The Demand-driven Information Market

Shiyang Huang, Dan Li, and Bohui Zhang *

This Draft: December 2020

* 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 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

One central topic in finance and economics is how information is produced and distributed to investors. Information providers play an important role in these processes. Examples of information providers include sell-side analysts, credit rating agents, media reporters, and alternative data, vendors. How these information providers produce and distribute information largely shape financial markets. For example, financial economists argue that the information markets contribute to market frenzies/investor herding behavior (Veldkamp, 2006a), return co-movement (Veldkamp, 2006b), and informational inefficiency (Dugast and Foucault, 2018; Huang, Xiong, and Yang, 2020). An important feature of the information market is that information production is not free and information providers only produce costly information when there is a high demand for such information. Although this feature is intuitive, its empirical evidence is relatively limited. The goal of our paper is to fill this gap.

The challenge of testing how information providers cater to investor demand lies in how to precisely identify investors' information demand. The innovation of our paper is to take advantage of investment constraints on institutional investors. Due to regulations, contract constraints, or agency considerations in the asset management industry, institutional investors, such as mutual funds and pension funds, always face constraints on their investments (Cao, Han, and Wang, 2017; Almazan, Brown, Carlson, and Chapman, 2004). As shown in Cao, Han, and Wang (2017), when mutual funds already overweight (underweight) a stock, they may not buy (sell) more of the stock even if they receive positive (negative) news about the stock. In this sense, when mutual funds already overweight (underweight) a stock, the provision of bad (good) news of the stock is more valuable than good (bad) news. Correspondingly, such investment constraints would incentivize asymmetric patterns in information production. That is, when mutual funds already overweight (underweight) a stock, information intermediaries (e.g., media) are more likely to dig the downside (upside) of the stock and thus produce more negative (positive) information.

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 years between the first quarter of 2000 and the fourth quarter of 2016. News articles from RavenPack are commonly used by institutional investors 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. For each firm in each quarter, we first calculate both the fraction of negative news and the average bad news sentiment scores over a quarter. We then follow Cao, Han, and Wang (2017) and use two steps to measure the extent to which institutions are subject to overinvestment. In the first step, for each stock in each institution's portfolio, if the stock's weight in the portfolio of this institution is larger than the corresponding weight in a market capitalization-weighted portfolio, we define this institution as overweighting this stock. In the second step, for each stock in each quarter, we compute the fraction of institutions that overweight this stock and denote this fraction as the extent of investment constraints (dubbed by overweight ratio). To alleviate the effect of firm characteristics, we also regress the measure of investment constraints on several firm characteristics (e.g., firm size, past stock returns, book-to-market ratio) and obtain the regression residuals as an alternative measure of investment constraint (dubbed by residual overweight ratio).

We first justify that the above two measures indeed capture institutions' investment constraints. We find that, when institutions overweight a stock in their portfolios, they tend to sell the stock or are less likely to increase the holding in the stock in the subsequent quarter, and vice versa. These results are largely consistent with Cao, Han, and Wang (2017).

Based on these two measures of investment constraints, we examine how investors' information demand affects information production among media. As we argue, investment constraints induce institutions to pay asymmetric attention to positive and negative news, leading to asymmetric patterns in information production. Specifically, when most institutions overweight a stock, they value negative news more than positive news, and thus the media strategically caters to institutional preference and produces the coverage on negative stories. We find strong evidence to support our argument. That is, there is a positive and strong association between the measures of overinvestment and the fraction of bad news (or the bad news sentiment score) in the subsequent quarter. 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 2.4% (3%) standard deviation higher level of the fraction of negative news (bad news sentiment score).

We are aware of the endogeneity concern. In particular, some unobservable firm characteristics might 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 address the endogeneity concern, 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 massive outflows from funds of implicated fund families while funds not implicated benefited from this scandal and experienced capital inflows. For example, Kisin (2011) estimates that implicated families all together lost about 14.1% of their capital within one year or two. Intuitively, the capital outflow and inflow arising from the scandal should result in an exogenous change in the overweight ratio and residual overweight ratio. Following the two-stage instrumental variable approach of Anton and Polk (2014), we document a consistently positive relation between investor constraints and the fraction of negative news (or the bad news sentiment score).

To corroborate our argument, we implement a series of additional tests. First, we conduct a cross-sectional analysis to strengthen our argument that the media produces information to meet investors' information 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 firms nearby, it is more costly for media reporters to visit and collect information from distant firms. To cover the high cost of producing information about distant firms, media reports have stronger incentives to cater to investors' demand for information. Therefore, we expect that the investment constraints have stronger effects on information production among firms that are far from the media. We indeed find supporting evidence for this argument in our data.

Following similar intuition, we expect that investment constraints matter more for information production among firms whose information is more complex. To test this argument, we split our firms into two groups. The first group of firms operate the business in one single segment, while the second group of firms operate the business in multiple segments. Intuitively, firms operating in multiple segments are more complicated to analyze and the information production cost is higher. We indeed find that investment constraints have stronger effects on information production among firms with multi-segment business.

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 for this type of news. One obvious example of this kind of news is the firms' earnings announcements, which could be reproduced by the media easily. Therefore, we expect that institutions' investment constraints should not generate asymmetric media coverage of earnings announcements for the good and bad news. Indeed, we show that the media is indifferent between reporting positive and negative earnings news, which is consistent with our conceptual framework that the information production cost plays a key role in generating the asymmetric pattern in demand-driven information production.

Third, we explore the asset pricing implication from 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 and 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.

Fourth, we rule out the effect of firm fundamentals on our empirical findings. One could argue that investment constraints are negatively associated with firm fundamentals, which also lead to negative media news. To address this concern, we examine the association between firm's investment constraints and the subsequent fundamental performance. Strikingly, we find that firms' high investment constraints are associated with high fundamental performance.

Our paper contributes to three strands of 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 skewness in market returns has long been an important puzzle. Our study contributes to this literature by showing that 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 type of indirect price impact from institutions' trading potential.

The remainder of the paper is organized as follows. In Section 2, we review the related literature. In Section 3, we describe the construction of sample, the measurement of news sentiment and investment constraints, and the sample characteristics. In Section 4, we examine the relationship between investment constraints and news sentiment, and we exploit an identification strategy based on the 2003 mutual fund scandal to establish the causality. In Section 5, we carry out further studies. Finally, we provide concluding remarks in Section 6.

2. Literature review

Our study is closely related to three strands of literature. The first strand of literature is about the information market or information production. The theoretical work includes Veldkamp (2006a, b), Veldkamp and Wolfers (2007), Huang, Xiong and Yang (2018, 2020), and Dugast and Foucault (2018), and 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) argue that investor demand is higher for information about firms whose fundamentals help price not only their stocks but also the related stocks. Based on this argument, Hameed, Morck, Shen, and Yeung (2015) find that analysts disproportionally follow firms whose fundamentals correlated more with their industry peers. Our study complements this strand of literature by showing that investment constraints affect investor demand for 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, Cao and Chen 1997; Dumas, Fleming and Whaley, 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 increases 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 on 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., La Porta, Lakonishok, Shleifer, and Vishny, 1997; Chan, Chen, and Lakonishok, 2002; Cohen, Gompers, and Vuolteenaho, 2002; Cremers and Petajisto, 2009; Lewellen, 2011), fund flows (Lou, 2012; Huang, Xiang and Song, 2019), and investment constraints (Cao, Han, and Wang, 2017). The most related paper is Cao, Han, and Wang (2017), which shows 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 that investment constraints affect information production and then generate asymmetric patterns in stock returns.

3. Data and variable construction

This section describes the construction of the 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 database. Stocks with share prices below $5 and stocks in the lowest market capitalization decile based on NYSE breakpoints as of the end of the calendar year are excluded. Institutional equity holding data comes from Thomson Reuters Institutional Holdings (13F) database.

News data is 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 use the Dow Jones Edition of the RavenPack, 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 score based on how new or how novel a news story is, a relevance score to indicate how strongly related the firm is to the news story, and a sentiment score that reflects the potential market impact of the 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 assigns 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 the relevance score, a higher value indicates greater relevance. A sentiment score of 50 indicates neutral

[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 to obtain the maximum period coverage.

sentiment, and values above (below) 50 suggest positive (negative) sentiment. We will discuss it in more detail in the subsequent section.

The historical archive of the RavenPack database dates to January 2000. As a result, our sample period ranges from the first quarter of 2000 through the fourth quarter of 2016. 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.

3.2. Measures of news sentiment

For each firm at each quarter, we construct two measures of news sentiment using RavenPack data. The first measure captures the fraction of bad news (*PctBadNews*), and the second measure is an overall bad news sentiment score (*BadNewsScore*), which is zero minus the average sentiment score. 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 greater relevance. RavenPack takes into consideration multiple factors to determine the relevance score, 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. 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 a relevance score of less than 100 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 a positive (negative) sentiment, the higher (lower) the score, the greater market reaction a news article is expected to induce. RavenPack's algorithm relies on both an 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, for which financial experts classify news stories as having either positive or negative financial impact and determine the extent of the impact. The strength component relies on its proprietary natural language processing software that takes into consideration the use of emotionally charged language, and the software is also 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 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 confirms 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 a negative adjusted sentiment score

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

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. For easy interpretation, we consider *BadNewsScore* which is zero minus *NewsScore*.

3.3. Measures of investment constraints

We follow Cao, Han, and Wang (2017) and construct two proxies of investment constraints. Unlike individual investors, institutional investors often face a variety of constraints that limit their ability to invest and their positions in certain stocks, due to a combination of regulatory provisions, contractual arrangements, and investment strategies. There are two types of important constraints. The first is related to the diversification requirements. For example, mutual funds are required to meet various diversification requirements to be able to pass through gains to shareholders and avoid double taxation. Pension funds are required to divest investments 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. Larger tracking errors could lead to the termination of contracts or even financial penalties. These constraints make it difficult for institutional investors to deviate from their benchmarks. In this sense, when institutions already overweight (underweight) a stock, they are reluctant to add to (reduce) positions in the stock even when there is good (bad) news about the firm.

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

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

We compute and use the fraction of institutions that overweight a particular stock to proxy for investment constraints. We term this measure as *overweight ratio* or *OR*.

The second measure is the residual overweight ratio after controlling for firm characteristics that might be related to institutional holdings. At each quarter, we regress the overweight ratio on size, book-to-market ratio, stock returns over the previous 12 months, and a dummy for the S&P500 index membership. The residual from the regression is our second measure of investment constraints. We term the second measure as *residual overweight ratio* or *Residual-OR*.

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. *MOM* is the past 12-month returns. *S&P500* is a dummy variable which equals one for the S&P500 index constituent.

Table 1 reports the summary statistics of our sample. As shown in Panel A, on average, a firm has a *BadNewsScore* of −0.079, consistent with the literature that media coverage is in general positive. There is also substantial variation in *PctBadNews*: the mean is 0.267, with the 5th percentile equalling 0 and the 95th percentile equalling 60%. Institutions tend to overweight the stocks in our sample, with the mean residual overweight ratios of 0.004.

Panel B reports the Pearson correlation matrix. *OR* and *Residual-OR* is positively correlated. The correlation coefficient is 0.872, suggesting the two measures capture a different aspect of investment constraints. More importantly, at the first glance, there are positive correlations between the production of bad news (measured by *BadNewsScore/PctBadNews*) and investment constraints (measured by *OR/Residual-OR*). This suggests that when institutional investors have already overweighted a particular stock, the media, as important information producers, produces more bad news.

[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. For example, investment constraints limit institutions' ability to increase the position in a stock when they have already overweighted it, which leads them to pay more attention to negative news. Media strategically caters to institutional preference and covers more negative stories as a result. To make this argument clear, we develop a suggestive model in the Online Appendix. In the suggestive model, there is one stock with payoff following a binary distribution (e.g., good state vs. bad state). The information seller can choose to produce information about the good state and the bad state. Meanwhile, only investors that already overweight the stock are willing to pay for the bad signal, and only investors that already underweight the stock are willing to pay for the good signal. As shown in the model, when investors that already overweight the stock dominate, the information seller focuses on the bad state of the stock and only produce an informative signal about the bad state. In contrast, when investors that already underweight the stock dominate, the information seller focuses on the good state of the stock and only produce an informative signal about the good state.

We test our argument 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 production of bad news. Third, we use the mutual fund scandal of 2003 to pin down the causal relation between investment constraints and information production.

4.1. Investment constraints and changes in the institutional ownership

We first examine the effect of investment constraints on changes in institutional holdings. At each quarter, we sort all stocks equally into five groups based on *OR* or *Residual-OR*. We then calculate the change in *OR* or *Residual-OR* in the following quarter. Table 2 presents the time-series average change for each group. The changes in overweight ratio or residual overweight ratio decrease monotonically from the top group to the bottom group, consistent with the findings in Cao, Han, and Wang (2017). For example, the difference in *Residual-OR* between stocks in the top and bottom *Residual-OR* groups is 2.25%, significant at the 1% level. These results suggest that when 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:

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

where *i* indexes firm and *t* indexes time. *BadNewsScore* and *PctBadNews* are the two measures of bad news. *OR* and *Residual-OR* measure investment constraints. We are interested in the coefficient *b* as it captures the effect of investment constraints on the production of bad news. Intuitively, if the media caters to institutional demand for information, the coefficient *b* should be positive and significant. Vector *X* represents a set of control variables, including *Size*, *B/M*, *ROA*, *MOM*, 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 regression specification (1). The dependent variables are *BadNewsScore* 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 *Residual-OR* in columns (3) and (4), respectively. Across all regression specifications, the estimated coefficient *b* is positive and significant, suggesting that a higher overweight ratio or residual overweight ratio is associated with more bad news. The effect of the overweight ratio (residual overweight ratio) on the production of bad news is also economically significant. For example, a one-standard-deviation increase in investment constraints measured by *OR* is associated with a 2.7% (1.5%) standard deviation higher level of *BadNewsScore* (*PctBadNews*).

[Insert Table 3 About Here]

4.3. Mutual fund scandal: Instrumental variable regression

Although the results in Table 3 suggest a positive relation between investment constraints and the production of bad news, we can't rule out the possibility that there might exist some unobservable firm-specific factors that drive changes in both institutional ownership and news tones. In this section, we exploit an identification strategy based on the mutual fund scandal of 2003 to address the causality issue. On September 3, 2003, New York State Attorney General issued a complaint against a hedge fund, Canary Capital Partners, for engaging in illegal trading behaviours including extensive market timing and late trading with several mutual funds. The scandal kept unfolding. Until the end of 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 behaviour (McCabe, 2009). The scandal triggered massive outflows 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 that 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 overweight ratio, but they are unrelated to firm fundamentals and its news coverage. This setting allows us to draw inferences 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 before 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 Thomson-Reuters Mutual Fund Holdings database and 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 *Residual-OR*) 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 in regression specification (1). We control for industry fixed effects and year-quarter fixed effects, and cluster standard errors at the firm level for all tests. Table 4 reports the results. As shown in Panel A (for the first stage of the 2SLS instrumental variable regression), $RATIO_{200309}$ positively and significantly predict both *OR* and *Residual-OR*. In Panel B (for the second stage of the 2SLS instrumental variable regressions), we re-estimate the regression models in equation (1) by replacing the measure of investment constraints (*OR* or *Residual-OR*) with the predicted value from the first stage ($\widehat{OR}$ and $Resi\widehat{dual} - OR$). We find that both $\widehat{OR}$ and $Resi\widehat{dual} - OR$ still positively and significantly predict *BadNewsScore* and *PctBadNews*. These results demonstrate that there exists a causal effect of investment constraints on news tones.

[Insert Table 4 About Here]

5. Additional tests

To corroborate our evidence, we conduct several additional tests in this section. First, we conduct two cross-sectional tests to strengthen our argument that media strategically produce information to cater to investors' demand for information. In the first test, we focus on the information production cost, which is proxied by the geographic distance between the firm's headquarter location and Dow Jones offices. In the second test, we exploit the heterogeneities of firms in their information environment. We further carry out a placebo test using the earnings announcements. Last, we discuss some asset pricing implications from this demand-driven 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 the 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 production has a high fixed cost, whether or how the information sellers produce information largely 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 rationale, 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 distant firms, it is more convenient for media reporters to visit and collect information from firms nearby. To calculate the distance between firms' headquarters and Dow Jones offices, we obtain the 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 has eight offices, we focus on the minimum distance from a firm's headquarter to one of the eight Dow Jones offices. We classify all firms into two groups: firms close to Dow Jones offices, and firms far from Dow Jones offices. For a firm whose minimum distance from a firm's headquarter to one of the eight Dow Jones offices is in the bottom decile of the sample, we classify it into the group of firms close to Dow Jones offices. Otherwise, we classify the firm into the group of firms far from Dow Jones offices. After that, we run the regression specification (1) for each group separately.

Table 5 reports the results. Panel A is for firms close to Dow Jones offices, and Panel B is for firms far from Dow Jones offices. As shown in Panel A, on firms close to Dow Jones offices, neither *OR* nor *Residual-OR* can significantly predict *BadNewsScore* and *PctBadNews*. In contrast, from firms far from Dow Jones offices, both *OR* nor *Residual-OR* can positively and significantly predict *BadNewsScore* and *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. Cross-sectional studies: The role of information complexity

Following the similar argument in Section 5.1, we expect that investment constraints play a more important role in information production among firms with more complex information environment as the information production cost on these firms tend to be higher. We take several steps to test this argument. First, we follow Duchin, Matsusaka and Oguzhan (2010) and classify all firms into the group of single-segment firms and the group of multi-segment firms. We then repeat the specification of the regression (1) for each group separately. Table 6 reports the results.

As shown in Panel A, on firms operating the business in one single segment, neither *OR* nor *Residual-OR* can significantly predict *BadNewsScore* and *PctBadNews*. In contrast, on firms having business in multiple segments both *OR* nor *Residual-OR* can positively and significantly predict *BadNewsScore* and *PctBadNews*. Briefly, the results in Table 6 suggest that investment constraints have stronger effects on news production when the information environment is more complex.

[Insert Table 6 About Here]

5.3 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 of existing 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:

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

where $Bad_{i,t}$ is a dummy variable to indicate whether firm *i* announces a negative earnings surprises at quarter *t*. In these regressions, the dependent variable, $\mathrm{ND\_EA}_{i,t}$, 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. By construction, $\mathrm{ND\_EA}_{i,t}$ captures the media coverage of earnings announcements. $High\ OR_{i,t-1}(High\ Residual - OR_{i,t-1})$ is a dummy variable indicating whether the firms in the top tercile of overweight ratio. 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 three 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 the stock price. The second measure of SUE is 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. The third measure of SUE is actual earnings before extraordinary items minus the median of analyst forecasts in the 90 days before the earnings announcement, scaled by the stock price. After calculating SUE, the dummy variable $Bad_{i,t}$ equals 1 for earnings announcements with negative standardized earnings surprise (SUE<0), and 0 otherwise.

Table 7 reports the results of 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 negative, which contradicts our main results in Table 3.

[Insert Table 7 About Here]

5.3 Asset pricing implications: Stock return asymmetry

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 of investors has 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 patterns of information provision, one natural asset pricing implication is that investment constraints are associated with asymmetric patterns in stock returns.

To examine the relationship 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 return 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:

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

where $Asymmetric\ Measure$ includes the three measures: Skewness, NCSKEW and DUVOL. Following Chen, Hong and Stein (2001), the control variables include the lagged asymmetric measures, the book-to-market ratio in the previous quarter, stock return volatility in the previous quarter, the market-adjusted cumulative returns in previous three quarters, and the return on assets (ROA) in quarter $t-1$, and the S&P500 membership. We also include the year-quarter fixed effects and cluster the standard errors at the firm level.

Table 8 reports the results. In all model specifications, we find that investment 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 -3.68% decreases in the stock return skewness. For comparison, the skewness ranges from −61.3% at the 25th percentile to 43.2% at the 75th percentile.

[Insert Table 8 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, we find that our results are robust to Fama-MacBeth regressions. Second, Table A2 in Internet Appendix shows that the market indeed reacts strongly to the media news announcements, which suggests that the effect of investment constraints on stock return asymmetric comes from its effect on asymmetric patterns of media news reports.[3]

[3] In untabulated results, we find that there is no significant association between investment constraints and the asymmetric patterns of stock returns after excluding the announcements of media news.

Our results on stock return asymmetry has important asset pricing implications. It is well-known and puzzling that aggregate stock market returns are asymmetrically distributed. More importantly, the stock market is always subject to crashes. 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 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 works, explain this pattern through the joint effect of short-sale constraints and disagreement (Chen, Hong, and Stein, 2001; Hong and Stein, 2003), and 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 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 investors overreact to negative news ex-ante) or investors' misinterpreting 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 firm’s investment constraints and the subsequent fundamental performance. Strikingly, we find that high investment constraints are associated with high fundamental performance, measured by ROA or

ROE (see Table 9). 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 negative news (e.g., some investors overreact to negative news ex-ante) or investors misinterpret negative information.

[Insert Table 9 About Here]

6. Conclusion

We directly test the argument that information providers care about investor demand for information and selectively produce information when there is a high demand for such information. The challenge of testing how information providers cater to investor demand lies in how to precisely identify investors' information demand. The innovation of our paper is to take advantage of investment constraints on institutional investors. Institutions are subject to a variety of trading constraints as a combination of law, contractual arrangement, and investment strategy. That is, institutions cannot 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. Similarly, institutions cannot keep selling a underweighted stock even when there is bad news. Such constraints lead institutional investors to be more attentive to negative (positive) news for stocks overweighed (underweighted) in their holdings, which incentivizes information intermediaries, such as media, to focus more on negative (positive) news.

We find strong and consistent evidence to support the argument that the media caters to institutional investors by producing more negative news for stocks overweighed by institutions

[4] Cao, Han and Wang (2017) also find similar return predictions of investment constraints.

using data from RavenPack. The further test suggests the negative relation between institutional investment constraints and news sentiment is not due to worsening fundamentals. Using the mutual fund scandal of 2003 as a natural experiment, we confirm the causal relationship between investment constraints and asymmetric information production. 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 or among firms with business in multiple segments.

Our results have important asset pricing implications. We find that that through the effect on information production, the fraction of institutions that overweight stocks 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.

Almazan, A., Brown, K. C., Carlson, M., Chapman, D. A., 2004. Why constrain your mutual fund manager?. Journal of Financial Economics, 73, 289-321.

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.

Core, J. E., Guay, W., Larcker, D. F., 2008. The power of the pen and executive compensation. Journal of Financial Economics, 88, 1-25.

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

Duchin, R., Matsusaka, J. G., Ozbas, O., 2010. When are outside directors effective?. Journal of Financial Economics, 96, 195-214.

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.

Hartzell, J. C., Starks, L. T., 2003. Institutional investors and executive compensation. Journal of Finance, 58, 2351-2374.

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

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

Huang, S., Xiong. Y., Yang, L., 2020, Skill Acquisition and Data 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.

McCabe, P.E., 2009. The economics of the mutual fund trading scandal. Working paper.

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

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

von Beschwitz, B., Keim, D.B., Massa, M., 2018. First to "read" the news: New analytics and algorithmic trading. Working Paper.

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 (BadNewsScore), the ratio of bad news reported over the number of news reported (PctBadNews), overweight ratio (OR), residual overweight ratio (Residual-OR), the logarithm of market capitalization (Size), book-to-market ratio (B/M), return-on-asset ratio (ROA), past 12-month returns (MOM), and S&P500 membership dummy (S&P500). This table reports the number of observations (N), mean, standard deviation (Std), and $5^{th}/25^{th}/50^{th}/75^{th}/95^{th}$ percentile values (P5/P25/Median/P75/P95) of the variables. The sample period is from 2000Q1 to 2016Q4. Panel B reports the Pearson correlation matrix for the variables.

Panel A. Statistics Summary

Variable	N	Mean	Std	P5	P25	Median	P75	P95
BadNewsScore	138,514	–0.079	0.127	–0.304	–0.151	–0.075	–0.002	0.127
PctBadNews	138,514	0.267	0.183	0.000	0.136	0.250	0.379	0.600
OR	138,514	0.487	0.113	0.276	0.417	0.503	0.569	0.647
Residual−OR	138,514	0.004	0.097	–0.186	–0.047	0.020	0.072	0.137
Size	138,514	7.413	1.482	5.321	6.332	7.209	8.290	10.221
B/M	138,514	0.564	0.401	0.107	0.283	0.477	0.736	1.320
ROA	138,514	0.007	0.033	–0.049	0.002	0.010	0.021	0.045
MOM	138,514	0.149	0.469	–0.511	–0.123	0.100	0.340	0.968
S&P500	138,514	0.207	0.405	0.000	0.000	0.000	0.000	1.000

Panel B: Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9
1. BadNewsScore	1.000								
2. BadNewsPct	0.761	1.000							
3. OR	0.032	–0.023	1.000						
4. Residual OR	0.002	–0.003	0.872	1.000					
5. Size	–0.055	0.055	–0.399	–0.027	1.000				
5. B/M	0.072	0.016	0.038	0.004	–0.267	1.000			
6. ROA	–0.220	–0.156	0.072	0.143	0.231	–0.138	1.000		
7. MOM	–0.147	–0.071	0.091	0.0149	0.142	–0.328	0.179	1.000	
8. S&P500	–0.040	0.044	–0.421	–0.020	0.712	–0.087	0.125	–0.026	1.000

Table 2: Investment Constraints and Institutional Trading

This table reports the time-series mean of the change in the overweight ratio (OR) or a residual overweight ratio (Residual-OR) in quarter t for portfolios sorted on OR or Residual-OR in quarter t–1, respectively. Institutional trading activity for stocks is measured by the change in OR or Residual-OR. At the end of each quarter, we sort the stocks into five groups based on OR and Residual-OR, respectively. We then calculate the average change in OR and the average change in Residual-OR during the next quarter following the measurement of investment constraints. The row "Low–High" reports the differences in the trading activity measures between the bottom OR (Residual-OR) group and the top OR (Residual-OR) group. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Change in OR	Change in Residual-OR
	Sorted on OR in quarter t–1	Sorted on Residual-OR in quarter t–1
Low	0.013	0.010
P2	0.005	0.00
P3	0.002	0.001
P4	–0.002	–0.002
High	–0.01	–0.015
Low–High	0.029***	0.025***

Table 3. News Sentiment and Investment Constraints

This table examines the relation between investment constraints and news sentiment. Investment constraint is measured by the overweight ratio (OR) and residual overweight ratio (Residual-OR), and news sentiment is measured by bad news score (BadNewsScore) and the ratio of bad news reported over the number of news reported (PctBadNews). The dependent variables, $\mathrm{BadNewsScore_t}$ and $\mathrm{PctBadNews_t}$, represent the average of news score and percentage of bad news for each firm in quarter *t*, respectively. The key independent variables, $\mathrm{OR_{t-1}}$ and $\mathrm{Residual-OR_{t-1}}$, are the overweight ratio and residual overweight ratio 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 the book value of equity to the market value of equity at the previous year-end. $\mathrm{ROA_{t-1}}$ is the return on assets in quarter *t*–1. $\mathrm{MOM_{t-1}}$ is the past 12-month return. $\mathrm{S\&P500}_{t-1}$ equals 1 if the firm is included in S&P 500 index in quarter *t*–1. Standard errors are clustered by firms. The *t*-statistics are reported in the parentheses. ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

	(1)	(2)	(3)	(4)
DepVar:	$\mathrm{BadNewsScore_t}$	$\mathrm{PctBadNews_t}$	$\mathrm{BadNewsScore_t}$	$\mathrm{PctBadNews_t}$
$\mathrm{OR_{t-1}}$	0.030***	0.024**		
	(4.43)	(2.39)		
$\mathrm{Residual-OR_{t-1}}$			0.033***	0.036***
			(4.91)	(3.59)
$\mathrm{Size_{t-1}}$	–0.002**	0.011***	–0.002***	0.011***
	(–2.27)	(11.22)	(–3.29)	(11.05)
$\mathrm{(B/M)_{t-1}}$	0.006***	0.015***	0.006***	0.014***
	(3.76)	(5.25)	(3.65)	(5.15)
$\mathrm{ROA_{t-1}}$	–0.821***	–0.990***	–0.822***	–0.995***
	(–43.42)	(–36.27)	(–43.50)	(–36.48)
$\mathrm{MOM_{t-1}}$	–0.030***	–0.020***	–0.029***	–0.019***
	(–25.86)	(–12.03)	(–25.38)	(–11.79)
$\mathrm{S\&P500}_{t-1}$	0.001	0.002	–0.001	0.001
	(0.40)	(0.73)	(–0.40)	(0.33)
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
No. Obs.	138,514	138,514	138,514	138,514
Adj. R^2	0.132	0.090	0.132	0.090

Table 4. Mutual Funds Scandal (IV Regression)

This table reports the results using a two-stage instrumental variable (IV) approach. Following Anton and Polk (2014), we utilize the 2003 mutual fund scandal as a plausibly exogenous shock to institutional overweight constraint for identification. 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 the first stage, we predict the variable overweight ratio (OR) and residual overweight ratio (Residual-OR) 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 Residual OR ($\widehat{\mathrm{Residual}} - \mathrm{OR}$) to forecast the news score (BadNewsScore) 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. The *t*-statistics are reported in the parentheses are ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

Panel A: First-Stage		
DepVar:	OR	Residual-OR
RATIO_{200309}	0.184***	0.183***
	(6.78)	(6.81)
Size_{t-1}	–0.034***	–0.007**
	(–12.11)	(–2.34)
$(\mathrm{B/M})_{t-1}$	0.023***	0.022***
	(3.09)	(3.07)
ROA_{t-1}	0.410***	0.399***
	(6.77)	(6.65)
MOM_{t-1}	0.042***	0.004
	(12.52)	(1.11)
$\mathrm{S\&P500}_{t-1}$	–0.046***	0.005
	(–6.93)	(0.79)
Year-quarter FE	YES	YES
Industry FE	YES	YES
No. Obs.	22,372	22,372

Panel B: Second Stage				
DepVar:	$BadNewsScore_t$	$PctBadNews_t$	$BadNewsScore_t$	$PctBadNews_t$
$\widehat{OR}$	0.166**	0.361***		
	(1.96)	(2.67)		
$\widehat{Residual\_OR}$			0.167**	0.364***
			(1.96)	(2.67)
$Size_{t-1}$	0.001	0.014***	–0.004**	0.004*
	(0.31)	(3.02)	(–2.39)	(1.70)
$(B/M)_{t-1}$	0.001	0.010	0.001	0.011
	(0.15)	(1.14)	(0.15)	(1.15)
ROA_{t-1}	–1.001***	–1.086***	–1.000***	–1.083***
	(–14.90)	(–10.80)	(–14.95)	(–10.83)
MOM_{t-1}	–0.055***	–0.040***	–0.049***	–0.026***
	(–11.10)	(–5.41)	(–13.65)	(–5.19)
$S\&P500_{t-1}$	0.012**	–0.014	0.003	–0.033***
	(1.98)	(–1.49)	(0.81)	(–4.87)
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
No. Obs.	22,372	22,372	22,372	22,372
Adj. R^2	0.082	0.035	0.082	0.034

Table 5. News Sentiment and Distance to Dow Jones Offices

This table presents the results for the relation between a firm's proximity to Dow Jones Offices and news sentiment. We obtain the firm's headquarter location from Compustat quarterly. The zip code-level location for the eight Dow Jones offices on the US mainland is 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 1 if Mindis of observation is equal to or smaller than the 10th percentiles of Mindis, otherwise, D equals 0. We split the firms into "close to Dow Jones Office" and "far from Dow Jones Office" groups based on their values of D. We estimate the relation between investment constraints and news sentiment within the two groups of firms separately, and the results are reported in Panel A and B, respectively. Other variables are constructed as the same in Table 3. Year-quarter and industry fixed effects are applied. The *t*-statistics are reported in the parentheses are ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

Panel A: Firms Close to Dow Jones Office				
	(1)	(2)	(3)	(4)
DepVar:	$\mathrm{BadNewsScore}_t$	$\mathrm{PctBadNews}_t$	$\mathrm{BadNewsScore}_t$	$\mathrm{PctBadNews}_t$
OR_{t-1}	-0.022	-0.072**		
	(-1.01)	(-2.04)		
$\mathrm{Residual} - \mathrm{OR}_{t-1}$			-0.019	-0.065*
			(-0.86)	(-1.85)
Size_{t-1}	-0.006***	0.008**	-0.005***	0.009***
	(-2.70)	(2.20)	(-2.62)	(2.84)
$(\mathrm{B/M})_{t-1}$	-0.005	0.000	-0.005	0.001
	(-0.97)	(0.06)	(-0.93)	(0.12)
ROA_{t-1}	-0.704***	-0.807***	-0.706***	-0.811***
	(-13.44)	(-9.93)	(-13.47)	(-9.99)
MOM_{t-1}	-0.025***	-0.017***	-0.026***	-0.019***
	(-7.21)	(-3.03)	(-7.61)	(-3.49)
$\mathrm{S\&P500}_{t-1}$	0.001	-0.008	0.002	-0.004
	(0.12)	(-0.77)	(0.33)	(-0.40)
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
No. Obs.	13,003	13,003	13,003	13,003
Adj. R^2	0.113	0.070	0.113	0.069

Panel B: Firms Far from Dow Jones Office				
	(1)	(2)	(3)	(4)
DepVar:	$BadNewsScore_t$	$PctBadNews_t$	$BadNewsScore_t$	$PctBadNews_t$
OR_{t-1}	0.033***	0.032***		
	(4.69)	(3.13)		
$Residual - OR_{t-1}$			0.036***	0.046***
			(5.17)	(4.38)
$Size_{t-1}$	-0.001*	0.011***	-0.002***	0.011***
	(-1.87)	(10.58)	(-2.92)	(10.14)
$(B/M)_{t-1}$	0.007***	0.016***	0.007***	0.015***
	(4.04)	(5.25)	(3.91)	(5.12)
ROA_{t-1}	-0.831***	-1.011***	-0.832***	-1.016***
	(-40.74)	(-34.81)	(-40.81)	(-35.02)
MOM_{t-1}	-0.030***	-0.020***	-0.029***	-0.019***
	(-24.40)	(-11.30)	(-23.84)	(-10.92)
$S\&P500_{t-1}$	0.001	0.004	-0.001	0.003
	(0.53)	(1.27)	(-0.29)	(0.78)
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
No. Obs.	124,590	124,590	124,590	124,590
Adj. R^2	0.134	0.093	0.135	0.093

Table 6. Subsample Analysis: Single-segment Firms vs. Multi-segment Firms

This table presents the results for the subsample analysis of Table 3. In each year, we split the sample into single-segment and multi-segment firms. The single-segment firms operate in one industry, while multi-segment firms operate in multiple industries. The dependent variables are $\mathrm{BadNewsScore_t}$ and $\mathrm{PctBadNews_t}$, which represent the average of news score and percentage of bad news for each firm in quarter *t*, respectively. OR_{t-1} and Residual-OR_{t-1} are overweight ratio and residual overweight ratio 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 the book value of equity to the market value of equity at previous year-end. $\mathrm{ROA_{t-1}}$ is the return on assets in quarter *t*–1. $\mathrm{MOM_{t-1}}$ is the past 12-month return. $\mathrm{S\&P500}_{t-1}$ equals 1 if the firm is in S&P 500 in quarter *t*–1. Standard errors are clustered by firms. Year-quarter and industry fixed effects are applied. The *t*-statistics are reported in the parentheses. ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

Panel A: Single-segment Firms				
	(1)	(2)	(3)	(4)
DepVar:	$\mathrm{BadNewsScore_t}$	$\mathrm{PctBadNews_t}$	$\mathrm{BadNewsScore_t}$	$\mathrm{PctBadNews_t}$
$\mathrm{OR_{t-1}}$	–0.017	0.031		
	(–1.47)	(1.45)		
$\mathrm{Residual-OR_{t-1}}$			–0.020*	0.034
			(–1.65)	(1.58)
$\mathrm{Size_{t-1}}$	–0.008***	0.019***	–0.007***	0.018***
	(–6.44)	(9.19)	(–6.23)	(9.31)
$\mathrm{(B/M)_{t-1}}$	–0.009***	–0.009	–0.009***	–0.009
	(–2.81)	(–1.47)	(–2.80)	(–1.48)
$\mathrm{ROA_{t-1}}$	–0.377***	–0.511***	–0.377***	–0.512***
	(–15.25)	(–12.07)	(–15.22)	(–12.11)
$\mathrm{MOM_{t-1}}$	–0.012***	–0.005	–0.012***	–0.004
	(–6.48)	(–1.50)	(–6.71)	(–1.30)
$\mathrm{S\&P500}_{t-1}$	–0.004	0.002	–0.003	0.000
	(–0.91)	(0.27)	(–0.72)	(0.06)
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
No. Obs.	18,550	18,550	18,550	18,550
Adj. R^2	0.096	0.067	0.097	0.067

Panel B: Multi-segment Firms				
	(1)	(2)	(3)	(4)
DepVar:	BadNewsScore_t	PctBadNews_t	BadNewsScore_t	PctBadNews_t
OR_{t-1}	0.031**	0.104***		
	(2.34)	(4.50)		
$\text{Residual} - \text{OR}_{t-1}$			0.029**	0.107***
			(2.18)	(4.52)
Size_{t-1}	–0.008***	0.013***	–0.009***	0.010***
	(–5.82)	(5.63)	(–6.38)	(4.36)
$(\text{B/M})_{t-1}$	–0.008**	–0.016**	–0.008**	–0.016**
	(–2.23)	(–2.38)	(–2.23)	(–2.38)
ROA_{t-1}	–0.575***	–0.681***	–0.574***	–0.681***
	(–11.04)	(–8.15)	(–11.04)	(–8.17)
MOM_{t-1}	–0.022***	–0.014***	–0.022***	–0.011**
	(–7.98)	(–2.89)	(–7.75)	(–2.34)
$\text{S\&P500}_{t-1}$	–0.003	0.022***	–0.004	0.017**
	(–0.66)	(3.02)	(–1.00)	(2.46)
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
No. Obs.	12,854	12,854	12,854	12,854
Adj. R^2	0.115	0.068	0.115	0.068

Table 7. Dissemination of Earnings Announcements News (Placebo Test)

This table reports the results for the regressions of dissemination of earnings announcements news on the overweight ratio (OR) and residual overweight ratio (Residual-OR). 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 which equals 1 for earnings announcements with negative standardized earnings surprise ($SUE_{t-1}<0$), and equals 0 otherwise. For Model (1) and (4), we follow Livnat and Mendenhall (2006) to define SUE as actual earnings before extraordinary items minus the reported earnings for the same quarter of the prior year, scaled by the stock price. Specifically,

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

For Model (2) and (5), 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 a 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 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

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

For Model (3) and (6), we follow Livnat and Mendenhall (2006) to define SUE as actual earnings before extraordinary items minus the median of analyst forecasts in the 90 days before the earnings announcement, scaled by the stock price. OR_{t-1} and Residual-OR_{t-1} are overweight ratio and residual overweight ratio in the quarter *t*–1. In each quarter, we rank firms into terciles based on their OR and Residual-OR. The dummy variables, High OR and High Residual-OR, indicate the firms in the top and bottom tercile, respectively. We drop the sample in the middle tercile. $\mathrm{Size_{t-1}}$ 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 the book value of equity to market value of equity at previous year-end. $\mathrm{ROA_{t-1}}$ is the return on assets in quarter *t*–1. $\mathrm{S\&P500}_{t-1}$ equals 1 if the firm is in S&P 500 in quarter *t*–1. Standard errors are clustered by firms. Year-quarter and industry fixed effects are applied. The sample period is between the first quarter of 2000 and the fourth quarter of 2016. The *t*-statistics are reported in the parentheses. ***, **, * 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	(5) ND_EA_t	(6) ND_EA_t
$High\ OR_{t-1}$	0.052***	0.052***	0.087***			
	(3.64)	(3.58)	(5.98)			
$High\ Residual - OR_{t-1}$				0.043***	0.042***	0.061***
				(3.27)	(3.16)	(4.56)
$High\ OR_{t-1} \times Bad_t$	–0.035***	–0.035***	–0.113***			
	(–3.19)	(–2.99)	(–7.23)			
$High\ Residual - OR_{t-1} \times Bad_t$				–0.036***	–0.034***	–0.080***
				(–3.67)	(–3.29)	(–5.96)
Bad_t	0.032***	0.032***	0.012*	0.038***	0.037***	0.016*
	(5.63)	(5.51)	(1.65)	(5.77)	(5.49)	(1.95)
$Size_{t-1}$	0.202***	0.202***	0.198***	0.205***	0.205***	0.202***
	(27.31)	(27.32)	(26.63)	(27.18)	(27.19)	(26.59)
$(B/M)_{t-1}$	0.097***	0.097***	0.098***	0.098***	0.097***	0.099***
	(5.83)	(5.82)	(5.99)	(5.88)	(5.87)	(6.03)
ROA_{t-1}	–1.027***	–1.025***	–1.064***	–1.023***	–1.021***	–1.065***
	(–8.05)	(–8.04)	(–8.37)	(–7.94)	(–7.94)	(–8.26)
Mom_{t-1}	–0.055***	–0.055***	–0.058***	–0.058***	–0.058***	–0.062***
	(–7.57)	(–7.53)	(–8.05)	(–8.03)	(–8.00)	(–8.59)
$S\&P500_{t-1}$	0.255***	0.255***	0.252***	0.258***	0.258***	0.259***
	(10.65)	(10.66)	(10.60)	(10.87)	(10.88)	(10.92)
Year-quarter FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
No. Obs.	74,188	74,188	74,188	74,188	74,188	74,188
Adj. R^2	0.656	0.656	0.657	0.656	0.656	0.657

Table 8. Forecasting Crash Risk

This table examines the relation between the overweight ratio (OR), residual overweight ratio (Residual-OR), 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, 2001). $\mathrm{Skewness_t}$ 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 $\mathrm{DUVOL_{t-1}}$ have lagged measures of skewness in quarter *t*–1. OR_{t-1} and Residual-OR_{t-1} are overweight ratios and residual overweight ratio in quarter *t*–1. We also include various control variables: $\mathrm{Sigma_{t-1}}$ is the standard deviation of daily market-adjusted log returns in quarter *t*–1. $\mathrm{Dturnover_{t-1}}$ is the average monthly turnover in 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 the market value of equity at previous year-end. $\mathrm{ROA_{t-1}}$ is the return on assets in quarter *t*–1. S&P500 is the S&P500 index membership dummy in quarter *t*–1. Year-quarter fixed effects are included, and standard errors are clustered by firms. The *t*-statistics are reported in the parentheses. ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

DepVar:	$\mathrm{Skewness_t}$	$\mathrm{NCSKEW_t}$	$\mathrm{DUVOL_t}$	$\mathrm{Skewness_t}$	$\mathrm{NCSKEW_t}$	$\mathrm{DUVOL_t}$
$\mathrm{OR_{t-1}}$	–0.326***	0.365***	0.169***			
	(–9.58)	(9.13)	(8.43)			
$\mathrm{Residual - OR_{t-1}}$				–0.318***	0.344***	0.155***
				(–9.19)	(8.48)	(7.62)
$\mathrm{Skewness_{t-1}}$	0.017***			0.017***		
	(5.22)			(5.26)		
$\mathrm{NCSKEW_{t-1}}$		0.008**			0.008**	
		(2.48)			(2.48)	
$\mathrm{DUVOL_{t-1}}$			0.023***			0.023***
			(7.24)			(7.27)
Year-quarter FE	YES	YES	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES	YES	YES
No. Obs.	132.291	132.291	132.291	132.291	132.291	132.291
Adj. R^2	0.0143	0.0132	0.0272	0.0144	0.0133	0.0273

Table 9. Investment Constraints and Firm Fundamentals

This table examines the relation between investment constraints and firm fundamentals. The dependent variables are firms' ROA and ROE, which represent the firms' fundamental values in quarter *t*, respectively. The key independent variables, OR_{t-1} and Residual-$OR_{t-1,}$ are the overweight ratio and residual overweight ratio in quarter *t*–1. Control variables follow the definition in Table 3. Standard errors are clustered by firms. Year-quarter fixed effects are included. The *t*-statistics are reported in the parentheses. ***, **, * denote 1%, 5%, and 10% significant levels, respectively.

DepVar:	ROA_t	ROE_t	ROA_t	ROE_t
OR_{t-1}	0.047***	0.084***		
	(16.29)	(13.90)		
$Residual - OR_{t-1}$			0.047***	0.083***
			(15.86)	(13.57)
$Size_{t-1}$	0.006***	0.013***	0.005***	0.011***
	(19.34)	(19.73)	(16.75)	(17.44)
$(B/M)_{t-1}$	–0.008***	–0.023***	–0.008***	–0.024***
	(–10.87)	(–12.03)	(–11.28)	(–12.29)
MOM_{t-1}	0.007***	0.015***	0.009***	0.018***
	(17.13)	(15.38)	(20.59)	(17.85)
$S\&P500_{t-1}$	–0.001	0.000	–0.004***	–0.004**
	(–1.60)	(0.23)	(–4.95)	(–2.23)
Year-quarter FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
No. Obs.	135,174	135,174	135,174	135,174
Adj. R^2	0.176	0.147	0.175	0.146

Table A1: Determinants of Overweight Ratio

This table reports the regression results on the determinants of overweight ratio (OR). OR is defined as the fraction 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 ratio 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 the corresponding *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 at the end of the previous calendar year. Throughout the rest of the analyses, the overweight ratio (OR) refers to the residual of the overweight-ratio regression as in column 3 of Table A1.

	(1)	(2)	(3)
DepVar:	$\mathrm{OR_t}$	$\mathrm{OR_t}$	$\mathrm{OR_t}$
Constant	0.762***	0.686***	0.700***
	(59.53)	(49.88)	(49.50)
$\mathrm{Size_t}$	–0.037***	–0.025***	–0.027***
	(–26.38)	(–14.89)	(–16.04)
$\mathrm{(B/M)_t}$		–0.060***	–0.053***
		(–22.81)	(–21.00)
$\mathrm{MOM_t}$			–0.001
			(–0.77)
$\mathrm{S\&P500}_t$			0.027***
			(10.64)
No. Obs.	160725	160725	160725
Adj. R^2	0.222	0.243	0.257

Table A2. 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 (BadNewsScore) for each of the stocks and sort the firms into five groups based on the average sentiment score. We next calculate the average abnormal returns for the event day (CAR(0)), 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 CAR(0) and CAR (0,1) for the five portfolios, where the abnormal returns are adjusted using the returns on the Fama-French 2×3 portfolios (see Fama and French (1993)). 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	CAR(0)	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	CAR(0)	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***

Online Appendix

A Model of

Demand-driven Information Market

1 A Model of Demand-driven Information Market

Our main hypothesis is that information production is not free and information providers only produce costly information when there is a high demand for such information. For example, investors have investment constraint, probably due to institutional restrictions. When mutual funds already overweight (underweight) a stock, they may not buy (sell) more of the stock even if they receive positive (negative) news about the stock. In this sense, when mutual funds already overweight (underweight) a stock, the provision of bad (good) news of the stock is more valuable than good (bad) news. We argue that such investment constraints would incentivize asymmetric patterns in information production. That is, when mutual funds already overweight (underweight) a stock, information intermediaries (e.g., media) are more likely to dig the downside (upside) of the stock and thus produce more negative (positive) information. To illustrate the economic intuition for the hypothesis in the paper, we develop a suggestive model in this section.

In this model, there is one risky asset, a continuum of investors, and one information seller. The asset payoff is denoted as v, and v follows a binary distribution with Δ (> 0)and $-\Delta$ with equal probabilities. The information seller accesses to the set of binary-signal information structure and chooses sell the information to investors, who can then trade on the information. Following Yang (2020), the information sellers observe a signals $s \in \{0,1\}$ parameterized by measurable function $m : \{-\Delta,\Delta\} \to [0,1]$. m measures the accuracy of the signal. Specifically, m_1 is the probability of observing signal $s = 1$ if the true state v is Δ (and so $1-m_1$ is the probability of observing signal $s = 0$). m_2 is the probability of observing signal $s = 0$ if the true state v is $-\Delta$ (and so $1-m_2$ is the probability of observing signal $s = 0$). In short, m_1 and m_2 are the probability that the signal equals the true value.

Here, the quantity of information gained through $m(\cdot)$ equals the difference between the information seller's prior entropy and expected posterior entropy as follows:

$$\begin{aligned} I(m) &= \frac{1}{2}g(m(\Delta)) + \frac{1}{2}g(m(-\Delta)) \\ &\quad -g(\frac{1}{2}m(\Delta) + \frac{1}{2}m(-\Delta)) \end{aligned} \tag{1.1}$$

where

$$g(x) = x \cdot \ln(x) + (1-x) \cdot \ln(1-x).$$

In short, $I(m)$ measure the uncertainty reduction by the information structure m. Following Yang (2020), simplifying $I(m)$yields:

$$\begin{aligned} I(m) \quad = \quad & \frac{1}{2}m_1 \cdot \ln(m_1) + \frac{1}{2}(1-m_1) \cdot \ln(1-m_1) \\ & + \frac{1}{2}m_2 \cdot \ln(m_2) + \frac{1}{2}(1-m_2) \cdot \ln(1-m_2) \\ & - (\frac{1}{2} + \frac{1}{2}m_1 - \frac{1}{2}m_2) \cdot \ln(\frac{1}{2} + \frac{1}{2}m_1 - \frac{1}{2}m_2) \\ & - (\frac{1}{2} - \frac{1}{2}m_1 + \frac{1}{2}m_2) \cdot \ln(\frac{1}{2} - \frac{1}{2}m_1 + \frac{1}{2}m_2) \end{aligned} \tag{1.2}$$

We turn to discuss investors' demand for information. We assume that α fraction of investors have already overweighed the risky asset and can not buy but can sell the risky asset, and $1-\alpha$ fraction of investors have already underweighed the risky asset and cannot sell but can buy the risky asset. The overweight can be due to the benchmark concern. For example, mutual funds track the benchmark and have restrictions to buy (sell) the stock when they overweight (underweight) the stock relative to the benchmark Cao, Han, and Wang (2016). When one individual investor trades on the signal provided by the information seller, he or she needs to pay 1 to the information seller. In this sense, when the information seller reports her signal $s=0$, only investors that have overweighed the risky asset are willing to buy the signal. When the information seller reports her signal $s=1$, only investors that have underweighed the risky asset are willing to buy the signal.

Considering the demand of the information the information chooses the information structure (i.e., m_1 and m_2 to maximize the expected information-sales profit. The optimization problem is characterized as follows:

$$\max_{m_1,m_2} E\left[\mathbf{I}_{s=0} \cdot \alpha + \mathbf{I}_{s=1} \cdot (1-\alpha)\right] \tag{1.3}$$

subject to:

$$I(m) \leq \mu, 0.5 \leq m_1 \leq 1, 0.5 \leq m_2 \leq 1,$$

where $\mathbf{I}_{s=0}$ is the indicator function where $s=0$, and $\mathbf{I}_{s=1}$ is the indicator function where $s=1$.

Noted that μ describes the information capacity, and it is a positive and finite constant. As we can see, if the signals are pure noise ($m_1 = 0.5$ and $m_2 = 0.5$), the uncertainty reduction, $I(m)$, is zero. Since $\mathbf{I}_{s=0}$ is the unconditional probability that $s = 0$ and $\mathbf{I}_{s=1}$ is the unconditional probability that $s = 1$, the optimization problem is equivalent to

$$\max_{m_1,m_2} prob(s=0)\cdot\alpha + prob(s=1)\cdot(1-\alpha). \tag{1.4}$$

Given $m_1$and m_2, we have:

$$prob(s=0) = \frac{1}{2}(1-m_1) + \frac{1}{2}(1-m_2),$$

$$prob(s=1) = \frac{1}{2}m_1 + \frac{1}{2}m_2.$$

Thus, the information seller's optimization problem is equivalent to

$$\max_{m_1,m_2} \frac{1}{2} + \frac{1}{2}(m_1 - m_2)(1-2\alpha)$$

subject to:

$$\begin{aligned} I(m) \quad &= \quad \frac{1}{2}m_1 \cdot \ln(m_1) + \frac{1}{2}(1-m_1)\cdot\ln(1-m_1) \\ &\quad\quad +\frac{1}{2}m_2 \cdot \ln(m_2) + \frac{1}{2}(1-m_2)\cdot\ln(1-m_2) \\ &\quad\quad -(\frac{1}{2} + \frac{1}{2}m_1 - \frac{1}{2}m_2)\cdot\ln(\frac{1}{2} + \frac{1}{2}m_1 - \frac{1}{2}m_2) \\ &\quad\quad -(1 - \frac{1}{2}m_1 + \frac{1}{2}m_2)\cdot\ln(1 - \frac{1}{2}m_1 + \frac{1}{2}m_2) \\ &\leq \quad \mu \end{aligned}$$

We can characterize the equilibrium as follows. First, as shown in Lemma 1, we know that $I(m)$ is always increasing with m_1 and m_2. This is intuitive as $I(m)$ measures the uncertainty reduced by the information structure and m_1 and m_2 measure the accuracy of the signals.

LEMMA **1.** *$I(m)$ is increasing with m_1 and m_2.*

As we can see from information seller's optimization problem, information seller needs to consider the demand of signal $s = 1$ and signal of $s = 0$. Intuitively, when the population of investors that have overweighed the risky asset dominates, the information seller prefers to produce signal of $s = 0$ as only these investors are willing to pay for signal of $s = 0$. Otherwise, the information seller prefers to produce signal of $s = 1$. In short, the demand of information determines information seller's information production. We summarize information seller's information production in the following proposition.

PROPOSITION **1.** *The information seller's optimal choices as follows:*

(i) when $\alpha > \frac{1}{2}$, the information seller sets $m_1^* = \frac{1}{2}$ and $m_2^* = m^*(> 0.5)$;

(ii) when $\alpha < \frac{1}{2}$, the information seller sets $m_1^* = m^*$ and $m_2^* = \frac{1}{2}(> 0.5)$.

Proposition 1 clearly shows that the information seller's optimal choices on the signal structure depends on the demand of information. When the demand of the signal of $s = 0$ is high (i.e., $\alpha > \frac{1}{2}$), the information seller focuses on the signal about the bad state ($v = -\Delta$) and totaly ignores the information about the good state ($v = \Delta$). That is, when $\theta = \Delta$, the signal s is a pure noise and takes 1 or 0 with equal probability.

Similarly, when the demand of the signal $s = 1$ is high (i.e., $\alpha < \frac{1}{2}$), the information seller focuses on the signal about the good state ($v = \Delta$) and totaly ignores the information about the bad state ($v = -\Delta$). That is, when $v = -\Delta$, the signal s is a pure noise and takes 1 or 0 with equal probability.

References

Yang, Ming, 2020. Optimality of Debt under Flexible Information Acquisition. Review of Economics Studies, 78(1), 487-536.

Cao, Jie, Bing Han, and Qinghai Wang, 2016. Institutional Investment Constraints and Stock Prices. Journal of Financial and Quantitative Analysis, 52, 465-489.