# Crash Beliefs From Investor Surveys

William N. Goetzmann

*Yale School of Management, Yale University*

Dasol Kim

*Weatherhead School of Management, Case Western Reserve University*

Robert J. Shiller

*Yale University*

Draft: March 19, 2016

*Please do not quote without permission*

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 the factors that influence investor responses and test the role of media influence. We find evidence consistent with an availability bias. Recent market declines and adverse market events made salient by the financial press are associated with higher subjective crash probabilities. Non-market-related, rare disasters are also associated with higher subjective crash probabilities.

Keywords: Crash Beliefs, Availability Bias, Investor Surveys

JEL: G00, G11, G23, E03, G02

Acknowledgements: We thank 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 Brian Knutson, Alan Moreira, Tyler Muir, John Scruggs, 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. Please direct correspondence to: william.goetzmann@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,[1] time-varying market premiums,[2] cross-sectional differences in asset returns,[3] the volatility smile,[4] and investor choice.[5] 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.[6]

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 investors. 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 19th, 1987 [ -22.61%] or October 28th 1929 [-12.82%]. 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."

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.

---

[1] Cf. Reitz (1988), Barro (2006), Berkman et. al. (2011) and Welch (2015), Santa-Clara and Yan (2010).
[2] Cf. Gabaix (2008), Wachter (2013), Tsai and Wachter (2015) and Manela and Moreira (2016).
[3] Cf. Gao and Song (2015)
[4] Cf. Bollerslev and Todorov (2011), Bates (2000)
[5] Cf. Guerrero et.al. (2015).
[6] Cf. Jackwerth et. al. (1996), Seo and Wachter (2013) and Ross (2015).

In particular, we test for the incremental effects of positive vs. negative crash-related terms about the market on the days prior to the survey. We find evidence that the financial press mediates investor crash beliefs asymmetrically. Articles with "crash" related terms are associated with higher crash probability assessments, but articles with "boom" related terms are not. We explore the question of whether this association operates through the availability heuristic, and extent to which affect plays a role.

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.[7] 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 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. Finally, we also find evidence consistent with an availability bias when examining

[7] Cf. Barberis (2013).

the crash probabilities of investors who recently experienced exogenous rare events; in this case, moderate earthquakes.

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 (2000) 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 the Shiller survey 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.[8] 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-valence 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 valence; 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.

[8] 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.

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.[9]

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 declines 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. Of course, it is possible that selection or survivorship has biased the empirical estimate downwards. However, the frequency of a major one-day crash would need to be ten times that observed in the US data, and have resulted in non-survival in order to arrive at a conditional sample whose unconditional probability of a crash is consistent with the subjective probability estimates in the Shiller survey.

*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 is 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

[9] Cf. Gabaix (2008), Wachter (2013), Bollerslev and Todorov (2011).

influence media articles and investor heuristics. Market volatility implied by the VIX is obtained from FRED.

*2.3. Media Data*

We used 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 January 24 & 31 of 2016. We searched articles containing words and phrases associated with a stock market boom or a stock market crash.

The terms “stock market boom” and “stock market crash” came into widespread use in American English in the 1920’s. Before 1924, there were virtually no instances of these terms in the Google Ngram corpus of books published in America. This coincides closely with the emergence of widespread stock market investing in the United States. The frequency of the use of both terms rose rapidly from 1929 to a local peak in 1933, doubtless due to the crash of 1929. Their frequencies were more or less stable until the1987-1990 period when the use of the term “stock market crash” more than doubled in frequency and then declined – with some variation – until 2008, which is the terminal date for the corpus. 2003-2004 saw a local maximum for the term “stock market boom” but the average ratio of the two terms is about 7:1 – with “stock market crash” the more prevalent. While there are potential synonyms for “crash” and “boom”, and constructing a variable through topic modeling or other latent semantic extraction techniques has potential, our approach in this paper is to focus on the term “crash” and what we take to be its logical antonym. We also augment the crash/boom pairing with more general positive and negative terms such as “good/bad” and “good news/bad news.” These terms are less specific descriptors of the market and are more moderate in valence, but they increase our sample size.

Because of the potential for data-snooping, all searches are listed in Table 1, and the terms we use for analysis are identified. Although some of the terms in the table are only tangentially related to the current study, we have retained them for completeness. In certain cases, we searched on a term like “market crash” and then discovered that many articles were about other kinds of markets. We then re-ran the search with the added term “stock market” but retained the unconditional results for completeness. In addition, we intended to test some predictions about the relation between negative events and causality. Although not the topic of this study, we include these for completeness.

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. 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 (1) negative news is more prevalent in our sample of crash and boom related terms, and (2) the availability bias – if it exists – should be asymmetric. Negative events and terms should have a greater effect on probability assessments than positive events and terms.

Table 1 summarizes the results of the ProQuest search. Of some interest is the higher number of articles containing the words “good news” [15,372] as opposed to “bad news” [10,751]. This contrasts to the presumption that the news generally has a negativity bias. However, when we condition on the additional term “stock market,” this ratio decreases [2,342 versus 2,182] and is not statistically significantly different from the fraction of positive DJIA days [52%].

## 3. Empirical Results

### *3.1. Summary Statistics*

Table 2 displays the variable descriptions and summary statistics. The interquartile range of the stock indices are comparable, through the overall range for the NYSE-AMEX-Nasdaq-Arca 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 19% respectively indicating a positive skew.

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

We begin by examining the relationship between returns and the valence and subject matter of WSJ articles on the following day. As a preview of the results, we show that negative returns in the prior day(s) are associated with significantly higher negative article counts, and positive returns are associated with significantly higher positive article counts, although the positive results are somewhat weaker.[10] There are significant coefficients on volatility, signed extreme returns, prior month returns, and positive/negative article counts.

[10] Negative article count refers to the number of articles containing the term “stock market” and a negative valence term or phrase, such as “crash” , “market crash”, “bad” or “bad news.” Positive article count is define analogously. Positive and negative valence terms are indicated in Table 2.

We estimate the following specification:

$$Count_t = \beta_1 r_{t-1} + \beta_2 r_{t-2} + \beta_3 r_{t-31,t-2} + \beta_4 \sigma_{t-31,t-2} + \beta_5 Count^{+}_{t-31,t-2} + \beta_6 Count^{-}_{t-31,t-2} + FE(Day\ of\ Week, Month) + \varepsilon_t \tag{1}$$

where $Count_t$ is defined in three different ways: the natural log of one plus the number of articles classified as positive on date t ($Count_t^{+}$), the natural log of one plus the number of articles classified as negative on date t ($Count_t^{-}$), and the difference between the natural logs of one plus the number of articles classified as positive and negative on date t (Net$Count_t$). We define $r_{t-1}$ as the market return on the prior day, $\sigma_{t-31,t-2}$ as the daily volatility estimated over the period t-31 to t-2. In alternative specifications, we replace $r_{t-1}$ with $D_{r(t-1)<10\%}$, a dummy variable for whether prior day return is in the bottom sample decile, and $D_{r(t-1)>90\%}$, a dummy variable for whether the prior day return is in the top sample decile. 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, the S&P500 and the DJIA. In all models, we include day-of-the-week and month fixed effects and cluster errors daily.

Table 3 displays the results. Columns (1) and (2) report results for net counts, Columns (3) and (4) report results for negative article counts and Columns (5) and (6) report results for positive article counts.

Across the specifications, the previous day return is positively associated with the net article counts, 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, using the terms we selected to search on. Columns (2), (4) and (6) present similar results using dummies based upon whether the previous day returns are in the top or bottom sample decile. We include this specification to address the possibility that only unusual market moves are deemed newsworthy. This specification also mitigates the effect of outliers driving the return-based results – i.e., a few extreme market moves accounting for the significance of the coefficients. Only the low return dummy is significant for the net article counts and negative article count models, while only the high return dummy is significant for the positive article count models. The estimates are generally larger for the negative than positive article counts, consistent with asymmetry in the association, although the difference is not statistically significant. The regressions also indicate an association between prior month returns and prior month volatility and prior month counts of articles containing positive vs. negative terms. The coefficients on these variables are all consistent with the media responding to market trends and with the selected terms used in the analysis as meaningful measures of media valence. Volatility is associated with an increased frequency of both negative and positively toned articles about the stock market and there is a temporal dependence in valence. This temporal dependence is consistent with temporal dependence in volatility. The results are not specific to the use of any particular market measure.

To test whether these results are due to the general association between the "good news/bad news" and "stock market" vs. the "crash/boom" pair, we run the analysis separately for the "crash" vs. "boom" pair. The results [not reported] using this subset were insignificant. While this may be due to low power, perhaps the specific the terms "crash" and "boom" are not commonly used by the press to characterize daily stock market trends.

Next, we examine the relationship between media responses and past market returns. The motivation for this is that journalists themselves may be influenced more by negative market returns and thus more prone to focusing attention on them. Table 4 displays the results. For each set of returns variables, future article counts are projected onto the return variable, while using the same control variables of Table 3. We separate positive and negative media responses and regress counts on past markets returns for up to six lags.

The association between returns and negative media responses appear to persist for up to seven days subsequently, while the association of returns with positive article counts subsides within one or two days. The coefficients attenuate as the article count variable is measured further into the future. When using a lag of beyond six days, the results become statistically insignificant [not reported].

ProQuest also provides information about whether the article appears on the front page or elsewhere in the newspaper. This allows us to test whether negative events are accorded greater prominence by the media. Table 5 reports regressions of counts on prior day returns, breaking out front-page from vs. non-front-page placement. The dependent variables count the number of negative and positive articles that appear on the front page of any section (Front) or not (NotFront).

Low return days have a strong positive association with the number of negative articles appearing on the front page, while positive return days have a strong positive association with the number of articles appearing in other pages. 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.

### *3.3. Crash Probabilities*

We next test whether media valence 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: negative media valence during market downturns and positive valence during market upswings. Given that the survey results 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).

We estimate the following specification:

$$\pi_{i,t} = \beta_1 r_t \times Count_t^- + \beta_2 r_t \times Count_t^+ + \beta_3 r_t + \beta_4 Count_t^- + \beta_5 Count_t^+ + \beta_6 r_{t-30,t-1} + \beta_7 \sigma_{t-30,t-1} + \beta_8 VIX_{t-1} + \beta_9 Crash_{t-30,t-1} + FE(Day\ of\ Week, Month) + \varepsilon_{i,t} \quad (2)$$

where the dependent variable $\pi_{i,t}$ is the probability assessment of investor *i* at time *t*. The explanatory variables are as specified above, and include returns, media valence, and interaction terms between returns and media valence on the same day the respondent filled out the survey. We include control variables related to returns, and the average crash probabilities of other investors over the previous 30 days, as well as measures of backward-looking and forward-looking market volatility: VIX and daily market volatility over the prior 30 days. In all models, we include day-of-the-week and month fixed effects and cluster errors daily.

Table 6 presents the results. 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 returns variables are based upon NYSE/AMEX/Nasdaq/Arca (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 media valence during market downturns, while the same does not hold for positive media valence during market upswings. The (two-tailed) null is rejected at the 10% level for the two broader indexes, and at the 5% level for the DJIA.[11] In contrast, negative media valence is not significant for institutional investors – the coefficient signs are similar but the magnitudes are less. After controlling for media valence, returns for the prior day do not have a significant coefficient, nor do negative article counts by themselves. The significance seems to be confined to the interaction term. In contrast, positive news valence is associated with a decrease in crash probabilities for both sets of investors; irrespective of the daily return.

[11] 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.

*3.4 Robustness Checks*

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 valence 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 valence interaction terms with same day returns, two additional interaction terms are added between media valence and returns over the previous 30 days. The results are similar but slightly stronger for the same day return interaction terms. The interaction terms associated with previous 30 day returns are statistically insignificant. Furthermore, the coefficients are considerably smaller in absolute magnitude than those for the same day returns.

Table 8 checks whether the interaction with previous month's media valence is significant. The coefficient same day negative media valence during market downturns remain significant, while negative media valence over the previous month during market downturns are insignificant. In contrast, positive media valence over the same day during market upswings significantly decreases crash probabilities of individual investors, but not institutional investors. In other words, when controlling for past media valence, we obtain similar results for positive media valence as well as negative media valence – at least in one specification. Additionally, the significance of uninteracted, same day, positive media valence decreases substantially after controlling for past media valence.

We next consider whether the significance of media valence varies according to the page placement of the related articles. The availability heuristic predicts that a market return accompanied by a front-page news article will have relatively more influence on the subject's probability assessment. Table 9 separates the media valence terms according to whether they are front page articles or not. We find that the association between negative media valence and subjective probabilities during market downturns is driven by front page articles – negative articles that are not on the front page do not have a significant coefficient. Positive media valence during market upswings also has a significant coefficient, albeit weaker, although these results are concentrated in positive articles that are not on the front page. This may reflect a choice by the paper to accentuate the negative by moving it to the front page.

Table 9 breaks out results by general vs. specific valence terms. In particular, we test whether the article counts for the "crash/boom" (specific) antonym pair influences investor probability assessments, compared to the terms such as "good" or "good news". We would expect stronger results from the higher valence "crash" and "boom" terms, used more commonly to describe the stock market or the economy, and associated in the American English Google Books corpus with the 1929 and 1987 events referenced in the question. In the discussion of Table 3, we noted that the use of these terms was not associated with past returns, or even with extreme returns. In Table 9, we find that the significance of media interaction with past returns is confined to these high valence, or specific, terms. When the word "crash" or "boom" is frequently used in articles about the stock market, it seems to be highly relevant to investor probability assessment of a future crash.

In prior work we documented 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.

*3.5 Exogenous Rare Events*

One drawback thus far in the analysis is that the media use of the negative valence terms, the stock market and the individual probability assessment 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 day 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. Ferreira and Karali (2015) show that, despite this link, stock markets react little to earthquakes. To address this possibility we distinguish between weak magnitudes, or earthquakes with a magnitude of between 2.5 up to 5.5, and strong magnitudes, or earthquakes with a magnitude of 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 2007 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, while 0.2% experienced a strong earthquake.

Table 11 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 strong magnitude earthquakes. The coefficients on the strong magnitude earthquakes are generally larger, though are not statistically significant – perhaps due to the small number of such event in the sample.

As a robustness check, we compare the effect related to the timing of the earthquakes. We expect the effects to be pronounced in the time period shortly after the event, and attenuate over time. Table 12 presents the results. The effects are not significant two and three months after the event, only for the most recent 30 days. Again, the results only hold for individual investors.

## 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 role potentially 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 counts of articles containing positive or negative valence terms of the same sign. We find some evidence of asymmetry in reporting. Extreme negative returns are more likely to be followed by higher counts of articles with negative valence terms. Negative returns appear to influence the counts articles with negative valence terms for several days, whereas the association with article counts containing positive valence terms is confined to a single day. Front page placement of articles containing negative valence terms is more likely than for those with positive valence terms. 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 valence 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 appears to make a difference for negative valence articles. The high valence term “crash” appears to drive the main results.

Finally, we use local earthquakes as an 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 on 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 of current information, conditional upon negative outcomes or valence, 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., 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, p.hhu063.

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.

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.

Loewenstein, G.F., Weber, E.U., Hsee, C.K. and Welch, N., 2001. Risk as feelings. *Psychological bulletin*, *127*(2), p.267.

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.

Manela, A. and Moreira, A., 2013. News implied volatility and disaster concerns. Forthcoming, Journal of Financial Economics.

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.

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., 2000. Measuring bubble expectations and investor confidence. Journal of Psychology and Financial Markets, 1, pp. 49-60.

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.

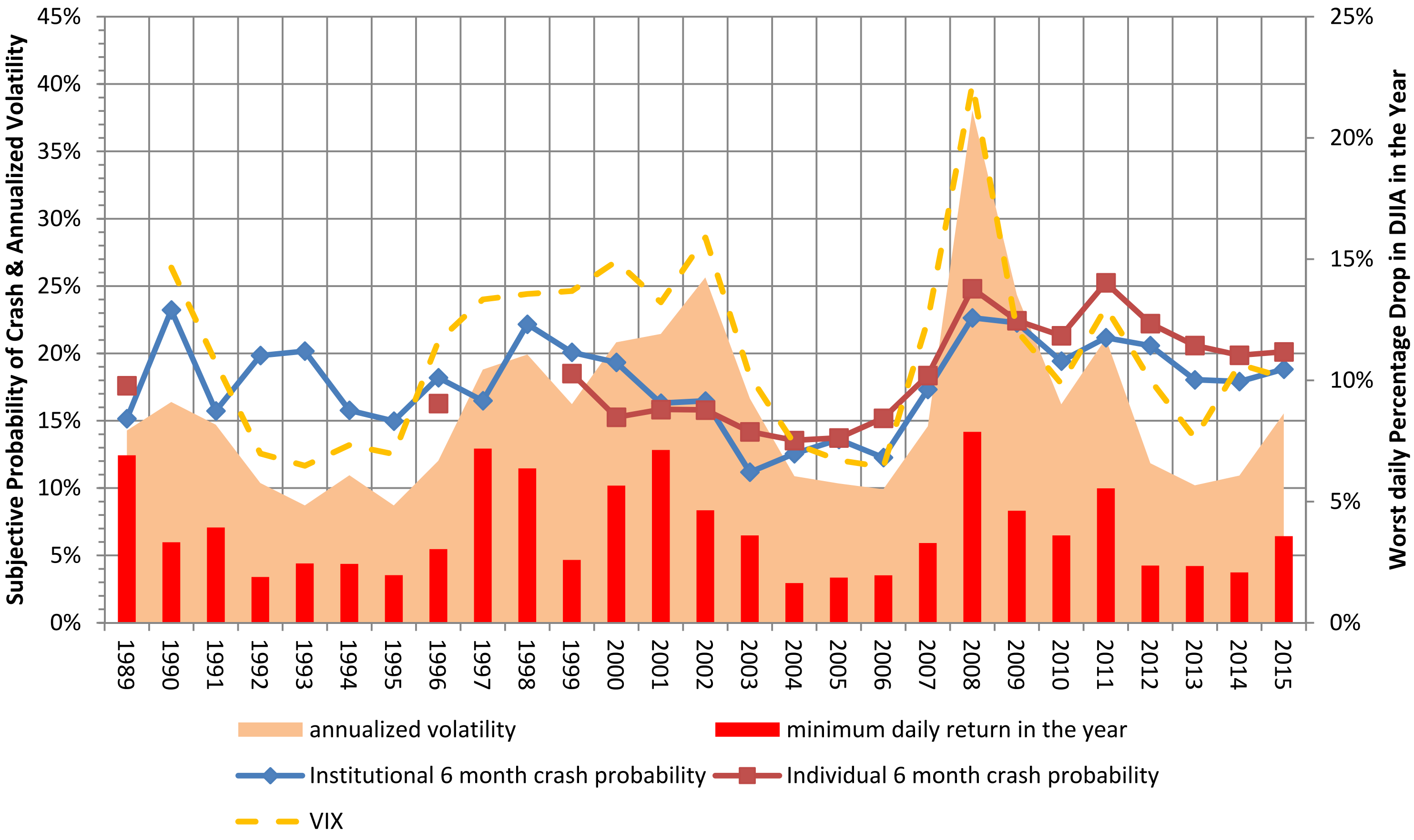
Wachter, J.A., 2013. Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?. The Journal of Finance, 68(3), pp.987-1035.

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

.



## Table 1: Article Keyword, Tone and Sample Frequency

There are 8,396 unique articles. Article counts by tone for each day are based upon unique articles (e.g., not double counted if same article has same keyword).

| Keyword | Number of Articles | Positive/Negative |
|---|---|---|
| bad_news_AND_stock_market | 2182 | Negative |
| market_crash_AND_stock_market | 2268 | Negative |
| stock_market_crash | 1991 | Negative |
| bad_news_AND_stock_market_AND_cause | 195 | Negative |
| market_crash_AND_stock_market_AND_cause | 74 | Negative |
| stock_market_crash_AND_cause | 182 | Negative |
| good_news_AND_stock_market | 2342 | Positive |
| market_boom_AND_stock_market | 322 | Positive |
| stock_market_boom | 275 | Positive |
| good_news_AND_stock_market_AND_cause | 219 | Positive |
| market_boom_AND_stock_market_AND_cause | 18 | Positive |
| stock_market_boom_AND_cause | 14 | Positive |
| good | 192,884 | |
| bad | 80,183 | |
| good_news | 15,372 | |
| bad_news | 10,751 | |
| 1929_crash | 131 | |
| 1987_crash | 690 | |
| caused_the_boom | 5 | |
| caused_the_crash | 88 | |
| crash_of_1929 | 293 | |
| crash_of_1987 | 248 | |
| John_Law | 54 | |
| market_boom | 453 | |
| market_crash | 2272 | |
| market_drop | 442 | |
| market_rise | 127 | |
| market_up | 915 | |
| market_down | 859 | |
| market_up_AND_stock_market | 358 | |
| market_down_AND_stock_market | 462 | |
| stock_market_advance | 20 | |
| stock_market_decline | 314 | |
| stock_market_down | 124 | |
| stock_market_up | 103 | |
| stocks_up | 612 | |
| stocks_down | 750 | |
| stocks_up_AND_stock_market | 276 | |
| stocks_down_AND_stock_market | 368 | |

## Table 2: 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 Description

| **Variable Name** | **Description** | **Source** |
|---|---|---|
| Returns Variables: | | |
| r (t-1) | Total return on date t-1 based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA index. | CRSP |
| r (t-31,t-2) | Total cumulative return from date t-31 to date t-2 based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA index. | CRSP |
| $\sigma$ (t-31,t-2) | Daily returns volatility from dates t-31 to date t-2 based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA index. | CRSP |
| Media Variables: | | |
| $Count^{+}$ (t) | The natural log of one plus the number of articles classified to have positive valence on date t. | WSJ |
| $Count^{-}$ (t) | The natural log of one plus the number of articles classified to have negative valence on date t. | WSJ |
| NetCount(t) | The difference between $Count^{+}$ (t) and $Count^{-}$ (t) | WSJ |
| $Count^{+}$ (t-31,t-2) | The natural log of one plus the average number of articles classified to have positive valence from date t-31 to date t-2. | WSJ |
| $Count^{-}$ (t-31,t-2) | The natural log of one plus the average number of articles classified to have negative valence from date t-31 to date t-2. | WSJ |
| Survey Variables: | | |
| $\pi$(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 date 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 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **Mean** | **StDev** | **Min** | **Q1** | **Median** | **Q3** | **Max** |
| Returns Variables: | | | | | | | | |
| r (t-1) (All) | 3430 | 0.000 | 0.012 | -0.090 | -0.005 | 0.001 | 0.006 | 0.115 |
| r (t-31,t-2) (All) | 3430 | 0.009 | 0.058 | -0.320 | -0.018 | 0.016 | 0.044 | 0.288 |
| $\sigma$ (t-31,t-2) (All) | 3430 | 0.010 | 0.006 | 0.003 | 0.006 | 0.008 | 0.011 | 0.050 |
| r (t-1) (SP500) | 3430 | 0.000 | 0.013 | -0.090 | -0.005 | 0.000 | 0.006 | 0.116 |
| r (t-31,t-2) (SP500) | 3430 | 0.007 | 0.056 | -0.309 | -0.020 | 0.013 | 0.041 | 0.274 |
| $\sigma$ (t-31,t-2) (SP500) | 3430 | 0.010 | 0.006 | 0.003 | 0.007 | 0.009 | 0.012 | 0.051 |
| r (t-1) (DJIA) | 3430 | 0.000 | 0.012 | -0.079 | -0.005 | 0.000 | 0.006 | 0.111 |
| r (t-31,t-2) (DJIA) | 3430 | 0.008 | 0.055 | -0.279 | -0.019 | 0.014 | 0.042 | 0.233 |
| $\sigma$ (t-31,t-2) (DJIA) | 3430 | 0.010 | 0.006 | 0.003 | 0.007 | 0.008 | 0.012 | 0.049 |
| Media Variables: | | | | | | | | |
| $Count^{+}$(t) | 3430 | 0.355 | 0.410 | 0.000 | 0.000 | 0.000 | 0.693 | 1.946 |
| $Count^{-}$ (t) | 3430 | 0.535 | 0.471 | 0.000 | 0.000 | 0.693 | 0.693 | 2.197 |
| NetCount(t) | 3430 | -0.355 | 1.226 | -7.000 | -1.000 | 0.000 | 1.000 | 4.000 |
| $Count^{+}$(t-31,t-2) | 3430 | 0.558 | 0.284 | 0.000 | 0.333 | 0.524 | 0.750 | 1.737 |
| $Count^{-}$ (t-31,t-2) | 3430 | 0.915 | 0.493 | 0.067 | 0.571 | 0.789 | 1.105 | 2.870 |
| Survey Variables: | | | | | | | | |
| $\pi$(i,t) | 9953 | 0.194 | 0.199 | 0.000 | 0.050 | 0.100 | 0.250 | 0.999 |
| $\pi$(t-30,t-1) | 9953 | 0.185 | 0.059 | 0.000 | 0.149 | 0.182 | 0.217 | 0.775 |
| Institutional | 9953 | 0.569 | 0.495 | 0.000 | 0.000 | 1.000 | 1.000 | 1.000 |

## Table 3: Media Valence and Previous Day Returns

The table displays the results from OLS regression models where the dependent variable is the difference between the natural log of one plus the number of positive versus negative articles (NetCount); natural log of one plus the number of negative articles ($Count^{-}$); and natural log of one plus the number of positive articles ($Count^{+}$). The results are displayed separately where the returns variables are based upon All (NYSE/AMEX/Nasdaq/Arca) (Panel A), S&P 500 (Panel B) or DJIA (Panel C) indices. Robust standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as ***, **, and *, respectively.

Panel A: All Returns

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dependent Variable: | NetCount(t) | NetCount(t) | $Count^{-}$ (t) | $Count^{-}$ (t) | $Count^{+}$ (t) | $Count^{+}$ (t) |
| r (t-1) | 3.239*** | | -2.012*** | | 1.227** | |
| | (0.857) | | (0.696) | | (0.593) | |
| D(Rk(r (t-1))<10%) | | -0.084** | | 0.095*** | | 0.011 |
| | | (0.035) | | (0.026) | | (0.023) |
| D(Rk(r (t-1))>90%) | | 0.053 | | 0.008 | | 0.061** |
| | | (0.034) | | (0.025) | | (0.025) |
| r (t-2) | 0.991 | 0.867 | -0.684 | -0.513 | 0.306 | 0.355 |
| | (0.901) | (0.903) | (0.705) | (0.702) | (0.625) | (0.634) |
| r (t-31,t-2) | 0.404* | 0.374* | -0.584*** | -0.536*** | -0.179 | -0.162 |
| | (0.211) | (0.213) | (0.154) | (0.154) | (0.145) | (0.146) |
| σ (t-31,t-2) | 4.277** | 4.431** | -2.749* | -3.765** | 1.528 | 0.666 |
| | (1.962) | (2.033) | (1.462) | (1.524) | (1.352) | (1.416) |
| $Count^{+}$ (t-31,t-2) | 0.628*** | 0.633*** | -0.099** | -0.107** | 0.529*** | 0.526*** |
| | (0.061) | (0.061) | (0.045) | (0.045) | (0.040) | (0.040) |
| $Count^{-}$ (t-31,t-2) | -0.737*** | -0.736*** | 0.677*** | 0.680*** | -0.060** | -0.056* |
| | (0.045) | (0.045) | (0.034) | (0.034) | (0.030) | (0.030) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| N | 3430 | 3430 | 3430 | 3430 | 3430 | 3430 |
| Adjusted $R^2$ | 12.17% | 12.03% | 13.32% | 13.40% | 6.60% | 6.63% |

| Panel B: S&P 500 Returns | | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Dependent Variable: | NetCount(t) | NetCount(t) | $Count^{-}$ (t) | $Count^{-}$ (t) | $Count^{+}$ (t) | $Count^{+}$ (t) |
| r (t-1) | 3.133*** | | -1.963*** | | 1.170** | |
| | (0.842) | | (0.691) | | (0.585) | |
| D(Rk(r (t-1))<10%) | | -0.087** | | 0.095*** | | 0.009 |
| | | (0.035) | | (0.026) | | (0.023) |
| D(Rk(r (t-1))>90%) | | 0.027 | | 0.022 | | 0.048** |
| | | (0.034) | | (0.025) | | (0.024) |
| r (t-2) | 1.165 | 0.964 | -0.781 | -0.579 | 0.384 | 0.385 |
| | (0.882) | (0.884) | (0.697) | (0.689) | (0.620) | (0.620) |
| r (t-31,t-2) | 0.320 | 0.273 | -0.511*** | -0.451*** | -0.192 | -0.178 |
| | (0.219) | (0.219) | (0.160) | (0.160) | (0.150) | (0.150) |
| σ (t-31,t-2) | 3.143 | 3.604* | -1.703 | -2.856* | 1.440 | 0.747 |
| | (1.952) | (2.036) | (1.468) | (1.536) | (1.359) | (1.437) |
| $Count^{+}$ (t-31,t-2) | 0.633*** | 0.640*** | -0.103** | -0.108** | 0.531*** | 0.532*** |
| | (0.061) | (0.061) | (0.045) | (0.045) | (0.040) | (0.040) |
| $Count^{-}$ (t-31,t-2) | -0.740*** | -0.739*** | 0.679*** | 0.681*** | -0.061** | -0.058* |
| | (0.046) | (0.046) | (0.035) | (0.035) | (0.030) | (0.030) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| N | 3430 | 3430 | 3430 | 3430 | 3430 | 3430 |
| Adjusted $R^2$ | 12.10% | 11.92% | 13.20% | 13.29% | 6.59% | 6.56% |

| Panel C: DJIA Returns | | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Dependent Variable: | NetCount(t) | NetCount(t) | $Count^{-}$ (t) | $Count^{-}$ (t) | $Count^{+}$ (t) | $Count^{+}$ (t) |
| r (t-1) | 2.951*** | | -2.064*** | | 0.888 | |
| | (0.871) | | (0.740) | | (0.621) | |
| D(Rk(r (t-1))<10%) | | -0.074** | | 0.083*** | | 0.009 |
| | | (0.035) | | (0.026) | | (0.022) |
| D(Rk(r (t-1))>90%) | | 0.021 | | 0.020 | | 0.042* |
| | | (0.035) | | (0.025) | | (0.025) |
| r (t-2) | 1.460 | 1.265 | -1.205 | -0.999 | 0.255 | 0.266 |
| | (0.930) | (0.930) | (0.739) | (0.735) | (0.655) | (0.664) |
| r (t-31,t-2) | 0.297 | 0.255 | -0.417** | -0.357** | -0.120 | -0.102 |
| | (0.223) | (0.226) | (0.163) | (0.163) | (0.152) | (0.152) |
| σ (t-31,t-2) | 2.374 | 2.816 | -0.867 | -1.947 | 1.508 | 0.869 |
| | (2.120) | (2.200) | (1.605) | (1.678) | (1.493) | (1.552) |
| $Count^{+}$ (t-31,t-2) | 0.640*** | 0.643*** | -0.106** | -0.110** | 0.533*** | 0.534*** |
| | (0.061) | (0.061) | (0.045) | (0.045) | (0.040) | (0.040) |
| $Count^{-}$ (t-31,t-2) | -0.739*** | -0.738*** | 0.679*** | 0.681*** | -0.059* | -0.056* |
| | (0.046) | (0.046) | (0.035) | (0.035) | (0.030) | (0.030) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| N | 3430 | 3430 | 3430 | 3430 | 3430 | 3430 |
| Adjusted $R^2$ | 12.03% | 11.86% | 13.17% | 13.16% | 6.47% | 6.47% |

## Table 4: Persistence of Media Valence

The table displays the results from OLS regression models where the dependent variable is the natural log of one plus the number of negative articles (Panel A) and natural log of one plus the number of positive articles (Panel B). The results are displayed separately where the returns variables are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. All regressions include the control variables of Table 4. Robust standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as ***, **, and *, respectively.

Panel A: Negative Tone Article Counts by Day

| Dependent Variable: | (1) Count $^-$ (t+1) | (2) Count $^-$ (t+2) | (3) Count $^-$ (t+3) | (4) Count $^-$ (t+4) | (5) Count $^-$ (t+5) | (6) Count $^-$ (t+6) |
|---|---|---|---|---|---|---|
| All Returns | | | | | | |
| r (t-1) | -2.012*** | -1.267* | -1.583** | -1.997*** | -1.029 | -1.458** |
| | (0.696) | (0.700) | (0.665) | (0.653) | (0.627) | (0.657) |
| S&P 500 Returns | | | | | | |
| r (t-1) | -1.963*** | -1.258* | -1.439** | -1.923*** | -1.061* | -1.337** |
| | (0.691) | (0.691) | (0.660) | (0.647) | (0.628) | (0.649) |
| DJIA Returns | | | | | | |
| r (t-1) | -2.064*** | -1.619** | -1.525** | -1.987*** | -1.141* | -1.000 |
| | (0.740) | (0.736) | (0.703) | (0.692) | (0.667) | (0.690) |

Panel B: Positive Tone Article Counts by Day

| Dependent Variable: | (1) Count $^+$ (t+1) | (2) Count $^+$ (t+2) | (3) Count $^+$ (t+3) | (4) Count $^+$ (t+4) | (5) Count $^+$ (t+5) | (6) Count $^+$ (t+6) |
|---|---|---|---|---|---|---|
| All Returns | | | | | | |
| r (t-1) | 1.227** | 0.163 | 0.564 | -0.856 | -0.318 | 0.037 |
| | (0.593) | (0.628) | (0.593) | (0.629) | (0.588) | (0.619) |
| S&P 500 Returns | | | | | | |
| r (t-1) | 1.170** | 0.219 | 0.556 | -0.822 | -0.228 | 0.025 |
| | (0.585) | (0.627) | (0.585) | (0.618) | (0.586) | (0.624) |
| DJIA Returns | | | | | | |
| r (t-1) | 0.888 | 0.172 | 0.570 | -0.715 | -0.062 | 0.104 |
| | (0.621) | (0.661) | (0.612) | (0.644) | (0.622) | (0.614) |

## Table 5: Article Placement

The table displays the results from OLS regression models where the dependent variable is the natural log of one plus the number of negative valence articles that placed on the first page of any section (Count $^{Front,-}$) or not (Count $^{NotFront,-}$), and the natural log of one plus the number of positive valence articles that placed on the first page of any section (Count $^{Front,+}$) or not (Count $^{NotFront,+}$). The results are displayed separately where the returns variables are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500, or DJIA indices. Robust standard errors are displayed in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted as ***, **, and *, respectively.

| Dependent Variable: | (1)<br>Count $^{Front,-}$ (t) | (2)<br>Count $^{NotFront,-}$ (t) | (3)<br>Count $^{Front,+}$ (t) | (4)<br>Count $^{NotFront,+}$ (t) |
|---|---|---|---|---|
| All Returns | | | | |
| r (t-1) | -2.428*** | -0.237 | 0.271 | 1.028** |
| | (0.553) | (0.623) | (0.460) | (0.487) |
| S&P 500 Returns | | | | |
| r (t-1) | -2.284*** | -0.334 | 0.238 | 1.009** |
| | (0.546) | (0.618) | (0.457) | (0.478) |
| DJIA Returns | | | | |
| r (t-1) | -2.456*** | -0.369 | 0.077 | 0.794 |
| | (0.588) | (0.660) | (0.479) | (0.502) |

**Table 6: Crash Probabilities and Media Valence**

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 returns variables (r) are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Robust standard errors clustered on the date level 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) × Count $^{-}$ (t) | -0.877* | -0.154 | -0.904* | -0.223 | -1.160** | -0.170 |
| | (0.519) | (0.453) | (0.522) | (0.438) | (0.553) | (0.459) |
| r (t) × Count $^{+}$ (t) | -0.978 | 0.263 | -0.753 | 0.249 | -0.339 | 0.272 |
| | (0.776) | (0.611) | (0.784) | (0.580) | (0.808) | (0.619) |
| r (t) | 0.160 | -0.186 | 0.149 | -0.175 | 0.058 | -0.256 |
| | (0.348) | (0.305) | (0.347) | (0.302) | (0.361) | (0.320) |
| Count $^{-}$ (t) | 0.004 | 0.000 | 0.004 | 0.000 | 0.004 | 0.001 |
| | (0.008) | (0.005) | (0.008) | (-0.006) | (0.008) | (0.005) |
| Count $^{+}$ (t) | -0.030*** | -0.017*** | -0.030*** | -0.016** | -0.030*** | -0.016** |
| | (0.008) | (0.006) | (0.008) | (0.006) | (0.008) | (0.006) |
| r (t-30,t-1) | -0.133** | -0.033 | -0.127* | -0.025 | -0.146** | 0.000 |
| | (0.062) | (0.054) | (0.066) | (0.057) | (0.066) | (0.057) |
| $\sigma$ (t-30,t-1) | 0.354 | -1.561* | 0.020 | -1.782* | -1.214 | -2.000** |
| | (1.105) | (0.892) | (0.993) | (0.914) | (1.084) | (0.935) |
| VIX (t-1) | 0.001 | 0.002*** | 0.001* | 0.003*** | 0.002*** | 0.003*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| $\pi$ (t-30,t-1) | 0.302*** | 0.290*** | 0.309*** | 0.289*** | 0.310*** | 0.286*** |
| | (0.064) | (0.054) | (0.063) | (0.054) | (0.062) | (0.054) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| N | 4286 | 5667 | 4286 | 5667 | 4286 | 5667 |
| Adjusted $R^2$ | 2.21% | 1.74% | 2.17% | 1.75% | 2.23% | 1.77% |

## Table 7: Salience of Past Returns

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 returns variables (r) are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Control variables of Table 6 are included in all the models but not reported. Robust standard errors clustered on the date level 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) × Count $^-$ (t) | -0.974* | -0.225 | -1.036** | -0.305 | -1.219** | -0.250 |
| | (0.505) | (0.460) | (0.503) | (0.442) | (0.535) | (0.462) |
| r (t) × Count $^+$ (t) | -1.184 | 0.091 | -0.993 | 0.075 | -0.554 | 0.157 |
| | (0.774) | (0.609) | (0.770) | (0.578) | (0.803) | (0.627) |
| r (t-30,t-1) × Count $^-$ (t) | -0.175 | -0.063 | -0.203 | -0.094 | -0.235* | -0.140 |
| | (0.121) | (0.089) | (0.132) | (0.094) | (0.140) | (0.096) |
| r (t-30,t-1) × Count $^+$ (t) | -0.131 | -0.185 | -0.174 | -0.175 | -0.065 | -0.078 |
| | (0.159) | (0.116) | (0.174) | (0.124) | (0.176) | (0.124) |
| r (t) | 0.208 | -0.169 | 0.209 | -0.151 | 0.099 | -0.237 |
| | (0.347) | (0.302) | (0.343) | (0.297) | (0.367) | (0.320) |
| r (t-30,t-1) | -0.047 | 0.032 | -0.026 | 0.050 | -0.061 | 0.073 |
| | (0.073) | (0.068) | (0.078) | (0.071) | (0.081) | (0.073) |
| Count $^-$ (t) | 0.005 | -0.001 | 0.004 | 0.000 | 0.005 | 0.001 |
| | (0.008) | (0.005) | (0.008) | (0.006) | (0.008) | (0.006) |
| Count $^+$ (t) | -0.028*** | -0.015** | -0.029*** | -0.016** | -0.029*** | -0.016** |
| | (0.009) | (0.007) | (0.009) | (0.006) | (0.009) | (0.007) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| Control Variables | YES | YES | YES | YES | YES | YES |
| N | 4286 | 5667 | 4286 | 5667 | 4286 | 5667 |
| Adjusted $R^2$ | 2.25% | 1.77% | 2.24% | 1.78% | 2.26% | 1.78% |

**Table 8: Salience of Past News**

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 returns variables (r) are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Control variables of Table 6 are included in all the models but not reported. Robust standard errors clustered on the date level 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) × Count $^-$ (t) | -1.063** | -0.141 | -1.062** | -0.231 | -1.156** | -0.183 |
| | (0.532) | (0.471) | (0.531) | (0.452) | (0.561) | (0.494) |
| r (t) × Count $^+$ (t) | -1.570** | 0.308 | -1.380* | 0.305 | -0.982 | 0.330 |
| | (0.737) | (0.616) | (0.742) | (0.587) | (0.792) | (0.623) |
| r (t) × Count $^-$ (t-30,t-1) | 0.540 | -0.008 | 0.519 | 0.047 | 0.246 | 0.058 |
| | (0.491) | (0.388) | (0.489) | (0.390) | (0.534) | (0.418) |
| r (t) × Count $^+$ (t-30,t-1) | 0.480 | -0.053 | 0.385 | -0.144 | 0.476 | -0.162 |
| | (0.565) | (0.485) | (0.558) | (0.480) | (0.588) | (0.505) |
| r (t) | -2.145* | -0.065 | -1.893 | 0.007 | -1.524 | -0.053 |
| | (1.198) | (1.078) | (1.191) | (0.713) | (1.270) | (1.062) |
| Count $^-$ (t) | 0.011 | 0.001 | 0.011 | 0.001 | 0.011 | 0.001 |
| | (0.008) | (0.006) | (0.008) | (0.006) | (0.008) | (0.006) |
| Count $^+$ (t) | -0.016* | -0.009 | -0.016** | -0.009 | -0.016** | -0.009 |
| | (0.008) | (0.006) | (0.008) | (0.006) | (0.008) | (0.006) |
| Count $^-$ (t-30,t-1) | -0.012* | 0.004 | -0.013* | 0.004 | -0.012 | 0.006 |
| | (0.007) | (0.005) | (0.007) | (0.005) | (0.007) | (0.005) |
| Count $^+$ (t-30,t-1) | -0.042*** | -0.024*** | -0.042*** | -0.024*** | -0.041*** | -0.025*** |
| | (0.007) | (0.005) | (0.007) | (0.005) | (0.007) | (0.005) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| Control Variables | YES | YES | YES | YES | YES | YES |
| N | 4286 | 5667 | 4286 | 5667 | 4286 | 5667 |
| Adjusted $R^2$ | 3.62% | 2.06% | 3.57% | 2.08% | 3.50% | 2.10% |

## Table 9: Salience of 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 returns variables (r) are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Count $^{Front}$ indicates the daily counts of the subset of negative (-) or positive (+) valence articles that are placed on the first page of any section. Count $^{NotFront}$ indicates the daily counts of the subset of negative (-) or positive (+) valence articles that are not placed on the first page of any section. Control variables of Table 6 are included in all the models but not reported. Robust standard errors clustered on the date level 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) × Count $^{-,Front}$ (t) | -1.357* | -0.883 | -1.610** | -0.877 | -1.947** | -0.611 |
| | (0.798) | (0.701) | (0.781) | (0.665) | (0.821) | (0.679) |
| r (t) × Count $^{-,NotFront}$ (t) | -0.642 | 0.742 | -0.649 | 0.636 | -0.923 | 0.542 |
| | (0.629) | (0.549) | (0.607) | (0.508) | (0.600) | (0.506) |
| r (t) × Count $^{+,Front}$ (t) | -0.262 | -0.580 | 0.065 | -0.619 | 0.550 | -0.688 |
| | (0.937) | (0.784) | (0.924) | (0.737) | (0.916) | (0.740) |
| r (t) × Count $^{+,NotFront}$ (t) | -1.762* | 0.725 | -1.741* | 0.748 | -1.511 | 0.955 |
| | (0.913) | (0.748) | (0.907) | (0.720) | (0.962) | (0.764) |
| r (t) | 0.168 | -0.209 | 0.166 | -0.210 | 0.076 | -0.303 |
| | (0.343) | (0.298) | (0.339) | (0.295) | (0.363) | (0.316) |
| Count $^{-,Front}$ (t) | 0.004 | 0.003 | 0.004 | 0.003 | 0.003 | 0.004 |
| | (0.012) | (0.007) | (0.012) | (0.007) | (0.011) | (0.008) |
| Count $^{-,NotFront}$ (t) | 0.005 | -0.004 | 0.005 | -0.003 | 0.005 | -0.002 |
| | (0.010) | (0.007) | (0.010) | (0.007) | (0.010) | (0.007) |
| Count $^{+,Front}$ (t) | -0.029** | -0.023*** | -0.030** | -0.023** | -0.029** | -0.023** |
| | (0.012) | (0.009) | (0.012) | (0.009) | (0.012) | (0.009) |
| Count $^{+,NotFront}$ (t) | -0.024** | -0.007 | -0.024** | -0.007 | -0.023** | -0.008 |
| | (0.011) | (0.008) | (0.011) | (0.008) | (0.011) | (0.008) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| Control Variables | YES | YES | YES | YES | YES | YES |
| N | 4286 | 5667 | 4286 | 5667 | 4286 | 5667 |
| Adjusted $R^2$ | 2.15% | 1.77% | 2.14% | 1.78% | 2.20% | 1.78% |

## Table 10: Generic versus Specific Valence Terms

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 returns variables (r) are based upon all (NYSE/AMEX/Nasdaq/Arca), S&P 500 or DJIA indices. Count $^{Specific}$ indicates the daily counts of the subset of negative (-) or positive (+) valence articles that uses specific terms. Count $^{Generic}$ indicates the daily counts of the subset of negative (-) or positive (+) valence articles that use general terms. Control variables of Table 6 are included in all the models but not reported. Robust standard errors clustered on the date level 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) × Count $^{-,Specific}$ (t) | -1.451** | -0.282 | -1.305* | -0.335 | -1.393** | -0.230 |
| | (0.688) | (0.513) | (0.673) | (0.465) | (0.686) | (0.490) |
| r (t) × Count $^{-,Generic}$ (t) | -0.157 | -0.213 | -0.287 | -0.173 | -0.598 | -0.113 |
| | (0.581) | (0.533) | (0.552) | (0.508) | (0.575) | (0.540) |
| r (t) × Count $^{+,Specific}$ (t) | -0.815 | 0.800 | -0.641 | 0.945 | 0.208 | 0.998 |
| | (1.537) | (2.000) | (1.563) | (1.889) | (1.735) | (2.037) |
| r (t) × Count $^{+,Generic}$ (t) | -1.359* | 0.068 | -1.146 | 0.023 | -0.890 | -0.020 |
| | (0.768) | (0.564) | (0.769) | (0.584) | (0.795) | (0.677) |
| r (t) | 0.271 | -0.198 | 0.342 | -0.132 | 0.369 | -0.125 |
| | (0.362) | (0.295) | (0.343) | (0.282) | (0.346) | (0.285) |
| Count $^{-,Specific}$ (t) | 0.022* | 0.010 | 0.022* | 0.011 | 0.021* | 0.011 |
| | (0.012) | (0.007) | (0.012) | (0.007) | (0.012) | (0.007) |
| Count $^{-,Generic}$ (t) | -0.009 | -0.005 | -0.009 | -0.004 | -0.008 | -0.004 |
| | (0.009) | (0.007) | (0.009) | (0.007) | (0.009) | (0.007) |
| Count $^{+,Specific}$ (t) | -0.063*** | -0.028* | -0.063*** | -0.028* | -0.063*** | -0.027* |
| | (0.016) | (0.015) | (0.016) | (0.015) | (0.016) | (0.015) |
| Count $^{+,Generic}$ (t) | -0.016* | -0.009 | -0.016* | -0.009 | -0.016* | -0.009 |
| | (0.009) | (0.007) | (0.009) | (0.007) | (0.009) | (0.007) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| Control Variables | YES | YES | YES | YES | YES | YES |
| N | 4730 | 6253 | 4730 | 6253 | 4730 | 6253 |
| Adjusted $R^2$ | 2.17% | 1.32% | 2.12% | 1.33% | 2.14% | 1.33% |

## Table 11: 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 6 are also included, but not reported. The returns variables for the control variables are based upon the DJIA index. Robust standard errors clustered on the date level 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. | Indiv. | Indiv. | Inst. | Inst. | Inst. |
| Dependent Variable: | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ |
| Weak Magnitude (t-30,t) | 0.034** | | 0.032** | -0.013 | | -0.012 |
| | (0.014) | | (0.014) | (0.012) | | (0.013) |
| Strong Magnitude (t-30,t) | | 0.180 | 0.153 | | -0.034 | -0.023 |
| | | (0.160) | (0.159) | | (0.051) | (0.053) |
| Day of Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| Control Variables | YES | YES | YES | YES | YES | YES |
| N | 2961 | 2961 | 2961 | 3352 | 3352 | 3352 |
| Adjusted $R^2$ | 1.73% | 1.62% | 1.77% | 0.77% | 0.75% | 0.74% |

## Table 12: Timing of Earthquakes

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 from 0 to 30 days, from 31 to 60 days, and from 61 to 90 days prior to the response date. 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. Dummies associated with strong magnitudes are included in the regression models are also included in the models, but are not reported. Control variables and fixed effects from Table 6 are also included, but not reported. The returns variables for the control variables are based upon the DJIA index. Robust standard errors clustered on the date level 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. | Indiv. | Indiv. | Inst. | Inst. | Inst. |
| Dependent Variable: | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ | $\pi(i,t)$ |
| Weak Magnitude (t-30,t) | | | 0.057*** | | | -0.010 |
| | | | (0.021) | | | (0.020) |
| Weak Magnitude (t-60,t-31) | 0.007 | | -0.043 | -0.008 | | 0.025 |
| | (0.011) | | (0.031) | (0.010) | | (0.025) |
| Weak Magnitude (t-90,t-61) | | 0.008 | 0.015 | | -0.012 | -0.027 |
| | | (0.011) | (0.027) | | (0.010) | (0.019) |
| Strong Magnitude Controls | YES | YES | YES | YES | YES | YES |
| Day-of-Week FEs | YES | YES | YES | YES | YES | YES |
| Month FEs | YES | YES | YES | YES | YES | YES |
| Control Variables | YES | YES | YES | YES | YES | YES |
| N | 2961 | 2961 | 2961 | 3352 | 3352 | 3352 |
| Adjusted $R^2$ | 1.54% | 1.60% | 1.82% | 0.73% | 0.79% | 0.72% |