The Demand-driven Information Market

Shiyang Huang, Dan Li, and Bohui Zhang *

This Draft: February 2019

* Shiyang Huang is from School of Economics and Finance, The University of Hong Kong; Dan Li and Bohui Zhang are from School of Management and Economics, Shenzhen Finance Institute, and CUHK Business School, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), Longxiang Boulevard, Longgang District, Shenzhen, China, 518172. Authors' contact information: Shiyang Huang: huangsy@hku.hk, (852) 3917 8564; Dan Li: lidan@cuhk.edu.cn; Bohui Zhang: bohuizhang@cuhk.edu.cn, (86)75523518868.

The Demand-driven Information Market

Abstract

We hypothesize a demand-driven information market where information production is tailored by investors' investment constraints. Using a comprehensive data set of news releases and institutional equity holdings during the 2000–2016 period, we show that more negative (positive) news are produced for stocks overweighed (underweighted) by institutions. A natural experiment based on the 2003 mutual funds scandal confirms the negative relation between institutional investment constraints and news sentiment. The effect is more pronounced when the cost of information production is higher, especially when the distance between the information producer and a firm's headquarter is larger. The asymmetry in information production causes stock returns to display negative skewness, increasing the probability for overweighed stocks to experience large negative price movement in the future.

Keywords: News; Institutional investors; Investment constraints; Market efficiency; Skewness

JEL Code: G02; G10; G14

1. Introduction

The interaction between trading and information is the key in financial markets. Despite the extensive literature showing that information determines investors' investment decisions, there is limited evidence of the reverse causality from trading to information provision. Building on Veldkamp (2006 a, b)'s theoretical framework, we characterize the information market by investor demand and investigate whether and how information production is affected by the extent of investor demand.

Information is not a free good and has a high fixed cost of production. Due to the high information production cost, information is only produced when there is high demand for such information in the market. In the context of equity investment, if investors are subject to constraints to increase a stock's holdings, the provision of bad news on the stock seems to be more valuable than good news from these constraint investors' perspective. Correspondingly, such investment constraints would incentivize information intermediaries, such as the media, to produce negative news. As such, an asymmetric pattern in information production would be induced by investment constraints. We call this statement, the *demand-driven information market hypothesis*.

To test our hypothesis, we use a comprehensive corporate news coverage data set collected by RavenPack, along with institutional equity holding data from the Thomson Reuters Institutional Holdings (13F) Database. Our sample covers U.S. stocks listed on the New York Stock Exchange (NYSE), American Stock Exchange (Amex), and National Association of Securities Dealers Automated Quotations (Nasdaq) over the 17-year period between the first quarter of 2000 and the fourth quarter of 2016. News articles from RavenPack are commonly used by institutional and sophisticated individual investors. RavenPack quantifies the positive (or negative) information (i.e., news sentiment score) in each news article based on professional algorithms. For example, a news article on a corruption scandal involving a firm's executives is associated with a low news sentiment score, and a news article regarding the successful development of a firm's new product is associated with a high news sentiment score.

Our main analysis is conducted at a quarterly frequency. We first calculate the firm's both the number of negative news divided by the total number of news and the average news sentiment scores over a quarter. Then, we follow Cao, Han, and Wang (2017) to construct two proxies of investment constraints. Specifically, we consider an institution as overweighting stock i if the stock's weight in its portfolio is larger than the corresponding weight in a market capitalization-weighted portfolio, and compute the fraction of institutions that overweight a particular stock. We also calculate the fraction of shares held by institutions. We regress the two variables separately on several firm characteristics to avoid other influences, and obtain the regression residuals as our investment constraint measures.

We find that institutional investors are indeed subject to investment constraints. When institutions overweight a stock in their portfolios, they tend to sell the stock in the subsequent quarter, and vice versa. More importantly, such investment constraints faced by institutional investors lead them to pay more attention to negative news, and hence the media strategically cater to institutional preference and produce the coverage on negative stories. The effect is not only statistically significant but also economically relevant. For example, a one-standard-deviation increase in the overweight ratio is associated with a 2.61% (6.74%) standard deviation higher (lower) level of fraction of negative news (news sentiment score).

Endogeneity is an important consideration in our empirical tests because there might exist some unobservable firm characteristics that affect both institutional ownership and news sentiment. For example, Core, Guay, and Larcker (2008) show that negative press coverage is more severe among CEOs who have exercised more options, while Hartzell and Starks (2003) documents a positive relationship between institutional ownership concentration and the pay-for-performance sensitivity of executive compensation.

To substantiate our baseline findings, we implement an identification strategy based on the 2003 mutual fund scandal to mitigate the issue of endogeneity. On September 3, 2003, New York State Attorney General issued a complaint against a hedge fund, Canary Capital Partners, for engaging in illegal trading behaviors including extensive market timing and late trading with several mutual funds. The scandal triggered a massive outflow from funds of implicated fund families while funds not implicated benefited from this scandal and experienced a capital inflow. For example, Kisin (2011) estimates implicated families all together lost about 14.1% of their capital within one year or two. Ideally, the capital outflow and inflow arising from the scandal should result in an exogenous change in the overweight ratio and residual institutional ownership. We document a consistently negative relation between investor constraints and news sentiment.

In the final part of our study, we implement a series of additional tests to further enrich our main findings. First, we conduct the cross-sectional analysis to strengthen our argument that the media produce information to meet investors' demand. Specifically, we focus on the information production cost, which is proxied by the geographic distance between a firm's headquarters location and Dow Jones' eight offices. Intuitively, compared to distant firms, it is more convenient for media reporters to visit and collect information from firms nearby. We show that investment constraints have a stronger effect on news production when the information cost is higher.

Second, we carry out a placebo test using news production around earnings announcements. Presumably, when the information production cost for a type of news is close to zero, information production is less sensitive to the demand of this type of news. One obvious example about this kind of news is firms' earnings announcements, which could be reproduced by the media easily. Therefore, we expect that institutions' investment constraints should not generate an asymmetric media coverage of earnings announcements for good and bad news. Indeed, we show that the media is indifferent in between reporting positive and negative earnings news, which contrasts our main results that the demand side affects the media's incentives to produce and report positive or negative news.

Third, we explore the asset pricing implication from the demand-driven information production. Given the relation between investment constraints and the asymmetric pattern of information provision, the natural asset pricing implication is that investment constraints are associated with asymmetric patterns in stock returns. We follow Chen, Hong, and Stein (2001) and construct three return asymmetry measures. Stocks with high investment constraints are significantly associated with negative stock return skewness or tend to experience large negative price movements. The effect is economically significant as well. For example, a one-standard-deviation increase in investment constraints measured by the overweight ratio is associated with a 43.4% decrease in the stock return skewness.

Four, we rule out the fundamental explanation for our empirical findings. One tends to argue that investment constraints should be negatively associated with firm fundamentals, which could be the origin of negative media news. To address this possibility, we examine the association between investment constraints and the subsequent firm's fundamental performance. Strikingly, we find that high investment constraints are associated with firms' high fundamental performance. Last, we show the robustness of our baseline findings. For example, we use several alternative methods to define if one institution overweighs a stock in the portfolio, and also conduct our analysis for subsamples. Our results remain unchanged.

Our paper contributes to three strands of the literature. First, we add to the literature on information production. By highlighting the information market from a demand perspective, we show that investors' investment constraints affect their demand for specific information and thus have asymmetric effects on information production. Second, the negative asymmetry in market returns has long been an important puzzle. Our study contributes to this literature by suggesting that the demand-driven information production could also generate asymmetric stock returns. The third strand of literature is about the price impacts of institutional trading in financial markets. In addition to the direct trading impact, we document a price impact by institutions from their trading potential.

The remainder of the paper is organized as follows. We explain the sample construction for the news and investor constraint variables and describe the sample characteristics in Section 2. In Section 3, we examine the relation between investor constraints and news sentiment. In Section 4, we exploit an identification strategy based on the 2003 mutual fund scandal to establish causality. In Section 5, we carry out further studies. Finally, we provide concluding remarks in Section 6.

2. Literature review

Our study is closed related to three strands of literature. The first strand of literature is about information market or information production. While the theoretical work includes Veldkamp (2006a, b), Veldkamp and Wolfers (2007), Huang, Xiong and Yang (2018a, b), and Dugast and Foucault (2018), the empirical work includes Hameed, Morck, Shen and Yeung (2015) and Kadach and Schain (2016). The most related paper is Veldkamp (2006a). Veldkamp (2006a) models that profit-maximizing information intermediaries face a fixed information production cost and sell information to investors. As information is costly to discover but cheap to replicate, investor demand plays an important role in information production decisions. Based on Veldkamp (2006a), Hameed, Morck, Shen and Yeung (2015) argues that investor demand is higher for information about firms whose fundamentals help price not only their own stocks but also the stocks of related. Based on this argument, Hameed, Morck, Shen and Yeung (2015) find that analysts follow disproportionally firms whose fundamentals correlated more with their industry peers. Our study complements this strand literature by showing that investment constraints affect investor demand of information and have asymmetric effects on information production.

The second strand of literature is about asymmetric patterns in stock returns. The negative asymmetry in market returns has long been an important puzzle (Bates,1997; Bakshi et al., 1997; and Dumas et al., 1998.). There are some potential theories for this puzzle. One plausible theory is the leverage effect (Black, 1976; Christie, 1982), where a drop in prices increase operating and financial leverage and further increase the volatilities in stock returns. An alternative theory is the volatility feedback mechanism (Pindyck, 1984; French et al., 1987; Campbell and Hentschel, 1992). The recent theory is based short-sale constraints and disagreement (Chen, Hong and Stein, 2001; Hong and Stein, 2003). Our study contributes to this literature by suggesting that demand-based information production could also generate asymmetric stock returns.

The third strand of literature is about the price impacts of institutions in financial markets. It includes benchmarking or indexing (e,g., Lakonishok, Shleifer, and Vishny, 1997; Chan, Chen, and Lakonishok, 2002; Cohen, Gompers, and Vuolteenaho, 2002; Cremers and Petajisto, 2009), and Lewellen, 2011), fund flow (Lou, 2012; Huang, Xiang and Song, 2018), and investment constraints (Cao, Han and Wang, 2017). The most related paper is Cao, Han and Wang (2017), which show that investment constraints lead to price underreaction to news and stock return predictability. Although we also study investment constraints (e.g., overweight ratio) as Cao, Han and Wang (2017), our focus is fundamentally different from theirs. Specifically, we show investment constraints affect information production and then generate asymmetric patterns in stock returns.

3. Data and variable construction

This section describes the construction of sample, variables and methodologies.

3.1. Data and sample construction

Our sample construction starts with the U.S. common stocks listed on the New York Stock Exchange (NYSE), American Stock Exchange (Amex) and National Association of Securities Dealers Automated Quotations (NASDAQ). Stock returns and accounting data are obtained from the CRSP/Compustat merged database. Stocks with price below $5 and stocks in the lowest market capitalization decile based on NYSE breakpoints as of the end of the calendar year are excluded. Data on institutional equity holding come from the Thomson Reuters Institutional Holdings (13F) Database.

News data are obtained from the RavenPack News Analytics database. RavenPack is the leading data analytics provider that supplies real-time news analytics based on traditional media news, firms' press release, and social media feeds. For our analysis, we rely on its Dow Jones Edition, which consolidates relevant information from Down Jones Newswires, regional editions of the Wall Street Journal, Barron's and MarketWatch[1]. It includes more than 5000 employees around the world, and includes more than 2000 journalists in 58 countries. Press releases are removed since they don't constitute information production by journalists. For each news article, RavenPack utilizes its proprietary algorithm to determine its novelty, relevance and sentiment. Specifically, the algorithm first identifies the list of companies mentioned in the article. For each of the firms, it assigns a novelty scores based on how new or novel a news story is, a relevance score to indicate how strongly related a firm is to the news story, and a sentiment score that reflects the potential market impact of a news article. All three scores range between 0 and 100. For each of the firms identified, RevenPack's algorithm determines whether the particular article is the first news story in the sequence of similar events (novelty) and assign a novelty score between 0 and 100. A novelty score equals to 100 suggests a new story, and subsequent articles covering the same story are given a lower score. For relevance score, a higher value indicates greater relevance. A sentiment score of 50 indicates neutral sentiment, and values above (below) 50 suggest positive (negative) sentiment. We will discover it in more detail in the subsequent section.

The historical archive of the RavenPack database dates back to January 2000. As a result, our sample period ranges from the first quarter of 2000 through the fourth quarter of 2016. In order to have non-missing measures on news sentiment, we only consider firms with at least one article covered by RaevnPack throughout our sample period. Our final sample includes 130,504 firm-quarter observations.

[1] In addition to the Dow Jones Edition, Ravenpack also provides a Web Edition, a PR Edition and a Full Edition which is composed of all other Editions. The Web Edition contains articles from industry and business publishers, national and local news, blog sites, government and regulatory updates, starting from 2007. The PR Edition includes press releases and regulatory disclosures from 2004. We focus on the Dow Jones edition in order to obtain the maximum time period coverage.

3.2. Measures of news sentiment

For each firm and quarter, we construct two measures of news sentiment using RavenPack data. The first measure captures the fraction of bad news (*PctBadNew*), and the second measure is an overall sentiment score (*NewsScore*). We rely on the relevance score (*RELEVANCE*) to filter out unrelated news and utilize the Event Sentiment Score (*ESS*) to calculate the above two measures.

As discussed in the previous section, for each firm identified in a news article, RavenPack assigns a relevance score between 0 and 100 to indicate the role of the firm in the story. A higher score indicates a greater relevance. RavenPack takes into consideration multiple factors including where the firm is mentioned (headline, first paragraph, second paragraph, etc.), the number of times a firm is referenced, and how many firms are mentioned in the news story in order to determine the score. For example, a score of 100 suggests that an entity plays a key role in the article, and a score of 0 means that a firm is only passively mentioned in the story. Usually, a score of 90 or more indicates that a firm is referenced in the headline or main title, while firms referenced further in the story body are given a value below 90. We filter out news items with relevance score less than 100 in order to reduce the noise in the data[2].

The sentiment score intends to measure the potential market impact of a news article on a firm mentioned in the article, which ranges between 0 and 100 with 50 indicating neutral sentiment, values above (below) 50 indicating positive (negative) sentiment. For positive (negative) sentiment, the higher (lower) score, the greater market reaction a news article is expected to induce. RavenPack's algorithm relies on both expert consensus survey and a strength component consisting of a variety of factors to dynamically assign a score. Specifically, RavenPack builds up an extensive database of news stories, which are categorized by financial experts with extensive backgrounds in finance and economics as having positive or negative financial impact and to what degree. The strength component relies on its proprietary natural language processing software that takes into consideration the use of emotionally charged language, and capable of interpreting actual figures, estimates, ratings, revisions, magnitudes, and recommendations disclosed in news stories. In addition to the *ESS*, RavenPack also provides several other measures of sentiment based on alternative methodologies. These sentiment analytic results

[2] Some studies such as von Beschwitz, Keim and Massa (2018) use relevance score equal to 90 to filter the news data.

correlate highly with *ESS*, although they might differ for certain cases. For example, Composite Sentiment Score (*CSS*), a measure that combines traditional tagging, expert consensus and market response, agrees with ESS in about 95% of the cases. This provides comfort that our measure of sentiment score is not sensitive to the underlying classification method.

We standardize the sentiment score by subtracting 50 from *ESS* and scale it by 50, yielding an adjusted score that takes a value between -1 and 1. Our first measure, the fraction of bad news (*PctBadNews*), is calculated as the number of news with negative adjusted sentiment score divided by the total number of news. A higher value suggests that a firm has more bad news in a particular quarter. We take an average of the adjusted sentiment scores across all relevant news within a quarter to compute our second measure *NewsScore*, with -1 (1) indicating the most negative (positive) sentiment and 0 indicating neutral sentiment.

3.3. Measures of investment constraints

We follow Cao, Han and Wang (2017) to construct two proxies of investment constraints. Unlike individual investors, institutional investors often face a variety of constrains that limit their ability to invest and their positons in certain stocks, due to a combination of regulatory provisions, contractual arrangements, and investment strategies. There are two types of constraints that are particularly important. The first is related to the diversification requirements. For example, Mutual funds are required to meet various diversification requirements in order to be able to pass through gains to shareholders and avoid double taxation. Pension funds are required to divest investments in order to minimize the risk of large losses. Failing to comply with the diversification requirements runs the risk of civil lawsuits. The second constraint concerns tracking errors, which measures the divergence between a portfolio's performance and its benchmark's performance. Greater tracking errors could lead to termination of contracts or even financial penalties. These constraints make it difficult for institutional investors to deviate from their benchmarks. Moreover, for institutions that already overweight (underweight) a stock, this consideration makes them reluctant to add to (reduce) positions in the stock even when there is good (bad) news about the firm.

The first measure is based on each individual institution's holding. Assume an institution's portfolio comprises N_i shares of stocks, i=1 to m. Stock i's price is P_i, and market capitalization is M_i. An institution is considered to overweight stock *i* if the stock's weight in its portfolio is larger than the corresponding weight in a market capitalization-weighted portfolio:

$$\frac{P_i * N_i}{\sum_{j=1}^{k} P_j * N_j} > \frac{M_i}{\sum_{j=1}^{k} M_j} .$$

This allows us to compute the fraction of institutions that overweight a particular stock. In order to control for firm characteristics that might be related to holdings, for each quarter, we regress the above ratio on size, book-to-market ratio, stock returns over the previous 12 months and a dummy for the S&P500 index constituent. The residual from the regression is our first measure overweight ratio (*OR*). It captures the percentage of institutions that overweight a stock.

Our second measure is based on aggregate institutional ownership (*IO*), defined as the fraction of shares held by institutions. We regress *IO* on the same groups of controls to obtain the residual institutional ownership (*ResIO*) as our second measure.

3.4. Measures of control variables

We construct a list of controls that might be related to news sentiment. *Size* is the market capitalization calculated as the number of shares outstanding times stock price at the end of each quarter. *B/M* is the ratio of book value over market value at quarter end. *ROA* is the return on asset. *S&P500* is a dummy equal to 1 for the S&P500 index constituent.

Table 1 Panel A reports the summary statistics on the list of variables for the full sample. A typical firm has a *NewsScore* of 0.078, ranging between -0.257 and 0.420, consistent with the literature that media coverage is in general positive. There is also substantial variation in *PctBadnews*: the mean is 0.279, with the minimum equal to 0 and the maximum equal to 80%. For an average firm, 66.8% of the shares are held by institutional investors. Institutions tend to overweight the stocks in our sample, with the mean residual overweight ratios *OR* equal to 0.003 and mean *ResIO* equal to 0.006. The mean book-to-market ratio is 0.593, and the mean quarterly return on asset is 2.8%. Panel B reports the Pearson correlation matrix. *OR* and *ResIO* are positively correlated. The correlation coefficient is 0.661, suggesting the two measures capture different aspect of investment constraints. Our two news measures *NewsScore* and *PctBadnews* are negatively correlated, with the Pearson correlation coefficient equal to -0.771. We will employ both measures in the following empirical analysis to ensure the robustness of our results.

[Insert Table 1 About Here]

4. Main results: news sentiment and investment constraints

In this section, we test whether institutional investment constraints lead to asymmetry in information production by media. Investment constraints limit institutions' ability to keeping buying more stocks when they already overweight a stock, which leads them to pay more attention to negative news. Media strategically caters to institutional preference and covers more negative stories as a result. In this section, we test our hypotheses by performing the following analysis. First, we confirm institutional investors are indeed subject to investment constraints. Second, we examine the demand side effect on news production by focusing on the relation between investment constraints and the probability of bad news.

4.1. Investment constraints and changes in institutional ownership

We first examine the effect of investment constraints on changes in institutional holdings. For each quarter, we assign each stock to one of the five groups based on the quantile cutoffs of *OR* or *ResIO*, which I refer to as *OR* or *ResIO* quantiles. We then calculate the change in *OR* or *ResIO* for the following quarter. Table 2 presents the time-series average change for each of the quintile groups. The change in overweight ratio or residual institutional holding decreases monotonically, consistent with the findings in Cao, Han and Wang (2017). The difference in *ResIO* between stocks in the top and bottom *OR* (*ResIO*) quantile is 1.47% (3.31%), significant at 1% level. Moreover, both overweight ratio and residual institutional ownership exhibit pattern of mean reversion. If institutions overweight a stock in their portfolio, they tend to sell the stock in the subsequent quarter, and vice versa. These results provide strong support to our argument that institutions don't deviate largely from their benchmarks.

[Insert Table 2 About Here]

4.2. News sentiment and investment constraints

In this session, we examine the impact of investment constraints on news sentiment by estimating the following regression models:

$$NewsScore_{i,t+1}\left(PctBadNews_{i,t+1}\right) = a + b * OR_{i,t}\left(ResIO_{i,t}\right) + c * X_{i,t} + e_{i,t}, \quad (1)$$

where *i* indexes firm and *t* indexes time. *NewsScore* and *PctBadnews* are the two news sentiment measures. *OR* and *ResIO* measure investment constraints. We are interested in the coefficient *b* as it captures the effect of investment constraints on news sentiment. Vector X represents a set of control variables, including *Size*, *B/M*, *ROA* and the dummy variable *S&P500.* We control for industry fixed effect and year-quarter fixed effect, and cluster standard errors at the firm level for all tests.

Table 3 reports the results for estimating equation (1). The dependent variables are *NewsScore* in columns (1) and (3) and *PctBadnews* in columns (2) and (4), respectively. The key independent variables are *OR* in columns (1) and (2) and *ResIO* in Columns (3) and (4), respectively. The coefficients on *OR* in columns (1) and (3) are negative and significant at 1% level, and the coefficients on *OR* in columns (2) and (4) are positive and significant at 1% level, suggesting higher overweight ratio is associated with more bad news and lower sentiment score. Specifically, a one-standard-deviation increase in investment constraints measured by *OR* is associated with 2.61% (6.74%) standard deviation higher (lower) level of *NewsScore* (*PctBadnews*), suggesting that the impact of investment constraints is also economically significant. Columns (3) to (4) show similar results using *ResIO* as the dependent variable.

[Insert Table 3 About Here]

4.3. Mutual fund Scandal: Instrumental Variable Regression

Although the results in the previous section suggests a negative relation between investment constraints and news sentiment, it, however, can’t rule out the possibility there might exist some unobservable firm-specific factors that drives changes in both institutional ownership and news tones. In this section, we exploit an identification strategy based on the 2003 mutual fund scandal to establish causality. On September 3, 2003, New York State Attorney General issued a complaint against a hedge fund, Canary Capital Partners, for engaging in illegal trading behaviors including extensive market timing and late trading with several mutual funds. The scandal kept unfolding. Until the end of the 2006, at least 20 mutual fund families, which together managed 22% of industry assets in late 2003, negotiated a settlement with the Securities and Exchange Commission regarding allegations of abusive trading behavior (McCabe, 2009). The scandal triggered massive outflow from funds of the implicated fund families. For example, investors pulled $4.4 billion from Putnam Investments in the week ending November 5, 2003.

Kisin (2011) estimates implicated families all together lost about 14.1% of their capital within one year or two. On the other hand, funds not implicated benefited from it and experienced an increase in capital by nearly 12%. We argue the capital outflow and inflow arising from the scandal results in an exogenous change in overweight ratio and residual institutional ownership, which are unable to firm fundamentals and its news coverage. This setting allows us to draw inference about causal connections between investment constraints and news sentiment.

Specifically, we collect data on the implicated fund families from Stanford Law School Securities Class Action Clearinghouse, and follow Anton and Polk (2014) to estimate a 2SLS instrumental variable regression using observations from 3 years prior to the scandal (July 2000 to June 2003) and 3 years after the end of the scandal (January 2007 to December 2010). For this test, we require stocks to be covered by both the Thomson-Reuters Mutual Fund Holdings (13F) database and the Thomson-Reuters Institutional Holdings database as of the third quarter of 2003 to ensure consistency.

In the first stage, we regress the measure of investment constraints (*OR* or *ResIO*) on the instrumental variable $RATIO_{200309}$, which is the number of implicated funds that own the stock divided by the total number of institutional owners as of September 2003, and the same list of controls. We control for industry fixed effect and year-quarter fixed effect, and cluster standard errors at the firm level for all tests. The coefficient on $RATIO_{200309}$ on both regressions are negative and significant at 1% level. In the second stage, we re-estimate the regression models in Equation (1) by replacing the measure of investment constraints (*OR* or *ResIO*) with the predicted value from the first stage ($\widehat{OR}$ and $\widehat{ResIO}$). Table 4 presents the results of both the first-stage and second-stage regressions. The coefficient on $\widehat{OR}$ is still negative and significant at 1% level. Similar patterns can be found in the regressions with $\widehat{ResIO}$ as the regressor, although the significance level drops compared to the main results. These findings support our hypothesis there exists a causal effect of investment constraints on news tone.

[Insert Table 4 About Here]

5. Additional tests

We carry out further studies in this section. First, we do cross sectional studies to strengthen our argument that media strategically produce information to meet investors' demand. Specifically, we focus on the information production cost, which is proxied by the geographic distance between firm's headquarter location and Dow Jones offices. Second, we carry out a placebo test using the earnings announcements. Third, we discuss some asset pricing implications from this demand-based information production. For the asset pricing implications, we focus on the asymmetric patterns of the stock returns. Last, we discuss some alternative stories for our empirical findings.

5.1. Cross-sectional studies: the role of firm's proximity to Dow Jones Offices

As argued in the information production literature (e.g., Veldkamp, 2006a, b; Veldkamp and Wolfers, 2007), the information cost plays a key role. Specifically, because information discovery has a high fixed cost, whether or how the information sellers produce information crucially depends on the demand side (e.g., institutional investors in our main analysis). For example, when the information cost is higher, the information sellers care more about the demand side, which largely determines whether the information cost could be covered. Based on this intuition, we carry out cross-sectional studies with the information cost to strengthen our main argument.

To measure the information cost, we follow some recent studies (e.g., Da, Gurun, Li and Warachka, 2018; Bernstein, Giroud and Townsend, 2016) and use the firm's proximity to Dow Jones offices. Intuitively, compared to the distant firms, it is more convenient for the media reporters to visit and collect information from firms nearby. To calculate the distance between firms' headquarters and Dow Jones offices, we obtain firm's headquarter location from Compustat and the location information (street-level) for the eight Dow Jones offices from the Dow Jones Official website. Because Dow Jones have eight offices, we focus on the minimum distance from a firm's headquarter to one of the eight Dow Jones offices. We define a dummy variable, D, to indicate whether firms are close to Down Jones offices. Specifically, D equals to 1 if the distance is smaller than p35 percentiles of the whole sample, otherwise it equals to 0. To examine the relation between firm's proximity to Dow Jones offices and news sentiment, we estimate the following regression models:

$$NewsScore_{i,t+1}(PctBadNews_{i,t+1})$$
$$= a + b * OR_{i,t}(ResIO_{i,t}) + c * D_{i,t} + d * OR_{i,t}(ResIO_{i,t}) * D_{i,t} + f * X_{i,t} + e_{i,t}.$$

We are interested in the interaction between investment constraints and the distance as it captures the effect of the information cost. Table 5 reports the results for above estimating equations. As shown in Tale 5, while investment constraints negatively predict news score in columns (1) and (3), the coefficients on the interaction term between investment constraints and geographic distance are significantly positive. Meanwhile, while investment constraints positively predict the percentage of bad news in columns (2) and (4), the coefficients on the interaction term between investment constraints and geographic distance are significantly negatively. The effects of geographic distances are also economically significant. Specifically, for distant firms, a one-standard-deviation increase in investment constraints measured by OR is associated with 4.1% (7.9%) standard deviation higher (lower) level of NewsScore (PctBadnews). In contrast, for firms close to Dow Jones offices, a one-standard-deviation increase in investment constraints measured by OR is associated with 1.7% (1.9%) standard deviation higher (lower) level of NewsScore (PctBadnews). Briefly, the results in Table 5 suggest that investment constraints have stronger effects on news production when the information costs are higher.

[Insert Table 5 About Here]

5.2 Placebo test

Following the aforementioned argument that the information cost plays an important role, we carry out one placebo test in this section. Specifically, we focus on earnings announcements in the placebo test. Theoretically, when the information cost is close to zero, the information production does not depend on the demand side. This is corresponding to the scenarios when the information producers just reprint or reproduce the existing news. One obvious example about this type of news is firms' earnings announcements, which could be reproduced by media easily. Therefore, we expect that the institutions' investment constraints could not generate asymmetric media coverage of earnings announcements of good and bad news.

To carry out this placebo test, we estimate the following regression models:

$$ND\_EA_{i,t+1} = a + b * OR_{i,t}\left(ResIO_{i,t}\right) + c * Bad_{i,t} + d * OR_{i,t}\left(ResIO_{i,t}\right) * Bad_{i,t} + f * X_{i,t} + e_{i,t},$$

where $Bad_{i,t}$ is a dummy variable to indicate whether firm i announces a negative earnings news at quarter t. In these regressions, the dependent variable, $\text{ND\_EA}_{i,t+1}$, is the logarithm of the total number of earnings announcement news in quarter t+1, constructed by counting the new articles on RavenPack that are specific to each quarterly earnings announcement. By construction, $\mathrm{ND\_EA}_{i,t+1}$ captures the media coverage of earnings announcements. We are interested in the coefficient d as it captures the asymmetric effect of investment constraints on media coverage of positive and negative earnings news. As we argue before, since it costs media little to quote or reprint earnings announcements, the media would not selectively report positive or negative earnings news. Thus, we expect the coefficient d to be insignificant.

To identify positive and negative earnings news, we follow the literature (e.g., Jegadeesh and Livant, 2006) and construct two measures of standardized earnings surprise (SUE). More specifically, the first measure of SUE is actual earnings minus expected earnings, after excluding "special items" from Compustat data, scaled by stock price. The second measure of SUE as actual earnings minus expected earnings, scaled by the standard deviation of quarterly earnings growth. Expected earnings are estimated using a seasonal random walk with drift model. After calculating SUE, the dummy variable $Bad_{i,t}$ equals to 1 for earnings announcements with negative standardized earnings surprise (SUEt-1<0), and 0 otherwise.

Table 6 reports the results on this placebo test. While investment constraints are positively associated with media coverages in all specification, the coefficients of the interaction term between investment constraints and the indicator of negative news are not statistically significant. This suggests that the media is indifferent between reporting positive and negative earnings news, which contrasts our main results that the demand side affects the media's incentives to produce and report positive or negative news.

[Insert Table 6 About Here]

5.3 Asset pricing implications: stock return asymmetric

The previous analysis shows that investors' investment constraints play an important role in shaping information production, particularly on the asymmetric pattern in information provision (positive vs. negative). For example, when a large population investors have already overweighed some stocks in their portfolio (proxied by a high overweight ratio), the media selectively chooses to produce/provide negative information to cover the production costs. Given the relation between investment constraints and asymmetric pattern of information provision, the natural asset pricing implication is that investment constraints are associated with asymmetric patterns in stock returns.

To examine the relation between investment constraints and the asymmetric patterns of stock returns, we follow Chen, Hong and Stein (2001) and construct three firm-quarter level measures of return asymmetry: Skewness, NCSKEW and DUVOL. Skewness is the total skewness, which is the skewness of daily log returns in one specific quarter; NCSKEW is calculated by taking the negative of the third moment of daily market-adjusted log returns, and dividing it by the standard deviation of daily market-adjusted log returns raised to the third power in one specific quarter; DUVOL is the log of the ratio of down-day to up-day standard deviation, measured using (daily) market-adjusted log returns in one specific quarter. After that, we estimate the following regression models:

$$Asymmetric\ Measure_{i,t+1} = a + b * OR_{i,t}\left(ResIO_{i,t}\right) + c * X_{i,t} + e_{i,t},$$

where $Asymmetric\ Measure$ includes the three measures: Skewness, NCSKEW and DUVOL. Following Chen, Hong and Stein (2001), the control variables include the lag asymmetric measures, the book-to-market ratio in the previous quarter, stock return volatility in the previous quarter, the previous three quarters' market-adjusted cumulative return, and the return on assets (ROA) in quarter t-1, and the S&P500 membership. We also include year-quarter fixed effect and cluster the standard error at the firm level.

Table 7 reports the results. In all model specifications, we find that investments constraints negatively forecast the stock return skewness. Specifically, stocks with high investment constraints are significantly associated with negative stock return skewness or tend to experience large negative price movements. The effect is economically significant as well. For example, a one-standard-deviation increase in investment constraint measured by *OR* is associated with 43.4% decreases in the stock return skewness. For comparison, the skewness ranges from -61.3% at the 25^{th} percentile to 43.2% at the 75^{th} percentile.

[Insert Table 7 About Here]

We also carry out some further studies to show that our results are robust and are indeed driven by the asymmetric patterns of media news. First, Table A3 in Internet Appendix shows that our results are robust to Fama-Macbeth regressions. Second, Table A4 in Internet Appendix shows that market indeed react strongly to the media news announcements, which suggests that the effect of investment constraints on stock return asymmetric is driven by its effect on asymmetric patterns of media news reports.[3]

Our results on stock return asymmetry has important asset pricing implications. It is well-know and puzzling that aggregate stock market returns are asymmetrically distributed. More importantly, the stock market is more prone to melt down than to melt up. For example, nine of the ten biggest one-day movements in the S&P 500 since 1947 were declines. Or, a large literature documents that market returns exhibit negative skewness, or a closely related property, ‘‘asymmetric volatility’’ – a tendency for volatility to go up with negative returns (e.g., Bates, 1997; Bakshi et al., 1997; Dumas et al., 1998; Chen, Hong and Stein, 2001). Some recent work, including short-sale constraints and disagreement (Chen, Hong and Stein, 2001; Hong and Stein, 2003), volatility feedback mechanism (Pindyck, 1984; French et al., 1987; Campbell and Hentschel, 1992), our study complements to the literature by proposing an alternative channel for the puzzling asymmetric return patterns.

5.4. Firm fundamentals

While the previous empirical findings, taken at face value, are consistent with the argument that the information providers cater to the demand side in order to cover the fixed information production cost, there is one alternative way to think about the evidence. Specifically, the effect of investment constraints on asymmetric patterns of media news could be driven by investors' constraining trading on negative news (e.g., some other investors over react to negative news ex ante) or investors' mis-interpreting negative information. If this is the case, investment constraints should be negatively associated with firm fundamentals, which could be the origin of negative media news. To address this possibility, we examine the association between investment constraints and the subsequent firm's fundamental performance. Strikingly, we find that high investment constraints are associated with high fundamental performance, measured by ROA or ROE (see Table 8). In untabulated results, we also find that high investment constraints are associated with positive stock returns in the subsequent quarters.[4] In a summary, these results are inconsistent with the alternative channel that investors implement constraining trading on

[3] In untabulated results, there are no significant association between investment constraints and the asymmetric patterns of stock returns after excluding the announcements of media news.
[4] Cao, Han and Wang (2017) also find similar return predictions of investment constrains.

negative news (e.g., some other investors over react to negative news ex ante) or investors mis-interpret negative information.

[Insert Table 8 About Here]

5.5. Robustness

In this section, we consider several robustness checks. First, we use alternative methods to define if one institution overweighs a stock in the portfolio. The method in our main analysis follows Cao, Han and Wang (2017) and defines that the institution overweighs a stock if its weight in the portfolio is higher than that corresponding weight in the value-weighted portfolio. In our robustness check, we consider alternative weights in the value-weighted portfolio to identify if the institution overweighs a stock. Second, we carry out sub-sample analysis. Specifically, we split our sample into three periods: 2000-2005, 2005-2009, and 2010-2016.

Table 9 reports the robustness results. While Panel A reports the results for alternative definitions of overweigh, Panel B reports the results for sub-sample analysis. For the alternative definitions, we adopt alternative weights in the value-weighted portfolio to definite overweighting. In our main analysis, if the weight of stock *i* in one institution's portfolio is higher than the weight of stock *i* in the same market portfolio, then this institution overweights stock *i*. In Panel A1 – A4, we multiply 1.05, 1.1, 1.2 and 1.5 with the weight of stock *i* in the same market portfolio as the cutoffs to identify overweighting, respectively, and apply the same requirement to calculate overweight ratio as the main analysis. It is clear that both overweight ratio (OR) and residual institutional ownership (ResIO) still negatively predict news scores or the percentage of bad news. Meanwhile, Panel B1- B3 show that our results are quantitatively and qualitatively similar for different sample periods. Overall, Table 9 confirms the robustness of our main findings.

[Insert Table 9 About Here]

6. Conclusion

We study whether market friction arising from institutional investment constraints induces an asymmetric pattern in information production. Institutions are subject to a variety of constraints as a combination of law, contractual arrangement and investment strategy, resulting in a mean revision pattern in their stock holdings. Institutions are reluctant to keep buying a stock even when there is good news, if the stock's weight in their portfolios is already higher than a given benchmark. Such constraints lead institutional investors to be more attentive to negative news for stocks overweighed in their holdings, which incentivizes information intermediary, such as media especially financial press, to focus more on negative coverage.

We find strong and consistent evidence that media caters to institutional investors by producing more negative news for stocks overweighed by institutions using data from RavenPack. Further test suggests the negative relation between institutional investment constraints and news sentiment is not due to worsening fundamentals. Using the 2003 mutual funds scandal as a natural experiment, we confirm the relation is causal. The effect is more pronounced when the cost of information production is higher, especially when the distance between the information producer and a firm's headquarter is larger. The asymmetry in information production causes stock returns to display negative skewness, increasing the probability for overweighed stocks to experience large negative price movement in the future.

References

Anton, M., Polk, C., 2014. Connected stocks. Journal of Finance 69, 1099-1127.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

Bates, D.S., 1997. Post-87 Cash Fears in S & P 500 Futures Options. NBER working paper 5894.

Bernstein, S., Giroud, X., Townsend, R.R., 2016. The impact of venture capital monitoring. Journal of Finance 71, 1591-1622.

Black, F., 1976. Studies of stock price volatility changes. Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section, 177–181.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Cao, J., Han, B., Wang, Q., 2017. Institutional investment constraints and stock prices. Journal of Financial and Quantitative Analysis 52, 465-489.

Chan, L.K., Chen, H.L., Lakonishok, J., 2002. On mutual fund investment styles. Review of Financial Studies 15, 1407-1437.

Chen, J., Hong, H., Stein, J.C., 2001. Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. Journal of financial Economics 61, 345-381.

Christie, A.A., 1982. The stochastic behavior of common stock variances – value, leverage and interest rate effects. Journal of Financial Economics 10, 407–432.

Cohen, R.B., Gompers, P.A., Vuolteenaho, T., 2002. Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. Journal of Financial Economics 66, 409-462.

Cremers, K.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Review of Financial Studies 22, 3329-3365.

Da, Z., Gurun, U.G., Li, B., Warachka, M., 2018. Investment in a Smaller World: The Implications of Air Travel for Investors and Firms. Working paper

Dugast, J., Foucault, T., 2018. Data abundance and asset price informativeness. Journal of Financial Economics 130, 367-391.

Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. Journal of Finance 53, 2059–2106.

French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Hameed, A., Morck, R., Shen, J., Yeung, B., 2015. Information, analysts, and stock return comovement. Review of Financial Studies, 28, 3153-3187.

Hong, H., Stein, J.C., 2003. Differences of opinion, short-sales constraints, and market crashes. Review of Financial Studies 16, 487-525.

Huang, S., Yang, L., Xiong, Y., 2018a. Clientele, Information Sales, and Asset Prices. Information Sales, and Asset Prices. Working paper.

Huang, S., Yang, L., Xiong. Y., 2018b, Skill Acquisition and Information Sales, the University of Toronto, Working paper.

Huang, S., Song, Y., Xiang, H., 2019. Fragile Factor Premia. Working paper.

Jegadeesh, N., Livnat, J., 2006. Revenue surprises and stock returns. Journal of Accounting and Economics 41, 147-171.

Kadach, I., Schain, K., 2016. The Effect of Institutional Ownership on Analyst Coverage: An Instrumental Variable Approach. Working paper.

Kisin, R., 2011. The impact of mutual fund ownership on corporate investment: Evidence from a natural experiment. Working paper.

Koch, A., Ruenzi, S., Starks, L., 2016. Commonality in liquidity: a demand-side explanation. Review of Financial Studies 29, 1943-1974.

Lewellen, J., 2011. Institutional investors and the limits of arbitrage. Journal of Financial Economics 102, 62-80.

La Porta, R., Lakonishok, J., Shleifer, A., Vishny, R., 1997. Good news for value stocks: Further evidence on market efficiency. Journal of finance 52, 859-874.

Lou, D., 2012. A flow-based explanation for return predictability. Review of Financial Studies 25, 3457-3489.

Pindyck, R.S., 1984. Risk, inflation, and the stock market. American Economic Review 74, 334–351.

Veldkamp, L., Wolfers, J., 2007. Aggregate shocks or aggregate information? Costly information and business cycle comovement. Journal of Monetary Economics 54, 37-55.

Veldkamp, L., 2006. Information markets and the comovement of asset prices. Review of Economic Studies 73, 823-845.

Veldkamp, L., 2006. Media frenzies in markets for financial information. American Economic Review 96, 577-601.

von Beschwitz, B., Keim, D.B., Massa, M., 2018. First to "Read" the News: New Analytics and Algorithmic Trading, 1233.

Table 1: Descriptive Statistics

Panel A presents the summary statistics for the main variables used in our analysis. The variables include news scores for each firm/quarter (NewsScore), number of news reported (NobsNews) and ratio of number of bad news reported over number of news reported (PctBadNews), institutional holdings (IO), overweight ratio (OR), residual IO (ResIO), logarithm of market capitalization (Size), book-to-market ratio (B/M), return-on-asset ratio (ROA) and S&P500 membership dummy (S&P500). The table reports the number of observations (N), mean, median, standard deviation (STD), minimum (Min), maximum (Max), quartile (25% and 75%), the bottom/top 5% (5% and 95%) distribution, and the bottom/top 1% (1% and 99%) distribution of the variables. The sample period is from first quarter of 2000 through the fourth quarter of 2016. Panel B reports the Pearson correlation matrix for the variables. All the continuous variables are winsorized at 1% and 99% to reduce the effect of outliers on the analysis.

Panel A. Statistics Summary

Variable	N	Mean	Std	P5	P25	Median	P75	P95
NewsScore	130504	0.078	0.125	-0.123	0.002	0.074	0.150	0.301
PctBadnews	130504	0.279	0.179	0.000	0.153	0.266	0.392	0.600
IO	130504	0.668	0.238	0.191	0.530	0.714	0.848	1.000
OR	130504	0.003	0.099	-0.190	-0.050	0.018	0.070	0.139
ResIO	130504	0.006	0.201	-0.385	-0.120	0.035	0.156	0.282
Size	130504	7.34	1.49	5.21	6.25	7.13	8.23	10.15
B/M	130504	0.593	0.447	0.116	0.295	0.488	0.759	1.409
ROA	130504	0.028	0.035	-0.023	0.010	0.028	0.045	0.081
S&P500	130504	0.214	0.410	0.000	0.000	0.000	0.000	1.000

Panel B. Pearson correlation matrix.

Variable	1	2	3	4	5	6	7	8	9
1.NewsScore	1								
2.PctBadNews	-0.771	1							
3.IO	-0.086	0.072	1						
4.OR	0.004	0.010	0.566	1					
5.ResIO	-0.032	0.041	0.849	0.661	1				
6.Size	0.021	0.046	0.188	-0.002	-0.010	1			
7.B/M	-0.119	0.068	-0.045	-0.015	-0.015	-0.247	1		
8.ROA	0.174	-0.131	0.193	0.212	0.148	0.224	-0.206	1	
9.S&P500	0.008	0.040	0.083	-0.002	-0.010	0.716	-0.095	0.112	1

Table 2: Investment Constraints and Institutional Trading

This table reports the time-series mean of the change in OR or ResIO in quarter t for portfolios sorted on OR or ResIO in quarter t-1, respectively.institutional trading activity for stocks sorted on overweight ratio (OR) and residual institutional ownership (ResIO). At the end of each quarter, we sort the stocks into five groups based on OR and ResIO respectively. We then calculate the average change in institutional ownership (IO), average change in OR, and average change of ResIO during the next quarter following the measurement of investment constraints. The column "D (L, H)" reports the differences of the five trading measures between the Low OR (ResIO) stocks and High OR (ResIO) stocks. The sample period is from first quarter of 2000 through the fourth quarter of 2016.

	Sorted on OR in quarter t-1		Sorted on ResIO in quarter t-1	
	Change in OR	Change in ResIO	Change in OR	Change in ResIO
Low	1.066	0.763	0.276	1.372
p2	0.518	0.282	0.119	0.682
p3	0.243	0.066	0.025	0.289
p4	-0.340	-0.264	-0.058	-0.270
High	-1.542	-0.704	-0.415	-1.933
Low - High	2.607***	1.466***	0.691***	3.305***

Table 3. News Sentiment and Investment Constrains

This table examines the relation between overweight ratio (OR) and residual institutional ownership (ResIO) and news score (NewsScore), and ratio of number of bad news reported over number of news reported (PctBadNews). The dependent variables are $\mathrm{NewsScore_t}$ and $\mathrm{PctBadNews_t}$, which represent the average of news score and percentage of bad news for each firm in quarter t, respectively. $\mathrm{OR_{t-1}}/\mathrm{ResIO_{t-1}}$ are overweight ratio and residual institutional ownership in the quarter t-1. $\mathrm{Size_{t-1}}$ is the logarithm of market capitalization in quarter t-1, $\mathrm{(B/M)_{t-1}}$is the book-to-market ratio defined as the ratio of book value of equity to market value of equity in year t-1. $\mathrm{ROA_{t-1}}$ is the return on assets in quarter t-1. $\mathrm{S\&P500}_{t-1}$ equals to 1 if the firm is in S&P 500 in quarter t-1. Standard errors are clustered by firms. All the continuous variables are winsorized at 1% and 99% to reduce the effect of outliers on the analysis. The t-statistics are reported in the parentheses are ***, **, * denote 1%, 5%, and 10% significant levels, respectively. The sample period is from first quarter of 2000 through the fourth quarter of 2016.

DepVar	(1) $\mathrm{NewsScore_t}$	(2) $\mathrm{PctBadNews_t}$	(3) $\mathrm{NewsScore_t}$	(4) $\mathrm{PctBadNews_t}$
$\mathrm{OR_{t-1}}$	-0.033***	0.060***		
	(-4.67)	(5.76)		
$\mathrm{ResIO_{t-1}}$			-0.028***	0.041***
			(-7.89)	(7.81)
$\mathrm{Size_{t-1}}$	0.002***	0.009***	0.003***	0.009***
	(3.77)	(9.47)	(3.91)	(9.34)
$\mathrm{(B/M)_{t-1}}$	-0.028***	0.034***	-0.027***	0.033***
	(-17.58)	(14.22)	(-17.31)	(13.96)
$\mathrm{ROA_{t-1}}$	0.642***	-0.795***	0.646***	-0.793***
	(33.33)	(-28.82)	(33.93)	(-29.07)
$\mathrm{S\&P500}_{t-1}$	-0.009***	0.003	-0.010***	0.003
	(-4.65)	(0.90)	(-4.74)	(0.96)
Constant	0.143***	0.170***	0.142***	0.171***
	(22.61)	(18.06)	(22.59)	(18.17)
Observations	130,504	130,504	130,504	130,504
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.121	0.076	0.122	0.077

Table 4. Mutual funds scandal (IV Regression)

Following Anton and Polk (2014), we utilize 2003 mutual fund scandal as a plausibly exogenous shock to institutional overweight constraint for identification of our main result in table 4. The implicated funds experienced significant outflows beginning in the last quarter of 2003 and lasting through the end of 2006. We collect the data on funds implicated in the 2003 trading scandal from Stanford Law School Securities Class Action Clearinghouse, which provides information relating to the prosecution, defense and settlement of federal class action securities fraud litigation. In this test, we require stocks that are covered both by the Thomson-Reuters Mutual Fund Holdings database and the Thomson-Reuters Institutional Holdings database as of the third quarter of 2003. In the first stage, we predict the variable overweight ratio (OR) and residual institutional ownership (ResIO) with the RATIO_{200309}, which is the number of implicated owners divided by the number of all institutional owners as of September 2003 for each firm. The second stage of the regression uses the fitted OR ($\widehat{\mathrm{OR}}$) and the ResIO ($\widehat{\mathrm{ResIO}}$) to forecast the news score (NewsScore) and the ratio of the number of bad news reported over the number of news reported (PctBadNews). Panel A reports the results for the first-stage regression, and Panel B presents the results for the second-stage regression. All the continuous variables are winsorized at 1% and 99% to reduce the effect of outliers on the analysis. The t-statistics are reported in the parentheses are ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

First-Stage		
DepVar	OR	ResIO
RATIO_{200309}	-0.753***	-1.154***
	(-6.72)	(-4.34)
Size_{t-1}	-0.015***	-0.024***
	(-4.66)	(-3.56)
$(\mathrm{B/M})_{t-1}$	0.030***	0.083***
	(3.83)	(4.51)
ROA_{t-1}	0.761***	1.249***
	(11.61)	(7.85)
$\mathrm{S\&P500}_{t-1}$	-0.003	-0.003
	(-0.52)	(-0.21)
Constant	0.134***	0.179**
	(4.24)	(2.53)
Observations	22,025	22,025
Year-quarter FE	YES	YES
Industry FE	YES	YES
Cluster	YES	YES

	Second Stage			
DepVar	$\mathrm{NewsScore}_t$	$\mathrm{PctBadNews}_t$	$\mathrm{NewsScore}_t$	$\mathrm{PctBadNews}_t$
$\widehat{\mathrm{OR}}$	-0.502***	0.250*		
	(-4.26)	(1.78)		
$\widehat{\mathrm{ResIO}}$			-0.328***	0.163*
			(-3.50)	(1.74)
RATIO_{200309}				
Size_{t-1}	0.002	0.003	0.002	0.003
	(1.04)	(1.31)	(0.67)	(1.37)
$(\mathrm{B/M})_{t-1}$	-0.011	0.023**	0.001	0.017
	(-1.36)	(2.34)	(0.12)	(1.34)
ROA_{t-1}	0.896***	-0.651***	0.923***	-0.664***
	(8.62)	(-5.24)	(6.84)	(-4.90)
$\mathrm{S\&P500}_{t-1}$	-0.012**	-0.034***	-0.011*	-0.034***
	(-2.40)	(-5.60)	(-1.93)	(-5.54)
Constant	0.013	0.319***	0.005	0.323***
	(0.82)	(16.39)	(0.24)	(15.66)
Observations	22,025	22,025	22,025	22,025
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES

Table 5. News Sentiment and Distance to Dow Jones Offices

This table presents the results for the relation between firm's proximity to Dow Jones Offices and news sentiment. We obtain firm's headquarter location from Compustat quarterly. The street-level location for the eight Dow Jones offices in US mainland are from Dow Jones Official Website. Mindis (in miles) is the minimum distance from a firm's headquarter to one of the eight Dow Jones offices. D is a dummy variable which equals to 1 if Mindis is smaller than 35th percentiles of Mindis, otherwise D equals to 0. Other variables are constructed in the same way as before. All continuous variables are winsorized at 1% and 99% to reduce the outliers on the analysis.

	(1)	(2)	(3)	(4)
DepVar	$\mathrm{NewsScore}_t$	$\mathrm{PctBadNews}_t$	$\mathrm{NewsScore}_t$	$\mathrm{PctBadNews}_t$
OR_{t-1}	-0.041***	0.079***		
	(-4.70)	(6.39)		
ResIO_{t-1}			-0.036***	0.056***
			(-8.33)	(9.01)
D_{t-1}	-0.003**	0.006***	-0.003**	0.006***
	(-2.15)	(2.71)	(-2.04)	(2.64)
$\mathrm{OR}_{t-1} * \mathrm{D}_{t-1}$	0.024*	-0.060***		
	(1.72)	(-2.81)		
$\mathrm{ResIO}_{t-1} * \mathrm{D}_{t-1}$			0.024***	-0.046***
			(3.23)	(-4.15)
Size_{t-1}	0.003***	0.009***	0.003***	0.009***
	(4.00)	(9.06)	(4.12)	(8.99)
$(\mathrm{B/M})_{t-1}$	-0.028***	0.034***	-0.027***	0.033***
	(-17.51)	(14.18)	(-17.16)	(13.85)
ROA_{t-1}	0.638***	-0.785***	0.642***	-0.785***
	(33.05)	(-28.40)	(33.76)	(-28.81)
$\mathrm{S\&P500}_{t-1}$	-0.009***	0.003	-0.009***	0.003
	(-4.62)	(0.90)	(-4.64)	(0.85)
Constant	0.143***	0.170***	0.142***	0.171***
	(22.47)	(17.86)	(22.45)	(18.03)
Observations	130,504	130,504	130,504	130,504
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.1210	0.0764	0.1220	0.0776

Table 6. Dissemination of Earnings Announcements News (Placebo Test)

This table reports regressions of dissemination of earnings announcements news on overweight ratio (OR) and residual institutional ownership (ResIO). All variables are winsorized at the 1% and 99% levels to mitigate the effect of outliers. The dependent variable $\mathrm{ND\_EA_t}$ is the logarithm of the total number of earnings announcement news in quarter t, constructed by counting the new articles on Ravenpack that are specific to each quarterly earnings announcement. $\mathrm{Bad_{t-1}}$ is a dummy variable equal to 1 for earnings announcements with negative standardized earnings surprise (SUE_{t-1}<0), and 0 otherwise. For Model (1) and (2), we follow Livnat and Mendenhall (2006) to define SUE is as actual earnings minus expected earnings, after excluding "special items" from Compustat data, scaled by stock price. Expected earnings is defined as the reported earnings for the same quarter of the prior year. Specifically,

$$SUE(Compustat)_{i,t} = \frac{Q_{i,t} - Q_{i,t-4}}{P_{i,t}}$$

For Model (3) and (4), we follow Jegadeesh and Livant (2006) to define SUE is as actual earnings minus expected earnings, scaled by the standard deviation of quarterly earnings growth. Expected earnings are estimated using a seasonal random walk with drift model. Specifically,

$$Q_{i,t} = \partial_{i,t} + Q_{i,t-4} + \varepsilon_{i,t}$$

$$SUE(RW)_{i,t} = \frac{Q_{i,t} - E(Q_{i,t})}{\sigma_{i,t}}$$

Where $\sigma_{i,t}$ is the he standard deviation of quarterly earnings growth. we estimate the drift $\partial_{i,t}$ as follows:

$$\partial_{i,t} = \frac{\sum_{j=1}^{8} (Q_{i,t-j} - X_{i,t-j-4})}{8}$$

and

$$E\,(Q_{i,,t}) = \partial_{i,t} + Q_{i,\,t-4}$$

$\mathrm{OR_{t-1}}$ *and* $\mathrm{ResIO_{t-1}}$ are overweight ratio and residual institutional ownership in the quarter t-1. $\mathrm{Size_{t-1}}$ is the logarithm of market capitalization in quarter t-1, $\mathrm{(B/M)_{t-1}}$is the book-to-market ratio defined as the ratio of book value of equity to market value of equity in year t-1. $\mathrm{ROA_{t-1}}$ is the return on assets in quarter t-1. $\mathrm{S\&P500}_{t-1}$ equals to 1 if the firm is in S&P 500 in quarter t-1. All the continuous variables are winsorized at 1% and 99% to reduce the effect of outliers on the analysis. Standard errors are clustered by firms. Year-quarter and industry fixed effect are applied. The sample period is between first quarter of 2000 and fourth quarter of 2016. The robust *t*-statistics are reported in the parentheses are ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

DepVar	(1) ND_EA_t	(2) ND_EA_t	(3) ND_EA_t	(4) ND_EA_t
OR_{t-1}*Bad_{t-1}	0.056		-0.018	
	(1.33)		(-0.54)	
OR_{t-1}	0.253***		0.290***	
	(4.15)		(4.42)	
$ResIO_{t-1}$*Bad_{t-1}		0.008		-0.011
		(0.37)		(-0.69)
$ResIO_{t-1}$		0.226***		0.249***
		(7.22)		(7.30)
Bad_{t-1}	0.063***	0.061***	0.030***	0.031***
	(14.53)	(14.20)	(9.15)	(9.47)
$Size_{t-1}$	0.176***	0.175***	0.180***	0.180***
	(28.46)	(27.97)	(27.62)	(27.17)
$(B/M)_{t-1}$	0.087***	0.084***	0.099***	0.094***
	(6.64)	(6.39)	(6.96)	(6.64)
ROA_{t-1}	0.658***	0.628***	0.563***	0.531***
	(4.93)	(4.75)	(3.78)	(3.61)
$S\&P500_{t-1}$	0.252***	0.253***	0.246***	0.246***
	(12.52)	(12.64)	(11.88)	(11.99)
Constant	0.153***	0.159***	0.140**	0.144**
	(2.72)	(2.82)	(2.34)	(2.40)
Observations	111,821	111,821	103,486	103,486
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	Yes	YES	YES	YES
Adjusted R^2	0.670	0.671	0.657	0.659

Table 7. Forecasting Crash Risk

This table examines the relation between overweight ratio (OR), residual institutional ownership (ResIO) and three measures of skewness (Skewness, NCSKEW and DUVOL) using a panel data analysis with fixed effects. The dependent variables are $\mathrm{Skewness_t}$, $\mathrm{NCSKEW_t}$ and $\mathrm{DUVOL_t}$ (Chen, Hong and Stein, JFE2001). $\mathrm{Skewness_t}$ is the total skewness, which is the skewness of daily log returns in quarter t; $\mathrm{NCSKEW_t}$ is calculated by taking the negative of the third moment of daily market-adjusted log returns, and dividing it by the standard deviation of daily market-adjusted log returns raised to the third power in quarter t; $\mathrm{DUVOL_t}$ is the log of the ratio of down-day to up-day standard deviation, measured using (daily) market-adjusted log returns in quarter t. $\mathrm{Skewness_{t-1}}$,$\mathrm{NCSKEW_{t-1}}$ and $\mathrm{DUVOL_{t-1}}$ re three measures of skewness in quarter t-1. $\mathrm{OR_{t-1}}$ and $\mathrm{ResIO_{t-1}}$ are overweight ratio and residual institutional ownership in quarter t-1. $\mathrm{Sigma_{t-1}}$ is the standard deviation of (daily) market-adjusted log returns in quarter t-1. $\mathrm{Dturnover_{t-1}}$ is the average monthly turnover in the quarter t-1, detrended by a moving average of turnover in the prior 3 quarters. $\mathrm{Ret_{t-1}}$, $\mathrm{Ret_{t-2}}$ and $\mathrm{Ret_{t-3}}$ are the market-adjusted cumulative log return in the quarter t-1 through t-3. $\mathrm{(B/M)_{t-1}}$ is the book-to-market ratio defined as the ratio of book value of equity to market value of equity in the year t-1. $\mathrm{ROA_{t-1}}$ is the return on assets in quarter t-1. S&P500 is the membership dummy in quarter t-1. All the continuous variables are winsorized at 1% and 99% to reduce the effect of outliers on the analysis. We control for year-quarter fixed effect and cluster standard errors by firm.

DepVar	$Skewness_t$	$NCSKEW_t$	$DUVOL_t$	$Skewness_t$	$NCSKEW_t$	$DUVOL_t$
OR_{t-1}	-0.434***	0.507***	0.215***			
	(-13.87)	(13.84)	(11.63)			
$ResIO_{t-1}$				-0.094***	0.111***	0.038***
				(-6.14)	(6.24)	(4.26)
$Skewness_{t-1}$	0.017***			0.017***		
	(4.77)			(5.00)		
$NCSKEW_{t-1}$		0.012***			0.013***	
		(3.43)			(3.63)	
$DUVOL_{t-1}$			0.027***			0.027***
			(8.15)			(8.35)
$Sigma_{t-1}$	-1.575***	1.079***	-0.361**	-1.390***	0.863**	-0.441***
	(-5.32)	(3.24)	(-2.21)	(-4.65)	(2.57)	(-2.67)
$Dturnover_{t-1}$	13.433**	-8.084	1.132	10.362	-4.501	2.708
	(2.09)	(-1.07)	(0.32)	(1.61)	(-0.59)	(0.76)
RET_{t-1}	-0.299***	0.347***	0.263***	-0.303***	0.352***	0.266***
	(-18.09)	(18.01)	(27.68)	(-18.26)	(18.19)	(27.86)
RET_{t-2}	-0.179***	0.225***	0.143***	-0.184***	0.230***	0.145***
	(-11.04)	(11.95)	(15.90)	(-11.33)	(12.24)	(16.17)
RET_{t-3}	-0.169***	0.190***	0.117***	-0.173***	0.194***	0.119***
	(-10.84)	(10.58)	(13.80)	(-11.07)	(10.82)	(14.02)
$Size_{t-1}$	-0.042***	0.044***	0.020***	-0.040***	0.042***	0.019***
	(-11.99)	(10.56)	(9.51)	(-11.18)	(9.84)	(8.84)
$(B/M)_{t-1}$	-0.150***	0.176***	0.101***	-0.151***	0.178***	0.102***
	(-18.96)	(19.07)	(22.08)	(-18.96)	(19.05)	(22.03)
ROA_{t-1}	0.301***	-0.403***	-0.053	0.129	-0.203*	0.040
	(2.91)	(-3.29)	(-0.89)	(1.26)	(-1.68)	(0.67)
$S\&P500_{t-1}$	0.054***	-0.060***	-0.030***	0.052***	-0.057***	-0.029***
	(4.71)	(-4.17)	(-4.28)	(4.44)	(-3.94)	(-4.05)
Constant	0.706***	-0.650***	-0.376***	0.691***	-0.632***	-0.369***
	(21.08)	(-17.39)	(-19.68)	(20.44)	(-16.77)	(-19.10)
Observations	143,136	143,136	143,136	143,136	143,136	143,136
Year-quarter FE	YES	YES	YES	YES	YES	YES
Cluster	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.0202	0.0167	0.0313	0.0191	0.0157	0.0304

Table 8. Investment Constrains and Firm Fundamentals

This table examines the relation between overweight ratio (OR) and residual institutional ownership (ResIO) and firm fundamentals. The dependent variables are firms' ROA and ROE, which represent the firms' fundamental values in quarter t, respectively. OR_{t-1}/ResIO_{t-1} are overweight ratio and residual institutional ownership in the quarter t-1. Size_{t-1} is the logarithm of market capitalization in quarter t-1, $(\mathrm{B/M})_{t-1}$is the book-to-market ratio defined as the ratio of book value of equity to market value of equity in quarter t-1. Standard errors are clustered by firms. All the continuous variables are winsorized at 1% and 99% to reduce the effect of outliers on the analysis. The t-statistics are reported in the parentheses are ***, **, * denote 1%, 5%, and 10% significant levels, respectively. The sample period is from first quarter of 2000 through the fourth quarter of 2016.

DepVar	ROAt	ROAt	ROEt	ROEt
OR_{t-1}	0.068***		0.087***	
	(18.72)		(13.99)	
ResIO_{t-1}		0.023***		0.015***
		(12.24)		(4.67)
Size_{t-1}	0.005***	0.005***	0.012***	0.013***
	(15.49)	(15.18)	(20.16)	(19.98)
$(\mathrm{B/M})_{t-1}$	-0.015***	-0.016***	-0.032***	-0.032***
	(-19.55)	(-19.68)	(-20.20)	(-19.96)
sp500	-0.008***	-0.008***	-0.011***	-0.011***
	(-7.54)	(-7.44)	(-6.19)	(-6.23)
Constant	0.014***	0.015***	-0.023***	-0.023***
	(4.99)	(5.04)	(-4.26)	(-4.27)
Observations	128,544	128,544	128,544	128,544
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.192	0.174	0.128	0.120

Table 9.Robustness Check

This table summarize the results of the robustness regressions. In Panel A, we shrink the boundary for calculating overweight ratio (OR). In the main result (table 4), if the weight of stock *i* in one institution's portfolio is higher than the weight of stock *i* in the same market portfolio, then this institution overweights stock *i*. Now we multiply 1.05, 1.1, 1.2 and 1.5 with the weight of stock *i* in the same market portfolio, respectively, and apply the same requirement to calculate overweight ratio as before. Panel A.1. to Panel A.4. reports the results for different OR from the above four boundaries, respectively. ResIO is the same as before; In Panel B, we replicate the main analysis (Table 4) for different subsamples. Panel B.1. to Panel B. 3. reports the subsample results for three different periods, 2000 to 2004, 2005 to 2009 and 2010 to 2016, respectively.

Panel A: Alternative Measures of Investment Constraints				
Panel A1: Multiplier=1.05				
	(1)	(2)	(3)	(4)
DepVar	$\mathrm{NewsScore_t}$	$\mathrm{PctBadNews_t}$	$\mathrm{NewsScore_t}$	$\mathrm{PctBadNews_t}$
$\mathrm{OR_{t-1}}$	-0.034***	0.062***		
	(-4.68)	(5.86)		
$\mathrm{ResIO_{t-1}}$			-0.028***	0.041***
			(-7.89)	(7.81)
$\mathrm{Size_{t-1}}$	0.002***	0.009***	0.003***	0.009***
	(3.77)	(9.47)	(3.91)	(9.34)
$\mathrm{(B/M)_{t-1}}$	-0.028***	0.034***	-0.027***	0.033***
	(-17.58)	(14.21)	(-17.31)	(13.96)
$\mathrm{ROA_{t-1}}$	0.642***	-0.796***	0.646***	-0.793***
	(33.31)	(-28.83)	(33.93)	(-29.07)
$\mathrm{S\&P500}_{t-1}$	-0.009***	0.003	-0.010***	0.003
	(-4.65)	(0.90)	(-4.74)	(0.96)
Constant	0.143***	0.170***	0.142***	0.171***
	(22.61)	(18.07)	(22.59)	(18.17)
Observations	130,504	130,504	130,504	130,504
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.121	0.0760	0.122	0.0769
Panel A2: Multiplier=1.1				
	(1)	(2)	(3)	(4)
DepVar	$\mathrm{NewsScore_t}$	$\mathrm{PctBadNews_t}$	$\mathrm{NewsScore_t}$	$\mathrm{PctBadNews_t}$
$\mathrm{OR_{t-1}}$	-0.034***	0.063***		
	(-4.66)	(5.91)		
$\mathrm{ResIO_{t-1}}$			-0.028***	0.041***
			(-7.89)	(7.81)

$Size_{t-1}$	0.002***	0.009***	0.003***	0.009***
	(3.77)	(9.47)	(3.91)	(9.34)
$(B/M)_{t-1}$	-0.028***	0.034***	-0.027***	0.033***
	(-17.58)	(14.21)	(-17.31)	(13.96)
ROA_{t-1}	0.642***	-0.796***	0.646***	-0.793***
	(33.31)	(-28.85)	(33.93)	(-29.07)
$S\&P500_{t-1}$	-0.009***	0.003	-0.010***	0.003
	(-4.65)	(0.90)	(-4.74)	(0.96)
Constant	0.143***	0.170***	0.142***	0.171***
	(22.62)	(18.07)	(22.59)	(18.17)
Observations	130,504	130,504	130,504	130,504
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R2	0.121	0.0760	0.122	0.0769

Panel A3: Multiplier=1.2

	(1)	(2)	(3)	(4)
DepVar	$NewsScore_t$	$PctBadNews_t$	$NewsScore_t$	$PctBadNews_t$
OR_{t-1}	-0.034***	0.065***		
	(-4.62)	(6.01)		
$ResIO_{t-1}$			-0.028***	0.041***
			(-7.89)	(7.81)
$Size_{t-1}$	0.002***	0.009***	0.003***	0.009***
	(3.78)	(9.47)	(3.91)	(9.34)
$(B/M)_{t-1}$	-0.028***	0.034***	-0.027***	0.033***
	(-17.58)	(14.21)	(-17.31)	(13.96)
ROA_{t-1}	0.642***	-0.796***	0.646***	-0.793***
	(33.31)	(-28.88)	(33.93)	(-29.07)
$S\&P500_{t-1}$	-0.009***	0.003	-0.010***	0.003
	(-4.64)	(0.90)	(-4.74)	(0.96)
Constant	0.143***	0.170***	0.142***	0.171***
	(22.62)	(18.08)	(22.59)	(18.17)
Observations	130,504	130,504	130,504	130,504
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.121	0.0760	0.122	0.0769

Panel A3: Multiplier=1.5

	(1)	(2)	(3)	(4)
DepVar	$NewsScore_t$	$PctBadNews_t$	$NewsScore_t$	$PctBadNews_t$
OR_{t-1}	-0.036***	0.073***		
	(-4.59)	(6.46)		
$ResIO_{t-1}$			-0.028***	0.041***
			(-7.89)	(7.81)

Size_{t-1}	0.002***	0.009***	0.003***	0.009***
	(3.78)	(9.48)	(3.91)	(9.34)
$(\text{B/M})_{t-1}$	-0.028***	0.034***	-0.027***	0.033***
	(-17.58)	(14.20)	(-17.31)	(13.96)
ROA_{t-1}	0.641***	-0.797***	0.646***	-0.793***
	(33.33)	(-28.99)	(33.93)	(-29.07)
$\text{S\&P500}_{t-1}$	-0.009***	0.003	-0.010***	0.003
	(-4.63)	(0.88)	(-4.74)	(0.96)
Constant	0.143***	0.170***	0.142***	0.171***
	(22.62)	(18.10)	(22.59)	(18.17)
Observations	130,504	130,504	130,504	130,504
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.121	0.0761	0.122	0.0769

Pane B: Sub-Sample Analysis				
Panel B1: Sample Period=2000 to 2004				
	(1)	(2)	(3)	(4)
DepVar	$NewsScore_t$	$PctBadNews_t$	$NewsScore_t$	$PctBadNews_t$
OR_{t-1}	-0.045***	0.040***		
	(-4.09)	(2.65)		
$ResIO_{t-1}$			-0.035***	0.033***
			(-6.51)	(4.42)
$Size_{t-1}$	-0.003***	0.010***	-0.003***	0.010***
	(-3.29)	(7.64)	(-3.25)	(7.64)
$(B/M)_{t-1}$	-0.040***	0.052***	-0.039***	0.051***
	(-14.68)	(13.52)	(-14.40)	(13.32)
ROA_{t-1}	0.989***	-1.220***	1.000***	-1.233***
	(32.35)	(-29.05)	(33.03)	(-29.47)
$S\&P500_{t-1}$	-0.016***	-0.013***	-0.016***	-0.013***
	(-4.92)	(-2.91)	(-4.99)	(-2.88)
Constant	0.173***	0.166***	0.172***	0.167***
	(19.12)	(13.07)	(19.14)	(13.17)
Observations	45,007	45,007	45,007	45,007
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.135	0.102	0.137	0.103
Panel B2: Sample Period=2005-2009				
	(1)	(2)	(3)	(4)
VARIABLES	$NewsScore_t$	$PctBadNews_t$	$NewsScore_t$	$PctBadNews_t$
OR_{t-1}	-0.033***	0.082***		
	(-3.03)	(5.00)		
$ResIO_{t-1}$			-0.029***	0.052***
			(-5.63)	(6.66)
$Size_{t-1}$	0.005***	0.006***	0.005***	0.006***
	(4.79)	(3.58)	(4.95)	(3.45)
$(B/M)_{t-1}$	-0.030***	0.032***	-0.030***	0.032***
	(-13.28)	(9.24)	(-13.16)	(9.16)
ROA_{t-1}	0.605***	-0.695***	0.611***	-0.691***
	(19.66)	(-15.46)	(20.06)	(-15.57)
$S\&P500_{t-1}$	-0.021***	0.007	-0.021***	0.008
	(-6.18)	(1.43)	(-6.33)	(1.55)
Constant	0.013	0.296***	0.011	0.297***
	(1.42)	(20.83)	(1.27)	(20.98)
Observations	38,275	38,275	38,275	38,275
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES

Cluster	YES	YES	YES	YES
Adjusted R^2	0.0852	0.0484	0.0867	0.0499

Panel B3: Sample Period=2010-2016				
	(1)	(2)	(3)	(4)
VARIABLES	NewsScore_t	PctBadNews_t	NewsScore_t	PctBadNews_t
OR_{t-1}	-0.023***	0.076***		
	(-3.00)	(5.48)		
ResIO_{t-1}			-0.025***	0.053***
			(-5.74)	(6.88)
Size_{t-1}	0.006***	0.013***	0.006***	0.013***
	(7.22)	(9.20)	(7.30)	(9.04)
$(\text{B/M})_{t-1}$	-0.010***	0.011***	-0.010***	0.011***
	(-5.28)	(3.36)	(-5.10)	(3.27)
ROA_{t-1}	0.230***	-0.333***	0.227***	-0.314***
	(11.06)	(-9.32)	(11.14)	(-8.91)
$\text{S\&P500}_{t-1}$	0.005*	0.007	0.005*	0.007
	(1.78)	(1.57)	(1.77)	(1.57)
Constant	0.019***	0.189***	0.018**	0.190***
	(2.70)	(15.48)	(2.54)	(15.55)
Observations	47,222	47,222	47,222	47,222
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.0566	0.0567	0.0587	0.0587

Table A1: Determinants of Overweight Ratio

This table reports the results of regressions that study the determinants of a stock's overweight ratio, defined as the number of institutions that overweight a stock divided by the number of institutions that hold the stock. For each quarter between 1999 and 2016, we run cross-sectional regressions of the overweight ratios of stocks on a set of firm characteristics, including the natural logarithm of the market capitalization of the stock at the end of the previous quarter (SIZE), a dummy that takes a value of 1 for stocks that belong to the S&P 500 index (S&P500), book-to-market equity ratio as of the end of the previous year (B/M), and the stock return over the previous 12 months (MOM). We report the time-series averages of the estimated coefficients from the cross-sectional regressions and their t -statistics. The sample includes domestic common stocks traded on NYSE, AMEX, and NASDAQ, excluding REITs, closed-end funds, ADRs, and stocks that are priced below $5 or that rank in the lowest market-capitalization decile as the end of the previous calendar year. Throughout the rest of analysis, overweight ratio (OR) refers to the residual of the overweight-ratio regression as in model 3 of table 1.

DepVar	OR	OR	OR
Constant	0.762***	0.686***	0.700***
	(59.53)	(49.88)	(49.50)
Size	-0.037***	-0.025***	-0.027***
	(-26.38)	(-14.89)	(-16.04)
S&P500		-0.060***	-0.053***
		(-22.81)	(-21.00)
B/M			-0.001
			(-0.77)
MOM			0.027***
			(10.64)
Observations	160725	160725	160725
Adjusted R^2	0.222	0.243	0.257

Table A2. News Sentiment and Investment Constrains (Subsample)

This table replicates Table 3 in the main results using a subsample consisting of the fourth quarters of 2003 through the fourth quarter of 2006 (or 200304-200604). In order to check whether our main result is still hold for our causal relationship identification IV regression in table 6.

DepVar	(1) NewsScore_t	(2) PctBadNews_t	(3) NewsScore_t	(4) PctBadNews_t
OR_{t-1}	-0.041***	0.054***		
	(-2.87)	(2.64)		
ResIO_{t-1}			-0.028***	0.034***
			(-4.30)	(3.56)
Size_{t-1}	0.004***	0.000	0.004***	0.000
	(2.68)	(0.21)	(2.80)	(0.13)
$(\text{B/M})_{t-1}$	-0.031***	0.029***	-0.030***	0.028***
	(-6.08)	(4.04)	(-5.89)	(3.90)
ROA_{t-1}	0.553***	-0.487***	0.558***	-0.489***
	(13.72)	(-8.63)	(14.12)	(-8.76)
$\text{S\&P500}_{t-1}$	-0.014***	-0.030***	-0.015***	-0.029***
	(-3.74)	(-5.29)	(-3.82)	(-5.25)
Constant	0.023**	0.328***	0.021*	0.330***
	(2.04)	(19.42)	(1.89)	(19.45)
Observations	26,098	26,098	26,098	26,098
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Cluster	YES	YES	YES	YES
Adjusted R^2	0.057	0.044	0.058	0.045

Table A3. Forecasting Crash Risk: Fama-Macbeth Regression

This table examines the relation between overweight ratio (OR), residual institutional ownership (ResIO) and three measures of skewness (Skewness, NCSKEW and DUVOL) using a Fama-Macbeth regression. The dependent variables are $\mathrm{Skewness}_t$, NCSKEW_t and DUVOL_t (Chen, Hong and Stein, JFE2001). $\mathrm{Skewness}_t$ is the total skewness, which is the skewness of daily log returns in quarter t; NCSKEW_t is calculated by taking the negative of the third moment of daily market-adjusted log returns, and dividing it by the standard deviation of daily market-adjusted log returns raised to the third power in quarter t; DUVOL_t is the log of the ratio of down-day to up-day standard deviation, measured using (daily) market-adjusted log returns in quarter t. $\mathrm{Skewness}_{t-1}$,NCSKEW_{t-1} and DUVOL_{t-1} re three measures of skewness in quarter t-1. OR_{t-1} and ResIO_{t-1} are overweight ratio and residual institutional ownership in quarter t-1. Sigma_{t-1} is the standard deviation of (daily) market-adjusted log returns in quarter t-1. $\mathrm{Dturnover}_{t-1}$ is the average monthly turnover in the quarter t-1, detrended by a moving average of turnover in the prior 3 quarters. Ret_{t-1}, Ret_{t-2} and Ret_{t-3} are the market-adjusted cumulative log return in the quarter t-1 through t-3. $(\mathrm{B/M})_{t-1}$ is the book-to-market ratio defined as the ratio of book value of equity to market value of equity in the year t-1. ROA_{t-1} is the return on assets in quarter t-1. S&P500 is the membership dummy in quarter t-1. All the continuous variables are winsorized at 1% and 99% to reduce the effect of outliers on the analysis. The t-statistics are adjusted for heteroscedasticity and serial correlation by using Newey-West adjusted standard errors.

DepVar	$Skewness_t$	$NCSKEW_t$	$DUVOL_t$	$Skewness_t$	$NCSKEW_t$	$DUVOL_t$
OR_{t-1}	-0.446***	0.532***	0.240***			
	(-10.51)	(10.46)	(8.01)			
$ResIO_{t-1}$				-0.093***	0.106***	0.038**
				(-3.52)	(3.49)	(2.53)
$Skewness_{t-1}$	0.017***			0.018***		
	(3.63)			(3.85)		
$NCSKEW_{t-1}$		0.013***			0.013***	
		(2.86)			(3.05)	
$DUVOL_{t-1}$			0.028***			0.029***
			(6.84)			(7.06)
$Sigma_{t-1}$	-2.398***	3.626***	0.995*	-2.095**	3.248***	0.826
	(-2.70)	(3.24)	(1.80)	(-2.29)	(2.84)	(1.48)
$Dturnover_{t-1}$	17.166*	-21.152**	-6.001	13.062	-16.445	-3.662
	(1.97)	(-2.05)	(-1.24)	(1.48)	(-1.58)	(-0.74)
RET_{t-1}	-0.372***	0.424***	0.323***	-0.376***	0.428***	0.325***
	(-8.37)	(8.46)	(10.50)	(-8.51)	(8.63)	(10.64)
RET_{t-2}	-0.185***	0.228***	0.154***	-0.189***	0.233***	0.156***
	(-6.58)	(7.79)	(10.06)	(-6.77)	(8.11)	(10.53)
RET_{t-3}	-0.139***	0.154***	0.098***	-0.142***	0.159***	0.100***
	(-5.60)	(5.67)	(6.83)	(-5.78)	(5.87)	(7.07)
$Size_{t-1}$	-0.038***	0.043***	0.020***	-0.036***	0.040***	0.018***
	(-6.27)	(5.86)	(4.47)	(-5.93)	(5.52)	(4.22)
$(B/M)_{t-1}$	-0.191***	0.231***	0.126***	-0.193***	0.233***	0.127***
	(-9.58)	(9.94)	(10.18)	(-9.50)	(9.86)	(10.13)
ROA_{t-1}	0.411***	-0.510***	-0.106	0.233	-0.301*	-0.003
	(2.73)	(-2.94)	(-1.26)	(1.60)	(-1.81)	(-0.04)
$S\&P500_{t-1}$	0.040***	-0.042**	-0.022**	0.037**	-0.039**	-0.021**
	(2.68)	(-2.60)	(-2.52)	(2.50)	(-2.41)	(-2.37)
Observations	143,136	143,136	143,136	143,136	143,136	143,136
Adjusted R^2	0.015	0.015	0.024	0.014	0.014	0.023

Table A4. Market Reaction to News

This table reports market reaction for firms sorted on daily average news sentiment scores. News published after market close is treated as reported on the next day. For news articles published on a non trading day, the following trading day is considered as the event date. For each trading day, we compute the average news sentiment score (NewsScore) for each of the stock and sort the firms to five groups based on the average sentiment score. We next calculate the average abnormal returns for the event day (CAR0), and the average cumulative abnormal returns for the two-day window [0,1] (CAR(0,1)) for each of the five portfolios. Panel A reports the time-series mean of the CAR0 and CAR (0,1) for the five portfolios, where the abnormal returns are adjusted using the returns on the Fama-French 2*3 portfolios (2 size portfolios and 3 book-to-market ratios). Panel B reports the results for the five portfolios using DGTW-adjusted abnormal returns. The sample period is from 2000 to 2016.

Panel A: Fama-French Portfolio Adjusted Return		
News	CAR0	CAR(0,1)
Low	-0.009	-0.01
p2	-0.003	-0.003
p3	0.001	0.001
p4	0.006	0.007
High	0.013	0.014
High - Low	0.021***	0.024***
Panel B: DGTW-Adjusted Return		
News	CAR0	CAR(0,1)
Low	-0.008	-0.01
p2	-0.003	-0.003
p3	0.001	0.001
p4	0.006	0.007
High	0.012	0.014
High - Low	0.020***	0.023***