8.518)	(8.514)
ST_t					-0.087
t_{ST}					(-4.509)
LT_t					-0.043
t_{LT}					(-1.412)
$\overline{R}^2$	0.007	0.018	0.019	0.048	0.057
$Days$	5241	5241	5241	5241	5241

Table 12: Latent Information of Firm-Specific News Sentiment Factor Risk Premium Testing

This table shows daily risk-adjusted returns (a) from a zero-cost portfolio constructed based on $EmotionVsFact_{i;t} * Sentiment_{i;t}$ ($EFSENT_{i;t}$) as a proxy of latent information from news sentiment factors, to capture the potential effects of genuine information or biased valuation about firm fundamentals for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalisation from the last month to split stocks into two portfolio sizes: small and big. Independently, I rank stocks based on $EFSENT_{i;t}$ at day t into three sentiment portfolios: emotional (factual) and optimistic (pessimistic) ($EP(FN)$) 30%, neutral (M) 40%, emotional (factual) and pessimistic (optimistic) ($EN(FP)$) 30%. The six interacted value-weighted portfolios respecting size and latent information from news sentiment: $EP(FN)$=S; $EP(FN)$=B; M=S; M=B; $EN(FP)$=S; $EN(FP)$=B sorting on the size and $EFSENT_{i;t}$ independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the stocks with news either containing the most emotional negative sentiment or the most fundamental positive sentiment 30% ($EN(FP)$=S; $(EN)FP$=B) and the average of short position in the stocks with either the most emotional positive sentiment or the most fundamental negative sentiment 30% ($EP(FN)$=S; $EP(FN)$=B). Each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. Panel A shows the Pearson correlation between the latent information of news sentiment portfolio return and conventional factors. Panel B presents the risk-adjusted return of the latent information of news sentiment zero-cost portfolio from models of CAPM, Fama–French three or five factors with Pastor and Stambaugh liquidity factor, momentum factor and short- and long-term reversal factors. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

Panel A Correlations Between Different Factors

	MKT_t	SMB_t	HML_t	RMW_t	CMA_t	UMD_t	ST_t	LT_t	$PSLIQ_t$
$EFSENT_t$	0.0164	-0.0023	-0.0004	-0.0228	-0.0124	0.0297	0.0386	0.0029	0.0013
MKT_t		0.0702	-0.0119	-0.4246	-0.3331	-0.2572	0.3547	-0.0837	0.0850
SMB_t			0.0523	-0.2985	0.0553	0.0286	0.0138	0.2834	0.0422
HML_t				0.0876	0.4831	-0.3438	-0.0973	0.4771	0.0976
RMW_t					0.2797	0.1508	-0.2453	-0.1613	0.0402
CMA_t						0.0651	-0.2835	0.5202	0.0252
UMD_t							-0.1260	0.0301	-0.0656
ST_t								-0.1380	0.0607
LT_t									-0.0286

Panel B Risk-Adjusted Latent Information of Firm-Specific News Sentiment Zero-Cost Portfolio Returns

	$EFSENT_t$	$CAPM$	$FF3$	$FF5$	$FF5+UMD$	$FF5+Full$
a	0.027	0.026	0.026	0.027	0.026	0.025
t_a	(2.859)	(2.831)	(2.840)	(2.941)	(2.867)	(2.721)
MKT_t		0.009	0.010	0.003	0.008	0.002
t_{MKT}		(0.831)	(0.816)	(0.275)	(0.648)	(0.182)
SMB_t			-0.004	-0.011	-0.015	-0.014
t_{SMB}			(-0.199)	(-0.548)	(-0.789)	(-0.726)
HML_t			0.000	0.005	0.028	0.030
t_{HML}			(-0.001)	(0.259)	(1.218)	(1.306)
RMW_t				-0.030	-0.036	-0.032
t_{RMW}				(-0.969)	(-1.162)	(-0.980)
CMA_t				-0.010	-0.025	-0.014
t_{CMA}				(-0.223)	(-0.548)	(-0.288)
$PSLIQ_t$			0.000	0.001	0.002	0.000
t_{PSLIQ}			(-0.001)	(0.084)	(0.130)	(0.035)
UMD_t					0.034	0.036
t_{UMD}					(2.193)	(2.391)
ST_t						0.029
t_{ST}						(1.749)
LT_t						-0.003
t_{LT}						(-0.100)
$\bar{R}^2$	0.001	0.000	0.000	0.000	0.001	0.002
$Days$	5241	5241	5241	5241	5241	5241

Table 13: Risk-Adjusted Firm-Specific News Sentiment Zero-Cost Portfolio Returns Controlling $EFSENT_{i;t}$.

This table shows daily risk-adjusted returns (a) from firm-specific news zero-cost portfolio for the sample period from 1998 to 2018. At the end of each day, I use NYSE breakpoints of market capitalization from the last month to split stocks into two portfolios sizes: small and big. Independently, I rank stocks based on day t news sentiment into three sentiment portfolios: pessimistic (N) 30%, neutral (M) 40%, optimistic (P) 30%. The six interacted value-weighted portfolios respecting size and news sentiment:N=S;N=B;M=S;M=B;;P=S;P=B sorting on the size and the news sentiment independently. The zero-cost portfolio to be tested is constructed by taking the average of long position in the two positive sentiment portfolios 30% (P=S;P=B) and the average of short position in the two negative sentiment portfolios 30% (N=S;N=B) each day and I calculate the next day $t+1$ value-weighted portfolio returns from this zero-cost trading strategy. By adding an additional pricing factor $EFSENT_{i;t}$ – an invented news factor capturing latent information in news sentiment such as genuine information or biased valuation about firm fundamentals – the table presents the risk-adjusted return of the news sentiment zero-cost portfolio from models of CAPM, Fama–French three or five factors with Pastor and Stambaugh liquidity factor, momentum factor and short- and long-term reversal factors. Newey–West Standard errors are robust to heteroskedasticity and twelve days of autocorrelation. The robust t-statistics are in parentheses.

	$Sentiment_t$	$CAPM$	$FF3$	$FF5$	$FF5+UMD$	$FF5+Full$
a	0.067	0.069	0.069	0.065	0.063	0.067
t_a	(6.567)	(6.758)	(6.929)	(6.554)	(6.331)	(6.808)
$EFSENT_t$	-0.063	-0.062	-0.061	-0.060	-0.066	-0.062
t_{EFSENT}	(-1.649)	(-1.611)	(-1.616)	(-1.601)	(-1.752)	(-1.663)
MKT_t		-0.065	-0.069	-0.031	-0.016	0.001
t_{MKT}		(-4.734)	(-5.272)	(-2.555)	(-1.324)	(0.048)
SMB_t			0.051	0.055	0.040	0.035
t_{SMB}			(1.953)	(2.252)	(1.665)	(1.375)
HML_t			-0.123	-0.202	-0.124	-0.133
t_{HML}			(-4.260)	(-7.380)	(-4.750)	(-4.989)
RMW_t			0.021	0.056	0.036	0.027
t_{RMW}			(1.156)	(1.619)	(1.056)	(0.732)
CMA_t				0.234	0.183	0.146
t_{CMA}				(5.351)	(4.327)	(3.143)
$PSLIQ_t$			0.021	0.018	0.020	0.024
t_{PSLIQ}			(1.151)	(1.035)	(1.181)	(1.436)
UMD_t					0.117	0.112
t_{UMD}					(6.826)	(6.848)
ST_t						-0.086
t_{ST}						(-4.595)
LT_t						0.018
t_{LT}						(0.562)
$\bar{R}^2$	0.011	0.014	0.027	0.038	0.055	0.064
$Days$	5241	5241	5241	5241	5241	5241