Negativity Bias in Attention Allocation: Retail Investors' Reaction to Stock Returns*

Isaac Hacamo† and Tomas Reyes‡

This Draft: November 2012

Abstract

We argue negative stock market performance attracts more attention from retail investors than comparable positive performance. Specifically, we test and confirm the hypothesis that retail investors pay more attention to negative than positive extreme returns. We present a measure of attention at the aggregate and company specific level using internet search volume from Google. These measures correlate with, but are different from, existing proxies of attention. Our empirical results strongly support that investors display a negativity bias in attention allocation with respect to extreme stock returns. Across all specifications, lagged negative extreme returns are stronger predictors, than positive extreme returns, of high attention at the stock and market level. We rule out that negative returns are stronger simply because they are more unusual or because negative and positive returns are not symmetrical events to stockholders.

*Special thanks to Ulrike Malmendier for continuous guidance and feedback. Thanks also to Stefano DellaVigna, Simon Gervais, Thomas Mertens, Atif Mian, Terrance Odean, Richard Stanton, Adam Szeidl, Paul Tetlock, Hal Varian, and Wei Xiong for helpful comments. We have also benefitted from comments by participants at Yale Whitebox Behavioral Conference, the Management Frontiers and Economic Issues Forum of Beihang SEM, the Haas Finance student seminar, and the Berkeley-Stanford student seminar.

†Finance Ph.D. Candidate, Haas School of Business, University of California, Berkeley. Email: isaac_hacamo@haas.berkeley.edu

‡Finance Ph.D. Candidate, Haas School of Business, University of California, Berkeley. Email: tomas_reyes@haas.berkeley.edu

1 Introduction

Hirshleifer et al. (2004) study how retail investors trade in response to earnings surprises. They show individuals are net buyers after negative and positive earning surprises, and the level of net purchases is far greater after extreme negative earnings surprises than after extreme favorable ones. They claim that these facts provide support for the hypothesis that individual investors cause post-earnings-announcement drift, but do not explain what may be causing this asymmetry. Similarly, Barber and Odean (2008) show that individual investors are net buyers of attention grabbing stocks, and average buy-sell imbalances are greater after negative return days than after positive return days. Possible explanations given for this asymmetry by the authors are Shefrin and Statman (1985)'s disposition effect (the preference for selling winners and holding losers) and the execution of limit orders, but not further tests are performed due to data unavailability. In this paper, we ask whether these irregularities may be explained by a negativity bias in attention allocation.

If this negativity bias exists, negative stock market performance will attract more attention than comparable positive performance. Since high attention is linked to buying, this asymmetry would explain why the level of net purchases or buy-sell imbalances is greater after extreme unfavorable earnings or negative returns than after extreme favorable earnings or positive returns.

Research in psychology supports that bad is stronger than good. Baumeister et al. (2001) argue that in most situations, negative events will produce larger, more consistent, or more intense consequences than comparable positive events. Anecdotally, human beings usually ask to hear the bad news first, and bad news sells more newspapers. Pratto and John (1991) test whether attentional resources are automatically directed away from the current task when inessential good or bad traits are present. They find that bad extraneous stimulus attract more attention in an automatic and non-intentional fashion than good stimulus. Similarly, Sheldon et al. (1996) study how long the impact of positive or negative everyday events last on a person's mood. They conclude negative information receives more processing time and contributes more to the creation of impressions than positive information.

In what follows, we relate this negative-positive attention asymmetry to stock market behavior. We argue that negative stock market performance draws more attention than comparable positive performance. Specifically, we measure performance in the stock market using stock returns and test the hypothesis that retail investors pay more attention to extreme negative returns than to extreme positive returns.

Attention and its allocation across tasks are difficult to measure directly. We measure attention in the stock market using internet search volume from Google. Search volume is a powerful tool to capture attention for two reasons. First, internet users commonly use a search engine to collect information, and Google is the preferred one; so search volume is representative of the general population's interest on a topic[1]. Second, while we search for a term online, we have to pay attention to it; therefore, search volume is a better and more direct proxy for attention than alternative measures used in the literature.

Search Volume Index (SVI) is available from Google Insights for Search [2]. SVI for a search term is the percentage of searches for that term during a week within a geographical region, scaled by its time-series maximum. Data is available from January 2004 to December 2010 for most common terms used by people while searching on Google. The top panel in Figure 1 shows SVI for the query "diet, twitter". The plots conform with intuition. The SVI for "diet" is seasonal and has no time trend. It drops by the end of each year, during the year holiday season; and, spikes at the beginning of the year, probably driven by new year's resolutions to shed some extra pounds. The SVI for "twitter" has a time trend, since the proportion of Google users that search for that term has increased over time. SVI is zero before twitter's launch in July 2006. Since then, twitter has gained popularity worldwide and awareness of the service has exploded as captured by highest SVI in 2009-2010.

Search volume data has been used in different fields, mainly exploiting its prediction power. Varian and Choi (2009) use data from Google to provide evidence that search volume can predict home sales, automotive sales, and tourism. Ginsberg et al. (2008) find that search data for terms related to influenza predict flu outbreaks weeks before CDC reports. On finance, Da et al. (2010) proxy for investor sentiment using search terms such as "job lost" or "recession", and show decreases in search volume lead to contemporaneous price increases and future return reversals. Da et al. (2009) use searches on tickers to proxy for attention in individual stocks and provide support for Barber and Odean (2008) price pressure hypothesis.

[1] In 2010, internet usage in the U.S. was 77% and Google's market share was 72%.

[2] http://www.google.com/insights/search

Throughout this paper we use three aggregate measures of attention in the stock market based on search volume from Google. All Retail Investors Attention (*AllInv*) is defined as the sum of SVI for terms as "stock market" and "best stock prices", which investors search when seeking general information about the stock market and price movements. New Investors Attention (*NewInv*) relates to search terms as "online brokerage account" and "best brokerage account", and proxies for people researching for discount brokers to open a new account. Old Investors Attention (*OldInv*) concerns to retail investors who already own a brokerage account and use Google to access its web and login. It is defined as the sum of SVI for terms such as "ameritrade" and "etrade". We also implement a measure of attention at the stock level. Following Da et al. (2009) we use SVI for ticker symbols (e.g., "YHOO" for Yahoo and "WMT" for WalMart) for the 100 largest companies in the S&P 500.

After constructing these measures, we proceed as follows. First, we study what drives SVI and how it relates to other indirect proxies for attention. We find aggregate U.S. level SVI measures have positive contemporaneous and lagged correlations with trading volume and volatility. When we explore the relationship between lagged returns and attention, we find that all and new investors display the greatest amount of attention for extreme positive and negative returns. And more importantly, lagged negative extreme returns are stronger predictors of attention in the stock market than positive extreme returns. Finally and consistent with our intuition, we find retail investors pay more attention to high market capitalization portfolios.

Second, we get SVI data from Google for each U.S. state, and construct within state versions of our attention variables. Likewise, we sort companies by state using location codes from Compustat. For each state and week, we construct a portfolio of high market capitalization firms located within the state, and another with firms located outside the state and analogous characteristics. Similar to the results of our time-series regressions, we find individual investors show a negativity bias and pay more attention to negative than positive extreme returns. Finally, we observe mixed evidence of investors paying more attention to firms located at their home state versus outside companies.

Third, we center on attention to specific stocks and its relationship with individual stock returns. We use SVI for ticker symbols and stock level market data to create a panel with the 100 largest companies in the S&P 500. We find similar patterns supporting a negativity bias hold at the individual stock level, even after controlling for the market-value weighted index and other proxies of attention.

Our empirical results strongly support investors display a negativity bias in attention allocation with respect to extreme stock returns. Across all specifications, lagged negative extreme returns are stronger predictors of retail investors attention than positive extreme returns. To the best of our knowledge, this is the first time such a pattern is documented for individual investors in the United States.

The paper is organized as follows. Section 2 describes data sources, how we determine the relevant search terms, and how we form attention variables based on search volume from Google. Section 3 compares our measures of attention to alternative proxies. Section 4 relates U.S., state, and company level measures of attention to extreme returns. Section 5 presents robustness checks. Section 6 concludes.

2 Data and Sample Construction

When users search in Google[3] they need to provide a search term. Examples of the most commonly used search terms for any date are available at Google Hot Searches [4]. Each day Google tracks the amount of searches for every term and their geographical origin. This times-series search volume data is formally called Search Volume Index (SVI).

Google makes available SVI for a search term or a query from its product Google Insights for Search. A query is a group of at most five search terms that can be entered into Google Insights for Search. SVI for a term is available from 2004 for different countries, states, cities, and counties; and, reflects the number of searches for that particular term (during a day or week and within a region), relative to the total number of searches done at Google (within that region and during that period). Formally, SVI is defined as:

$$SVI^j{}_{r,t} = \frac{SVT^j{}_{r,t}}{TSV_t \times MSV_{r,t}} = \frac{SVT^j{}_{r,t}}{TSV_t \times \max_{\{q,i\}}\{SVT^i{}_{r,q}/TSV_q\}} \tag{1}$$

[3]http://www.google.com

[4]http://www.google.com/trends/hottrends

$SVT^j{}_{r,t}$ is the total search volume for term j at period t within region r, TSV_t is the total search volume in Google at time t, and $MSV_{r,t}$ is the maximum of such ratios among all terms in the query and within the sample period. Search volume is divided by TSV_t to eliminate any trends that could be present due to a change in the amount of Google users, and divided by $MSV_{r,t}$ to scale the time series and not reveal the raw number of searches. Therefore, SVI for a search term is proportional to the percentage of searches for that term during a period of time and within a geographical region. Back to figure 1, note that the largest value for SVI in the plot is one, due to the scaling (division by $MSV_{r,t}$ in (1)). Since we included both terms (”diet” and ”twitter”) in the query, we can make comparisons across them; and say, for example, that ”twitter” has been searched more often than ”diet” since late 2008.

Search terms can be complex: ”diet plans and weight loss programs” and ”how to use twitter like a pro”, or simple: ”diet” and ”twitter”. Data is available for most simple terms used by people while searching on Google. Complex terms are less frequently used and usually are not reported in Google Insights for Search. Another limitation is Google computes SVI from a random subset of historical data, which varies from day to day. This sampling error adds additional noise to the data and should be accounted for. Since these samples are independent from day to day, we ask data for seven days in a row and use the average to address this problem. Another shortcoming is that Google allows users to ask for a maximum of (around) 200 queries a day. After hitting that threshold, the service is blocked and users need to wait 24 hours to download more data using the same IP address, which slows down the data acquisition process dramatically.

2.1 All Retail Investor’s Attention ($AllInv$)

To proxy for all retail investors’ attention in the stock market we use search volume for the list of terms presented in table the first column of 1. We arrive to this list by starting with a small set of terms:

$$\{\text{"}stock\ market\text{"},\ \text{"}stock\ prices\text{"},\ \text{"}best\ stocks\text{"}\} \tag{2}$$

We search on google each of these terms and obtain related searches recommended by

Google [5] (i.e., similar search terms that people use while searching for the original term). Since Google related searches is a reliable way to learn how users search, it helps us to make the final list presented in the table more objective and independent of the initial terms in (2).

This process of getting related searches originates a list of 60 terms, from which we drop out terms that are either company names (people may be searching for them for many other unrelated reasons), or very general (e.g., "fox news" and "cnn"), or unrelated (e.g., "online auctions" and "housing market"). With the remaining terms, we iterate one last time (get related terms and drop irrelevant ones) to get the final list shown in column 1 of table 1.

We then enter manually each of these terms (in queries formed by groups of five) into Google Insights for Search to find the one with the largest SVI value. This search term will be the initial term in each query we use to download data. We do this, to make sure Google scales each time-series using the same value for $MSV_{r,t}$ in equation (1), so we can aggregate them easily. For each query we collect weekly data for the U.S. and its states using a webcrawling program that inputs each term and geographical region into Google Insights for Search and downloads the SVI data into a CSV file.

The bottom panel in Figure 1 shows SVI for three of the terms presented in the first column of table 1: "stock quotes", "stock prices", and "best stocks". The plot shows search volumes are positively correlated. All three terms have a spike in late September 2008 after Lehman Brothers declared bankruptcy, which suggests negative events draw a good deal of attention. SVI for "stock quotes" and "stock prices" has decreased over time probably due to the growing popularity of websites (in which people can find similar information) such as Yahoo or Google finance, and show some year seasonality.

Unfortunately, Google does not return a valid SVI for some terms in some geographical regions. If a term is rarely searched in a state, Google Insights for Search may return only zeros or simply drop the term from the output. This is why after attempting to download all the data, we only have a total of 91,471 term-week-region data points, which is 39% of the maximum attainable. Finally and since we want a unique measure of retail investor's

[5] This is a standard feature provided by Google, available on the left-hand menu of the results page.

attention for each region, we aggregate the terms,

$$AllInv_{r,t} = \sum_{j \in J} SVI^{j}_{r,t} \ \forall r, t \tag{3}$$

where J represents all the search terms in column 1 of table 1.

The top panel in figure 3 shows data availability for *AllInv* across the forty-eight continuous U.S. states. Google returns no data for the states in white (i.e., DE, FL, KS, LA, MT, ND, SD, VT, WV, and WY). For the rest, the percentage of weeks with available data is provided in parenthesis. States with warmer colors have more data.

2.2 New Investor's Attention ($NewInv$)

Our second measure of attention in the stock market proxies for new investors, who are new to the stock market and seek information to open a brokerage account. This measure is based on search terms presented in the second column of table 1. To obtain this final list, we repeat the procedure described above starting with the following set of terms:

$$\{\text{"}\textit{online broker}\text{"}, \text{"}\textit{online trading}\text{"}, \text{"}\textit{stock broker}\text{"}\}$$

The top panel in figure 2 shows SVI for three of the final terms: "best online trading", "online broker", and "online stock trading". As in the bottom panel of figure 1, search volumes are for the U.S. level and are positively correlated. However, SVI related to new investors are more volatile, have no seasonality, and are less frequently searched. After downloading data, we have a total of 31,011 term-week-region points, which is only 17% of the maximum possible.

Next we aggregate SVI across terms to compute a combined measure of new investor's attention as in (3) but with J as the set of terms presented in column 2 of table 1. The bottom panel in figure 3 corroborates data for *NewInv* (across states) is scarcer than data for *AllInv*, and it is available only for 12 states (i.e., CA, FL, GA, IL, MA, MI, NJ, NY, NC, OH, PA, TX, VA, and WA).

2.2.1 New Accounts Opened

TD Ameritrade, a publicly traded online brokerage company, has to report, since 2007, the actual number of new accounts opened each quarter on forms 10-K or 10-Q. We use the brokerage data to validate the data obtained from google searches. The bottom panel in figure 2 shows how these quarterly numbers from TD Ameritrade relate to $NewInv$. $NewInv$ values are computed quarterly for the U.S. and re-scaled to have a maximum value of one. The plot shows SVI for new investor's attention behaves similarly to the real number of accounts open in TD Ameritrade, with positive correlation between the two data series. The plot supports the validity of our measure.

2.3 Old Investor's Attention ($OldInv$)

Our third measure is associated to old investors and proxies for individual investors that already own a brokerage account and use Google to obtain access to it. Search terms related with this measure are shown in column 3 of table 1. Each term in the list relates to one of the most popular online broker companies during 2010. Each of these companies may be searched in Google in various different ways (e.g., e*trade may be searched as: "etrade", "e*trade", "e-trade" or "e trade"). We manually enter each of this alternatives into Google Insights for Search and keep the most popular one (i.e., the one with the highest search volume). We aggregate SVI across terms as in (3) and present available data across states in the top panel of figure 4.

2.4 Stock Level Attention

Throughout this paper we also need measures of attention for company stocks. Following Da et al. (2009) we use SVI for ticker symbols. Using search volume for the company name to identify a stock is potentially problematic since people may be searching the company name for reasons unrelated to investing. Conversely, searching for a stock using its ticker is more precise and relates to people acquiring financial information about the company.

Since investors are more aware of big companies, we start analyzing tickers for companies in the S&P 500 index. We manually go through each ticker in search of ambiguities and drop 88 ticker symbols with generic meaning such as "A", "ALL", "BIG", and "CA". For the remaining 412 companies we compute market capitalization (as number of shares times price per share) and find its headquarter location. To keep the data collection manageable, we restrict our sample to the 100 largest ones measure by market capitalization. For each of them, we download regional (for the state where the headquarter is located) and U.S. level SVI data. Stock market data, such as returns and trading volume, are from the Center for Research in Security Prices (CRSP), and location information is obtained from Capital IQ Compustat.

3 SVI, Trading Volume, and Volatility

In this section, we study how SVI relates to indirect proxies for attention. One of the key variables in this paper is Abnormal Retail Investors' Attention ($AAllInv$), which is defined as:

$$\begin{aligned} AAllInv_{r,t} &= \log\left(\frac{AllInv_{r,t}}{\frac{1}{8}\sum_{q=1}^{8} AllInv_{r,t-q}}\right) \\ &= \log(AllInv_{r,t}) - \log[mean(AllInv_{r,t-1}, \ldots, AllInv_{r,t-8})] \end{aligned} \tag{4}$$

Intuitively, the mean over the past 8 weeks determines a reference level of attention. Therefore $AAllInv$ measures a change in interest with respect to the normal level in the recent past, and a large (low) value represents an relative increase (decrease) in attention. Analogously, we define Abnormal New Investors' Attention ($ANewInv$) and Abnormal Old Investors' Attention ($AOldInv$).

Next we study how these aggregate proxies for attention relate to other observable proxies that are likely to be linked with attention-grabbing events. For example, Barber and Odean (2008); Gervais et al. (2001); Hou et al. (2008) argue abnormally heavy volume is associated with information releases or large price moves that attract attention. Therefore, we use abnormal trading volume ($AVlm$) for a stock or portfolio p as an alternative proxy,

$$AVlm_{p,t} = \log(Vlm_{p,t}) - \log[mean(Vlm_{p,t-1}, \ldots, Vlm_{p,t-8})] \tag{5}$$

Similarly, De Long et al. (1990) claim noise trading following periods of extreme sentiment can create future volatility. Baker and Wurgler (2007) use the CBOE market volatility index (VIX), a popular measure of the implied volatility of S&P 500 index options, to proxy for aggregate market sentiment. Consequently, we use abnormal VIX ($AVIX$) as another alternative proxy for attention,

$$AVIX_t = \log(VIX_t) - \log[mean(VIX_{t-1}, \ldots, VIX_{t-8})]$$

Table 2 shows correlations among $AAllInv$, $ANewInv$, $AOldInv$, $AVlm$, and $AVIX$. SVI based measures of attention are computed using search volume at the U.S. level. $AVlm$ is calculated from (5) with volume (Vlm) from the New York Stock Exchange (NYSE) for a value-weighed portfolio formed by all stocks in CRSP.

Contemporaneous correlations among all variables are positive and range from 27% to 77%. They are all positive and considerably high. Our three measures of abnormal attention are highly correlated among themselves with correlations ranging from 63% to 77%, not surprising since all three types of investors should share common interests in aggregate market events. Also, measures of attention based on volume and volatility are imperfect, there are many changes in volume and volatility that are not attributable to new fundamental information about markets. Nevertheless, the correlation of $AAllInv$ with $AVIX$ is 58.9%, the notoriously high value increases the validity that $AAllInv$ as a measure of attention. Even though the correlation of $AAllInv$ with $AVlm$ is smaller, it is still 45%.

The correlation of $ANewInv$ with $AVlm$ and $AVIX$ is 25.5% and 44.5% respectively. We do not expect new investors to be as interested in market events as the whole pool of investors. Not surprisingly, our measure of old investors exhibits a smaller correlation with $AVIX$, one obvious explanation relies on the fact that, as opposed to the other two measures, investors are not seeking information when they use the search terms related with $AOldInv$, but instead using google to help them login in the brokerages websites that they can well have bookmarked in their web browsers.

All our three measures of attention correlate less with $AVlm$ than $AVIX$, suggesting that variation in price volatility is more correlated with market information that capture investors attention than volume variations, as many variations in volume are often related with institutional investors' liquidity motives.

4 SVI and Stock Returns

The core of the paper is presented in this section. We relate our abnormal attention measures with stock returns. Barber and Odean (2008); Yuan (2009) argue investors are likely to notice when stocks have extreme returns. Extreme positive or negative returns are often related to news about the firm. These news or the return itself should capture the attention of invertors. Specifically, we test the extent to which our attention measures respond to different stock returns.

We do this with three complementary but different specifications. First we explore the relationship between lagged returns and stock market attention at the U.S. and state level. Then we analyze if similar patterns are present at the company level.

4.1 U.S. Level

First we sort returns into quintiles, QU, (i.e., twenty percent partitions) and construct five level variables:

$$I_i Ret_{p,t} = \mathbf{1}\{Ret_{p,t} \in QU_i\}, \ \forall i \in \{1, \ldots, 5\} \tag{6}$$

and sensibility ones:

$$P_i Ret_{p,t} = Ret_{p,t} \times \mathbf{1}\{Ret_{p,t} \in QU_i\}, \ \forall i \in \{1, \ldots, 5\} \tag{7}$$

So $P_1 Ret_{p,t}$ is equal to $Ret_{p,t}$ if $Ret_{p,t}$ is one of the 20% smallest returns in portfolio p during the sample period, and zero otherwise. Therefore,

$$\sum_{i=1}^{5} P_i Ret_{p,t} = Ret_{p,t}$$

and by construction, $P_1 Ret_{p,t}$ will contain extreme negative returns and $P_5 Ret_{p,t}$ extreme positive returns.

To test our main hypothesis we run the following regression specification:

$$AAttention_t = \alpha + \sum_{i=1}^{5} \beta_i f(i,p,t) + \delta Q_{FE} + \varepsilon_t \tag{8}$$

where $f(i,p,t)$ is either $I_i Ret_{p,t}$ or $P_i Ret_{p,t}$, p distinguishes the portfolio of stocks, and Q_{FE} is a quarter fixed effect. In all regressions we use quarter fixed effects to control for major events that have happen during the period of analysis of our sample. We name the regression specification 'level regression' when $f(i,p,t) = I_i Ret_{p,t}$ and 'sensibility regression' when $f(i,p,t) = R_i Ret_{p,t}$. The level regression allows to test the change in attention when returns are extreme positive or negative, whereas the sensibility regression allows to test the change in attention for a change in returns when returns are positive or negative. We do not present a regression with both, $I_i Ret_{p,t}$ and $P_i Ret_{p,t}$, as explanatory variables because these variables are collinear which makes results difficult to interpret.

Tables 3 and 4 present results. Both tables present results for our three measures of attention, $AAttention_t$ is equal to $AAllInv_t$, $ANewInv$ and $AOldInv$ in columns 1, 2 and 3, respectively. $Ret_{p,t}$ is, in panels (a), the return on the value-weighed portfolio formed by all stocks in CRSP, and, in panels (b), is the return on the value-weighed portfolio of high market capitalization stocks (highest quartile).

Table 3 presents results for the sensibility regression. The first and second columns show all and new investors display the greatest amount of attention for a change in returns when returns are at the extreme level. Positive (negative) extreme returns have positive (negative) coefficients, so changes in returns towards the extreme grab investors' attention more intensely. In general, coefficients are increasing in quintiles of returns, and are more significant closer to the extremes (i.e., quintiles and rows 1 and 5). In panel (a) and when attention is measured with $AAllInv$, the coefficient on the lowest percentile is -0.442 and on the highest percentile is 0.108, both statistically significant. A change in lagged negative aggregate returns has a stronger impact, than a change in positive returns, in attracting attention to the stock market; this pattern also remains when we use $ANewInv$ as a measure of attention, the coefficients are -0.406 and 0.147 for the lowest and highest quintile respectively. New investors react less to changes in returns when this are at the negative extreme, and more positively when at the positive extreme, suggesting that new investors have a smaller negative bias than existing investors. Results for old investors in column 3 show a similar pattern, but negative extreme returns are not significant. When we compare panel (a) to (b) we see investors pay more attention to high market capitalization portfolios. This is not surprising, since larger firms are more familiar to individual investors than smaller firms, and thus more likely to capture their attention.

A Wald test for the hypotheses that the coefficients associated to the most negative and positive extreme returns are the same (in absolute value) is rejected with p-values 0.1% and 5.9% for columns 1 and 2 in panel (a), and 0.08% and 4.5% in panel (b). We cannot reject the hypotheses that both coefficients are the same for old investors in column 3 for neither panel. This is consistent with the ostrich effect documented by Karlsson et al. (2009), in which investors monitor they brokerage accounts more frequently after positive news than after negative news, which should counterbalance the negativity bias. A caveat to keep in mind is our measure for old investors may be a noisy proxy compared to Karlsson et al. (2009)'s measure.

Table 4 presents results for the level regression. All coefficients are positive and significant at %1 level. Reflecting that market events grab investors attention. The coefficients on the extreme quintiles are always higher than the coefficients in the intermediate quintiles, except in the regression where we use $AOldInv$ as a measure of investors attention. The results in both panels support the hypothesis that investors pay more attention to extreme returns, particularly negative ones. In panel (a), when we use $AAllInv$ as a measure of investors attention the regression coefficients in the lowest is 1.215 and in the highest quintile is 0.859, there is a 30% difference in all investors searches for an equal standard deviation variation in the two extremes of the return distribution. New investors show a much smaller difference between the coefficient in both extreme quintiles, again suggesting that new investors have a smaller negativity bias. The effect is flips around in the old investors regression which is is consistent with the ostrich effect documented by Karlsson et al. (2009).

4.2 State Level

We can run a panel regression similar to the time-series specification in (8), in which the cross-section is given by the U.S. states. Using state level SVI data, we construct $AAllInv$, $ANewInv$, and $AOldInv$ as defined in (4).

Next we sort stocks by state using company location codes from Compustat, which are used to identify a company's headquarter location. The bottom panel in figure 3 shows the geographical distribution of companies by state. States in white have less than 20 companies. For the rest, numbers in parenthesis represent the average number of companies located in that state relative to the maximum in any state, which is 817 in CA. For example, NY has $71\% \times 817 = 580$ and TX has $51\% \times 817 = 417$. States with warmer colors contain more companies.

For each state and week we construct a portfolio p^{in} of high market capitalization (highest quartile) companies within the state. In general, the number n^{in} of companies forming portfolio p will be different among states. Since the purpose of creating portfolios is to reduce noise, we drop states with very few firms (n^{in} smaller than 20).

Similarly, for each state and week we form a portfolio p^{out} of similar characteristics using only companies located outside the state. We select outside companies with market capitalizations in the same range than the companies forming the within state portfolio p^{in}. Overall, the number of firms in this outside portfolio is much larger than n^{in}. Every time this happens, we randomly drop just enough firms to let both portfolios with equal number of companies. Then we sort the returns on these portfolios into quintiles as in (7).

The first column in table 5 reports estimates of β^{in} and β^{out} in the regression,

$$\begin{aligned} AAllInv_{s,t} = \alpha + \sum_{i=1}^{5} \beta_i^{in} P_i Ret_{p_s^{in},t-1} + \sum_{i=1}^{5} \beta_i^{out} P_i Ret_{p_s^{out},t-1} \\ + \gamma Controls + \delta_1 M_{FE} + \delta_2 S_{FE} + \varepsilon_{s,t} \end{aligned} \tag{9}$$

Similarly, The first column in table 6 reports estimates from,

$$AAllInv_{s,t} = \alpha + \sum_{i=1}^{5} \beta_i^{in} I_i Ret_{p_s^{in},t-1} + \gamma Controls$$
$$+ \delta_1 M_{FE} + \delta_2 S_{FE} + \varepsilon_{s,t} \quad (10)$$

Sice $Ret_{p_s^{in},t-1}$ and $Ret_{p_s^{out},t-1}$ are highly correlated, we do not include $I_i Ret_{p_s^{out},t-1}$ in the later specification because of collinearity problems with $I_i Ret_{p_s^{in},t-1}$. Columns 2 and 3 of both tables show equivalent specifications but using $ANewInv$ and $OldInv$ as dependent variables respectively.

M_{FE} and S_{FE} are month and state fixed effects respectively. We use month instead of week dummies because the effect we are trying to capture is not purely cross sectional as shown in the previous section. Monthly state controls from St. Louis Fed are: (i) Coincident Economic Activity Index, to summarize current economic conditions; (ii) Leading Index, to predict the six-month growth rate of the state's coincident index; and (iii) Unemployment Rate. To account for correlations among different states in the same week and different weeks in the same state we double cluster using Petersen (2009) implementation of Cameron et al. (2006)'s procedure.

Similar to the time-series regression, columns 1 and 2 of tables 5 and 6 show all and new investors display the greatest amount of attention for extreme returns. In general, positive (negative) extreme returns have positive (negative) coefficients, and coefficients are increasing in quintiles of returns and are more significant closer to the extremes (i.e., rows 1 and 5 for the within state portfolio, and 6 and 10 for the outside portfolio). More importantly, lagged negative extreme returns have larger coefficients (in absolute value) and are more significant than positive extreme returns. This is consistent with our previous time-series results that support the existence of a negativity bias.[6]

Again, old investors is the group that exhibits the least negativity bias in the sensibility regression shown in table 5, measure by the difference in absolute values between the coefficients in rows 1 and 5. In row 5 we observe that positive within state extreme returns grab similar attention from all three types of investors, all coefficients are of similar magnitude.

[6]We also compare these results to those of the same specification but implemented with portfolios formed with low market capitalization firms, and find evidence of investors paying more attention to high market capitalization portfolios. Due to space constraints, tables are not included.

However, in the first row we can see that old (new) investors are the ones that pay less (more) attention after negative news. A Wald test for the hypotheses that the coefficients associated to the most negative and positive extreme returns for the within portfolio are the same in absolute value is rejected with 10% and 5% significance for columns 1 and 2 respectively, and cannot be rejected for column 3. Table 6 shows strong evidence of negativity bias for all investors, some evidence for old investors, and a similar but insignificant pattern for new investors.

Ivkovic and Weisbenner (2005); Zhu (2009) argue retail investors exploit local knowledge, and exhibit preferences for local investments. Columns 2 and 3 of table 5 suggest the presence of this home bias at the state level for new and old investors, since the absolute size and significance of the coefficients in extreme positive and negative returns of the within state portfolio (rows 1 and 5) are larger than the ones for the outside portfolio (rows 6 and 10). However, for the first column this only holds for extreme positive returns.

4.3 Company Level

In this section, we test if the negativity bias present at the aggregate U.S. and state level, documented in the previous sections, also exists for specific stocks. In the same spirit of our previous abnormal attention measures, we define $ATicker$ as:

$$ATicker_{c,t} = \log(Ticker_{c,t}) - \log[mean(Ticker_{c,t-1}, \ldots, Ticker_{c,t-8})]$$

$Ticker_{c,t}$ is search volume during week t for the ticker symbol associated to company c, and measures a change in attention with respect to its normal level. We use data for the 100 largest companies in the S&P 500 index, measured by market capitalization.

For each of these companies we download daily returns, trading volume, price, and number of shares from CRSP, and compute weekly values of these variables. Weekly returns are holding period returns from closing Friday to closing Friday, weekly trading volume is defined as the sum of daily trading volume during the week, and weekly market capitalization is price at the end of the week times number of shares at the end of the week.

Similar to what we did before, each week we sort weekly returns into quintiles (QU) as in (7), but in this case $Ret_{p,t}$ is the weekly return on company c at week t. We also compute abnormal trading volume from (5) where $Vlm_{p,t}$ is trading volume for company c at week t, and log market capitalization,

$$LMcap_{c,t} = \log(Mcap_{c,t})$$

Table 7 show results for different specifications of the following panel regression:

$$ATicker_{c,t} = \alpha + \sum_{i=1}^{5} \beta_i P_i Ret_{c,t-1} + \gamma Controls + \delta_1 M_{FE} + \delta_2 C_{FE} + \varepsilon_{c,t} \tag{11}$$

Similarly 8 show results for the regression:

$$ATicker_{c,t} = \alpha + \sum_{i=1}^{5} \beta_i I_i Ret_{c,t-1} + \gamma Controls + \delta_1 M_{FE} + \delta_2 C_{FE} + \varepsilon_{c,t} \tag{12}$$

M_{FE} and C_{FE} are month and company effects respectively. $Controls$ used in the second column are $VWRet$, $AVlm$ and $LMcap$. $VWRet$ are effectively five different variables; each of them is one quintile sort as in (7) of the market return, which is defined as the value weighed return of all stocks in CRSP. We include $VWRet$ to make sure our results are not driven simply by the correlation between company returns and the return in the market portfolio.

In parallel to what we find in section 3, tables 7 and 8 shows an increase in retail investors' attention, measured by searches in ticker symbols, is associated to extreme company returns during the previous week. Overall, coefficients are increasing in quintiles of returns and are more significant closer to the extremes. Across all specifications, lagged negative returns are stronger predictors of attention than positive returns.

Additionally, column 2 in both tables shows attention at the stock level is positive and significantly correlated with abnormal trading volume, which is also consistent with our previous results in section 3. In the same column, we do not find a strong relationship between attention and market capitalization; we believe this is not surprising since we are already controlling for market capitalization the way we choose the firms.

These results reinforce the main contribution of this paper. Investors display a negativity bias in attention allocation with respect to extreme stock returns. This relationship is robust to different specifications and holds at the aggregate and at the company specific level for large companies in the U.S.

5 Robustness Checks

5.1 Distribution of Returns

Positive and negative returns are not symmetrical events to stockholders. For example, it is possible (but unlikely) that investors get a large positive return of more than 100%. However, even in the worse of the crises, negative returns are always constrained to be lower (in absolute value) than 100%. Therefore, a valid concern is that this asymmetry, which causes skewness, may be driving our negativity bias result.

Another concern may be due to the sample period. As a consequence of data availability the sample period used in this paper overlaps with the 2008-2010 financial crisis. So it is possible that there are more negative than positive return outliers in our sample. These negative outliers could also be influencing the results.

Figure 5 shows histograms for the sets of returns used throughout this paper. The top left panel presents aggregate U.S. returns used in section 4.1, the middle left panel shows within state returns with high market capitalization used in section 4.2, and the bottom left panel displays returns for the 100 largest companies in the S&P 500 used in section 4.3. In general, the plots show that there are more negative than positive returns in our sample. As for outliers, within state and company returns seem fairly symmetric. However, U.S. returns present larger negative outliers (in absolute value) than positive ones.

Quantitatively, pooled aggregate returns in our sample are negatively skewed, -.72 and -.13 for U.S. and state level returns respectively. Conversely, pooled company level returns are positively skewed at .69. All samples have positive kurtosis: 10.86, 14.36, 19.96 for U.S., state, and company level respectively.

Quintile partitions of returns used in our regressions partially account for some of the previously mentioned problems. However, an alternative and complementary solution is to redistribute negative returns to replicate the distribution of positive returns, or vice-versa. More specifically, for each week and portfolio (or stock) we apply the following procedure:

- If the return is positive, keep it without modification.
- If the return is negative,
  (a) Get the time-series returns for the past 5 years associated with that portfolio (or stock).
  (b) Sort the previous time-series into two groups, one for positive and another for negate returns.
  (c) Find the percentile rank of the current (negative) return within the group of negative returns.
  (d) Find a (positive) return, within the group of positive returns, with percentile rank closest to the one in (c).
  (e) Replace the current (negative) return with the negative of the positive return from (d).

For example, for a given rolling window in which the current negative return is the most negative one, and the maximum positive return is x, we replace the current negative return by $-x$. So basically, this procedure will modify each negative return imposing the 5-year rolling window return distribution associated with it to be more symmetrical.[7]

The right-hand side of each panel in figure 5 presents histograms after the transformation just described. The densities for the positive range of returns (in the right-hand side of each plot) have not change since we assign new values only to negative returns and do not modify positive ones. Overall, histograms seem more symmetrical and balanced in terms of outliers. Quantitatively, all skewness values are now positive and larger: .74, .84, and .73 for U.S., state, and company level returns respectively; and kurtosis are smaller: 6.92, 13.43, and 18.66 in the previous order.

[7]We also try imposing the cross-sectional distribution (in the case of state and company level returns) to be more symmetrical, and results are similar.

To test the robustness of our results with respect to this transformation, we rerun all previous regressions after redistributing returns. For ease of presentation we include new results only at the company level in tables 9 and 10. In general, the tables show that the economic significance of the coefficients associated to extreme negative and positive returns is smaller (in absolute value) than before. Consequently, the average difference between the coefficients associated to the fifth and first quintile is also smaller but nonetheless it is still positive. Therefore, after this transformation the negativity bias in attention allocation is weaker, in the sense that the difference in attention is smaller than before, but the evidence for its existence remains strong.

6 Conclusion

Research in psychology supports negative events will produce larger, more consistent, or more intense consequences than comparable positive events. This negativity bias suggest negative information contributes more to the creation of impressions and attracts more attention in an automatic and non-intentional fashion than positive information.

In this paper, we relate this negative-positive attention asymmetry to stock market behavior. We argue negative stock market performance attracts more attention from retail investors than comparable positive performance. Specifically, we test the hypothesis that individual investors pay more attention to negative than positive extreme returns.

We measure attention using internet search volume from Google (SVI), a more direct proxy for attention than traditional measures as trading volume, volatility, etc. From SVI we construct aggregate measures of attention for different types of retail investors based on data from the U.S. as a whole and its states. We find aggregate U.S. level measures have positive contemporaneous and lagged correlations with indirect proxies for attention, and all and new investors exhibit the greatest amount of attention for extreme returns. Investors pay more attention to high, than low, market capitalization portfolios. And more importantly, lagged extreme negative returns are stronger predictors, than extreme positive returns, of high attention in the stock market. Using state level data, we find similar results and evidence of investors paying more attention to firms located at their home state than to outside companies.

Then we test if the negativity bias present at the U.S. and state level also exists at the company level. We measure attention to specific companies using SVI for their ticker symbols for a sample of the 100 largest companies in the S&P 500. We show the asymmetry in attention allocation remains. Individual investors pay more attention to extreme negative events, that affect companies, than to comparable positive events.

Overall, our empirical results strongly support investors display a negativity bias in attention allocation with respect to stock returns. Across all specifications, a change in lagged negative extreme returns generates a stronger increase in attention than a change in lagged positive extreme returns. We rule out that negative returns are stronger simply because they are more unusual, or because negative and positive returns are not symmetrical events to the holder in terms of their distribution or number and value of outliers.

References

Malcolm Baker and Jeffrey Wurgler. Investor sentiment in the stock market. (13189), June 2007.

Brad M. Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21(2):785–818, 2008.

R F Baumeister, Ellen Bratslavsky, and Kathleen D Vohs. Bad is stronger than good. *Review of General Psychology*, 5(4):323–370, 2001.

Colin A. Cameron, Jonah B. Gelbach, and Douglas L. Miller. Robust Inference with Multi-way Clustering. Working Paper 327, National Bureau of Economic Research, September 2006.

Zhi Da, Joseph Engelberg, and Pengjie Gao. In Search of Attention. *SSRN eLibrary*, 2009.

Zhi Da, Joseph Engelberg, and Pengjie Gao. The Sum of All FEARS: Investor Sentiment and Asset Prices. *SSRN eLibrary*, 2010.

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

Simon Gervais, Ron Kaniel, and Dan H. Mingelgrin. The high-volume return premium. *The Journal of Finance*, 56(3):877–919, 2001. ISSN 1540-6261.

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, November 2008. ISSN 0028-0836.

David Hirshleifer, James N. Myers, Linda A. Myers, and Siew Hong Teoh. Do individual investors drive post-earnings announcement drift? direct evidence from personal trades. Finance 0412003, EconWPA, December 2004.

Kewei Hou, Lin Peng, and Wei Xiong. A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum. *SSRN eLibrary*, 2008.

Zoran Ivkovic and Scott Weisbenner. Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, 60 (1):267–306, 2005. ISSN 1540-6261.

Niklas Karlsson, George Loewenstein, and Duane Seppi. The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, 38(2):95–115, 2009.

Whitney K. Newey and Kenneth D. West. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28(3):pp. 777–787, 1987. ISSN 00206598.

Mitchell A. Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435–480, 2009.

F Pratto and O P John. Automatic vigilance: the attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, 61(3):380–391, 1991.

Hersh Shefrin and Meir Statman. The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance*, 40(3):777–90, July 1985.

Kennon M. Sheldon, Richard Ryan, and Harry T. Reis. What makes for a good day? competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, 22(12):1270–1279, 1996.

Hal R. Varian and Hyunyoung Choi. Predicting the Present with Google Trends. *SSRN eLibrary*, 2009.

Yu Yuan. Attention and Trading. *SSRN eLibrary*, 2009.

Ning Zhu. The local bias of individual investors. Yale school of management working papers, Yale School of Management, 2009.

Figure 1: SVI Examples

The top panel shows SVI for the query "diet, twitter". The SVI for "diet" is seasonal and has no time trend. It drops by the end of each year, during the holiday season; and spikes at the beginning of the year, probably driven by new year's resolutions to shed extra pounds. The SVI for "twitter" has a time trend because the proportion of Google users that search for this terms has increased over time. SVI is zero before twitter's launch in July, 2006. Since then, twitter has gained popularity worldwide and awareness of the service has exploded as captured by the highest SVI in the last years. The bottom panel presents SVI for "stock quotes", "stock prices", and "best stocks". Search volumes are positively correlated. All three terms have a spike in late September 2008 after Lehman Brothers declared bankruptcy. SVI for "stock quotes" and "stock prices" has decreased over time, probably because of the growing popularity of websites such as Yahoo or Google finance, in which similar information is available.

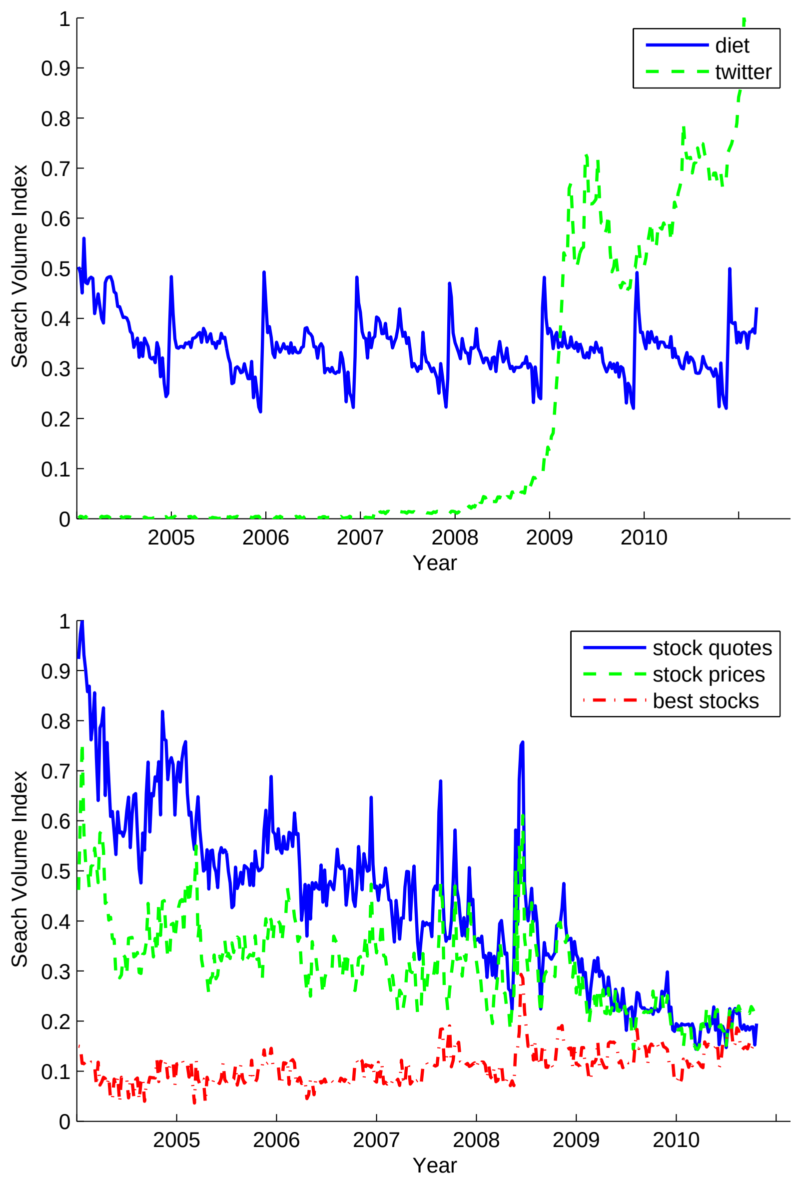
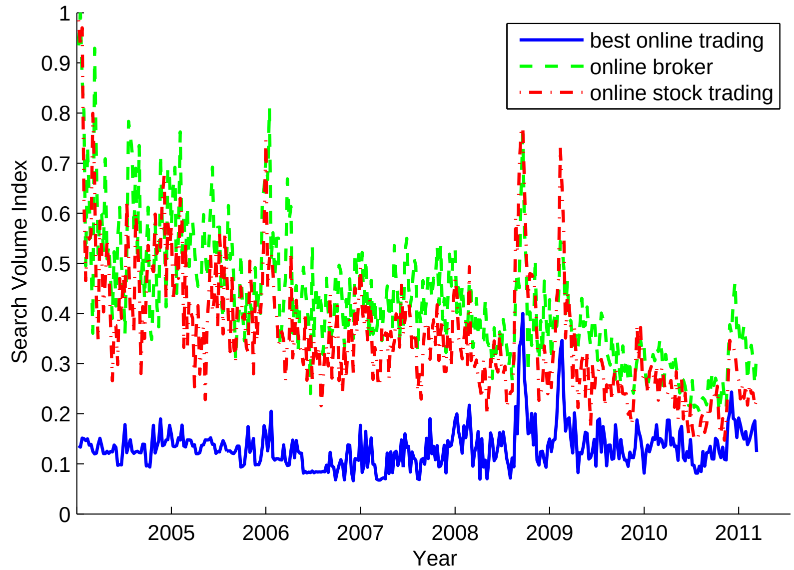


Figure 2: SVI and New Investors

The top panel shows SVI for three terms from column 2 of table 1: "best online trading", "online broker", and "online stock trading". Search volumes are positively correlated and have a spike in late September 2008. The bottom panel presents how quarterly new account opened in TD Ameritrade (reported in forms 10-K and 10-Q after 2007) relate to (quarterly) Aggregated SVI for new investors. Interest in opening brokerage accounts behaves similarly to the real number of accounts open in TD Ameritrade, which provides support for the validity of our measure.



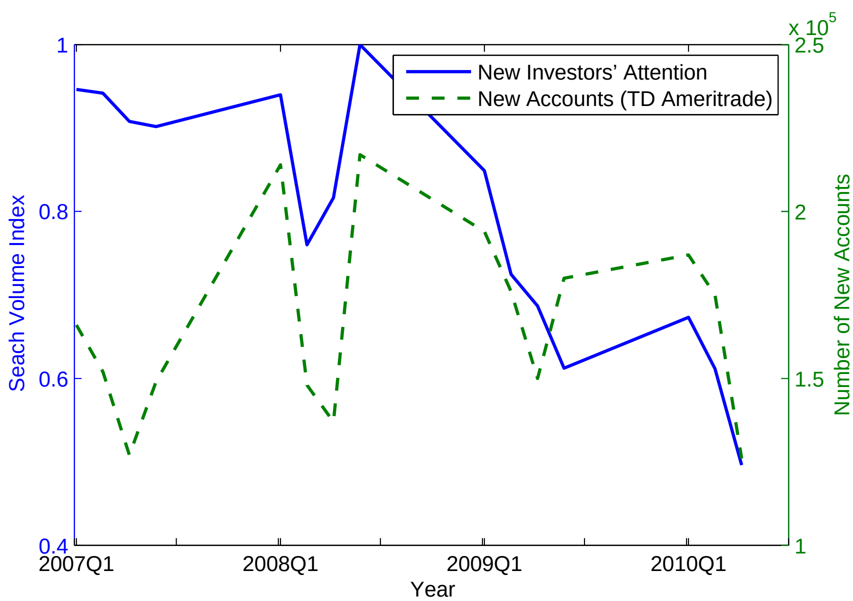


Figure 3: Geographical Distribution of Data

The top panel shows data availability for terms related to $AAllInv$ across the forty-eight continuous states. Google returns no data for the states in white (DE, FL, KS, LA, MT, ND, SD, VT, WV, and WY). For the rest, the percentage of weeks with available data is provided in parenthesis. States with warmer colors have more data. Similarly, the bottom panel shows data availability for terms related to $ANewInv$. Google returns no data for 36 out of 48 states.

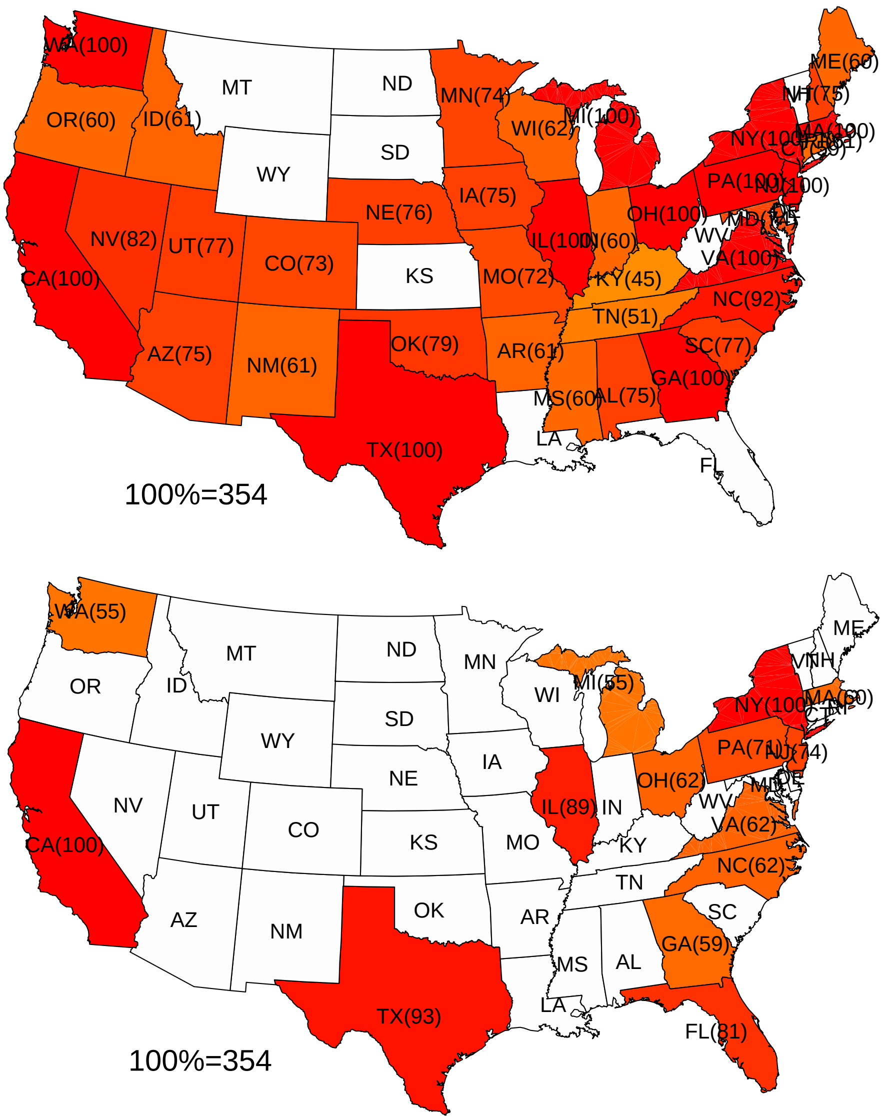


Figure 4: Geographical Distribution of Data

The top panel shows data availability for terms related to *OldInv* across the forty-eight continuous states. Google returns no data for the states in white. For the rest, the percentage of weeks with available data is provided in parenthesis. States with warmer colors have more data. The bottom panel shows the number of companies located in each state. Company location codes are from Compustat. Numbers in parenthesis are relative to the maximum number of companies, which is 817 in CA (followed by NY, with $71\% \times 817 = 580$).

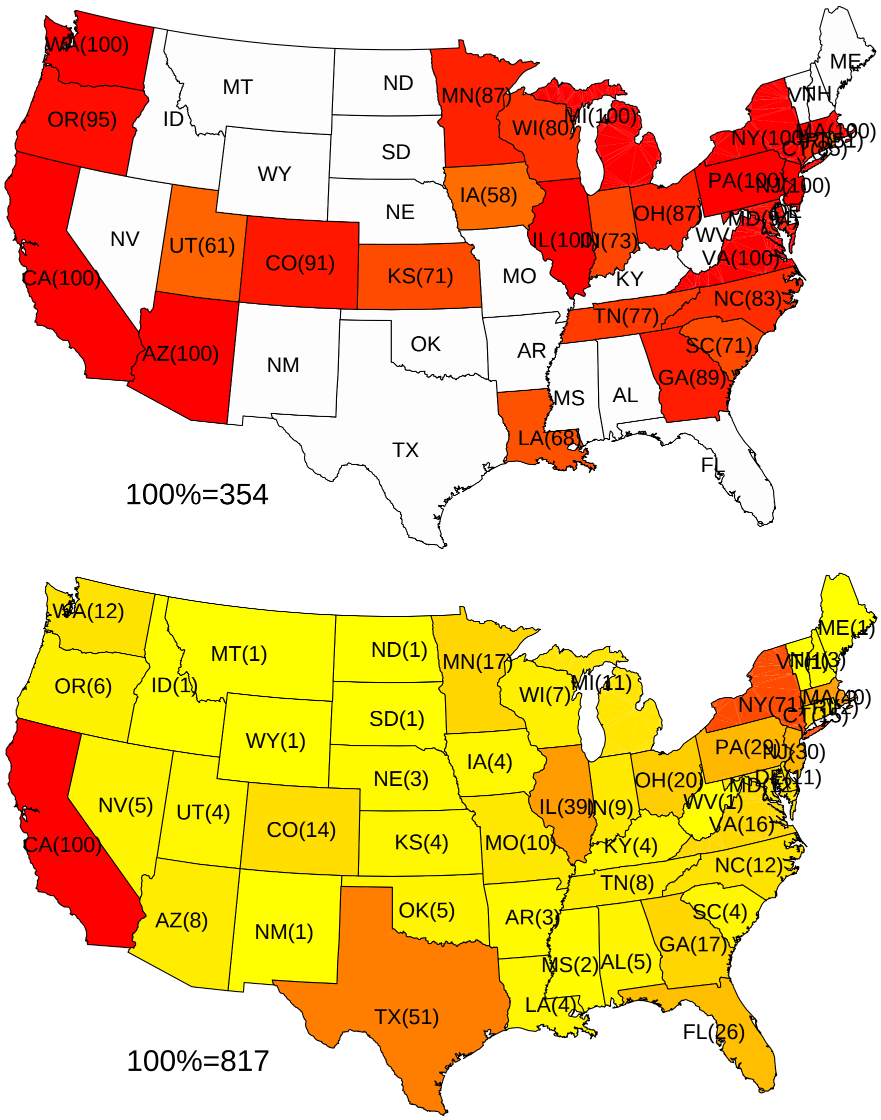


Figure 5: Distribution of Returns

The figure shows histograms for the sets of returns used throughout this paper: (i) top panel presents aggregate U.S. returns used in section 4.1; (ii) middle panel shows within state returns with high market capitalization used in section 4.2; and (iii) bottom panel displays returns for the 100 largest companies in the S&P 500 used in section 4.3. Histograms for original and transformed (using procedure in section 5.1) returns are shown at the left and right-hand side of each panel respectively.

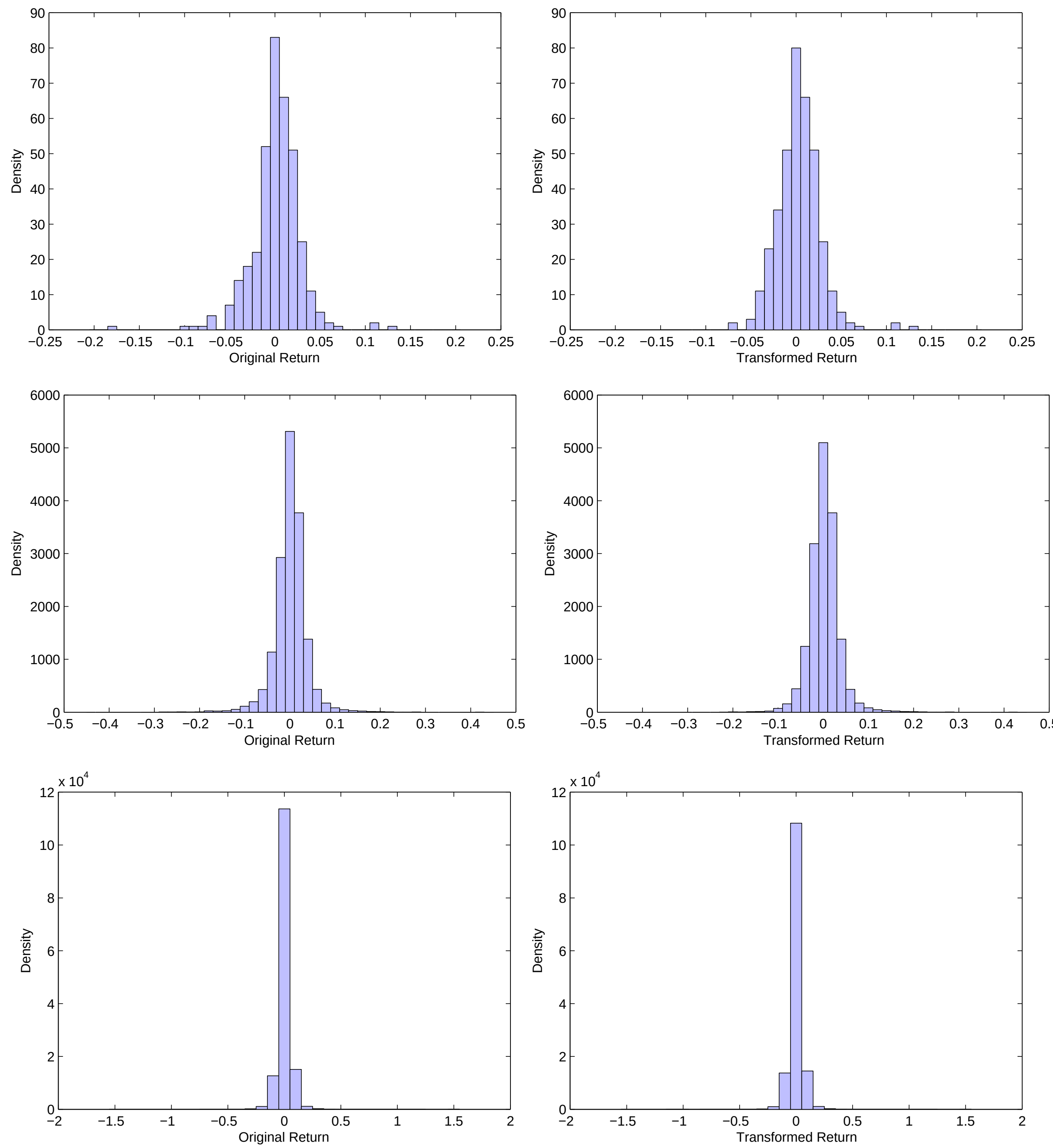


Table 1: List of Terms

Column 1 presents the list of terms used to proxy for All Retail Investors Attention (*AllInv*). We arrive to this list by starting with a small set of terms: "stock market", "stock prices", and "best stocks". We google each of these terms and obtain related searches recommended by Google. Then we drop terms that are either company names (people may be searching for them for many other unrelated reasons), or very general (i.e., online investment), or unrelated (i.e., online auctions). With the remaining terms, we iterate one last time (get related terms and drop irrelevant ones) to get the final list presented here. Similarly, column 2 and 3 show the list of terms used to proxy for New Investors Attention (*NewInv*) and Old Investor's Attention (*OldInv*) respectively.

(1) AllInv	(2) NewInv	(3) OldInv
best stocks	best online trading	ameritrade
dow jones	discount broker	charles schwab
good stocks	discount brokers	etrade
google finance	online broker	scottrade
hot stocks	online brokerage	sharebuilder
market watch	online brokers	
nasdaq	online investing	
stock market	online stock trading	
stock market news	online trading	
stock market today	stock broker	
stock prices		
stockquotes		
yahoo finance		

Table 2: Correlations

AllInv is Abnormal Retail Investor's Attention and measures general interest in the stock market, stock prices, and investment opportunities. *NewInv* is Abnormal New Investor's Attention and relates to people new to the stock market, who are searching information to open a brokerage account. *OldInv* is Abnormal Old Investor's Attention and concerns to retail investors who already own a brokerage account and use Google to find its website. *AAllInv*, *ANewInv* and *AOldInv* are computed using weekly search volume from Google at the aggregate U.S. level. $AVlm$ is Abnormal Trading Volume. Trading volume is from New York Stock Exchange for a value weighed portfolio formed by all stocks in CRSP. $AVIX$ is abnormal VIX, the CBOE market volatility index that measures the implied volatility of S&P 500 index options. *,** and *** represent significance at the 10%, 5% and 1% level. The sample period is from January 2004 and December 2010.

	$AAllInv_t$	$ANewInv_t$	$AOldInv_t$	$AVlm_t$	$AVIX_t$
$AAllInv_t$	1				
$ANewInv_t$	0.666***	1			
$AOldInv_t$	0.637***	0.774***	1		
$AVlm_t$	0.450***	0.255***	0.293***	1	
$AVIX_t$	0.589***	0.445***	0.288***	0.265***	1

Table 3: Aggregate Measures of Attention and Returns

The table reports estimates of β in the regression,

$$DepVar_t = \alpha + \sum_{i=1}^{5} \beta_i P_i Ret_{p,t-1} + \delta Q_{FE} + \varepsilon_t$$

Q_{FE} are quarter dummies. p is a value weighed portfolio formed by all stocks in CRSP in panel (a), and a value weighed portfolio of high market capitalization (highest quartile) firms in panel (b). Returns are sorted into quintiles (i.e., twenty percent partitions). So $P_1 Ret_{p,t}$ is equal to $Ret_{p,t}$ if $Ret_{p,t}$ is one of the 20% smallest returns in portfolio p during the sample period, and zero otherwise. Standard errors are computed using Newey and West (1987) with 3 lags to account for autocorrelation and heteroskedasticity. Variables are divided by their standard deviation so the regression coefficient on a variable can be interpreted as the effect of one standard deviation change on its value. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010.

Panel(a)

	(1)		(2)		(3)	
	$AAllInv_t$		$ANewInv_t$		$AOldInv_t$	
$P_1 Ret_{t-1}$	-0.442***	(-3.51)	-0.406**	(-2.33)	-0.206	(-1.23)
$P_2 Ret_{t-1}$	-0.0792*	(-1.73)	-0.0520	(-0.84)	-0.0310	(-0.55)
$P_3 Ret_{t-1}$	0.0796	(1.56)	0.106*	(1.66)	0.0687	(1.17)
$P_4 Ret_{t-1}$	0.0604	(1.34)	0.115**	(1.98)	0.103	(1.49)
$P_5 Ret_{t-1}$	0.108**	(2.39)	0.147**	(2.48)	0.158**	(2.37)
Q_{FE}	Yes		Yes		Yes	
Adj R-Squared	0.247		0.169		0.184	
Observations	345		345		345	

Panel(b)

	$AAllInv_t$		$ANewInv_t$		$AOldInv_t$	
$P_1 Ret_{mcap,t-1}$	-0.459***	(-3.61)	-0.419**	(-2.43)	-0.213	(-1.26)
$P_2 Ret_{mcap,t-1}$	-0.0848*	(-1.79)	-0.0512	(-0.80)	-0.0199	(-0.36)
$P_3 Ret_{mcap,t-1}$	0.103	(1.45)	0.124*	(1.86)	0.0724	(1.06)
$P_4 Ret_{mcap,t-1}$	0.0910*	(1.69)	0.142**	(2.33)	0.114	(1.59)
$P_5 Ret_{mcap,t-1}$	0.115**	(2.41)	0.149**	(2.47)	0.157**	(2.34)
Q_{FE}	Yes		Yes		Yes	
Adj R-Squared	0.253		0.177		0.188	
Observations	345		345		345	

Table 4: Aggregate Measures of Attention and Returns

The table reports estimates of β in the regression,

$$DepVar_t = \alpha + \sum_{i=1}^{5} \beta_i I_i Ret_{p,t-1} + \delta Q_{FE} + \varepsilon_t$$

Q_{FE} are quarter dummies. p is a value weighed portfolio formed by all stocks in CRSP in panel (a), and a value weighed portfolio of high market capitalization (highest quartile) firms in panel (b). $I_i Ret_{p,t}$ are indicator functions for returns sorted into quintiles (i.e., twenty percent partitions). So $I_1 Ret_{p,t}$ is equal to 1 if $Ret_{p,t}$ is one of the 20% smallest returns in portfolio p during the sample period, and zero otherwise. Standard errors are computed using Newey and West (1987) with 3 lags to account for autocorrelation and heteroskedasticity. Variables are divided by their standard deviation so the regression coefficient on a variable can be interpreted as the effect of one standard deviation change on its value. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010.

Panel(a)

	(1)		(2)		(3)	
	$AAllInv_t$		$ANewInv_t$		$AOldInv_t$	
I_1Ret_{t-1}	1.215***	(7.03)	0.875***	(5.06)	1.508***	(9.37)
I_2Ret_{t-1}	0.776***	(6.37)	0.514***	(3.75)	1.476***	(11.48)
I_3Ret_{t-1}	0.859***	(11.89)	0.713***	(9.07)	1.649***	(21.58)
I_4Ret_{t-1}	0.764***	(10.69)	0.723***	(9.46)	1.675***	(22.06)
I_5Ret_{t-1}	0.859***	(8.29)	0.828***	(6.95)	1.859***	(15.30)
Q_{FE}	Yes		Yes		Yes	
Adj R-Squared	0.149		0.0851		0.176	
Observations	345		345		345	

Panel(b)

	$AAllInv_t$		$ANewInv_t$		$AOldInv_t$	
$I_1Ret_{mcap,t-1}$	1.232***	(7.52)	0.873***	(5.21)	1.536***	(9.67)
$I_2Ret_{mcap,t-1}$	0.810***	(6.78)	0.486***	(3.50)	1.484***	(11.45)
$I_3Ret_{mcap,t-1}$	0.812***	(11.85)	0.694***	(9.29)	1.620***	(21.46)
$I_4Ret_{mcap,t-1}$	0.812***	(11.87)	0.742***	(10.26)	1.704***	(22.52)
$I_5Ret_{mcap,t-1}$	0.838***	(7.76)	0.773***	(6.23)	1.848***	(15.22)
Q_{FE}	Yes		Yes		Yes	
Adj R-Squared	0.150		0.0862		0.174	
Observations	345		345		345	

Table 5: State Level Measures of Attention and Returns

The table reports estimates of β in the regression,

$$DepVar_{s,t} = \alpha + \sum_{i=1}^{5} \beta_i P_i Ret_{p_s^{in},t-1} + \sum_{i=1}^{5} \beta_i P_i Ret_{p_s^{out},t-1}$$
$$+ \gamma Controls + \delta_1 M_{FE} + \delta_2 S_{FE} + \varepsilon_{s,t}$$

M_{FE} and S_{FE} are month and state fixed effects. $AAllInv_s$, $ANewInv_s$ and $AOldInv_s$ are based on search volume data for state s. Stocks are sorted by state using company location codes from Compustat. p^{in} is a portfolio of high market capitalization (highest quartile) companies within state s. p^{out} is a portfolio of similar characteristics formed by companies located outside state. Returns are sorted into quintiles. Monthly state controls are: (i) Coincident Economic Activity Index: summarizes current economic conditions; (ii) Leading Index: predicts the six-month growth rate of the states coincident index; and (iii) Unemployment Rate. To account for correlations among different states in the same week and different weeks in the same state we double cluster using Petersen (2009) implementation of Cameron et al. (2006)'s procedure. Variables are divided by their standard deviation. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010.

	$AAllInv_t$		$ANewInv_t$		$AOldInv_t$	
$P_1 Ret^{in}_{mcap,t-1}$	-0.0761***	(-3.96)	-0.102***	(-3.71)	-0.0644***	(-3.10)
$P_2 Ret^{in}_{mcap,t-1}$	0.00580	(0.59)	0.00410	(0.19)	-0.0134	(-0.98)
$P_3 Ret^{in}_{mcap,t-1}$	0.00627	(0.66)	0.0256	(1.11)	0.0138	(0.76)
$P_4 Ret^{in}_{mcap,t-1}$	0.000994	(0.10)	0.0453**	(2.35)	0.0277*	(1.74)
$P_5 Ret^{in}_{mcap,t-1}$	0.0427***	(3.07)	0.0458**	(2.03)	0.0455*	(1.90)
$P_1 Ret^{out}_{mcap,t-1}$	-0.0978***	(-3.85)	-0.0548**	(-2.11)	-0.0244	(-1.02)
$P_2 Ret^{out}_{mcap,t-1}$	0.00749	(0.61)	-0.0407**	(-2.11)	0.00197	(0.18)
$P_3 Ret^{out}_{mcap,t-1}$	0.0106	(0.90)	0.0384*	(1.70)	0.00769	(1.00)
$P_4 Ret^{out}_{mcap,t-1}$	-0.0141	(-1.00)	0.0256	(1.37)	0.0314*	(1.84)
$P_5 Ret^{out}_{mcap,t-1}$	-0.00516	(-0.27)	0.0206	(0.88)	0.0346*	(1.86)
$Controls$	Yes		Yes		Yes	
M_{FE}	Yes		Yes		Yes	
S_{FE}	Yes		Yes		Yes	
Adj R-Squared	0.392		0.174		0.188	
Observations	8089		3233		7013	

Table 6: State Level Measures of Attention and Returns

The table reports estimates of β in the regression,

$$DepVar_{s,t} = \alpha + \sum_{i=1}^{5} \beta_i I_i Ret_{p_s^{in},t-1} + \gamma Controls + \delta_1 M_{FE} + \delta_2 S_{FE} + \varepsilon_{s,t}$$

M_{FE} and S_{FE} are month and state fixed effects. $AAllInv_s$, $ANewInv_s$ and $AOldInv_s$ are based on search volume data for state s. Stocks are sorted by state using company location codes from Compustat. p^{in} is a portfolio of high market capitalization (highest quartile) companies within state s. $I_i Ret_{p,t}$ are indicator functions for returns sorted into quintiles. So $I_1 Ret_{p,t}$ is equal to 1 if $Ret_{p,t}$ is one of the 20% smallest returns in portfolio p during the sample period, and zero otherwise. Monthly state controls are: (i) Coincident Economic Activity Index: summarizes current economic conditions; (ii) Leading Index: predicts the six-month growth rate of the states coincident index; and (iii) Unemployment Rate. Standard error are computed using state level clustering. Variables are divided by their standard deviation. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010.

	(1) $AAllInv_t$		(2) $ANewInv_t$		(3) $AOldInv_t$	
$I_1Ret^{in}_{mcap,t-1}$	1.143**	(2.66)	0.753	(0.93)	1.167**	(2.30)
$I_2Ret^{in}_{mcap,t-1}$	0.920**	(2.09)	0.594	(0.72)	0.927	(1.74)
$I_3Ret^{in}_{mcap,t-1}$	0.953**	(2.16)	0.657	(0.79)	0.915	(1.75)
$I_4Ret^{in}_{mcap,t-1}$	0.917**	(2.07)	0.734	(0.91)	0.866	(1.63)
$I_5Ret^{in}_{mcap,t-1}$	0.965**	(2.22)	0.669	(0.81)	0.975*	(1.92)
$Controls$	Yes		Yes		Yes	
M_{FE}	Yes		Yes		Yes	
S_{FE}	Yes		Yes		Yes	
Adj R-Squared	0.385		0.173		0.384	
Observations	8089		3233		7013	

Table 7: Company Level Attention and Returns

The table reports estimates of β in the regression,

$$ATicker_{c,t} = \alpha + \sum_{i=1}^{5} \beta_i P_i Ret_{c,t-1} + \gamma Controls + \delta_1 M_{FE} + \delta_2 S_{FE} + \varepsilon_{c,t}$$

M_{FE} and C_{FE} are month and company fixed effects. $ATicker_{c,t}$ is abnormal search volume for the ticker of company c during week t. Weekly company returns are computed with data from CRSP and sorted into quintiles. Controls in column 3 are: (i) $AVlm$: Abnormal Trading Volume (from NYSE); (ii) $LMcap$: natural logarithm of the firms' market capitalization; and (iii) $VWRet$: quintiles of the market return (value weighed return of all stocks in CRSP). We double cluster using Petersen (2009) implementation of Cameron et al. (2006)'s procedure. Variables are divided by their standard deviation. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010 for the 100 largest companies in S&P 500.

	(1)		(2)	
P_1Ret_{t-1}	-0.0465***	(-3.95)	-0.0361***	(-2.98)
P_2Ret_{t-1}	0.000228	(0.03)	0.00636	(0.64)
P_3Ret_{t-1}	0.000138	(0.02)	0.00443	(0.55)
P_4Ret_{t-1}	0.0175*	(1.95)	0.0181*	(1.90)
P_5Ret_{t-1}	0.0340***	(3.32)	0.0215***	(2.70)
$AVlm_{t-1}$			0.133***	(3.45)
$LMcap_{t-1}$			-0.00776	(-0.19)
$VWRet_{t-1}$	No		Yes	
C_{FE}	Yes		Yes	
M_{FE}	Yes		Yes	
Adj R-Squared	0.0407		0.0445	
Observations	15618		15588	

Table 8: Company Level Attention and Returns

The table reports estimates of β in the regression,

$$ATicker_{c,t} = \alpha + \sum_{i=1}^{5} \beta_i I_i Ret_{c,t-1} + \gamma Controls + \delta_1 M_{FE} + \delta_2 C_{FE} + \varepsilon_{c,t}$$

M_{FE} and C_{FE} are month and company fixed effects. $ATicker_{c,t}$ is abnormal search volume for the ticker of company c during week t. Weekly company returns are computed with data from CRSP. $I_i Ret_{p,t}$ are indicator functions for returns sorted into quintiles. So $I_1 Ret_{p,t}$ is equal to 1 if $Ret_{p,t}$ is one of the 20% smallest returns in portfolio p during the sample period, and zero otherwise. Controls in column 3 are: (i) $AVlm$: Abnormal Trading Volume (from NYSE); (ii) $LMcap$: natural logarithm of the firms' market capitalization; and (iii) $VWRet$: quintiles of the market return (value weighed return of all stocks in CRSP). Standard error are computed using company level clustering. Variables are divided by their standard deviation. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010 for the 100 largest companies in S&P 500.

	(1)		(2)	
$I_1 Ret_{t-1}$	0.274***	(2.74)	0.394	(0.61)
$I_2 Ret_{t-1}$	0.199**	(2.29)	0.357	(0.56)
$I_3 Ret_{t-1}$	0.218**	(2.28)	0.375	(0.59)
$I_4 Ret_{t-1}$	0.239**	(2.52)	0.389	(0.61)
$I_5 Ret_{t-1}$	0.250***	(2.83)	0.390	(0.61)
$AVlm_{t-1}$			0.147***	(3.71)
$LMcap_{t-1}$			-0.0274	(-0.69)
$VWRet_{t-1}$	No		Yes	
S_{FE}	Yes		Yes	
M_{FE}	Yes		Yes	
Adj R-Squared	0.0429		0.0476	
Observations	15618		15588	

Table 9: Company Level Attention and Redistributed Returns

The table reports estimates of β in the regression,

$$ATicker_{c,t} = \alpha + \sum_{i=1}^{5} \beta_i P_i Ret_{c,t-1} + \gamma Controls + \delta_1 M_{FE} + \delta_2 C_{FE} + \varepsilon_{c,t}$$

M_{FE} and C_{FE} are month and state fixed effects. $ATicker_{c,t}$ is abnormal search volume for the ticker of company c during week t. Weekly company returns are computed with data from CRSP and sorted into quintiles. Controls in column 3 are: (i) $AVlm$: Abnormal Trading Volume (from NYSE); (ii) $LMcap$: natural logarithm of the firms' market capitalization; and (iii) $VWRet$: quintiles of the market return (value weighed return of all stocks in CRSP). Returns are redistributed to rule out that negative returns are stronger simply because they are more unusual or have stronger outliers. We double cluster using Petersen (2009) implementation of Cameron et al. (2006)'s procedure. Variables are divided by their standard deviation. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010 for the 100 largest companies in S&P 500.

	(1)		(2)	
P_1Ret_{t-1}	-0.0363***	(-3.37)	-0.0271**	(-2.51)
P_2Ret_{t-1}	0.000456	(0.06)	0.00310	(0.38)
P_3Ret_{t-1}	-0.000324	(-0.05)	0.00147	(0.16)
P_4Ret_{t-1}	0.0128	(1.27)	0.0111	(1.08)
P_5Ret_{t-1}	0.0335***	(3.07)	0.0191**	(2.27)
$AVlm_{t-1}$			0.136***	(3.36)
$LMcap_{t-1}$			-0.00959	(-0.21)
$VWRet_{t-1}$	No		Yes	
S_{FE}	Yes		Yes	
M_{FE}	Yes		Yes	
Adj R-Squared	0.0415		0.0453	
Observations	15177		15177	

Table 10: Company Level Attention and Redistributed Returns

The table reports estimates of β in the regression,

$$ATicker_{c,t} = \alpha + \sum_{i=1}^{5} \beta_i I_i Ret_{c,t-1} + \gamma Controls + \delta_1 M_{FE} + \delta_2 C_{FE} + \varepsilon_{c,t}$$

M_{FE} and C_{FE} are month and state fixed effects. $ATicker_{c,t}$ is abnormal search volume for the ticker of company c during week t. Weekly company returns are computed with data from CRSP. $I_i Ret_{p,t}$ are indicator functions for returns sorted into quintiles. So $I_1 Ret_{p,t}$ is equal to 1 if $Ret_{p,t}$ is one of the 20% smallest returns in portfolio p during the sample period, and zero otherwise. Controls in column 3 are: (i) $AVlm$: Abnormal Trading Volume (from NYSE); (ii) $LMcap$: natural logarithm of the firms' market capitalization; and (iii) $VWRet$: quintiles of the market return (value weighed return of all stocks in CRSP). Returns are redistributed to rule out that negative returns are stronger simply because they are more unusual or have stronger outliers. Standard error are computed using company level clustering. Variables are divided by their standard deviation. t-statistics are in parenthesis. *,** and *** represent significance at the 10%, 5% and 1% level. We use weekly data and the sample period is from January 2004 and December 2010 for the 100 largest companies in S&P 500.

	(1)		(2)	
$I_1 Ret_{t-1}$	0.128***	(3.52)	0.486	(0.67)
$I_2 Ret_{t-1}$	0.0640*	(1.98)	0.459	(0.63)
$I_3 Ret_{t-1}$	0.0659**	(2.25)	0.459	(0.64)
$I_4 Ret_{t-1}$	0.100***	(3.15)	0.487	(0.67)
$I_5 Ret_{t-1}$	0.108***	(3.56)	0.485	(0.68)
$AVlm_{t-1}$			0.143***	(3.50)
$LMcap_{t-1}$			-0.0238	(-0.55)
$VWRet_{t-1}$	No		Yes	
S_{FE}	Yes		Yes	
M_{FE}	Yes		Yes	
Adj R-Squared	0.0431		0.0474	
Observations	15177		15177	