

# AFFECT, MEDIA AND EARTHQUAKES: DETERMINANTS OF CRASH BELIEFS FROM INVESTOR SURVEYS\*

William N. Goetzmann<sup>a</sup>

Dasol Kim<sup>b</sup>

Robert J. Shiller<sup>c</sup>

December 8, 2017

**Abstract:** Historical data suggest that the base rate for a severe, single-day stock market crash is relatively low. Surveys of individual and institutional investors, conducted regularly over a 26-year period in the United States, show that they assess the probability to be much higher. We examine factors influencing investor responses and test the role of media influence, finding evidence consistent with an availability bias. Adverse market events made salient by financial press are associated with higher subjective crash probabilities. Exogenous shocks related to earthquakes are also associated with higher probabilities. Finally, subjective crash probabilities are negatively associated with mutual fund flows.

**Keywords:** Crash Beliefs, Availability Bias, Investor Surveys

**JEL codes:** G00, G11, G23, E03, G02

---

\* We thank the U.S. National Science Foundation, Whitebox Advisors, and the International Center for Finance at the Yale School of Management for support with the survey data. We thank Leigh Ann Clark, Sumithra Sudhir and Minhua Wan for help with the data. We thank Christa Bouwman, Brian Knutson, Alan Moreira, Tyler Muir, John Scraggs, Hersh Shefrin and Paul Slovic for their suggestions. We thank participants in the JOIM (2016) conference on the Past, Present and Future of Behavioral Finance, and participants in the Columbia program for Financial Studies (2016) News and Finance Conference. The authors take responsibility for all errors. Dasol Kim started the paper prior to joining the Office of Financial Research; the opinions expressed are those of the authors and do not necessarily reflect those of the Office of Financial Research or the U.S. Department of the Treasury. All errors are ours alone.

<sup>a</sup> **William N. Goetzmann; Yale School of Management, Yale University.** Email: [william.goetzmann@yale.edu](mailto:william.goetzmann@yale.edu)

<sup>b</sup> **Dasol Kim; Office of Financial Research, U.S. Department of Treasury.** Email: [dasol.kim@ofr.treasury.gov](mailto:dasol.kim@ofr.treasury.gov)

<sup>c</sup> **Robert J. Shiller (corresponding author); Yale University.** Address: Economics Department, Yale University, Box 208281, New Haven, Connecticut 06520; Email: [robert.shiller@yale.edu](mailto:robert.shiller@yale.edu)

## 1. Introduction

Disaster risk and concerns about severe stock market crashes are the subject of considerable recent research. Rare disaster concerns are relevant to the equity premium puzzle,<sup>1</sup> time-varying market premiums,<sup>2</sup> cross-sectional differences in asset returns,<sup>3</sup> the volatility smile,<sup>4</sup> and investor choice.<sup>5</sup> Despite their potential importance, rare disaster concerns are difficult to empirically quantify. Probabilities about extreme events are usually inferred from asset prices, and disentangling probabilities from risk preferences presents problems.<sup>6</sup>

In this paper, we turn to a different source of information about rare crash probabilities. Since 1989, Robert Shiller has been surveying individual and institutional investor attitudes. One question in the survey asks respondents to estimate the probability that a severe crash will occur over the next six months. The definition of a crash is specific: a drop in the U.S. stock market on the scale of October 19<sup>th</sup>, 1987 [-22.61%] or October 28<sup>th</sup> 1929 [-12.82%].<sup>7</sup> This definition is particularly relevant to the jump tail risk literature. Bollerslev and Todorov (2011) and Bollerslev, Todorov and Xu (2015) argue that a significant component of priced tail risk is attributable to investor fears about a near-instantaneous crash similar to the one-day drops of 1929 and 1987. A key question in this work and related literature is whether asset prices reflect probabilities or preferences. As Ross (2015) puts it, “State prices are the product of risk aversion—the pricing kernel—and the natural probability distribution.”

---

<sup>1</sup> Cf. Reitz (1988), Barro (2006), Berkman et. al. (2011) and Welch (2015), Santa-Clara and Yan (2010).

<sup>2</sup> Cf. Gabaix (2008), Wachter (2013), Tsai and Wachter (2015) and Manela and Moreira (2016).

<sup>3</sup> Cf. Gao and Song (2015)

<sup>4</sup> Cf. Bollerslev and Todorov (2011), Bates (2000)

<sup>5</sup> Cf. Guerrero et.al. (2015).

<sup>6</sup> Cf. Jackwerth et. al. (1996), Seo and Wachter (2013) and Ross (2015).

<sup>7</sup> The figures are based upon the one-day percentage change in the Dow Jones Industrial Average index from the close the day before to close.

We use the Shiller survey data to examine the magnitude of crash probabilities reported by individual and institutional investors. We find evidence that the average, subjective probability of an extreme, one-day crash on the scale of 1987 or 1929 [i.e., greater than 12.82%] to be an order of magnitude larger than would be implied by the historical frequency of such events in the U.S. market. Over the 1989-2015 period, the mean and median probability assessments of a one-day crash were 19% and 10%, respectively. To the extent that this rare crash risk fear is priced, our analysis suggests that it may function through extreme probability assessments rather than through risk aversion.

We find that crash probabilities vary significantly through time and are correlated to measures of jump risk such as the VIX and the occurrence of extreme negative returns. We also test behavioral hypotheses about whether investor priors are subject to the influence of the media.

In particular, we test for the incremental effects of media sentiment measured from articles about the market on the days prior to the survey. Positive and negative sentiment are measured for each article using sentiment analysis, and the information is aggregated across articles to capture media sentiment on a particular day. We find evidence that the financial press mediates investor crash beliefs, and does so asymmetrically. Articles with negative sentiment are associated with higher crash probability assessments following market downturns, but articles with positive sentiment have no effect. We explore the question of whether this association operates through the availability heuristic, and extent to which affect plays a role. We find evidence consistent with an availability bias when examining the crash probabilities of investors who recently experienced exogenous rare events. Specifically, we study a case where the direction of causality can be clearly stated, moderate earthquakes in the respondents' locality, and where the rational national market response practically cannot be significant.

The availability heuristic (cf. Tversky and Kahneman (1973) & Kahneman and Tversky (1982)) is the tendency to use easily recalled events to estimate the probability of an event occurring. Subjects prone to the availability heuristic “bias” their probability beliefs by giving more weight to “top-of-mind” data. Tversky and Kahneman (1973) tests show that it is possible to induce this bias through priming or framing. Studies of the availability heuristic have mostly focused on stock price reactions to information. Akhtar et. al. (2013) document an asymmetric response of stock prices to the release of consumer sentiment news. They report evidence consistent with the availability heuristic – inferring shifts in probability assessment from asset price changes. Kliger and Kudryavtsev (2010) likewise rely on the asymmetry implied by negativity bias to test the availability heuristic. They find that stock price reactions to analyst upgrades are weaker on days of large market moves. Taking a different tack, Nofsinger and Varma (2013) use investor decisions to test for the availability bias. They argue that the investor tendency to repurchase a stock previously held is evidence of reliance on the availability heuristic. The contribution of our study to research on the availability heuristic in finance is that we directly test its relationship to probability estimates; the setting in which the hypothesis was originally formulated by Tversky and Kahneman (1973).

The availability heuristic is particularly pertinent to investment decision-making because probability assessment of events – for example, the likelihood of tail risk events – affects investor allocations to risky assets.<sup>8</sup> If investors give too much weight to recent market events – perhaps because they look at recent investment outcomes – this may cause them to incorrectly estimate the probability of a crash. By the same token, the media may frame recent events through selective

---

<sup>8</sup> Cf. Barberis (2013).

reporting – emphasizing negative outcomes and thus making them more available when a subject is asked to assess the probabilities of a related event.

We find evidence that investors use recent market performance to estimate probabilities about a crash. We also find that the press makes negative market returns relatively more salient and this is associated with individual investor probability assessments of a crash. This latter mechanism is consistent with Barber and Odean (2008), Engelberg and Parsons (2011), Kräussl and Mirgorodskaya (2014), Yuan (2015) and other research documenting evidence that the news plays an important role in focusing investor attention and influencing behavior. Moreover, we provide supportive evidence that investor crash beliefs have meaningful consequences on aggregate portfolio allocations. The survey-based crash probabilities are negatively associated with future flows to equity-based mutual funds. We use plausibly exogenous variation in the crash probabilities based upon differences between individual and institutional investors that are unlikely to be driven by fundamental factors.

Johnson and Tversky (1983) observe that “judgments about risk... seldom occur in an emotionally neutral context.” They find that emotion induced by brief reports about negative events have a major effect on probability assessments. This has come to be termed the affect heuristic (Slovic et. al. (2007)). Paul Slovic and co-authors, as well as other researchers (cf. Keller et. al. (2006)) have explored how emotions – particularly fear and dread – influence probabilities assessment. It is interesting from a finance perspective that their research identifies an inverse relationship between expectations of risk and reward – the opposite implied by standard financial models. The affect heuristic is similar to the “risk-as-feelings” model proposed by Lowenstein et. al. (2001), who propose that risk is perceived experientially and is therefore subject to the broad

variety of factors known to influence emotions, including vividness of outcomes, personal experience and mood. The affect and availability models are not necessarily mutually exclusive.

The balance of the paper is organized as follows. Section 2 describes the data used in the analysis. Section 3 presents the empirical findings and a number of robustness tests. Section 4 concludes.

## **2. Data**

### *2.1. Survey Data*

Robert Shiller's Stock Market Confidence Indices are based on survey data collected continuously since 1989; semi-annually for a decade and then monthly by the International Center for Finance at the Yale School of Management since July, 2001. Shiller (2000a) describes the indexes constructed from these surveys and compares them to other sentiment indicators and studies their dynamics in the aggregate. In this paper we use the disaggregated survey responses that are used to construct the indexes. About 300 questionnaires each month are mailed to individuals identified by a market survey firm as high-net-worth investors and institutional investors. They may fill it in when they wish, but they are asked to mark the date on which they complete the survey. It is not a longitudinal survey – each month comprises a different sample of respondents with the sampling goal of 20 to 50 responses by each of the two types – individual and institutional. There is existing research that uses data from the Shiller surveys. Greenwood and Shleifer (2015) find that the Shiller monthly investor confidence index is well-correlated to several other investor surveys and to mutual fund flows. Barone-Adesi et. al. (2015) estimate behavioral pricing kernels from market data and find them to correlate well to investor sentiment

surveys, including indexes constructed from the Shiller survey data used in this paper. Goetzmann et. al. (2014) use the institutional investor responses from a telephone version of the survey about beliefs in market mispricing in order to study variation in investor mood. Their results are consistent with evidence derived from a different dataset of investor trading behavior.

In the current study, we use responses to the survey question:

*“What do you think is the probability of a catastrophic stock market crash in the U. S., like that of October 28, 1929 or October 19, 1987, in the next six months, including the case that a crash occurred in the other countries and spreads to the U.S.? (An answer of 0% means that it cannot happen, an answer of 100% means it is sure to happen.)*

*Probability in U. S.: \_\_\_\_\_%*

The phrasing of this question has not been significantly altered during the sample period we examine.<sup>9</sup> Thus it has the advantage of consistency throughout a period of 26 years, during which time the stock market, the macro-economy and the financial system has experienced considerable variation. In addition to the responses to the questions, survey participants provided the date on which the questionnaire was completed. Information about the ZIP codes of the respondents is readily available from 2007. The combined sample contains 9,953 responses.

One issue to consider is that the phrasing of the question may make a crash salient and lead to a heightened probability assessment. The term “catastrophic” and the highlighting of the two crash dates may themselves trigger a response biased towards higher probability. The high-

---

<sup>9</sup> This wording has remained the same since 1994. Prior to 1994, the question is phrased as: “What do you think is the probability of a catastrophic stock market crash, like that of October 28, 1929 or October 19, 1987, in the next six months?” Only approximately 10% of the observations used in the analysis are associated with the earlier wording. The results are not sensitive to the exclusion of these observations.

sentiment term “catastrophic” could make investors more pessimistic. By the same token, highlighting two crashes out of a century or more of data could trigger an availability heuristic. There are several other questions in the surveys – some with positive and some with negative sentiment; all about the stock market. These may also prime an investor response. These stimuli make it potentially difficult to identify the marginal influence of news articles on probability assessments. Another feature of the question is that it relies partially on a narrative about an event occurring in other countries and spreading to the U.S. Experimental evidence suggests that people rely on numerical and narrative evidence in assessing probabilities, and the relative degree of reliance may depend on numeracy (Dieckmann et. al. 2009).

Given the affective and narrative features of the question, prior research suggests that we should find cross-sectional differences among respondents based on their numerical sophistication and perhaps other factors. If the high base-line probability assessments are due solely to framing factors within the questionnaire, this would suggest that direct priming may be a source of extreme bias about the probabilities – an interesting fact in itself. Later in the paper, we return to the issue of whether responses may be attributable to affect, availability or an interaction of both tendencies.

Figure 1 graphs the average annual probabilities for the individual and institutional respondents. It also shows a set of additional variables: the annualized volatility of the daily DJIA, the largest negative return in each year (represented as a positive number on the figure) and the VIX implied volatility. The individual and institutional means are relatively similar. Crash probabilities were higher in the period 1997-2003 and 2007-2011. These periods also correspond to higher realized volatility, implied volatility and most extreme one-day DJIA percentage declines. These trends suggest that the probability assessments change with factors associated with extreme market declines. Not shown in the figure are probabilities inferred from historical



market performance. It is well known that a log-normal model is not appropriate to estimate the probability of an extreme decline. The average daily standard deviation of the DJIA is about 1% and the two crashes of interest are 12 times and 20 times the daily standard deviation. This has motivated the use of mixed jump processes to describe stock market moves.<sup>10</sup>

A simple approach to estimating a baseline probability is to use the historical frequency of such events. Under the assumption of an i.i.d. distribution of daily returns, and using the number of trading days since October 23, 1929 through December 31, 1988 [taking the most conservative bounds] gives an average probability of an extreme crash over a six-month horizon of 1.7%. This decline to approximately 1% when the entire history of the DJIA is used. The average reported crash probability from the Shiller surveys is thus more than 10 times the conservative estimate.

## 2.2. *Market Data*

For stock market data, we use daily data on the Dow Jones Industrial Average, the S&P 500 and a value-weighted index of the NYSE-AMEX-Nasdaq-Arca universe. The daily returns of each index are used to empirically measure market volatility and the occurrence of extreme events. We also use the returns to the indices on and before the day that the questionnaires are completed as a control for market trends that jointly influence media articles and investor heuristics. Market volatility implied by the VIX is obtained from Federal Reserve Economic Data (FRED).

---

<sup>10</sup> Cf. Gabaix (2008), Wachter (2013), Bollerslev and Todorov (2011).

### 2.3. *Media Data*

We use ProQuest to search the Eastern Edition of the Wall Street Journal [WSJ] for the period of the questionnaire sample: 1989-2015. This is the only edition available on ProQuest for that period. We presume that it corresponds reasonably well to the national edition. Data were collected in the weeks of May 17 through June 11 of 2016. We searched articles containing words and phrases associated with the stock market, yielding a total of 133,496 articles with a word count of at least 250 words.<sup>11</sup>

Garcia (2014) documents a significant asymmetry in media reportage of past market returns – negative outcomes are reported more frequently in certain Wall Street Journal columns.<sup>12</sup> This is consistent with evidence that both animals and humans are conditioned to give stronger weight to negative things, experiences and events (cf. Baumeister et. al., 2001 and Rozin and Royzman, 2001). Negative experiences engage greater cognitive effort (Ito et. al., 1998), have greater influence in evaluations (Ito et. al., 1998), are more likely to be taken as valid (Hilbig, 2009) increase arousal, and enhance the memory and comprehension of the event (Grabe and Kammhawi, 2006). These prior results lead us to expect that the availability bias – if it exists – should be asymmetric. Negative events should have a greater effect on probability assessments than positive events.

---

<sup>11</sup> The Proquest search term used to identify the articles is: “(stock NEAR/5 market) OR SU(stock) OR SU(securities)”. We did not use broader search terms, such “SU(markets)”, because they yielded articles on other asset markets, such as for bonds and commodities.

<sup>12</sup> Garcia (2014) focuses on financial market columns from the New York Times and the Wall Street Journal from 1905 to 2007. These columns may not necessarily appear on the front page, where articles are more likely to be viewed by readers. For example, the “Abreast of the Market” column in the Wall Street Journal, which is used in Garcia (2014), appears on the front page of a section 40.8% of the time over our sample period.

For each article, sentiment is defined as a function of the number of positive and negative words.<sup>13</sup> Media sentiment is measured by aggregating the sentiment of articles for each date. We classify words using the General Inquirer (GI) – a lexicon of positive and negative words that is widely used in text sentiment analysis. A number of prior studies use the GI to construct media sentiment measures for newspaper articles.<sup>14</sup> Other approaches include constructing lexica tailored to specific types of source documents. For example, Loughran and McDonald (2011) manually code common terms found in SEC regulatory filings into valence classes, as some terms may have different connotations in financial documents compared to other documents. Our main results are not sensitive to the choice of the word list.<sup>15</sup>

We follow Loughran and McDonald (2011) in weighting word frequencies using the “tf-idf (term frequency-inverse document frequency)” method which accounts for a word’s relative prevalence within and across documents. Using un-weighted word counts does not control for the fact that some words are simply more common than others (cf. Manning, Raghavan, and Schütze (2008)).<sup>16</sup> Our use of the tf-idf weighting scheme is motivated by Loughran and McDonald’s

---

<sup>13</sup> The General Inquirer assigns approximately 10,000 words to 26 major and 182 minor categories, or tags. It aggregates categories from the Harvard IV-4 dictionary, Lasswell value dictionary, a social cognition dictionary from Semin and Fiedler (1988). The two largest categories are the positive and negative valence classes: the positive list includes 1,915 words, while the negative list includes 2,291 words.

<sup>14</sup> These studies include Tetlock (2007), Engelberg (2010), and Garcia (2013). There are a large number of studies that apply sentiment analysis to firm regulatory disclosures. Loughran and McDonald (2016) provides an overview of this literature.

<sup>15</sup> In assessing the robustness of the choice of word lists, we focused on alternatives that are derived from different types of document sources. Loughran and McDonald (2011) manually codes words that appear in SEC regulatory filings for positive (354 words) and negative (2,355 words) categories, as well as others. Hu and Liu (2004) automates term coding for valence categories (2,006 positive and 4,783 negative words) using a large collection of customer reviews. An earlier version of this paper used article counts containing a few specific words such as “crash” and “boom.” These results and the full list of search terms are available in the prior working paper of the same name.

<sup>16</sup> Specifically, the weighted word frequency for word  $w$  appearing in article  $a$  is calculated as the product of the log-scaled word frequency and the log-inverse document frequency:  $\frac{1+\ln(n_{w,a})}{1+\ln(\bar{n}_a)} \ln\left(\frac{A}{df_w}\right)$ , where  $n_{w,a}$  is the frequency for word  $w$  in article  $a$ ,  $\bar{n}_a$  is the average frequency for all words appearing in article  $a$ ,  $A$  is the total number of articles used in the analysis, and  $df_w$  is the number of articles containing word  $w$ . Words that do not appear in the article ( $n_{w,a} = 0$ ) are assigned a value of zero. See Loughran and McDonald (2016) for a complete discussion and survey of methods using word frequency in textual analysis to measure sentiment.

(2011) finding “...that this approach [tf-idf] produces regressions with better fit than the approaches using simple proportions.”

The first set of media sentiment measures relates to daily counts of articles coded as negative or positive. For each article, we calculate a score based upon the difference between weighted positive and negative word frequencies. To mitigate the influence of articles with larger word counts, the difference is scaled by the sum of the weighted positive and negative word counts.<sup>17</sup> Articles are coded as negative and positive based upon whether their score is in the bottom and top 10<sup>th</sup> percentile over all the articles used in the analysis, respectively. Finally, we count the number of articles that are coded as negative (*Negative*) and positive (*Positive*) for each date.

The second set is based upon a score that aggregates information across articles for each date. Articles counts may be useful for distinguishing between the effect on positive versus negative counts. However, the *Positive* and *Negative* measures are based upon the extreme article-level scores, and may not account for information from articles with non-extreme values. We define *Sentiment* as the difference between the weighted positive and negative frequencies across all articles for each date, scaled by the sum of the weighted positive and negative word frequencies across all articles for each date.<sup>18</sup> For the main analysis, we consider both sets of measures, though we focus mainly on the *Sentiment* measure.

---

<sup>17</sup> Normalization to adjust for document size is common in sentiment analysis. The term weighting scheme we use mitigates document size effects to some extent, though it may be inadequate for some cases. We use the sum of the weighted positive and negative word frequencies to simplify comparisons between positive and negative sentiment, ignoring neutral sentiment. However, the results are not sensitive to normalizing based upon total word counts, which includes neutral sentiment terms.

<sup>18</sup> We also consider a number of other specifications: the weighted negative word count divided by the total weighted word count; the difference between the positive and negative word count divided by the total weighted word count; and the difference between the natural log of one plus the weight positive word count and the natural log of one plus the weighted negative word count. The results are not sensitive to these alternative specifications.

Figure 2 graphs the average daily *Sentiment* and the average daily number of articles related to the stock market for each year in the sample period. The sentiment scores reach their peak during the dot-com boom and just before the financial crisis, while the average daily number of articles peaks during the dot-com boom and during the financial crisis. The correlation between *Sentiment* and article counts is close to zero over the entire sample period. Nonetheless, we control for total article counts in all the regression models.

### **3. Empirical Results**

#### *3.1. Summary Statistics*

Table 1 displays the variable descriptions and summary statistics. The interquartile range of the stock indices are comparable, through the overall range for the CRSP-VW and S&P 500 indices are slightly larger than that of the DJIA index. The mean and median of the subjective probabilities are reported. They are 10% and 18.5%, respectively, indicating a positive skew.

#### *3.2. Media Responses to Market Events*

We begin by examining asymmetry in the effects on media sentiment of positive and negative market events. As a preview of the results, we show that negative returns in the prior day are associated with significantly higher negative article counts, and positive returns are associated with significantly higher positive article counts, although the positive article count results are smaller in absolute magnitude. We find consistent results when examining the *Sentiment* directly.

There are significant coefficients on volatility, extreme returns, prior month returns, and past media sentiment.

We estimate the following econometric specification:

$$V_t = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-2} + \alpha_3 r_{t-30,t-3} + \alpha_4 \sigma_{t-30,t-3} + \alpha_5 \text{Sentiment}_{t-30,t-1} + \alpha_6 \ln(1 + \#Articles)_t + FE(\text{Day of Week, Month})_t + \varepsilon_t \quad (1)$$

where  $V_t$  represent the media sentiment measures. The analysis examines the two sets of measures: article counts based upon sentiment and sentiment scores.

We focus on the effects of the market return on the prior day ( $r_{t-1}$ ). Regular trading hours on major exchanges end at 4 PM Eastern Standard Time, so that news about market events on date  $t-1$  is likely to appear in print on date  $t$ . The return variables are specified so that the calculation periods are not overlapping. We control for other market or fundamental factors related to market returns on date  $t-1$  that may also influence media sentiment. Specifically, we control for the market return on two days prior ( $r_{t-2}$ ), the market return over the period  $t-30$  to  $t-3$  ( $r_{t-30,t-3}$ ), the daily volatility estimated over the period  $t-30$  to  $t-3$  ( $\sigma_{t-30,t-3}$ ), the average daily media sentiment over the period  $t-30$  to  $t-1$  ( $\text{Sentiment}_{t-30,t-1}$ ), and the natural log of the total number of articles that are related to the stock market on date  $t$  ( $\ln(1 + \#Articles)_t$ ). In all models, we include day-of-the-week and month fixed effects. Robust standard errors that account for heteroskedasticity and autocorrelation in the residuals with optimal lag length (Newey and West (1994)) are used to calculate the test statistics.

In alternative specifications, we replace  $r_{t-1}$  in equation (1) with  $D_{(rk(r(t-1)) < 10\%)}$ , a dummy variable for whether the prior day return is in the bottom sample decile, and

$D_{(rk(r(t-1))>9\%)}$ , a dummy variable for whether the prior day return is in the top sample decile. By construction, the sample frequency of negative and positive market events is the same. This allows us to directly assess asymmetry in the effect of negative and positive market events on media sentiment. To alleviate data-mining concerns, we estimate the regression separately where the market returns are proxied by the value-weighted index of NYSE/AMEX/Nasdaq/Arca (CRSP-VW), the Standard and Poor's 500 (S&P 500) and the Dow Jones Industrial Average (DJIA).

To start, we focus on the positive and negative article counts.<sup>19</sup> Table 2 displays the results. Columns (1), (3) and (5) report results for *Negative*, while Columns (2), (4) and (6) report results for *Positive*. The model fit across all the specifications range from 44.0% to 44.9%, suggesting that the media have an independent influence on the market, beyond just mechanically reporting past returns. Across the different return variable specifications, the previous day return is negatively associated with negative article counts and positively associated with positive article counts. This is consistent with the news reporting the direction of the prior-day market return. The estimates are considerably generally larger in absolute magnitude for the negative than positive article counts, consistent with asymmetry in the association. The regressions also indicate an association between returns two days prior and prior month returns, though only for negative article counts. Prior month volatility has a positive association and prior month *Sentiment* has a negative association with negative article counts, and these relationships are opposite to those for positive article counts. The coefficients on these variables are all consistent with the media responding to market trends and with the sentiment scores calculated on the newspaper articles

---

<sup>19</sup> The range of *Positive* and *Negative* are non-negative, potentially raising concerns on the consistency of the OLS regression model estimates. To address this issue, the tests are re-estimated using tobit models that account for the range of the dependent variable. Table A.1 displays the results. The model estimates are slightly larger in magnitude compared to those in Table 2. The results provide evidence that the findings are not sensitive to the model specification.

used in the analysis as meaningful measures of media sentiment. The results are not specific to the use of any particular market return measure.

We next turn our attention to tests that use *Sentiment* as the dependent variable.<sup>20</sup> Table 3 presents the results in Columns (1), (3) and (5).<sup>21</sup> The results are consistent with those in Table 2. Previous day returns are negatively associated with *Sentiment*. Columns (2), (4) and (6) present similar results using dummies based upon whether the previous day return is in the top or bottom sample decile. We include this specification to directly assess asymmetry in effects of negative and positive prior day returns on media sentiment. It also addresses the possibility that only unusual market moves are deemed newsworthy. Moreover, this specification mitigates the effect of outliers driving the return-based results – i.e., a few extreme market moves accounting for the significance of the coefficients. The low return dummy is significant and its estimate is considerably larger than that of the high return dummy. The marginal effects of the low return dummy on *Sentiment* account for 36.5% to 39.4% of the overall sample variation for *Sentiment*, which is economically meaningful.<sup>22</sup> In contrast, the marginal effects of the high return dummy account for 4.2% to 7.0% of the overall sample variation for *Sentiment*. The differences in the coefficients are also significant across the specifications, consistent with asymmetry in the return association with media sentiment. Similar results obtain when also controlling for expected returns volatility (see Table A.2).

---

<sup>20</sup> It is possible that the *Sentiment* measure may incorporate reader responses, such as reader letters to the editor, to prior day stock performance. However, analogous measure that only incorporate front page articles, which is unlikely to feature reader content, or lead paragraphs yield similar results.

<sup>21</sup> The range of *Sentiment* is between -1 and 1, potentially raising concerns on the consistency of the OLS regression model estimates. To address this issue, the tests are re-estimated using tobit models that account for the range of the dependent variable. Table A.1 displays the results. The model estimates are virtually identical to those in Table 3. The results provide evidence that the findings are not sensitive to the model specification.

<sup>22</sup> For example, the low return dummy coefficient in Model (6) of Table 3 (-0.028) divided by the sample standard deviation for *Sentiment* from Table 1 (0.071) is -0.394 ( $= -0.028 / 0.071$ ).



Next, we examine the persistence of the relationship between media responses and market returns. Table 4 displays the results. For each set of return variables, future *Sentiment* is projected onto the return variable, while using the same control variables in Table 3, for up to ten lags. We only present the results for the extreme return dummies in order to focus on the effects of negative and positive shocks. Consistent results (untabulated) obtain when using total market return instead of a dummy. The association between low returns and media sentiment appears to persist for at least ten days subsequently, while the association of high returns with media sentiment subsides within one day. However, when compared to the results of Table 3, the effects are relatively pronounced for the date immediately following the market event.

We are also able to identify the lead paragraph of each article and whether the article appears on the front page or elsewhere in the newspaper. This allows us to further test whether negative events are accorded greater prominence by the media. By controlling for sentiment of content from non-lead paragraphs or articles that do not appear on the front page, these tests also better account for the effects of fundamental factors than those in Table 3. We assume that fundamental factors are likely to influence sentiment within each article or across articles on the same date similarly. To that end, we reconstruct *Sentiment* based upon whether (i) the text appears in the lead paragraph ( $Sentiment^{Lead}$ ) or (ii) whether the article appears on the front page ( $Sentiment^{Front}$ ). In the  $Sentiment^{Lead}$  models, we also include the sentiment scores based upon text from non-lead articles to assess whether the return variables have any remaining explanatory power. Similarly, in the  $Sentiment^{Front}$  models, we also include the sentiment scores based upon articles that are not on the front page. Significance in the return variables would suggest that the market return has an effect on editorial decisions about sections of the newspaper that are likely to be more widely viewed.

Table 5 reports the results. Controlling for the sentiment of non-lead paragraphs, prior day returns have a strong negative association with the sentiment of lead paragraphs. As before, the effects are considerably larger for negative versus positive return days, and the difference in the absolute magnitudes are statistically significant. This difference may be due to a negativity bias by the media, or it may reflect a recognition that negative news is more engaging to readers, and/or relevant to investors – and thus will sell more papers. Importantly, these incentives may have an effect on investor beliefs beyond what would be implied by fundamentals. Similar results obtain for articles placed on the front page, controlling for the sentiment of articles not placed on the front page. We argue that the newspaper’s readership pays closer attention to lead paragraphs and front page articles. We test whether these associations affect the survey responses to questions about crash beliefs in the next section.

A related question is prior day returns determine whether articles on the stock market are more likely to have negative sentiment and appear on the front page. To address this question, we perform tests on the article-level. We estimate probit regression models where the dependent variable is a dummy associated with whether the article is coded as negative and appears on the front page. The key explanatory variable is the prior day return, and the control variables are identical to those in Table 5. Table A.3 displays the results. Column (1) shows that the coefficient for prior day returns is negative and statistically significant at the 1% level. When restricting the sample to only front page articles, we find similar results (Column (2)). The results suggest that articles with negative sentiment are more likely to appear on the front page following negative return days.

Finally, we examine whether articles with negative sentiment are more likely to appear on the front page over articles with positive sentiment. We estimate probit regression models where

the dependent variable is a dummy associated with whether the article appears on the front page. The key explanatory variables are dummies based upon whether the article is coded as negative or positive, and the control variables are identical to those in Table 5. Table A.4 displays the results. The results show that negative articles are 3.2 percent more likely to appear on the front page (Column (1)), while positive articles are 18.8 percent less likely to appear on the front page (Column (2)), compared to articles that are neither coded as negative or positive, or neutral sentiment articles. The results are comparable when including both in the same model (Column (3)). What is interesting is that the asymmetry in the effects – the likelihood that positive articles do not appear on the front page is greater than that of negative articles to appear in terms of absolute magnitude. The results suggest that front page coverage skews towards articles with negative or neutral sentiment.

### *3.3. Crash Probabilities*

In this section, we test whether media sentiment is a factor influencing the crash probabilities from the survey. The availability bias predicts that investors will overweight recent information in forming crash probabilities. We consider a set of events that may be particularly salient to investors: media sentiment during market downturns and upswings. Given that surveys are sent to both individual and institutional investors, we further expect the effects of availability to be more pronounced for unsophisticated investors (e.g., individual investors).

Before turning to the tests, we consider the internal validity of the crash probabilities with the other survey responses. One potential concern is that the crash probabilities may reflect severity due to the wording of the question rather than their actual beliefs. This would suggest

inconsistency between the crash probabilities and the responses to other questions where the wording is unlikely to be confused with severity. We assess this concern in two ways.

First, we estimate a model that examines whether the crash probabilities can be explained by responses to the questions related to stock market valuation. The key explanatory variables include the expected percentage change in the DJIA over the next six months (*Expected Returns*); a dummy based upon whether stock prices when compared to true fundamental value is too high ( $D_{TooHighValuation}$ ); a dummy based upon whether stock prices when compared to true fundamental value is too low ( $D_{TooLowValuation}$ ); a dummy based upon whether the investor is inclined to sell stocks overall ( $D_{Sell}$ ); and a dummy based upon whether the investor is inclined to buy stocks overall ( $D_{Buy}$ ).<sup>23</sup> Table A.5 displays the results. We estimate the model separately for individual (Panel A) and institutional (Panel B) investors. The explanatory variables are mostly significant, and their coefficients are signed in a manner consistent with the crash probability responses. The magnitudes are comparable across the individual and institutional investors subsamples. The results suggest internal validity for the crash probabilities.

Second, we directly test whether specific words in the survey question influences the crash probability responses. Different versions of the survey question that alter keywords that may potentially prime severity in the question phrasing are randomly presented to 500 subjects using Amazon's Mechanical Turk.<sup>24</sup> We estimate a regression model where the dependent variable is

---

<sup>23</sup> The explanatory variables are constructed from the following questions: (i) "How much of a change in percentage terms do you expect in 6 months [for the Dow Jones Industrial Average]?", (ii) "Stock prices in the United States, when compared with measures of true fundamental value or sensible investment value, are [too high, too low, or about right].", and (iii) "Are you inclined to buy stocks overall, or sell stocks overall, or hold steady?"

<sup>24</sup> Specifically, we consider four versions of the survey question: the original question; one that excludes the words "catastrophic" and replaces the word "crash" with "decline"; one that replaces "like that of October 28, 1929 or October 19, 1987" with "of 20%"; and one that excludes the words "including the case that a crash occurred in the other countries and spreads to the U.S.".

the crash probability response and the explanatory variables are dummies associated with keywords, along with date fixed effects.<sup>25</sup> We do not find any significant differences in the crash probabilities across the different versions.<sup>26</sup> Overall, these results provide evidence that the crash probability responses are unlikely to be driven by confusion over the wording of the survey question.

We assess the influence of media sentiment on investor crash probabilities by estimating the following econometric specification:

$$\pi_{i,t} = \beta_1 r_{t-1} \times V_t + \beta_2 r_{t-1} + \beta_3 V_t + \beta_4 r_{t-30,t-2} + \beta_5 \sigma_{t-30,t-2} + \beta_6 \pi_{t-30,t-1} + \beta_7 \text{Sentiment}_{t-30,t-1} + \beta_8 \ln(1 + \#Articles)_t + FE(\text{Day of Week, Month})_t + \xi_{i,t} \quad (2)$$

where the dependent variable  $\pi_{i,t}$  is the probability assessment of investor  $i$  at time  $t$ .

The tests focus on the effects of media coverage of market returns from the previous day on investor responses on the same date the investor fills out the survey. The explanatory variables are as specified above, and include prior day returns; media sentiment; and the interaction term between returns and media sentiment. We focus on the interaction term. Media sentiment may be negative in some cases due to reasons unrelated to market conditions, so that media sentiment during market downturns and upswings may be more informative. Moreover, investors may be more likely to associate market events with their crash beliefs when media coverage is congruently valanced.

---

<sup>25</sup> The survey was conducted over three dates: July 14, July 15, and July 16, 2016.

<sup>26</sup> In particular, the crash probabilities for the versions that exclude the words “catastrophic” and replaces the word “crash” with “decline” (*estimate* = 0.007, *t-value* = 1.22); replace “like that of October 28, 1929 or October 19, 1987” with “of 20%” (*estimate* = 0.013, *t-value* = 0.57); and exclude the words “including the case that a crash occurred in the other countries and spreads to the U.S.” (*estimate* = 0.010, *t-value* = 0.63) are not significantly different from those of the original question.

We also include control variables related to media sentiment over the past month, past month returns, the average crash probabilities of other investors of the same type over the previous 30 days, past month market volatility, and the total number of articles published over the past month. All models include day-of-the-week and month fixed effects. Robust standard errors clustered by date level are used to calculate the test statistics.

Table 6 presents the results when using the positive and negative articles for the media sentiment terms. The odd-numbered columns report the results for the individual investor subsample, while the even-numbered columns report those for the institutional investor subsample. The results are reported separately where the return variables are based upon the CRSP-VW (Columns (1) and (2)), S&P 500 (Columns (3) and (4)), and DJIA (Columns (5) and (6)) indices for completeness.

The results indicate that, regardless of the market index used, individual investors' responses are significantly associated with negative article counts during market downturns. Only the interaction term for the negative article counts is statistically significant. In contrast, the interaction terms are not significant for institutional investors. In terms of magnitudes, the interaction term effects are approximately five times larger for individual investors. The marginal effects of a 5% decline in the market and a standard deviation increase in the number of negative articles range 4.98% to 5.11% (0.82% to 1.04%) based upon the interaction terms across the individual (institutional) investor specifications. Positive article counts are not significantly associated with crash probabilities for either set of investors, while negative article counts are significantly associated with higher crash probabilities for only institutional investors. These results suggest that negative sentiment is only significant following market downturns. After controlling for the media sentiment terms, returns for the prior day do not have significant

coefficients. Past month returns are significant only for individual investors, where lower returns are associated with higher crash probability estimates.

The analysis next examines the results when using *Sentiment* for the media sentiment terms. Table 7 displays the results, and is formatted similarly to Table 6. Consistent with the results for the article counts, the interaction term for *Sentiment* is negative and statistically significant only for individual investors – negative sentiment during market downturns is associated with higher crash probabilities. The (two-tailed) null is rejected at the 10% level for the two broader indexes, and at the 5% level for the DJIA.<sup>27</sup>

To further test whether the variation in the probability estimates may be attributable to the availability heuristic, we augment the regression models of Table 7 to assess whether similar effects for media sentiment can be found following market rallies and declines over the previous 30 days. If investors use an availability heuristic, they are more likely to have the most recent return “top of mind”. In addition to the media sentiment interaction terms with same day returns, an interaction term is added between media sentiment and returns over the previous 30 days. We also control for interactions between news over the past month and the return variables. Media sentiment over the past month is likely to capture recent trends in fundamentals. Two additional interaction terms are added between past month media sentiment and returns over the previous day and month.

The results of Table 8 are similar but stronger for the same day return interaction terms. Again, the coefficients are only significant for individual investors. The interaction terms

---

<sup>27</sup> A potential explanation for the difference in statistical significance across the indexes is that the Wall Street Journal is a Dow Jones publication and the periodical that created and maintains the Dow Jones Industrial Average. This may incline WSJ reporters to write about the dynamics of the DJIA index as opposed to others.

associated with previous 30-day returns are statistically insignificant, and their coefficients are considerably smaller in absolute magnitude than those for the previous day returns. In contrast, coefficients for the interaction terms between past month media sentiment and past day returns are negative and statistically significant only for individual investors. The coefficients for the interaction terms between past month media sentiment and past month returns are positive and statistically significant only for institutional investors.

We next consider whether the significance of media sentiment varies according to sections that are more likely to attract attention. The availability heuristic would predict that the impact of recent market returns on the respondent's probability assessment would be more salient when mentioned in sections that the respondent is more likely to pay attention to. We decompose *Sentiment* according to whether (i) it is based upon lead paragraphs ( $Sentiment^{Lead}$ ) or other paragraphs ( $Sentiment^{NotLead}$ ), and (ii) it is based upon front page articles ( $Sentiment^{Front}$ ) or other articles ( $Sentiment^{NotFront}$ ). We expect the results to be pronounced for the media sentiment measures when constructed using lead paragraphs and front page articles.

Table 9 presents the results. Panel A displays the results for  $Sentiment^{Lead}$  and  $Sentiment^{NotLead}$ . Panel B displays the results for  $Sentiment^{Front}$  and  $Sentiment^{NotFront}$ . We find that the association between media sentiment and subjective probabilities during market downturns is driven by lead paragraphs and front page articles – interaction terms between past day returns and non-lead paragraphs, or non-front page articles, do not have a significant coefficient.

Prior work (Goetzmann et. al., 2014) documents the relationship between unusual cloud cover and investor forecasts of the stock market. Weather – for the most part -- is orthogonal to market conditions but has been shown to affect mood. To test whether the associations we have



found may be due to a general shift in mood related to market decline and media reporting, as opposed to conditioning on the market and the press, we tested whether the cloud cover variable had a significant association with subjective crash probability. The results (unreported) were insignificant.

### *3.4. Exogenous Rare Events*

One drawback thus far in the analysis is that media sentiment measures, stock market returns and individual probability assessments of a future crash may be jointly influenced by a common unidentified variable. This could be an economic event or condition that raises risk. We have included volatility measures in our regressions to capture this, but there may be other relevant variables we have omitted. In an efficient market, the price level of the stock market itself should capture value-relevant information, and thus the prior day's market return represents an adjustment to any potentially important but unidentified information. It still may be the case that the media interpretation of prior returns might reasonably convey information pertinent to crash probability assessment. Ideally, we would like a variable that puts the notion of a disaster "top of mind" but is orthogonal to the economy and stock market.

In this section, we examine the relationship between earthquakes and investor probability assessments. We exploit the ZIP code location of a subset of the Shiller survey respondents to identify regional events that plausibly make rare disasters more cognitively available. Since we focus on crash probabilities for the aggregate market, a moderate regional earthquake is unlikely to have economic relevance for a future stock market crash. Specifically, we use the occurrence of earthquakes whose epicenter is within 30 miles of the investor. The timing of earthquakes are

exogenous to current market conditions, but should be salient to individuals located close to the epicenter given that the earthquakes can be physically detected. Other studies have found that the realization of a low probability event increases subjective probabilities of the event occurring again, but also increases the subjective probabilities of other, unrelated events (Johnson and Tversky, 1983). We therefore expect earthquake events to induce overestimation of the likelihood of market crashes. This would support the role of affect in influencing the reliance on an availability heuristic.

Earthquakes of stronger magnitudes may have direct effects on economic conditions, which in turn can affect stock market conditions. However, the empirical evidence on this link is mixed. Ferreira and Karali (2015) show that stock markets do not exhibit a significant reaction within five days of strong earthquakes. In contrast, Shiller (2000b) points to the Tokyo stock market reaction to the earthquake in Kobe, Japan on January 17, 1995, which measured at 7.2 on the Richter scale. The Nikkei index fell by 8 percent overall within ten days of the earthquake, though significant price movements did not transpire until after one week. This market reaction far exceeded an official estimate, which was eventually set at approximately \$100 billion. While circumstantial, the delay in and magnitude of the market reaction to the Kobe earthquake suggest a mix of fundamental and sentiment factors.

To address potential confounding effects of the economic impact of stronger earthquakes, we distinguish between weak magnitude, or earthquakes with a magnitude between 2.5 up to 5.5, and strong magnitude, or earthquakes with a magnitude above 5.5. The cutoffs are based upon information from the United State Geological Services [USGS], which classifies earthquakes with magnitudes above 2.5 as physically detectable, and earthquakes with magnitudes above 5.5 as inflicting at least minor damage to buildings and other structures.

Earthquake data from 1900 to 2015 is collected from the USGS, and includes dates, magnitudes, and locations of each event. We match the earthquake data to the investor survey data using the centroid of the ZIP code location available for some of the survey respondents. From 2007, the survey includes the ZIP code of most of the survey respondents. Approximately 7.99% of the survey respondents experienced a weak earthquake within a 30-mile radius, while 0.2% experienced a strong earthquake.

To assess statistical significance, we construct the bootstrapped distribution of the model estimates by creating pseudo-events based generated using historical frequencies of earthquakes for each location from 1900-2006. The  $p$ -values for the model estimates based upon the actual events are estimated using 1,000 sets of pseudo-events.

Table 10 presents the results. Columns (1) through (3) display the results for the individual investor subsample, while Columns (4) through (6) present the results for the institutional investor subsample. We find that weak magnitude earthquakes have a positive and significant association with investor crash probabilities, but only for individual investors. The results are robust even after controlling for historical earthquake frequency for the investor's location, strong magnitude earthquakes, and the market-related control variables from Table 7. The coefficients on the strong magnitude earthquakes are generally statistically significant – perhaps due to the small number of such event in the sample.

### *3.5. Aggregate Fund Flows*

We next turn our focus to investor behavior by examining how subjective crash probabilities affect aggregate flows to mutual funds. While the results from the previous sections

show that financial media sentiment is strongly associated with investor beliefs related to stock market crash probabilities, it is unclear how the crash probabilities in turn affect investor decisions.

To investigate, we obtain daily data on mutual fund flows from TrimTabs from 2003 through 2015. We restrict the sample to U.S. Equity and U.S. Bond funds. We define daily percentage fund flows for each category group as the aggregate net fund flows scaled by the previous day's aggregate total net assets. The sample period coincides with the period when the surveys are administered with greater regularity, particularly for individual investors. Additional details about the data can be found in Goetzmann, Massa, and Rouwenhorst (2000).

We start by constructing a daily time-series of the investor crash probabilities ( $\pi$ ) from the survey. Given that some dates only have a single survey respondent, rolling, one-week moving averages are used to construct  $\pi$  with the requirement that there be at least five survey respondents over the estimation window.<sup>28</sup> For periods when this requirement cannot be satisfied, the last available value of  $\pi$  is used. We distinguish crash probability estimates by investor type: the moving average when only surveys for individual ( $\pi^{\text{Indiv.}}$ ) or institutional ( $\pi^{\text{Inst.}}$ ) investors are used for the crash probability measure.

There may be unobservable common factors related to fundamentals unaccounted for by our tests that may affect the crash probabilities of both investor types, potentially biasing the tests. If these factors affect both investor types similarly, then the difference in crash probabilities between these investor types should help mitigate the influence of omitted factors. As such, we also use the difference between  $\pi_t^{\text{Indiv.}}$  and  $\pi_t^{\text{Inst.}}$  ( $\pi^{\text{Indiv.-Inst.}}$ ) for the tests. In other words, while the

---

<sup>28</sup> The results are not sensitive to using alternative estimation windows, including a three day or a two week window.

crash probabilities of either investor type may be confounded by other fundamental factors, the difference is likely to be exogenous.

Table 11 presents the results. The dependent variable in each specification is the daily percentage fund flows for either equity (Panel A) or bond (Panel B) funds. For each panel, ten different models are estimated for each crash probability measure, or 30 models for all three measures. Columns (1) through (10) display the results when taking the forward lag of the dependent variable from one to ten days, respectively.<sup>29</sup> The same control variables and fixed effects used in the time-series regression models from Table 2 are included in all the models.

For equity fund flows, we find that only individual investor crash probabilities significantly predict future flows. While statistically insignificant for dates immediately following the estimation window, the coefficients for crash probabilities of individual investors are negative and statistically significant from date  $t+4$  and on. The magnitudes are largest for date  $t+6$  before declining. In contrast, the coefficients for crash probabilities of institutional investors are statistically insignificant for all periods considered. The results for the probability differences are similar to those for individual investors: the coefficients are all negative and are statistically significant from date  $t+5$  before becoming insignificant for date  $t+8$ . The results suggest that investors represented in the data may not immediately respond to changes in crash beliefs, but rather may respond gradually over time.

For bond fund flows, we find that both individual and institutional investor crash probabilities significantly predict future flows. The probability differences, however, is statistically insignificant for all periods. The coefficients for the crash probabilities of both

---

<sup>29</sup> Only trading days are considered given that these are the only days that the flow data is available.

investor types are positive and statistically significant for various periods considered, though the results appear stronger for institutional investors. While these results may be suggestive of investor reactions to higher stock market crash probabilities, the probability differences are statistically insignificant.

#### **4. Conclusion and Discussion**

Considerable experimental work has demonstrated how subjective probability assessments can be manipulated by priming or framing. The explicit questions in Robert Shiller's Investor Confidence Surveys afford an opportunity to examine factors that influence probability assessments about rare stock market crashes. These probability assessments are potentially important because they may determine such critical things as stock market participation, the demand for insurance against crashes and, to the extent that the investors surveyed are representative of marginal investors, perhaps even the equity premium. In this paper we summarized nearly 10,000 individual and institutional probability assessments of a specific kind of market crash over the period from 1989 through 2015.

We find that the crash probabilities are quite high – unreasonably high given the incidence of such events in U.S. capital market history. Our results may contribute to the literature about rare disasters. The reported probabilities are consistent with the parameterization in Barro (2006) for crashes of a 25% magnitude which identify the conditions for an equity risk premium in the 7% range and with the parametrization in Wachter (2013) for a Sharpe ratio in excess of 1.

The main focus of the paper is a test of the availability bias, with particular attention to the potential role played by the financial press in accentuating awareness of negative market outcomes.

Consistent with the news reporting past events, prior day returns are associated with media sentiment. We find some evidence of asymmetry in reporting. Extreme negative returns have a greater effect on media sentiment and extreme positive returns. Market returns appear to influence negative media sentiment terms for several days, whereas the association with positive media sentiment is confined to a single day. The effects on sentiment remains when focusing on sections of the newspaper (lead paragraphs or front page articles) that are more likely to be viewed by readers, and after controlling for the sentiment of sections that are less likely to be viewed. These findings are generally consistent with a negative bias in the financial media. It is also consistent with negative news being potentially more relevant to investors than positive news. There is considerable evidence that negative news garners more attention and reflection. Therefore, the asymmetry may be a response to rational reader demand.

Turning to the questionnaire results, we find that the coincidence of negative sentiment news and a negative market return is associated with a higher probability assessment of a future crash by individual investors. The evidence for institutional investors is statistically insignificant. We perform a number of robustness checks that provide more color to these results. Front-page placement and lead paragraphs appear to drive the main results. We also show a strong negative association between the crash probability assessments of individual investors with future flows to equity-based mutual funds.

Finally, we use local earthquakes as a priming instrument for rare event availability. We find that recent earthquakes in the immediate vicinity of the respondent are associated with a higher probability assessment of a “catastrophic” stock market crash. This lends credence to the hypothesis that the availability heuristic plays a role in subjective probability assessment about a crash. Our results are consistent with the findings of Da et. al. (2015) who find a correlation

between high-frequency measures of investor sentiment – in their case internet search terms – and investor capital flows. Our evidence also support the findings of Tetlock (2007), Engleberg and Parsons (2011), Kräussl and Mirgorodskaya (2014) and Yuan (2015) – all of whom document significant media influence on investor behavior and asset returns. Our distinctive contribution to this literature is the use of an explicit subjective probability assessment of a crash.

Our findings about asymmetric adjustment of beliefs to positive vs. negative stimuli coincide well with the Kuhnen (2014) findings that negative outcomes are more likely to cause subjects to update beliefs. One interesting effect that would attenuate the results we document is selective attention. Sicherman et. al. (2015) and Karlsson et. al. (2009) show that investors are less prone to check their investment accounts when the market has declined or is volatile. We might expect subjects who avoid current information, conditional upon negative outcomes or sentiment, to weight them relatively less in probability assessments. Perhaps this effect explains the significance of the media interaction term we document – absent the financial press calling attention to the potential for a crash conditional on a market decline, investors would ignore it.



## References

- Akhtar, S., Faff, R., Oliver, B. and Subrahmanyam, A., 2012. Stock salience and the asymmetric market effect of consumer sentiment news. *Journal of Banking & Finance*, 36(12), pp.3289-3301.
- Barber, B.M. and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21(2), pp.785-818.
- Barberis, N., 2013. The psychology of tail events: Progress and challenges. *The American Economic Review*, 103(3), pp.611-616.
- Barone-Adesi, G., Mancini, L. and Shefrin, H., 2015. Sentiment, Risk Aversion, and Time Preference. *Swiss Finance Institute Research Paper*, (12-21).
- Bates, D.S., 2000. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1), pp.181-238.
- Baumeister, R.F., Bratslavsky, E., Finkenauer, C. and Vohs, K.D., 2001. Bad is stronger than good. *Review of general psychology*, 5(4), p.323.
- Berkman, H., Jacobsen, B. and Lee, J.B., 2011. Time-varying rare disaster risk and stock returns. *Journal of Financial Economics*, 101(2), pp.313-332.
- Bollerslev, T. and Todorov, V., 2011. Tails, fears, and risk premia. *The Journal of Finance*, 66(6), pp.2165-2211.
- Bollerslev, T., Todorov, V. and Xu, L., 2015. Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1), pp.113-134.
- Engelberg, J.E. and Parsons, C.A., 2011. The causal impact of media in financial markets. *The Journal of Finance*, 66(1), pp.67-97.
- Da, Z., Engelberg, J. and Gao, P., 2015. The sum of all fears investor sentiment and asset prices. *Review of Financial Studies*, 28(1), pp.1-32.
- Dieckmann, N.F., Slovic, P. and Peters, E.M., 2009. The use of narrative evidence and explicit likelihood by decisionmakers varying in numeracy. *Risk Analysis*, 29(10), pp.1473-1488.
- Ferreira, S. and Karali, B., 2015. Do earthquakes shake stock markets?. *PloS one*, 10(7), p.e0133319.
- Finucane, M.L., Alhakami, A., Slovic, P. and Johnson, S.M., 2000. The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making*, 13(1), p.1.
- Folkes, V.S., 1988. The availability heuristic and perceived risk. *Journal of Consumer research*, pp.13-23.
- Gabaix, X., 2008. Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance (No. w13724). *National Bureau of Economic Research*.

- Gao, G. and Song, Z., 2015. Rare disaster concerns everywhere. Available at SSRN 2606345.
- Garcia, D., 2013. Sentiment during recessions. *The Journal of Finance*, 68(3), pp.1267-1300.
- Garcia, D., 2014. The kinks of financial journalism. Available at SSRN 2517591.
- Goetzmann, W.N., Kim, D., Kumar, A. and Wang, Q., 2014. Weather-induced mood, institutional investors, and stock returns. *Review of Financial Studies*, 28(1), pp. 73-111.
- Goetzmann, W., Massa, M., and Rouwenhorst, K., 2000. Behavioral Factors in Mutual Fund Flows. *Yale School of Management*.
- Greenwood, R. and Shleifer, A., 2014. Expectations of returns and expected returns. *Review of Financial Studies*, 27(3), pp.714-746.
- Guerrero, F.L., Stone, G.R. and Sundali, J.A., 2012. Fear in Asset Allocation During and After Stock Market Crashes An Experiment in Behavioral Finance. *Finance & Economics*, 1, p.1.
- Hilbig, B.E., 2009. Sad, thus true: Negativity bias in judgments of truth. *Journal of Experimental Social Psychology*, 45(4), pp.983-986.
- Hu, M. and Liu, B., 2004, August. Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168-177.
- Ito, T.A., Larsen, J.T., Smith, N.K. and Cacioppo, J.T., 1998. Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of personality and social psychology*, 75(4), p.887.
- Jackwerth, Jens Carsten, and Mark Rubinstein. "Recovering probability distributions from option prices." *The Journal of Finance*, 51, no. 5 (1996): 1611-1631.
- Johnson, E.J. and Tversky, A., 1983. Affect, generalization, and the perception of risk. *Journal of personality and social psychology*, 45(1), p.20.
- Keller, C., Siegrist, M. and Gutscher, H., 2006. The role of the affect and availability heuristics in risk communication. *Risk analysis*, 26(3), pp.631-639.
- Kliger, D. and Kudryavtsev, A., 2010. The availability heuristic and investors' reaction to company-specific events. *The Journal of Behavioral Finance*, 11(1), pp.50-65.
- Kräussl, R. and Mirgorodskaya, E., 2014. News media sentiment and investor behavior (No. 492). *Center for Financial Studies (CFS)*.
- Kuhnen, C.M., 2014. Asymmetric learning from financial information. *The Journal of Finance*, 70(5), pp. 2029-2062.
- Loewenstein, G.F., Weber, E.U., Hsee, C.K. and Welch, N., 2001. Risk as feelings. *Psychological bulletin*, 127(2), p.267.

- Loughran, T. and McDonald, B., 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-K's. *The Journal of Finance*, 66, pp. 35-65.
- Loughran, T. and McDonald, B., 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), pp.1187-1230.
- Manela, A. and Moreira, A., 2013. News implied volatility and disaster concerns. Forthcoming, *Journal of Financial Economics*.
- Manning, C.D., Raghavan, P. and Schutze, H., 2008. An information to information retrieval. *Cambridge University Press*, Cambridge.
- Newey, W.K. and West, K.D., 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4), pp.631-653.
- Nofsinger, J.R. and Varma, A., 2013. Availability, recency, and sophistication in the repurchasing behavior of retail investors. *Journal of Banking & Finance*, 37(7), pp.2572-2585.
- Ross, S., 2015. The recovery theorem. *The Journal of Finance*, 70(2), pp.615-648.
- Rozin, P. and Royzman, E.B., 2001. Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, 5(4), pp.296-320.
- Santa-Clara, P. and Yan, S., 2010. Crashes, volatility, and the equity premium: Lessons from S&P 500 options. *The Review of Economics and Statistics*, 92(2), pp.435-451.
- Semin, G.R. and Fiedler, K., 1988. The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of personality and Social Psychology*, 54(4), p.558.
- Seo, S.B. and Wachter, J.A., 2013. Option prices in a model with stochastic disaster risk (No. w19611). *National Bureau of Economic Research*.
- Shiller, R.J., 1990. Speculative prices and popular models. *The Journal of Economic Perspectives*, pp.55-65.
- Shiller, R. J., 2000a. Measuring bubble expectations and investor confidence. *Journal of Psychology and Financial Markets*, 1, pp. 49-60.
- Shiller, R. J., 2000b. Irrational Exuberance. *Princeton University Press*, Princeton.
- Slovic, P. and Peters, E., 2006. Risk perception and affect. *Current directions in psychological science*, 15(6), pp.322-325.
- Slovic, P., Finucane, M.L., Peters, E. and MacGregor, D.G., 2007. The affect heuristic. *European journal of operational research*, 177(3), pp.1333-1352.
- Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), pp.1139-1168.

Tsai, J. and Wachter, J.A., 2015. Disaster risk and its implications for asset pricing (No. w20926). *National Bureau of Economic Research*.

Tversky, A. and Kahneman, D., 1973. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), pp.207-232.

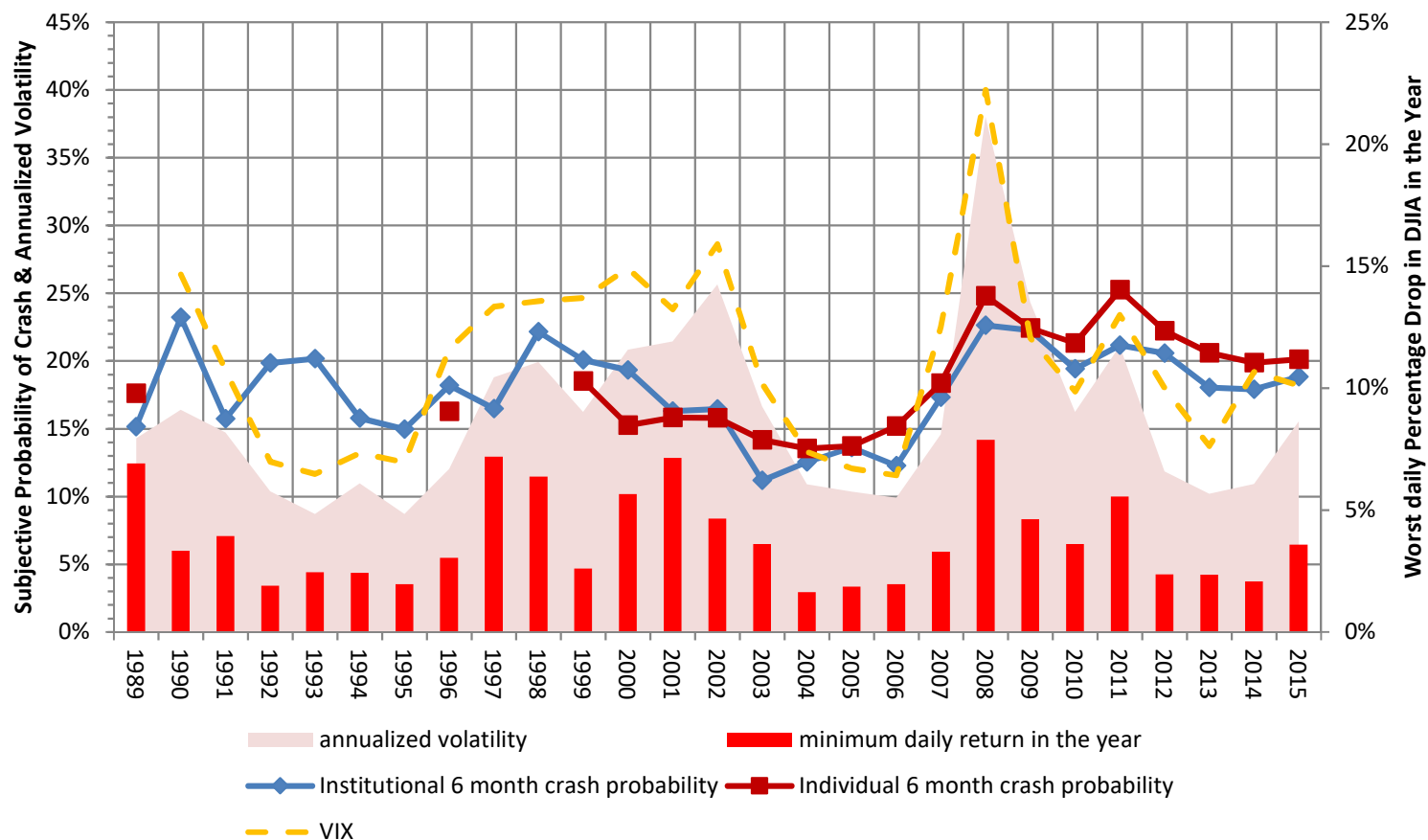
Yuan, Y., 2015. Market-wide attention, trading, and stock returns. *Journal of Financial Economics*, 116(3), pp. 548-564.

Wachter, J.A., 2013. Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?. *The Journal of Finance*, 68(3), pp.987-1035.

Welch, I., 2015. The Importance of Disaster Risk. Available at SSRN 2596909.

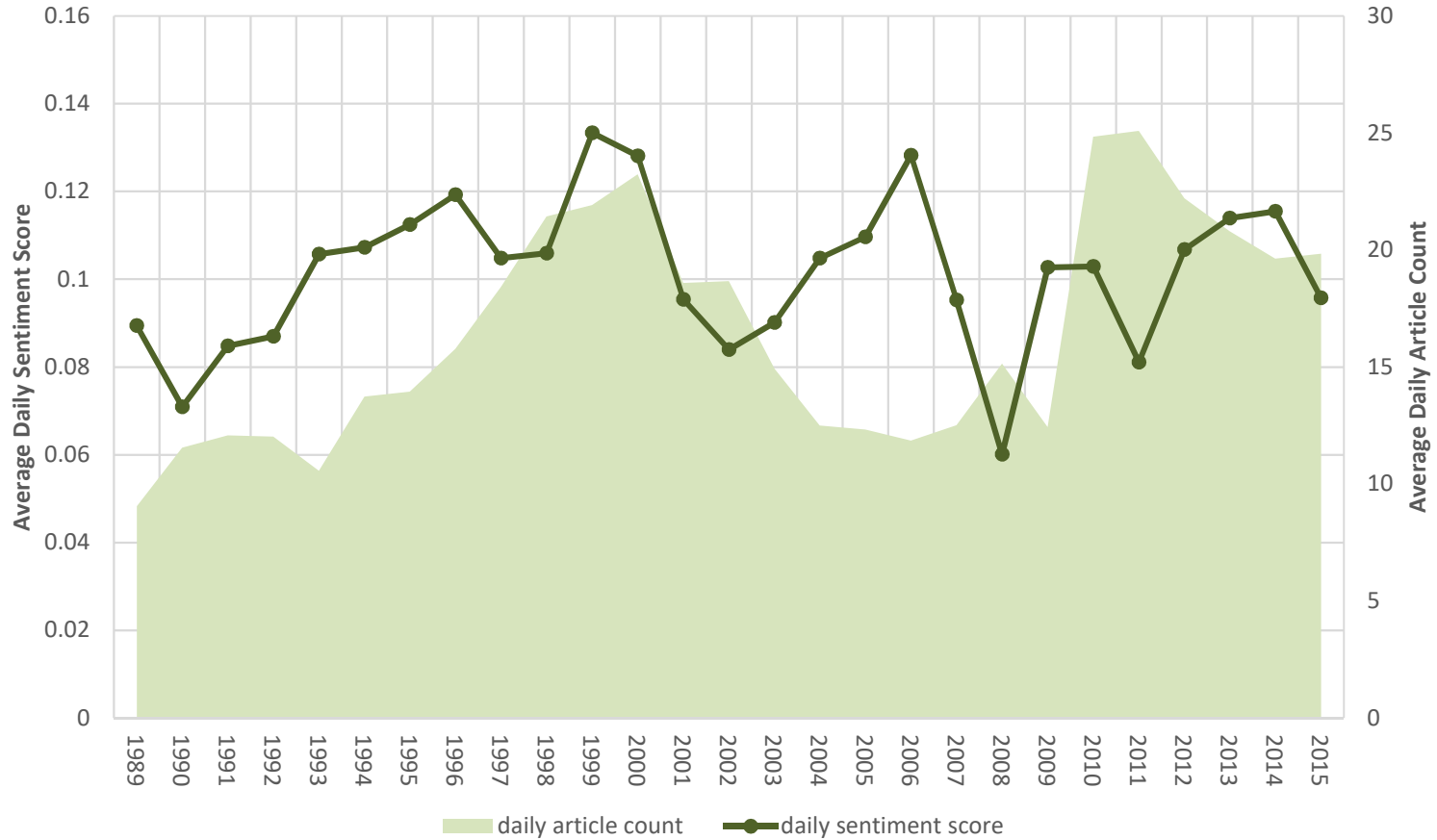
**Figure 1: Average Annual Crash Probabilities from 1989-2015**

This figure displays the average annual probabilities from 1989-2015 for the individual and institutional survey respondents of a crash in the next six months on the scale of 10/19/1987 or 10/28/1929. Also displayed are the annualized volatility of the daily DJIA, the largest negative return in each year (represented as a positive number on the figure) and the VIX implied volatility



**Figure 2: Average Sentiment Scores and Article Counts from 1989-2015**

This figure displays the average sentiment scores from 1989-2015 calculated annually based upon newspaper articles related to the stock market. Also displayed are average daily number of newspaper articles related to the stock market calculated annually.



**Table 1: Summary Statistics**

The table displays variable descriptions and summary statistics of the key variables used in the analysis. The variables are collected from the Center for Research on Security Prices (CRSP), the Wall Street Journal – Eastern Edition (WSJ), or the survey data from Robert Shiller’s Investor Confidence Surveys (ICS). The variables are divided based upon its source.

Panel A: Variable Descriptions		
Variable Name	Description	Source
<u>Return variables:</u>		
Return (t-1)	Total return on date $t-1$ based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA index.	CRSP
Return (a,b)	Total cumulative return from date $a$ to $b$ based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA index.	CRSP
Volatility (a,b)	Daily returns volatility from dates $a$ to $b$ based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA index..	CRSP
<u>Media Variables:</u>		
Negative (t)	The number of articles on date $t$ with article-level sentiment scores in the bottom 10th sample percentile.	WSJ
Positive (t)	The number of articles on date $t$ with article-level sentiment scores in the top 10th sample percentile.	WSJ
Sentiment (t)	The difference between weighted positive and negative word counts, scaled by the total count of weighted positive and negative words, across newspaper articles on date $t$ .	WSJ
Sentiment (t-30,t-1)	The average daily difference between term-weighted positive and negative word counts, scaled by the total count of positive and negative words, across newspaper articles from date $t-31$ to $t-1$ .	WSJ
#Articles (t)	The number of articles on date $t$ that are related to the stock market.	WSJ
<u>Survey Variables:</u>		
$\pi$ (i,t)	The crash probability reported by the survey respondent on date $t$ .	ICS
$\pi$ (t-30,t-1)	The average crash probability reported by survey respondents from dates $t-30$ to date $t-1$ for the same investor type.	ICS
Institutional	Dummy that takes value 1 if the survey respondent is an institutional investor, and zero otherwise.	ICS

Panel B: Summary Statistics

	<b>N</b>	<b>Mean</b>	<b>StDev</b>	<b>Min</b>	<b>P25</b>	<b>Median</b>	<b>P75</b>	<b>Max</b>
<u>Return variables:</u>								
CRSPVW Return (t-1)	9859	0.000	0.011	-0.090	-0.004	0.001	0.005	0.115
CRSPVW Volatility (t-30,t-3)	9859	0.010	0.006	0.003	0.006	0.008	0.011	0.050
SP500 Return (t-1)	9859	0.000	0.011	-0.090	-0.005	0.001	0.006	0.116
SP500 Volatility (t-30,t-3)	9859	0.010	0.006	0.003	0.006	0.008	0.011	0.051
DJIA Return (t-1)	9859	0.000	0.011	-0.079	-0.005	0.001	0.006	0.111
DJIA Volatility (t-30,t-3)	9859	0.009	0.005	0.003	0.006	0.008	0.011	0.049
<u>Media Variables:</u>								
Positive (t) (without log trans.)	9859	1.354	1.665	0.000	0.000	1.000	2.000	24.000
Negative (t) (without log trans.)	9859	1.354	1.673	0.000	0.000	1.000	2.000	19.000
Sentiment (t)	9859	0.083	0.071	-0.337	0.013	0.085	0.132	0.594
Sentiment (t-30,t-1)	9859	0.083	0.026	0.007	0.065	0.081	0.100	0.159
#Articles (t)	9859	13.537	10.074	0.000	6.000	13.000	19.000	106.000
<u>Survey Variables:</u>								
$\pi$ (i,t)	10983	0.189	0.194	0.000	0.050	0.100	0.250	0.999
$\pi$ (t-30,t-1)	10983	0.188	0.060	0.000	0.152	0.184	0.218	0.775
Institutional	10983	0.569	0.495	0.000	0.000	1.000	1.000	1.000



**Table 2: Article Counts and Previous Day Returns**

The table displays the results from OLS regression models where the dependent variables are daily article counts based upon sentiment scores. *Negative* is the natural log of one plus the number of articles that have sentiment scores in the bottom 10<sup>th</sup> sample percentile. *Positive* is the natural log of one plus the number of articles that have sentiment scores in the top 10<sup>th</sup> sample percentile. The results are displayed separately where the return variables are based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Newey-West standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Return variables: Dependent Variable:	(1) CRSP-VW Negative (t)	(2) CRSP-VW Positive (t)	(3) S&P500 Negative (t)	(4) S&P500 Positive (t)	(5) DJIA Negative (t)	(6) DJIA Positive (t)
r (t-1)	-7.532*** (0.447)	2.353*** (0.448)	-7.280*** (0.438)	2.159*** (0.438)	-7.554*** (0.456)	2.270*** (0.456)
r (t-2)	-3.239*** (0.437)	0.795* (0.438)	-3.323*** (0.431)	0.717* (0.431)	-3.426*** (0.449)	0.512 (0.450)
r (t-30,t-3)	-0.801*** (0.116)	0.114 (0.114)	-0.772*** (0.120)	0.167 (0.117)	-0.764*** (0.122)	0.135 (0.119)
$\sigma$ (t-30,t-3)	4.378*** (1.044)	-5.071*** (1.020)	4.523*** (1.060)	-5.096*** (1.034)	5.075*** (1.169)	-5.835*** (1.141)
Sentiment (t-30,t-1)	-1.076*** (0.231)	2.533*** (0.226)	-1.097*** (0.234)	2.454*** (0.229)	-1.070*** (0.237)	2.406*** (0.232)
log(1+#Articles (t))	0.343*** (0.007)	0.314*** (0.007)	0.345*** (0.007)	0.313*** (0.007)	0.346*** (0.007)	0.312*** (0.007)
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	9859	9859	9859	9859	9859	9859
Adjusted R <sup>2</sup>	44.92%	44.01%	44.03%	44.73%	44.78%	44.00%

**Table 3: Media Sentiment and Previous Day Returns**

The table displays the results from OLS regression models where the dependent variable is the sentiment scores across newspaper articles related to the stock market (*Sentiment*). The results are displayed separately where the return variables are based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Newey-West standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Return variables: Dependent Variable:	(1) CRSP-VW Sentiment (t)	(2) CRSP-VW Sentiment (t)	(3) S&P500 Sentiment (t)	(4) S&P500 Sentiment (t)	(5) DJIA Sentiment (t)	(6) DJIA Sentiment (t)
$r(t-1)$	0.820*** (0.057)		0.781*** (0.056)		0.812*** (0.058)	
$D_{(rk(r(t-1))<10\%)}$		-0.026*** (0.002)		-0.027*** (0.002)		-0.028*** (0.002)
$D_{(rk(r(t-1))>90\%)}$		0.005*** (0.002)		0.004** (0.002)		0.003 (0.002)
$r(t-2)$	0.430*** (0.056)	0.391*** (0.056)	0.428*** (0.054)	0.379*** (0.055)	0.435*** (0.057)	0.374*** (0.057)
$r(t-30,t-3)$	0.096*** (0.015)	0.084*** (0.015)	0.095*** (0.015)	0.083*** (0.015)	0.095*** (0.015)	0.081*** (0.015)
$\sigma(t-30,t-3)$	-0.560*** (0.132)	-0.357*** (0.137)	-0.593*** (0.135)	-0.376*** (0.138)	-0.660*** (0.149)	-0.413*** (0.151)
Sentiment (t-30,t-1)	0.413*** (0.029)	0.414*** (0.029)	0.412*** (0.030)	0.406*** (0.030)	0.408*** (0.030)	0.399*** (0.030)
$\log(1+\#\text{Articles}(t))$	0.031*** (0.001)	0.031*** (0.001)	0.031*** (0.001)	0.031*** (0.001)	0.030*** (0.001)	0.031*** (0.001)
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	9859	9859	9859	9859	9859	9859
Adjusted R <sup>2</sup>	32.21%	31.95%	31.97%	31.98%	32.04%	31.88%

**Table 4: Persistence of Media Sentiment**

The table displays the results from OLS regression models where the dependent variable is the sentiment scores for up to seven trading days after the returns event. All regressions include the control variables of Table 3. The results are displayed separately where the return variables are based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Newey-West standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Dependent Variable:	(1) Sentiment (t+1)	(2) Sentiment (t+2)	(3) Sentiment (t+3)	(4) Sentiment (t+4)	(5) Sentiment (t+5)	(6) Sentiment (t+6)	(7) Sentiment (t+7)	(8) Sentiment (t+8)	(9) Sentiment (t+9)	(10) Sentiment (t+10)
<u>CRSP-VW Returns</u>										
$D_{(rk(r(t-1)) < 10\%)}$	-0.014*** (0.002)	-0.010*** (0.002)	-0.012*** (0.002)	-0.010*** (0.002)	-0.006*** (0.002)	-0.008*** (0.002)	-0.012*** (0.002)	-0.010*** (0.002)	-0.006*** (0.002)	-0.008*** (0.002)
$D_{(rk(r(t-1)) > 90\%)}$	0.002 (0.002)	0.002 (0.002)	0.004* (0.002)	0.003 (0.002)	0.007*** (0.002)	0.003 (0.002)	0.001 (0.002)	-0.003 (0.002)	0.002 (0.002)	0.003 (0.002)
<u>S&amp;P 500 Returns</u>										
$D_{(rk(r(t-1)) < 10\%)}$	-0.015*** (0.002)	-0.009*** (0.002)	-0.010*** (0.002)	-0.008*** (0.002)	-0.007*** (0.002)	-0.009*** (0.002)	-0.012*** (0.002)	-0.010*** (0.002)	-0.005** (0.002)	-0.006*** (0.002)
$D_{(rk(r(t-1)) > 90\%)}$	0.003 (0.002)	0.002 (0.002)	0.004* (0.002)	0.002 (0.002)	0.006** (0.002)	0.000 (0.002)	-0.001 (0.002)	-0.006*** (0.002)	0.000 (0.002)	0.001 (0.002)
<u>DJIA Returns</u>										
$D_{(rk(r(t-1)) < 10\%)}$	-0.014*** (0.002)	-0.009*** (0.002)	-0.013*** (0.002)	-0.011*** (0.002)	-0.007*** (0.002)	-0.010*** (0.002)	-0.013*** (0.002)	-0.011*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)
$D_{(rk(r(t-1)) > 90\%)}$	0.002 (0.002)	0.001 (0.002)	0.002 (0.002)	0.000 (0.002)	0.003 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.005** (0.002)	0.002 (0.002)	0.004* (0.002)

**Table 5: Lead Paragraph and Article Placement**

The table displays the results from OLS regression models where the dependent variable is the sentiment scores for only lead paragraphs across all articles ( $Sentiment^{Lead}$ ) or only articles appearing on the front page ( $Sentiment^{Front}$ ). All regressions include the control variables of Table 3. Columns (1) and (2) also control for the sentiment scores for non-lead paragraphs. Columns (3) and (4) also control for the sentiment scores for articles not appearing on the front page. The results are displayed separately where the return variables are based upon CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA indices. Newey-West standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Dependent Variable:	(1) Sentiment <sup>Lead</sup> (t)	(2) Sentiment <sup>Lead</sup> (t)	(3) Sentiment <sup>Front</sup> (t)	(4) Sentiment <sup>Front</sup> (t)
<u>CRSP-VW Returns</u>				
r (t-1)	1.122*** (0.110)		0.944*** (0.087)	
D <sub>(rk(r (t-1))&lt;10%)</sub>		-0.037*** (0.004)		-0.027*** (0.003)
D <sub>(rk(r (t-1))&gt;90%)</sub>		0.007* (0.004)		0.004 (0.003)
<u>S&amp;P 500 Returns</u>				
r (t-1)	1.071*** (0.108)		0.907*** (0.086)	
D <sub>(rk(r (t-1))&lt;10%)</sub>		-0.037*** (0.004)		-0.028*** (0.003)
D <sub>(rk(r (t-1))&gt;90%)</sub>		0.007* (0.004)		0.007** (0.003)
<u>DJIA Returns</u>				
r (t-1)	1.110*** (0.112)		0.956*** (0.089)	
D <sub>(rk(r (t-1))&lt;10%)</sub>		-0.038*** (0.004)		-0.027*** (0.003)
D <sub>(rk(r (t-1))&gt;90%)</sub>		0.005 (0.004)		0.007*** (0.003)

**Table 6: Crash Probabilities and Article Counts**

The table displays the results from OLS regression models where the dependent variables are the investor crash probabilities ( $\pi$ ). The results are displayed separately for individual (*Indiv*) and institutional (*Inst*) investors, and where the return variables are based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Standard errors are clustered on the date level and are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Investor Subsample:	Indiv.	Inst.	Indiv.	Inst.	Indiv.	Inst.
Return Variable:	All	All	S&P500	S&P500	DJIA	DJIA
Dependent Variable:	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$
$r(t-1) \times \text{Negative}(t)$	-1.012*** (0.363)	-0.168 (0.295)	-0.939*** (0.365)	-0.152 (0.276)	-1.040*** (0.394)	-0.212 (0.283)
$r(t-1) \times \text{Positive}(t)$	0.206 (0.429)	0.239 (0.356)	0.333 (0.432)	0.189 (0.343)	0.476 (0.462)	0.310 (0.364)
Negative(t)	0.006 (0.006)	0.009** (0.004)	0.007 (0.006)	0.009** (0.004)	0.007 (0.006)	0.009** (0.004)
Positive(t)	-0.001 (0.005)	0.001 (0.005)	-0.001 (0.005)	0.001 (0.005)	-0.002 (0.006)	0.001 (0.004)
Sentiment(t-30,t-1)	0.036 (0.126)	-0.024 (0.098)	0.037 (0.128)	-0.021 (0.099)	0.028 (0.128)	-0.034 (0.102)
$r(t-1)$	0.068 (0.262)	-0.204 (0.217)	0.121 (0.258)	-0.186 (0.212)	0.136 (0.278)	-0.219 (0.221)
$r(t-30,t-2)$	-0.174*** (0.061)	-0.036 (0.052)	-0.179*** (0.064)	-0.031 (0.055)	-0.195*** (0.065)	0.006 (0.056)
$\sigma(t-30,t-2)$	1.725*** (0.548)	0.951** (0.485)	1.627*** (0.546)	1.006** (0.486)	1.505*** (0.581)	1.245** (0.527)
$\pi(t-30,t-1)$	0.223*** (0.062)	0.251*** (0.056)	0.228*** (0.061)	0.251*** (0.056)	0.242*** (0.061)	0.250*** (0.056)
$\log(1+\#\text{Articles}(t))$	0.019*** (0.007)	0.001 (0.007)	0.019*** (0.007)	0.001 (0.007)	0.019*** (0.007)	0.001 (0.007)
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	4730	6253	4730	6253	4730	6253
Adjusted R <sup>2</sup>	1.99%	1.22%	1.96%	1.21%	1.92%	1.23%

**Table 7: Crash Probabilities and Media Sentiment**

The table displays the results from OLS regression models where the dependent variables are the investor crash probabilities ( $\pi$ ). The results are displayed separately for individual (*Indiv.*) and institutional (*Inst.*) investors, and where the return variables are based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Standard errors are clustered on the date level and are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Investor Subsample:	Indiv.	Inst.	Indiv.	Inst.	Indiv.	Inst.
Return variable:	All	All	S&P500	S&P500	DJIA	DJIA
Dependent Variable:	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$
$r(t-1) \times \text{Sentiment}(t)$	6.881* (3.932)	0.482 (3.440)	7.223* (3.925)	-0.344 (3.132)	8.832** (4.267)	0.102 (3.396)
Sentiment(t)	-0.036 (0.053)	-0.053 (0.041)	-0.038 (0.053)	-0.055 (0.042)	-0.039 (0.053)	-0.056 (0.041)
Sentiment(t-30,t-1)	0.014 (0.131)	-0.017 (0.100)	0.018 (0.131)	-0.015 (0.100)	0.009 (0.131)	-0.025 (0.104)
$r(t-1)$	-0.683* (0.363)	-0.336 (0.280)	-0.667* (0.359)	-0.260 (0.257)	-0.820** (0.390)	-0.358 (0.271)
$r(t-30,t-2)$	-0.163*** (0.062)	-0.038 (0.052)	-0.172*** (0.065)	-0.034 (0.055)	-0.191*** (0.066)	0.001 (0.062)
$\sigma(t-30,t-2)$	1.605*** (0.552)	0.972** (0.493)	1.522*** (0.548)	1.008** (0.494)	1.380** (0.580)	1.253** (0.535)
$\pi(t-30,t-1)$	0.228*** (0.061)	0.246*** (0.057)	0.235*** (0.061)	0.247*** (0.057)	0.250*** (0.061)	0.246*** (0.058)
$\log(1+\#\text{Articles}(t))$	0.022*** (0.006)	0.009 (0.005)	0.023*** (0.006)	0.009 (0.005)	0.022*** (0.006)	0.009* (0.006)
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	4730	6253	4730	6253	4730	6253
Adjusted R <sup>2</sup>	1.87%	1.20%	1.91%	1.20%	1.92%	1.20%

**Table 8: Past Month Returns and Media Sentiment**

The table displays the results from OLS regression models where the dependent variables are the investor crash probabilities ( $\pi$ ). The results are displayed separately for individual (*Indiv.*) and institutional (*Inst.*) investors, and where the return variables ( $r$ ) are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Control variables of Table 7 are included in all the models but not all are reported. Standard errors are clustered on the date level and are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Investor Subsample:	Indiv.	Inst.	Indiv.	Inst.	Indiv.	Inst.
Return Variable:	All	All	S&P500	S&P500	DJIA	DJIA
Dependent Variable:	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$
$r(t-1) \times \text{Sentiment}(t)$	9.965** (4.390)	1.667 (3.473)	10.318** (4.335)	0.836 (3.343)	12.208*** (4.678)	1.564 (3.475)
$r(t-30,t-2) \times \text{Sentiment}(t)$	0.092 (0.923)	0.402 (0.718)	0.145 (0.964)	0.386 (0.773)	-0.262 (1.047)	0.610 (0.792)
$r(t-1) \times \text{Sentiment}(t-30,t-1)$	-15.586* (8.906)	-7.655 (7.432)	-15.244* (8.661)	-7.221 (7.079)	-14.769* (8.791)	-8.894 (7.231)
$r(t-30,t-2) \times \text{Sentiment}(t-30,t-1)$	1.121 (1.967)	3.277** (1.614)	1.184 (2.041)	3.317** (1.684)	1.779 (1.977)	3.685** (1.698)
Control Variables	YES	YES	YES	YES	YES	YES
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	4730	6253	4730	6253	4730	6253
Adjusted R <sup>2</sup>	1.95%	1.28%	1.93%	1.26%	1.89%	1.31%

**Table 9: Salience of Lead Paragraph and Article Placement**

The table displays the results from OLS regression models where the dependent variables are the investor crash probabilities ( $\pi$ ). The results are displayed separately for individual (*Indiv*) and institutional (*Inst*) investors, and where the return variables are based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Control variables of Table 7 are included in all the models but not all are reported. Standard errors are clustered on the date level and are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Panel A: Lead Paragraph						
	(1)	(2)	(3)	(4)	(5)	(6)
Investor Subsample:	Indiv.	Inst.	Indiv.	Inst.	Indiv.	Inst.
Return Variable:	All	All	S&P500	S&P500	DJIA	DJIA
Dependent Variable:	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$
$r(t-1) \times \text{Sentiment}^{\text{Lead}}(t)$	4.451** (2.130)	0.093 (1.556)	4.545** (2.206)	-0.017 (1.737)	5.726** (2.357)	0.150 (1.668)
$r(t-1) \times \text{Sentiment}^{\text{NotLead}}(t)$	1.366 (3.254)	-0.696 (2.576)	1.440 (3.200)	-1.276 (2.408)	2.520 (3.360)	-0.688 (2.648)
$\text{Sentiment}^{\text{Lead}}(t)$	-0.037 (0.030)	-0.027 (0.022)	-0.039 (0.030)	-0.027 (0.022)	-0.042 (0.030)	-0.026 (0.022)
$\text{Sentiment}^{\text{NotLead}}(t)$	0.070** (0.032)	0.023 (0.027)	0.069** (0.032)	0.022 (0.027)	0.069** (0.032)	0.021 (0.026)
Control Variables	YES	YES	YES	YES	YES	YES
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	4730	6253	4730	6253	4730	6253
Adjusted R <sup>2</sup>	2.01%	1.18%	2.03%	1.18%	2.04%	1.19%
Panel B: Article Placement						
	(1)	(2)	(3)	(4)	(5)	(6)
Investor Subsample:	Indiv.	Inst.	Indiv.	Inst.	Indiv.	Inst.
Return Variable:	All	All	S&P500	S&P500	DJIA	DJIA
Dependent Variable:	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$
$r(t-1) \times \text{Sentiment}^{\text{Front}}(t)$	4.485* (2.373)	2.810 (1.993)	4.952** (2.381)	2.533 (1.963)	5.099* (2.670)	3.006 (2.194)
$r(t-1) \times \text{Sentiment}^{\text{NotFront}}(t)$	2.174 (3.295)	-1.835 (3.059)	2.152 (3.310)	-2.172 (2.896)	3.330 (3.659)	-1.725 (2.975)
$\text{Sentiment}^{\text{Front}}(t)$	-0.012 (0.029)	-0.016 (0.024)	-0.012 (0.028)	-0.016 (0.024)	-0.013 (0.029)	-0.017 (0.024)
$\text{Sentiment}^{\text{NotFront}}(t)$	-0.049 (0.043)	-0.046 (0.035)	-0.052 (0.043)	-0.047 (0.035)	-0.051 (0.044)	-0.048 (0.035)
Control Variables	YES	YES	YES	YES	YES	YES
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	4730	6253	4730	6253	4730	6253
Adjusted R <sup>2</sup>	1.87%	1.21%	1.92%	1.20%	1.93%	1.21%



**Table 10: Earthquakes and Crash Probabilities**

The table displays the results from OLS regression models where the dependent variables are the investor crash probabilities ( $\pi$ ). The results are displayed separately for individual (Indiv) and institutional (Inst) investors. The key explanatory variables are dummies associated with whether the investor is located within a 30 mile radius of the epicenter of an earthquake that occurred within the past 30 days. Weak magnitudes are earthquakes with a magnitude of 2.5 up to 5.5. Strong magnitudes are earthquakes with a magnitude greater than 5.5. Control variables and fixed effects from Table 7 are included where indicated, but not reported. Historical frequency of earthquakes for each location are included where indicated, but not reported. The return variables for the control variables are based upon the DJIA index.  $p$ -values based upon bootstrapped distributions of the model estimates are displayed in brackets. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Investor Subsample:	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Indiv.	Indiv.	Indiv.	Inst.	Inst.	Inst.
	$\pi$ (i,t)	$\pi$ (i,t)	$\pi$ (i,t)	$\pi$ (i,t)	$\pi$ (i,t)	$\pi$ (i,t)
Weak Magnitude (t-30,t)	0.034** [0.036]	0.054** [0.035]	0.050** [0.045]	-0.009 [0.363]	-0.012 [0.513]	-0.013 [0.459]
Strong Magnitude (t-30,t)		-0.042 [0.421]	-0.030 [0.440]		-0.071 [0.336]	-0.078 [0.305]
Historical Earthquake Frequency	NO	YES	YES	NO	YES	YES
Control Variables	NO	NO	YES	NO	NO	YES
Day of Week FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	3087	3087	3087	3493	3493	3493
Adjusted R <sup>2</sup>	0.00%	0.00%	1.10%	0.10%	0.00%	0.50%

**Table 11: Aggregate Fund Flows and Crash Probabilities**

The dependent variable is %Flow, defined as the aggregate daily percentage mutual fund flow for either equity (Panel A) or bond (Panel B) funds. The investor crash probabilities ( $\pi$ ) are estimated as a one-week moving average when there are at least five survey respondents over the past week, or the last value available when there are less than five survey respondents over the past week. The crash probabilities are estimated separately for individual ( $\pi^{\text{Indiv.}}$ ) and institutional ( $\pi^{\text{Inst.}}$ ) investors. Control variables and fixed effects from Table 2 are included where indicated, but not reported. The return variables for the control variables are based upon the DJIA index. Newey-West standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Panel A: Equity Mutual Fund Flows										
Dependent Variable:	(1) %Flow (t+1)	(2) %Flow (t+2)	(3) %Flow (t+3)	(4) %Flow (t+4)	(5) %Flow (t+5)	(6) %Flow (t+6)	(7) %Flow (t+7)	(8) %Flow (t+8)	(9) %Flow (t+9)	(10) %Flow (t+10)
$\pi^{\text{Indiv.}}$ (t)	-0.032 (0.033)	-0.034 (0.033)	-0.053 (0.033)	-0.065* (0.034)	-0.065* (0.034)	-0.074** (0.034)	-0.065* (0.034)	-0.047 (0.034)	-0.048 (0.034)	-0.064* (0.035)
Control Variables and FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	3260	3259	3258	3257	3256	3255	3254	3253	3252	3251
Adjusted R <sup>2</sup>	10.20%	3.10%	2.90%	2.80%	2.80%	2.70%	2.60%	2.60%	2.60%	2.50%
$\pi^{\text{Inst.}}$ (t)	-0.010 (0.034)	-0.001 (0.035)	-0.005 (0.035)	-0.003 (0.036)	0.023 (0.036)	0.025 (0.036)	0.024 (0.036)	0.011 (0.036)	0.012 (0.036)	0.029 (0.037)
Control Variables and FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	3260	3259	3258	3257	3256	3255	3254	3253	3252	3251
Adjusted R <sup>2</sup>	10.20%	3.10%	2.80%	2.60%	2.60%	2.50%	2.50%	2.50%	2.50%	2.40%
$\pi^{\text{Indiv.}}$ (t) - $\pi^{\text{Inst.}}$ (t)	-0.012 (0.024)	-0.018 (0.025)	-0.027 (0.025)	-0.035 (0.025)	-0.047* (0.025)	-0.053** (0.025)	-0.048* (0.025)	-0.031 (0.025)	-0.033 (0.025)	-0.050* (0.026)
Control Variables and FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	3260	3259	3258	3257	3256	3255	3254	3253	3252	3251
Adjusted R <sup>2</sup>	10.20%	3.10%	2.80%	2.70%	2.80%	2.70%	2.60%	2.60%	2.60%	2.50%

Panel B: Bond Mutual Fund Flows										
Dependent Variable:	(1) %Flow (t+1)	(2) %Flow (t+2)	(3) %Flow (t+3)	(4) %Flow (t+4)	(5) %Flow (t+5)	(6) %Flow (t+6)	(7) %Flow (t+7)	(8) %Flow (t+8)	(9) %Flow (t+9)	(10) %Flow (t+10)
$\pi^{\text{Indiv.}}(t)$	0.082** (0.037)	0.072** (0.037)	0.064* (0.037)	0.061* (0.037)	0.044 (0.037)	0.056 (0.038)	0.069* (0.038)	0.062 (0.038)	0.060 (0.038)	0.058 (0.039)
Control Variables and FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	3260	3259	3258	3257	3256	3255	3254	3253	3252	3251
Adjusted R <sup>2</sup>	6.40%	6.20%	6.10%	5.20%	4.10%	3.90%	3.50%	3.60%	3.20%	3.00%
$\pi^{\text{Inst.}}(t)$	0.093** (0.040)	0.077* (0.040)	0.067* (0.040)	0.067* (0.041)	0.091** (0.041)	0.088** (0.041)	0.094** (0.041)	0.086** (0.041)	0.087** (0.042)	0.091** (0.042)
Control Variables and FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	3260	3259	3258	3257	3256	3255	3254	3253	3252	3251
Adjusted R <sup>2</sup>	6.50%	6.30%	6.10%	5.20%	4.50%	4.20%	3.70%	3.80%	3.40%	3.30%
$\pi^{\text{Indiv.}}(t) - \pi^{\text{Inst.}}(t)$	-0.002 (0.026)	0.001 (0.026)	0.001 (0.027)	-0.000 (0.027)	-0.022 (0.027)	-0.013 (0.027)	-0.009 (0.027)	-0.009 (0.027)	-0.011 (0.028)	-0.014 (0.028)
Control Variables and FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N	3260	3259	3258	3257	3256	3255	3254	3253	3252	3251
Adjusted R <sup>2</sup>	6.00%	5.90%	5.80%	4.90%	4.00%	3.70%	3.20%	3.40%	2.90%	2.80%

**Table A.1: Tobit Model Estimates**

The table displays the results from tobit regression models where the dependent variable is the daily article count based upon sentiment scores and the sentiment scores. The models based upon article counts are censored below at zero. The models based upon sentiment scores are censored below at -1 and above at 1. All regressions include the control variables of Table 2. Robust standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Return variables:	All Stocks	All Stocks	All Stocks	All Stocks	All Stocks	All Stocks
Dependent Variable:	Negative (t)	Positive (t)	Sentiment (t)	Negative (t)	Positive (t)	Sentiment (t)
$r(t-1)$	-10.188*** (0.629)	4.206*** (0.704)	0.820*** (0.055)			
$D_{(Rk(r(t-1))<10\%)}$				0.266*** (0.022)	-0.140*** (0.026)	-0.026*** (0.002)
$D_{(Rk(r(t-1))>90\%)}$				-0.148*** (0.027)	-0.012 (0.026)	0.005*** (0.002)
Control Variables	YES	YES	YES	YES	YES	YES
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	8046	8046	9859	8046	8046	9859

**Table A.2: Expected Returns Volatility Controls**

The table displays the results from OLS regression models where the dependent variable is the daily article count based upon sentiment scores and the sentiment scores. Expected returns volatility are the conditional expected volatilities from an EGARCH(1,1) model. All regressions include the control variables of Table 2. Newey-West standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Return variables:	All Stocks	All Stocks	All Stocks	All Stocks	All Stocks	All Stocks
Dependent Variable:	Negative (t)	Positive (t)	Sentiment (t)	Negative (t)	Positive (t)	Sentiment (t)
$r(t-1)$	-7.331*** (0.451)	2.347*** (0.450)	0.786*** (0.057)			
$D_{(Rk(r(t-1))<10\%)}$				0.193*** (0.018)	-0.063*** (0.018)	-0.023*** (0.002)
$D_{(Rk(r(t-1))>90\%)}$				-0.094*** (0.017)	-0.019 (0.017)	0.008*** (0.002)
$\ln(\text{Expected Returns Volatility})$	0.051*** (0.011)	-0.010 (0.011)	-0.008*** (0.001)	0.048*** (0.011)	-0.009 (0.012)	-0.031*** (0.001)
Control Variables	YES	YES	YES	YES	YES	YES
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	9670	9670	9670	9670	9670	9670
Adjusted R <sup>2</sup>	30.81%	32.30%	31.86%	31.48%	32.38%	32.12%

**Table A.3: Negative Front Page Articles and Past Returns**

The table displays the results from probit regression models where the dependent variable is a dummy based upon whether article  $j$  has negative sentiment and appears on the front page. An article is coded as negative if the sentiment score is in the bottom sample decile. The return variable is based upon the CRSP-VW (NYSE/AMEX/Nasdaq/Arca) index. Control variables and fixed effects from Table 2 are included where indicated, but not reported. Standard errors are clustered on the date level and are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

	(1)	(2)
Return variables:	All Stocks	All Stocks
Article Subsample:	All Articles	Only Front Page
Dependent Variable:	D <sub>Negative &amp; Front</sub> (j,t)	D <sub>Negative &amp; Front</sub> (j,t)
r (t-1)	-0.387*** (0.039)	-1.913*** (0.156)
Control Variables	YES	YES
Weekday FEs	YES	YES
Month FEs	YES	YES
N	133466	28272

**Table A.4: Sentiment and Article Placement**

The table displays the results from probit regression models where the dependent variable is a dummy based upon whether article  $j$  appears on the front page. An article is coded as negative (positive) if the sentiment score is in the bottom (top) sample decile. Control variables and fixed effects from Table 2 are included where indicated, but not reported. Standard errors are clustered on the date level and are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Dependent Variable:	(1) D <sub>Front</sub> (j,t)	(2) D <sub>Front</sub> (j,t)	(3) D <sub>Front</sub> (j,t)
Negative Article (j,t)	0.032*** (0.004)		0.017*** (0.004)
Positive Article (j,t)		-0.188*** (0.004)	-0.186*** (0.005)
Control Variables	YES	YES	YES
Weekday FEs	YES	YES	YES
Month FEs	YES	YES	YES
N	133466	133466	133466

**Table A.5: Internal Consistency of Investor Crash Probabilities**

The table displays the results from OLS regression models where the dependent variables are the investor crash probabilities ( $\pi$ ). The results are displayed separately for institutional and individual investors in Panels A and B, respectively. The explanatory variables include the expected percentage change in the DJIA over the next six months (Expected Returns); a dummy based upon whether stock prices when compared to true fundamental value is too high ( $D_{\text{TooHighValuation}}$ ); a dummy based upon whether stock prices when compared to true fundamental value is too low ( $D_{\text{TooLowValuation}}$ ); a dummy based upon whether the investor is inclined to sell stocks overall ( $D_{\text{Sell}}$ ); and a dummy based upon whether the investor is inclined to buy stocks overall ( $D_{\text{Buy}}$ ). Fixed effects based upon the month and day-of-week when the survey was filled out and month are included in the models, but are not reported. Standard errors are clustered on the date level and are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as \*\*\*, \*\*, and \*, respectively.

Panel A: Institutional Investors						
	(1)	(2)	(3)	(4)	(5)	(6)
Investor Subsample:	Inst.	Inst.	Inst.	Inst.	Inst.	Inst.
Dependent Variable:	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$
Expected Returns	-0.635*** (0.052)					-0.414*** (0.063)
$D_{\text{TooHighValuation}}$		0.109*** (0.005)				0.061*** (0.009)
$D_{\text{TooLowValuation}}$			-0.060*** (0.005)			0.001 (0.006)
$D_{\text{Sell}}$				0.165*** (0.011)		0.087*** (0.013)
$D_{\text{Buy}}$					-0.074*** (0.005)	-0.024*** (0.005)
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	5740	7709	7709	5394	5394	4024
Adjusted R <sup>2</sup>	9.30%	7.80%	1.50%	7.80%	4.00%	16.50%



Panel B: Individual Investors						
Investor Subsample:	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Ind.	Ind.	Ind.	Ind.	Ind.	Ind.
	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$	$\pi(i,t)$
Expected Returns	-0.522*** (0.012)					-0.429*** (0.114)
D <sub>TooHighValuation</sub>		0.084*** (0.006)				0.062*** (0.007)
D <sub>TooLowValuation</sub>			-0.029*** (0.007)			0.01 (0.007)
D <sub>Sell</sub>				0.153*** (0.012)		0.108*** (0.014)
D <sub>Buy</sub>					-0.060*** (0.005)	-0.023*** (0.006)
Weekday FEs	YES	YES	YES	YES	YES	YES
Month FEs	YES	YES	YES	YES	YES	YES
N	5197	6706	6706	6255	6255	4842
Adjusted R <sup>2</sup>	5.20%	3.90%	0.20%	4.30%	2.00%	11.60%