REVISITING THE CROSS-SECTION OF EXPECTED STOCK RETURNS: EVIDENCE FROM A TEXTUAL ANALYSIS OF BUY RECOMMENDATIONS

Hailiang Chen, Byoung-Hyoun Hwang, and Zhuozhen Peng[*]

This Draft: June 2023

Our paper examines sell-side analyst reports and online stock opinion articles, which recommend that investors buy stocks that, based on prior literature, trade at comparatively high prices and earn low future returns. We conduct textual analysis and test whether the justifications provided in these buy recommendations mostly (1) emphasize a stock's safe-haven quality, (2) indicate general investor exuberance, or (3) point to a specific preference for stocks with high upside potential. We find that the buy recommendations mostly emphasize stocks' upside potential. Our results suggest that non-traditional investor preferences play a material role in explaining the cross-section of expected stock returns.

JEL Classification: G11, G12, G14, G40.

Keywords: Cross-Section of Expected Stock Returns, Anomalies, Risk, Behavioral Finance, Textual Analysis.

1. Introduction

Over the past 50 years, the finance literature has produced a long list of firm characteristics that predict stock returns in the cross-section: As we sort stocks each month based on the firm characteristic, form decile portfolios, and compute the average raw return for each portfolio, we observe a monotonically increasing (or decreasing) pattern in portfolio returns. Most of the predictability comes from stocks on one end of the spectrum earning unusually low returns (e.g., Stambaugh, Yu, and Yuan, 2012). In the literature, these stocks are referred to as "short-leg securities" or as stocks residing "in the short leg" of an "anomaly."

There are at least three prominent explanations for why short-leg securities earn lower returns. One possibility is that short-leg securities are "safe-haven assets" that tend to do better when the states of the world are bad. It is rational behavior for risk-averse investors to pay high prices and accept low average returns for such stocks (e.g., Lucas, 1978; Breeden, 1979; Fama, 1998). An alternative, behavioral perspective is that investors tend to form overly optimistic beliefs about short-leg securities. Overly optimistic beliefs can arise from investors' extrapolating past cash-flow growth or stock market performance (e.g., Barberis, Shleifer, and Vishny, 1998; Barberis and Shleifer, 2003). They can also come from overconfidence (Daniel, Hirshleifer, and Subrahmanyam, 1998). If there are limits to arbitrage, overly optimistic beliefs generate overpricing. As overpricing becomes corrected, these stocks earn low future stock returns (e.g., Barberis and Thaler, 2003). A third perspective, also behavioral, is that investors have non-traditional preferences. Perhaps the most prominent non-traditional preference accords with cumulative prospect theory. Under cumulative prospect theory, people derive high utility from the possibility of extreme payoffs, even if the probability of receiving such payoffs is exceedingly small. Cumulative prospect theory explains why people "overpay" for lottery tickets. If investors believe that short-leg securities exhibit lottery-like features, it may also explain why investors overpay for stocks residing in the short leg of an anomaly (e.g., Barberis, 2018).

Gauging which of these three perspectives drives the unusually low returns of short-leg securities is challenging, as the patterns we observe are generally consistent with all three frameworks. To illustrate, consider the "asset-growth anomaly." Cooper, Gulen, and Schill (2008) show that stocks with high asset

growth subsequently earn unusually low average returns. One possibility is that investors perceive stocks with high asset growth as assets that provide insurance against bad states of the world and therefore are willing to pay higher prices and accept lower returns. Another possibility is that investors overreact to past growth and push prices up too high. As overpricing subsequently becomes corrected, these stocks earn unusually low returns. A final possibility is that investors believe that stocks with high asset growth are more likely to generate extreme payoffs. This perception causes skewness-loving investors to flock to high-asset-growth stocks and push the prices of these stocks above their fundamental values.

Our paper proposes a method that we believe could aid us in distinguishing between the three perspectives. Investors frequently express their opinions in writing. We posit that by parsing these texts and extracting the key considerations that drive their stock recommendations, we can make progress in better understanding which of the three frameworks best explains investors' buying decisions and the unusually low returns of short-leg securities. To continue with our above example, suppose we consider the texts of investors who express positive views of high-asset-growth stocks. Suppose further that in explaining why they like high-asset-growth stocks, investors primarily highlight the safe-haven quality of these stocks. In that case, we may infer that the relatively low returns on high-asset-growth stocks arise mostly because risk-averse investors view these stocks as safe-haven assets for which they are willing to pay high prices and accept low returns. Alternatively, suppose that investors are primarily engrossed by the stocks' seeming superiority and upward trajectory. In that case, we may infer that investors tend to become exuberant about high-asset-growth stocks, which can cause these stocks to become overpriced and subsequently earn low returns. Finally, suppose that investors primarily fixate on the perceived upside potential. In that case, we may infer that non-traditional investor preferences best explain the relatively high prices of and low future returns on high-asset-growth stocks.

To implement our method, we consider two large corpora of stock opinion articles. We first consider sell-side analyst reports. Sell-side analysts routinely issue written reports in which they synthesize their thoughts on the firms they cover. We compile the texts of 1,171,130 analyst reports written on 6,130 individual stocks in the United States (US). Our sample spans the period running from January 2006 through

October 2021. To the best of our knowledge, our study represents the largest and most comprehensive analysis of the written components of analyst reports in the literature.[1]

Our second source of stock opinion articles is Seeking Alpha (hereafter SA; http://seekingalpha.com). SA is a leading investments-related website in the US. Users can submit a stock opinion article for possible publication on the SA website. A team of editors curates these submissions. If their articles are deemed of adequate quality and published on the SA website, the authors receive income based on the article type and the number of page views their articles generate. SA reports that, as of March 2019, its website attracted more than 15 million unique visitors a month; its audience had an average household income of $321,302, 65% of whom traded at least once a month. We compile 140,412 SA articles written on 5,718 individual stocks trading in the US. Our SA sample spans the same time period as our analyst sample. While we view analyst reports as reflecting primarily institutional investors' opinions, we view SA articles as capturing mostly retail investors' perspectives.

We then proceed as follows: Chen and Zimmermann (2022) survey and replicate the literature and find that all but three of the 205 firm characteristics they consider strongly predict raw stock returns in the cross-section. We separately consider each of the 205 firm-level characteristics. For instance, each month, we examine stocks in the top-asset-growth decile. These are stocks that earn relatively low future returns. We parse all analyst- and SA buy recommendations and examine whether, in explaining why they like high-asset-growth stocks, analysts and SA contributors unusually fixate on the stocks' safe-haven quality, the stocks' general excellence, or the stocks' perceived upside potential. We conduct analog tests for the other 204 firm-level characteristics.

To assess whether investors fixate on a stock's safe-haven quality, a stock's general excellence, or a stock's perceived upside potential, we adopt a dictionary-based textual analysis approach. We create (1) a list of words that we think investors would use to highlight a stock's safe-haven status ("safety words"),

---

[1] Huang, Zang, and Zheng (2014) consider 363,952 analyst reports from 1995 to 2008. De Franco, Hope, Vyas, and Zhou (2015) consider 356,463 analyst reports from 2002 to 2009. Huang, Lehavy, Zang, and Zheng (2018) consider 476,633 analyst reports from 2003 to 2012.

(2) a list of words that investors would use to emphasize a stock's superiority ("exuberance words"), and (3) a list of words that investors would use to describe a stock's lottery-like features ("lottery words"). We then test whether words from any of these three lists appear unusually frequently in analysts' and SA contributors' buy recommendations for short-leg securities.

To create our three wordlists, we collaborate with a market research firm that conducts investor surveys for large financial institutions (CoreData Research). We recruit one hundred US-based institutional investors. 81% of these investors have more than $100 million in assets under management (AUM); 34% have AUM of $2.5 billion or more. 97% of the institutional investors in our survey have more than ten years of work experience. We ask the institutional investors to list five words that they would use to describe (1) "*a stock that, to you, is a 'safe-haven asset': a stock that does relatively well when times are bad*"; (2) "*a stock that has been doing well and that you expect will continue to do very well or, in general, a stock that you are very confident will earn above-normal returns*"; and (3) "*a stock that offers somewhat of a gamble: the stock will most likely not produce above-normal returns, but, if it does, the payoff will be enormous*."

For each question, we select the five most frequently mentioned terms. Our five safety terms are *conservative*, *defensive*, *protection*, *reliable*, and *stable*. Our five exuberance terms are *competitive*, *expanding*, *leader*, *outperformer*, and *strong*. Our five lottery terms are *gamble*, *potential*, *speculative*, *upside*, and *volatile*. We consider all possible word forms of the above terms that are meaningfully tied to the business realm, and we account for negation.

In additional analyses, we send an almost identical version of our institutional investor survey to 303 US retail investors and construct retail investor-based wordlists. We also consider exuberance terms that are rooted more in theory. The behavioral finance literature notes that investors can become exuberant about a stock (1) when they are overconfident in their beliefs and (2) when they extrapolate and assume that an actual or perceived trend will continue (Barberis, 2018). To capture exuberance rooted in overconfidence, we consider all sequences of four words in a text ("4-grams") and examine for each 4-gram whether a strong modal word appears jointly with a positive word (e.g., *definitely – achieve*). To

capture exuberance rooted in extrapolative tendencies, we consider all 4-grams and probe whether a "continuation term" appears jointly with a "growth term" (e.g., *continue – outperform*).

To discuss some of our findings at the firm-characteristic level, we find that analyst reports substantially more frequently use lottery words when discussing high-asset growth stocks than when discussing stocks that do not have high asset growth: the fraction of lottery words is 15% higher (*t*-statistic = 32.72). SA articles also use lottery words more frequently when discussing stocks with high asset growth (+9%, *t*-statistic = 6.65). We observe no reliable differences in the use of safety or exuberance words. We arrive at the same conclusion when we consider the retail investor wordlists and our more methodical, theory-driven exuberance terms. The abnormal use of lottery words in the buy recommendations for high-asset growth stocks, combined with the lack of reliable differences in the use of safety and exuberance words, suggests that investors particularly like the upside potential they see in these stocks. Coupled with investors' prospect theory preferences, this perception may help explain why stocks with high asset growth trade at comparatively high prices and earn low future returns.

We also find evidence for the relevance of the risk and irrational beliefs frameworks. For instance, we find that the buy recommendations for stocks with low operating leverage stand out for their unusually heavy use of safety words. We observe this pattern in both analyst reports and SA articles. We observe no abnormal use of exuberance or lottery words. Our results suggest that investors view unlevered stocks as particularly safe, which may explain why these stocks trade at comparatively high prices and earn low returns on average (Novy-Marx, 2011).

In comparison, we find that the buy recommendations for stocks with high returns over the past three years most notably stand out for their unusual reliance on exuberance words. We observe this pattern in both analyst reports and SA articles. Our results indicate that the poor future performance of stocks with high returns over the past three years (De Bondt and Thaler, 1985) is most congruent with overpricing due to overly optimistic beliefs.

For some of the 205 firm characteristics we consider, there are no analyst reports or SA articles for the stocks in the short leg either because there are very few stocks in the short leg or because the short-leg

securities represent microcap stocks, which sell-side analysts and SA contributors rarely cover. Our final analysis thus comprises 181 firm-level characteristics.

We find that analysts' rationales for liking short-leg securities are most consistent with the risk framework for 12 out of the 181 firm characteristics, or 7% of the time. That is, for 12 out of 181 anomalies, or 7% of the time, analyst buy recommendations for short-leg securities most notably stand out for their heavy reliance on safety words. Analysts' rationales for liking short-leg securities are most consistent with the irrational beliefs framework 17% of the time. That is, 17% of the time, analyst buy recommendations for short-leg securities most notably stand out for their heavy reliance on exuberance words. Finally, analysts' rationales are most consistent with the non-traditional preferences framework 58% of the time. That is, 58% of the time, analyst buy recommendations for short-leg securities most notably stand out for their heavy reliance on lottery words. The results are inconclusive 19% of the time, as we detect no abnormal use of either safety, exuberance, or lottery words. For SA articles, the corresponding fractions are 6% (risk framework), 10% (irrational beliefs framework), and 54% (non-traditional preferences framework). The results are inconclusive 29% of the time.

Overall, a comparison of the fractions across analyst reports and SA articles suggests that while all three frameworks explain components of the cross-section of expected stock returns, the majority of times, non-traditional investor preferences are the dominant force.

We arrive at the same conclusion when we consider our retail investor wordlists and the more methodical, theory-driven exuberance words. We also make similar observations when we consider only the "most important" firm characteristics, that is, firm characteristics for which the citation number of the paper documenting the corresponding predictability is in the top quartile. Our results do not change when we use only data generated since the corresponding academic paper has been published, and investors should be aware that the respective firm characteristics are associated with low future stock returns.

Our conclusion that non-traditional investor preferences play a material role in investors' buying decisions of short-leg securities hinges on two assumptions: (1) Our wordlists and dictionary-based

approach can capture beliefs and preferences. (2) In addition, the beliefs and preferences expressed in analyst reports and SA articles at least partially reflect those of the investor population.

To assess the validity of the first assumption, we draw from Bordalo, Gennaioli, LaPorta, and Shleifer (2019, 2022) and Nagel and Xu (2022), who provide strong evidence that the analysts issuing the highest long-term earnings growth forecasts have biased, overly optimistic beliefs due to extrapolation. We find that the analysts with the highest long-term earnings growth forecasts use substantially more exuberance words in their reports, suggesting that the use of our exuberance words, indeed, positively correlates with biased, overly optimistic beliefs due to extrapolation. In contrast, we observe no abnormal use of safety or lottery words. The lack of association indicates that the use of lottery words is tied more to non-traditional preferences than to biased beliefs.

Regarding the second assumption, we consider firm characteristics for which we have strong priors regarding why they predict returns. We then test whether the results based on our method agree with those priors. For instance, Bordalo, Gennaioli, LaPorta, and Shleifer (2019) provide compelling evidence that stocks with high earnings growth subsequently earn low returns because investors form overly optimistic beliefs and create overpricing. In line with this view, we find that the buy recommendations for stocks with high earnings growth are marked with an unusually heavy use of exuberance words. We observe no abnormal use of safety or lottery words. Similarly, it appears unlikely that the unusually low returns of stocks with high failure probability would be tied to overly optimistic beliefs. The fact that stocks with high failure probability have positively skewed returns and the possibility that investors exhibit cumulative prospect theory preferences offer a more reasonable explanation. Consistent with this notion, we find that the buy recommendations for stocks with high failure probability stand out for their unusually heavy use of lottery words. We observe no abnormal use of safety or exuberance words.

To more systematically evaluate the sensibility of our results, we draw from Barberis, Jin, and Wang (2021), who note that biased-belief-based anomalies should accrue a substantial portion of their returns around earnings announcements when mistaken beliefs are confronted with reality and become partially corrected. In contrast, if an anomaly is generated by non-traditional preferences, the returns should

not concentrate around earnings announcements. On a related note, Barberis (2018) notes that extrapolative tendencies should be more pronounced among larger firms, whereas skewness preferences should be more relevant for smaller firms.

We find that the more an anomaly earns its returns around earnings announcements and the larger the average short-leg security is, the more the corresponding buy recommendations stand out for their unusual reliance on exuberance words. In contrast, buy recommendations are significantly more likely to be marked by an unusual reliance on lottery words, the less an anomaly's returns concentrate around earnings announcements and the smaller the average short-leg security is.

We discuss important caveats and limitations of our method in the main body of the text. But given the above considerations, it appears plausible that our method of assigning anomalies to their most probable cause is not entirely futile and could therefore help us better understand the behavior of investors and the stock market.

The remainder of our paper is organized as follows. In Section 2, we situate our paper in the relevant literature streams. In Section 3, we discuss our data, wordlists, and key variables. We present the results from our main analysis and various sensitivity analyses in Sections 4 and 5. Sections 6 and 7 discuss important caveats and limitations of our method, and Section 8 concludes.

2. Literature Review and Contribution

Our paper relates to several large bodies of literature. First, our paper relates to the literature on the cross-section of expected stock returns and stock return anomalies. Our paper also relates to the behavioral finance literature, the textual analysis literature, and a growing body of work using surveys to better understand investors' decision-making.

2.1 The Cross-Section of Expected Stock Returns, Anomalies, and Behavioral Finance Literatures

The empirical asset-pricing literature has documented many cross-sectional stock-return predictabilities, that is, instances in which stocks on one end of a firm characteristic spectrum have reliably lower raw

returns than stocks on the other end of the spectrum. Recently, there has been increasing debate about whether these predictabilities are real or economically important.

Harvey, Liu, and Zhu (2016) argue that, once we account for the possibility that researchers run a series of tests and report only their most significant findings, most cross-sectional stock-return predictabilities are not real and do not hold out of sample. Linnainmaa and Roberts (2018) make a related argument. Hou, Xue, and Zhang (2020) propose that there is little predictability beyond the smaller and economically less meaningful stocks.

On the other hand, Chen (2021) conducts a thought experiment and argues that if the average stock-return patterns were not real, the amount of data mining researchers would have to engage in to generate the patterns documented in the literature is implausibly large.

McLean and Pontiff (2016) compare the difference in the average raw returns between long-leg- and short-leg securities reported in the "original" study to the long-short returns observed outside the original sample period and the long-short returns since publication. If the long-short returns are the product of data mining, they should drop significantly outside the original sample period. If the long-short returns are the product of mispricing, they should drop noticeably after being published and brought to arbitrageurs' attention. McLean and Pontiff find that the long-short returns are 26% weaker out of sample and an additional 32% weaker after publication. The findings of McLean and Pontiff suggest that while data mining plays a role, a substantial portion of the long-short returns is real. Jacobs and Müller (2020) and Jensen, Kelly, and Pedersen (2023) arrive at similar conclusions.

Our study builds on McLean and Pontiff (2016), Jacobs and Müller (2020), Chen (2021), and Jensen, Kelly, and Pedersen (2023) and assumes that the cross-sectional differences in average returns are real; short-leg securities are thus distinct not only in the minds of some econometricians but also in the minds of investors. Our finding that there are systematic differences in how sell-side analysts and SA contributors describe short-leg securities is consistent with this conjecture. Assuming that the cross-sectional differences are real, we contribute to the empirical asset-pricing literature by providing estimates of the extent to which the low returns of short-leg securities – a key feature of the cross-section of expected

stock returns – reflect differences in risk, mispricing induced by irrational beliefs, or mispricing induced by non-traditional preferences.

By comparing the relevance of irrational beliefs with that of non-traditional preferences, our study also adds to the behavioral finance literature. Irrational beliefs and non-traditional preferences represent the two primary departures from the traditional finance paradigm, and most research in behavioral finance falls into one of these categories (Barberis and Thaler, 2003; Barberis, 2018). To guide future research in behavioral finance, it is important that we know which of the two departures more comprehensively explains the low returns of short-leg securities. In a recent study, Moskowitz and Vasudevan (2022) analyze the sports-betting market and test whether the low returns associated with betting on underdogs come from irrational beliefs or non-traditional preferences. In line with our findings, the authors suggest "*that preferences for lottery-like payoffs, rather than incorrect beliefs, drive the lower returns to betting on risky underdogs versus safe favorites*" (abstract). Our study contributes to the literature by examining whether Moskowitz and Vasudevan's suggestion for the sports-betting market extends to the stock market.

2.2 The Textual Analysis and Investor Survey Literatures

By conducting a textual analysis of investors' thoughts and opinions, our paper naturally relates to the textual-analysis literature in finance. Unlike prior work, which measures the tone of a text to gauge *whether* investors like a particular stock (Antweiler and Frank, 2004; Das and Chen, 2007; Tetlock, 2007; Loughran and McDonald, 2011; Garcia, 2013), our study examines *why* investors like a particular stock and what the why tells us about investors' decision-making.

By recording investors' reasoning for investing in stocks, our paper also relates to recent studies that use surveys to understand investors' decision-making. Choi and Robertson (2020) survey US households and ask whether factors such as the "*concern that when I have to cut my spending, the stock market will tend to drop*" are important to households as they invest in the stock market. Relatedly, Chinco, Hartzmark, and Sussman (2022) survey investors to test whether the correlation between stock returns and consumption growth plays a material role in investors' decision-making. The authors find that "*only 11%*

reported thinking about consumption-growth correlations in a manner consistent with textbook theory"
(page 2186).

We add to the literature by proposing a complementary method to uncover what causes investors to make certain decisions. Investors constantly express their views and reasoning in various forms of text. By parsing these texts, we can obtain a unique perspective into investors' minds without continuously surveying them. Surveys have the advantage of enabling researchers to ask nuanced questions and provide relatively clean evidence. Our text-based approach has the advantage of being cheap and scalable. In the end, it appears to us that both the survey-based approach and the text-based method can play important roles in future research aimed at understanding investors' behavior.

3. Data and Variables

This section describes our key data sources and wordlists.

3.1 The Three Frameworks

To motivate our data and wordlists, we begin with a brief outline of the three frameworks that our wordlists intend to capture.

3.1.1 The Risk Framework

In the traditional finance paradigm, investors rationally pay a high price and accept low average returns for stocks that provide insurance against bad future states of the world. Examples of bad future states include low consumption growth (Lucas, 1978; Breeden, 1979), economic disaster (Barro, 2006), and the arrival of negative news about the expectations for and volatility of long-run consumption growth (Bansal and Yaron, 2004).

3.1.2 The Irrational Beliefs Framework

In contrast, behavioral finance contends that many of the observed stock market patterns reflect (1) mistaken investor beliefs coupled with limits to arbitrage, or (2) non-traditional preferences.

There are two reasons investors may systematically form incorrect beliefs. One reason is that people extrapolate: "*their estimate of the future value of a quantity is a positive function of the recent past values of that quantity*" (Barberis, 2018, p. 16). Greenwood and Shleifer (2014) provide survey evidence that investors extrapolate past stock market returns. Greenwood and Shleifer also find that the average belief of the surveyed investors negatively predicts future returns, suggesting that investors *over*-extrapolate and that there are limits to arbitrage. Theoretical models show that the extrapolation of past stock returns, coupled with frictions, can generate many cross-sectional return patterns such as momentum and long-run reversal (e.g., Barberis and Shleifer, 2003; Barberis, Greenwood, Jin, and Shleifer, 2015; 2018).

Investors may also extrapolate past growth in fundamentals. Barberis, Shleifer, and Vishny (1998), among others, show analytically that the extrapolation of past cash-flow growth, coupled with frictions, can generate cross-sectional differences in average stock returns.

The second reason investors may systematically form incorrect beliefs is that people are overconfident. When investors receive a private signal about a stock, they tend to overestimate the precision of the signal and overreact. Any subsequent event that aligns with the signal further boosts investors' confidence in the accuracy of their analysis. Any event that contradicts the signal is discounted as an exception. Daniel, Hirshleifer, and Subrahmanyam (1998) show analytically that such overconfidence, coupled with frictions, can help explain the cross-section of expected stock returns.

3.1.3 The Non-Traditional Preferences Framework

A second class of behavioral finance models contends that investors have non-traditional preferences. Most of these models are rooted in Kahneman and Tversky's prospect theory (1979; 1992). Prospect theory posits that investors' utility is defined over gains and losses ("reference dependence") and that losses weigh more heavily than gains ("loss aversion"). In addition, investors are risk-averse over moderate-probability gains

and risk-seeking over moderate-probability losses ("diminishing sensitivity"). Perhaps most strikingly, people overweight low-probability events at the tails of a distribution ("probability weighting"). A 1-in-1,000 chance of gaining $10,000 looks more attractive than earning a certain $10. Moreover, a 1-in-1,000 chance of losing $10,000 looks scarier than losing a certain $10. Probability weighting thus explains why people invest in both lotteries and insurance policies, a behavior that is difficult to capture under the traditional Expected Utility paradigm. As investors may perceive the return distributions of some financial assets to resemble those of lotteries, probability weighting could also explain investors' preferences for certain types of financial assets and the prices they are willing to pay.

While the irrational beliefs framework emphasizes general investor exuberance, probability weighting highlights one particular feature that investors like about a stock. Probability weighting can therefore explain why certain stocks, such as those with high failure probability, can become overpriced even when there is little ground for exuberance.[2]

3.2 Survey-Based Wordlists

To examine which of these frameworks most comprehensively explains investors' decision-making, we study investors' buy recommendations and the words they use.

3.2.1 Institutional Investors' Survey-Based Wordlists

Our first and primary set of wordlists is rooted in an online survey sent to 100 institutional investors. To reach institutional investors, we collaborate with CoreData Research (https://coredataresearch.com). CoreData Research is a market research firm that conducts investor surveys for large financial institutions.

Our subject pool comprises US-based wealth managers, mutual fund managers, pension fund managers, and hedge fund managers. We require that all managers actively invest in US stocks.

---

[2] Stocks with high failure probability have highly positive skewed returns.

As displayed in Online Appendix Figure A1, our online survey comprises questions regarding an institutional investor's age, gender, work experience, and AUM. Online Appendix Table A1 shows that 81% of the institutional investors in our sample report managing assets worth more than $100 million; 34% report having more than $2.5 billion in AUM. 97% of the institutional investors in our sample have more than ten years of work experience; 69% have more than twenty years of work experience.

Our key questions are as follows:

"*For each of the next three questions, please list up to five **nouns, verbs, or adjectives (NOT specific tickers, company names, industries, or product names/brands**) that you would use to:*

Q1. *Describe a stock that, to you, is a 'safe-haven asset': a stock that does relatively well when times are bad. If you would never invest in such a stock, **please leave everything blank** and simply move on to the next question.*

Q2. *Describe a stock that has been doing well and that you expect will continue to do very well or, in general, a stock that you are very confident will earn above-normal returns. If you would never invest in such a stock, **please leave everything blank** and simply move on to the next question.*

Q3. *Describe a stock that offers somewhat of a gamble: the stock will most likely not produce above-normal returns, but if it does, the payoff will be enormous. If you would never invest in such a stock, **please leave everything blank** and simply move on to the next question.*"

Our first question asks what words our survey participants would use to describe a stock that provides insurance against bad states of the world. Our second question elicits words of general excitement tied to extrapolation and overconfidence. Our third question asks what words our survey participants would use to describe the lottery-like features of a stock that will most likely not produce above-normal returns.

Skewness preference is only one aspect of prospect theory. We choose not to account for all four aspects of prospect theory as we believe it is challenging to elicit wordlists tied to prospect theory's other elements: diminishing sensitivity, loss aversion, and reference dependence. Since our analysis captures only one aspect of prospect-theory preferences (albeit a crucial one), we may interpret our results as a downward-biased estimate of prospect theory's explanatory power for the low returns of short-leg securities.[3]

---

[3] Barberis, Jin, and Wang (2021) show analytically that probability weighting and diminishing sensitivity can push model-implied returns in one direction, while loss aversion pushes those returns in the opposite direction. Their theory predicts that the former has an overall larger impact on model-implied returns.

Despite our instructions, some investors provide company names and industries. We delete these terms. We also remove references to investment strategies, such as "value strategy," as they may cause us to mechanically "explain" anomalies. We then select the five most frequently mentioned terms for each question. The five most frequent answers to Q1 are *conservative*, *defensive*, *protection*, *reliable*, and *stable*. The five most frequent answers to Q2 are *competitive*, *expanding*, *leader*, *outperformer*, and *strong*. The five most frequent answers to Q3 are *gamble*, *potential*, *speculative*, *upside*, and *volatile*.

We consider all possible word forms of our base terms, including plural forms, noun forms, verb forms, adjective forms, adverb forms, and verb conjugations.[4] For example, in addition to *stable*, our final list of safety words includes *stableness*, *stablenesses*, *stability*, *stabilities*, and *stably*. We delete word forms that are not meaningfully tied to the business realm. Continuing with the above example, while *stables* is a word form of *stable*, we do not include *stables* in our final wordlists as we do not deem *stables* meaningfully tied to the business realm.

In the end, we arrive at 18 safety words, and we count how often these safety words appear in a text. Similarly, we arrive at 19 exuberance words and 35 lottery words, and we count how heavily investors draw from these words as they explain why they like a particular stock. We provide a complete list of our safety, exuberance, and lottery words in Panel A of Table 1.

3.2.2 Retail Investors' Survey-Based Wordlists

To gauge the sensitivity of our results, we also create wordlists based on an online survey sent to 303 US-based retail investors. We recruit retail investors through Prolific (https://www.prolific.co). Prolific is a platform that enables researchers to recruit prescreened participants for online surveys and experiments. We require that participants be US residents, list English as their first language, and answer "*Yes*" to the following two questions: (1) "*Have you ever made investments (either personally or through your*

---

[4] We use a third-party software package in Python to generate the word forms (https://github.com/gutfeeling/word_forms).

employment) in the common stock or shares of a company?" and (2) "*Have you invested in any of the following types of investment in the past?—Stock market.*"

We display our online survey in Online Appendix Figure A2. Similar to institutional investors, we pose questions regarding a retail investor's age, gender, investment experience, and net investable assets. We also ask how frequently they check their investment accounts or discuss investments with family members, friends, or co-workers. We report the answers to these questions in Online Appendix Table A2.

Most importantly, we ask retail investors the same three questions we posed to institutional investors to generate our wordlists. The five most frequently listed safety terms by retail investors are *reliable*, *safe*, *secure*, *stable*, and *steady*. The five most frequently listed exuberance terms are *consistent*, *excellent*, *growth*, *innovative*, and *winner*. The five most frequently listed lottery terms are *exciting*, *gamble*, *potential*, *speculative*, and *volatile*.

Our final retail investor wordlists comprise all possible word forms of the above base terms. In total, we have 29 safety words, 42 exuberance words, and 40 lottery words. We report all retail investors' survey-based safety, exuberance, and lottery words in Panel B of Table 1.

As we discuss in Section 5.4, we consider variations of our wordlists. We also experiment with self-defined safety and lottery words and, most importantly, exuberance words that are rooted more in theory.

3.3 Stock Opinion Articles

We apply our lists of safety, exuberance, and lottery words to two types of stock opinions: analyst reports and SA stock opinion articles. We focus on reports and articles written about common shares that trade on the NYSE, AMEX, or NASDAQ and that were published between January 2006 and October 2021.

3.3.1 Sell-Side Analyst Reports

Our source of analyst reports is the Investext database. Investext provides research reports from a wide range of brokerages and research firms. We exclude brokerages and research firms that apply algorithms to

16

auto-generate articles (e.g., BuySellSignals Research, Sadif Investment Analytics). We also exclude brokerages and research firms whose articles never provide stock-level recommendations and, instead, focus on risk management and industry analyses (e.g., RiskMetrics Group). After these exclusions, we are left with 664 brokerages. We rank all brokerages by the total number of their reports and process the reports from the top 100 brokerages, which amounts to 98% of all analyst reports. The reason we focus on the top 100 brokerages is that extracting the text in the analyst reports (as well as other relevant information such as the analysts' overall recommendations) involves the labor-intensive task of tracking and recording each brokerage's report template over time, so that we can write the corresponding extraction codes.

Our final sample comprises 1,171,130 analyst reports. Of these, 690,036 represent buy recommendations, 440,719 represent hold recommendations, and 40,375 represent sell recommendations.

We expand contractions (e.g., *couldn't → could not*) and remove digits, punctuations, and special characters. We also remove standardized disclosure sections as well as tables and figures. For each analyst report, we have the report ID, date of publication, analyst name, brokerage name, stock ticker tagged to the article, overall recommendation, title, report text, and number of pages.[5]

For each article, we calculate the number of safety words scaled by the total number of words, *Safety [%]*. Similarly, we calculate the number of exuberance words scaled by the total number of words, *Exuberance [%]*, and the number of lottery words scaled by the total number of words, *Lottery [%]*. We account for simple negation in our calculations. Following Loughran and McDonald (2011), we take simple negation to be the observation of one of twelve words[6] (*no*, *not*, *none*, *neither*, *never*, *nobody*, *few*, *little*, *less*,[7] *low*, *hardly*, *rarely*) occurring within three words preceding a word from our wordlists.

---

[5] We find that in around 5% of cases the ticker that Investext assigns to an analyst report is inaccurate. This issue is particularly prevalent for analyst reports written on stocks that have been delisted as of the date of download. We correct Investext's misassignments in our sample.
[6] We take the first six words from Loughran and McDonald (2011) and add to them the ensuing six words.
[7] We exclude "a few" and "a little." Thus, while "little protection" (e.g., the stock offers little protection) is categorized as a negation of "protection," "a little protection" (e.g., the stock offers a little protection) is not.

In the end, we find that the average *Safety [%]*, *Exuberance [%],* and *Lottery [%]* across all analyst buy recommendations are 0.062%, 0.367%, and 0.246%, respectively; the average report length is 1,121 words.

Figure 1 provides word clouds of the safety, exuberance, and lottery words in our analyst buy recommendations. Examples of the most relevant safety words include *conservative*, *stable*, and *protection*. Examples of the most relevant exuberance words include *strong*, *expansion*, and *leading*. Examples of the most relevant lottery words include *potential*, *upside*, and *volatility*.

To provide readers with a sense of what analyst reports with a high *Safety [%]*, *Exuberance [%],* and *Lottery [%]* look like, we show in Online Appendix Figure A3 the first page of an analyst report with a high *Safety [%]*, the first page of an analyst report with a high *Exuberance [%]* and the first page of an analyst report with a high *Lottery [%]*, respectively. The red boxes in Online Appendix Figure A3 also illustrate which sections of the analyst reports we extract for parsing.

The first report in Online Appendix Figure A3 contains 1,184 words, of which six are safety words. The report has no exuberance words and no lottery words. In this report, the analyst recommends that investors buy RBC Bearings (ROLL), an industrial products company. The key feature driving the analyst's buy recommendation is the company's safe-haven quality: "*ROLL remains one of the highest quality names on our list given the more defensive nature of the company's primary end markets along with its consistent execution, enviable margin profile and net debt negative balance sheet*." The report concludes, "*Overall, we think ROLL is a company investors should want to own, and it becomes a particularly attractive story in times of uncertainty given its defensive nature*."

The second report contains 2,198 words, of which 27 are exuberance words. The article has no safety words and no lottery words. The analyst recommends that investors buy PPG Industries (PPG), a company in the chemicals industry. The key reason for this bullish view is that "*PPG has now delivered 15 straight quarters of record adjusted EPS. Given our belief that PPG will continue to post solid double-digit earnings growth for at least the next few years, we are raising our 2014E EPS to $9.60 (was $9.20) and*

2015E EPS to $11.00 (was $10.40)." The analysts conclude, "*we are reinforcing our BUY rating on shares of PPG and increasing our PT to $220 (was $215) as we have confidence in our estimates going forward.*"

The third report contains 1,948 words, of which 15 are lottery words. The report has no safety words and no exuberance words. The analysts recommend that investors buy Lone Pine Resources (LPR), an oil and gas company. Unlike in the second report, the analysts are not uniformly positive and emphasize the uncertainty that accompanies any investment in LPR. At the same time, the analysts highlight the upside potential and conclude: "*We see near-term headwinds from investor perceptions of high exposure to natural gas as well as from reduced production and increased cost guidance, but we believe the shares will find appeal among more risk-seeking investors as well as investors comfortable with a management team that is still establishing a track record with a newly-independent company.*"

3.3.2 SA Articles

Our second source of stock opinions is SA. We download all articles published in the "stock ideas" section of the SA website (https://seekingalpha.com/stock-ideas). We focus on single-ticker articles that have at least 50 words. For each article, we have the article ID, date of publication, author name, stock ticker tagged to the article,[8] whether the article is tagged as a "long idea" or a "short idea," the title, and the main text.

As with analyst reports, we expand contractions (e.g., *couldn't* → *could not*) and remove digits, punctuations, special characters, tables, and figures. Our sample comprises 140,412 articles. Of these, 72,027 are tagged as long ideas ("buy recommendations"), 9,459 are tagged as short ideas ("sell recommendations"), and 58,926 are untagged.

The average *Safety [%]*, *Exuberance [%],* and *Lottery [%]* across SA long ideas after accounting for simple negation are 0.057%, 0.312%, and 0.205%, respectively; the average article length is 1,213 words.

---

[8] Sometimes, the stock ticker tagged to an article by SA does not match the focal firm's ticker as of the article publication time. For instance, SA assigns the ticker AMZN to the following article—https://seekingalpha.com/article/1171961-double-digit-growth-rates-make-whole-foods-market-a-buy—even though the article is about Whole Foods. The reason for this discrepancy is that Amazon had acquired Whole Foods as of our download time, and the ticker WFM no longer exists. In such cases, we reassign the original ticker to the article so that we can properly merge it with our cross-sectional stock-return predictability data.

Figure 2 displays word clouds of the safety, exuberance, and lottery words in SA long ideas. Examples of the most relevant safety words include *stable*, *steady,* and *safe*. Examples of the most relevant exuberance words include *growth*, *growing*, and *excellent*. Examples of the most relevant lottery words include *potential*, *potentially*, and *volatility*.

Online Appendix Figure A4 displays the beginning paragraphs of a SA article with a high *Safety [%]*, the beginning paragraphs of a SA article with a high *Exuberance [%],* and the beginning paragraphs of a SA article with a high *Lottery [%]*.

The first article in Online Appendix Figure A4 contains 621 words, of which four are safety words. The article has no exuberance words and no lottery words. The author recommends that investors buy Northwestern Corporation (NWE), a utility company. The author argues that "*The stock isn't cheap, but you are paying a fair price in exchange for stability*."

The second article contains 2,155 words, of which 39 are exuberance words. The article has no safety words and no lottery words. The author recommends that investors buy Ansys (ANSS), a computer software company. The main reason for the author's buy recommendation is Ansys's seeming superiority and its growth, which the author projects to continue: "*The company is a best-in-class leader in its niche industry and consistently maintains a double-digit revenue growth rate combined with industry-leading operating margin. I expect from the company to scale its business… That can drive further shareholder value-creation by achieving Target 2020 double-digit organic revenue growth rate, together with maintaining best-in-class operating margins*."

The third article contains 769 words, of which six are lottery words. The article has no safety words and no exuberance words. The author recommends that investors buy Magnum Hunter Resources, an oil and gas producer. Although the author is concerned about Magnum Hunter Resources' long-term future ("*Magnum Hunter's fundamentals remain quite messy due to its large debts and high fixed payment costs*"), he still thinks investors should consider buying as, at "*$0.50, Magnum Hunter appears to offer some potential as a purely speculative play for monetizing its assets.*" The author also notes the upside potential tied to a possible short squeeze.

4. Main Analyses and Results

The above sample reports and articles show that stories rooted in risk, extrapolation, and upside potential all exist in the real world. In this section, we examine which of these three story types appears most pervasively for stocks residing in the short leg.

Chen and Zimmermann (2022) survey the literature and arrive at a list of 205 "*clear and likely predictors*" of *raw* stock returns in the US. For each of these predictors, the authors form long-short portfolios that they rebalance each month. The authors then test the null hypothesis that the mean monthly long-short portfolio return is zero. The authors reject the null hypothesis for all but three predictors. The authors make all 205 firm characteristics except for price, size, and past one-month returns available for download on their website https://www.openassetpricing.com/data.[9] The dataset contains, for each PERMNO and year-month, the corresponding firm characteristic signed such that, based on prior literature, a higher value predicts higher returns. We first download the dataset. We then reinsert price, size, and past one-month returns, and restrict our analysis to common shares that trade on the NYSE, AMEX, or NASDAQ from January 2006 through October 2021.

We adopt the following procedure separately for each of the 205 firm-level characteristics: For each month t, we rank stocks based on firm characteristic i. The stocks in the bottom decile represent the short-leg securities. We hereafter refer to reports and articles recommending that investors buy these short-leg securities as "short-leg recommendations." We refer to all other buy recommendations as "other recommendations."

Prior literature suggests that short-leg securities trade at comparatively high prices and correspondingly earn low future returns. The comparatively high prices must arise because there are investors who hold bullish views of the respective stocks. To understand the source of these bullish views, we parse all the bullish views of the short-leg securities and test whether these short-leg recommendations unusually fixate on the stocks' safe-haven quality, their seeming superiority, or their upside potential.

---

[9] This information pertains to their "April 2021 Data Release," which includes data until June 2022 (portfolios are formed and firm characteristics are assigned to stocks annually).

To this end, we compute the average *Safety [%]* across analysts' short-leg recommendations. To assess whether the use of safety words in the short-leg recommendations is abnormally high, we also compute the average *Safety [%]* across analysts' other recommendations. We then calculate the difference between the former and the latter (on a relative basis) and test whether the difference is positive and statistically significant at the 5% level:

$$\Delta Safety[\%] = \frac{\overline{Safety[\%]}_{short-leg\ recommendations} - \overline{Safety[\%]}_{other\ recommendations}}{\overline{Safety[\%]}_{other\ recommendations}} \tag{1}$$

Similarly, we compute the average *Exuberance [%]* and *Lottery [%]* across analysts' short-leg recommendations and the average *Exuberance [%]* and *Lottery [%]* across analysts' other recommendations. We then test whether Δ*Exuberance [%]* and Δ*Lottery [%]* are positive and statistically significant at the 5% level. We repeat the process for SA articles.

For some firm characteristics, there are no analyst or SA-short-leg recommendations. Our final analysis thus comprises 181 firm characteristics.

Suppose that, for firm characteristic *i*, the difference in the fraction of safety words was positive and significant. In other words, suppose that in explaining why investors like stocks on the "short-spectrum" of a firm characteristic, they unusually frequently draw from safety words. In that case, we would label investors' rationales for liking these stocks as "consistent with the risk framework." Similarly, suppose that the average *Exuberance [%]* or the average *Lottery [%]* across the short-leg recommendations were abnormally high. In those cases, we would label investors' rationales for liking the short-leg securities as "consistent with the irrational beliefs framework" or as "consistent with the non-traditional preferences framework," respectively.

For some firm characteristics *i*, we observe abnormally high uses of words from more than one wordlist. Investors' rationale for liking these short-leg securities is thus consistent with more than one

framework. In these cases, we record which difference is the largest, economically speaking. We refer to the framework with the largest difference as the "*most* consistent" framework.[10]

Finally, for some firm characteristics i, we detect no statistically significant difference in the occurrence of either safety, exuberance, or lottery words. We label such results "inconclusive."

In Table 2, we report our findings aggregated across all 181 firm characteristics for the institutional investor wordlists (Panel A) and the retail investor wordlists (Panel B).

We first describe our Panel A results. When we apply our institutional investor wordlists and parse analysts' explanations of why they like stocks that reside in the short leg of a specific anomaly, we find that in 16 out of the 181 cases, or 9% of the time (= 16 / 181), the short-leg recommendations use significantly more safety words. That is, the short-leg recommendations unusually frequently tell a safe-haven story.

For 53 out of the 181 cases, or 29% of the time, the short-leg recommendations are marked by an abnormally high use of exuberance words. That is, analysts unusually frequently tell stories of continuous growth or positive developments that they are certain will come to fruition.

For 124 out of the 181 cases, or 69% of the time, the short-leg recommendations use abnormally many lottery words. That is, the short-leg recommendations stand out in their unusually heavy emphasis on the stocks' upside potential.

In a nutshell, we find that analysts' rationales for liking short-leg securities are consistent with the risk framework 9% of the time, the irrational beliefs framework 29% of the time, and the non-traditional preferences framework 69% of the time.

The fractions of times analysts' rationales are *most* consistent with the risk, irrational beliefs, or the non-traditional preferences frameworks are 7%, 17%, and 58%, respectively. The remaining 19% of the time, the results are inconclusive.

The patterns are similar for SA articles. Investors' rationales for liking short-leg securities are consistent with the risk framework 6% of the time, the irrational beliefs framework 15% of the time, and

---

[10] If investors' rationale for liking a particular set of short-leg securities is consistent with one framework only, that framework naturally is also the most consistent framework.

the non-traditional preferences framework 58% of the time. The fractions of times investors' rationales are *most* consistent with the risk, irrational beliefs, or non-traditional preferences frameworks are 6%, 10%, and 54%, respectively. The remaining 29% of the time, the results are inconclusive.

Panel B of Table 2 presents the results based on the retail investor wordlists. The results are similar to those in Panel A. For analyst reports, the fractions of times that investors' rationales are most consistent with the risk, irrational beliefs, and non-traditional preferences frameworks are 6%, 14%, and 66%, respectively. For SA articles, the corresponding fractions are 8%, 13%, and 57%, respectively.

Overall, our results in Table 2 suggest that all three frameworks can explain components of the cross-section of expected stock returns. Since non-traditional preferences offer the most consistent explanation the majority of the time—irrespective of whether we consider the institutional investor wordlists, the retail investor wordlists, analyst reports, or SA articles—it appears that non-traditional preferences play a dominant role.

4.1 List of "Anomalies" Consistent with the Risk Framework

We next discuss our findings at the firm-characteristic level. Online Appendix Table A3 reports the findings separately for each of the 181 firm characteristics.[11] Since Online Appendix Table A3 can be overbearing, in Table 3, we also report the findings separately for the 22 anomalies that Barberis, Jin, and Wang (2021) consider in their test of whether prospect theory can explain stock market anomalies.[12] To keep the table sizes manageable, we present only the findings based on the institutional investors' survey-based wordlists. The results are similar for the retail investors' survey-based wordlists (results available upon request).

Our results show that when comparing the analyst buy recommendations for stocks with low operating leverage with those for stocks that do not have low operating leverage (Novy-Marx, 2011), the fraction of safety words is 53% higher for stocks with low operating leverage (*t*-statistic = 39.76). Among

---

[11] In Online Appendix Figure A5 we also plot our findings separately for each firm characteristic.
[12] There are differences in how Barberis, Jin and Wang (2021) and Chen and Zimmerman (2022) name and construct some their anomaly variables. For each of the 23 anomalies that Barberis et al. consider, we select the firm characteristic out of the 181 characteristics that Chen and Zimmerman consider with the best match. For one anomaly that Barberis et al. consider ("Capital Gains Overhang"), we were unable to find a match.

SA long ideas, the fraction is 27% higher (*t*-statistic = 5.90). We observe no reliable differences in the use of exuberance or lottery words. The abnormally high use of safety words in the buy recommendations for stocks with low operating leverage suggests that investors like the safety they see in these stocks. Coupled with investors' risk aversion, this view may explain why these stocks trade at comparatively high prices and earn relatively low returns on average.

We also observe abnormally high uses of safety words for stocks with other measures of low leverage and high debt capacity (Fama and French, 1992; Hahn and Lee, 2009), as well as stocks with low market beta (Fama and MacBeth, 1973), suggesting that investors view these stocks as safe-haven assets.

4.2 List of "Anomalies" Consistent with the Irrational Beliefs Framework

We observe abnormally high uses of exuberance words in the buy recommendations for stocks with high returns over the past three years (De Bondt and Thaler, 1985), high intangible returns (Daniel and Titman, 2006), high R&D (Chan, Lakonishok, and Sougiannis, 2001), high valuation ratios (Basu, 1977; Loughran and Wellman, 2011), high price levels (Blume and Husic, 1973), large market capitalizations (Banz, 1981), and high trading volumes (Brennan, Chordia, and Subrahmanyam, 1998).

The unusually heavy reliance on exuberance words suggests that investors frequently get overly excited about stocks with the above characteristics. Coupled with short-sale constraints, this view may explain why these stocks earn such poor returns going forward.

4.3 List of "Anomalies" Consistent with the Non-Traditional Preferences Framework

Finally, we observe abnormally high uses of lottery words in the buy recommendations for stocks with high forms of volatility or analyst forecast dispersion (Diether, Malloy, and Scherbina, 2002; Ali, Hwang, and Trombley, 2003; Ang, Hodrick, Xing, and Zhang, 2006), high forms of skewness or co-skewness (Harvey and Siddique, 2000; Ang, Chen, and Xing, 2006; Bali, Cakici, and Whitelaw, 2011; Bali, Engle, and Murray, 2016), high growth in assets and financing (e.g., Richardson, Sloan, Soliman, and Tuna, 2005; Daniel and Titman, 2006; Soliman, 2008; Cooper, Gulen, and Schill, 2008; Pontiff and Woodgate, 2008),

high accruals (e.g., Sloan, 1996; Xie, 2001; Richardson, Sloan, Soliman, and Tuna, 2005), recently listed stocks (Ritter, 1991), stocks with low forms of profitability (Haugen and Baker, 1996; Fama and French, 2006; Novy-Marx, 2013), stocks with high failure probability and high short interest (e.g., Dichev, 1998; Dechow, Hutton, Meulbroek, and Sloan, 2001) and stocks with poor recent returns (e.g., Jegadeesh, 1990; Jegadeesh and Titman, 1993; Novy-Marx, 2012).

The unusually heavy use of lottery words in the buy recommendations for the above types of stocks suggests that many investors particularly like the upside they see in these stocks. Coupled with prospect theory preferences, this view may explain why these stocks trade at comparatively high prices and earn low returns on average.

In general, our results suggest that analysts and SA contributors tend to see upside potential in the same types of stocks. Put differently, when analysts use unusually many lottery words as they describe why they like a particular type of stock, SA contributors also make abnormally heavy use of lottery words as they describe the same type of stock. The correlation of the abnormal fractions of lottery words across the 181 firm characteristics between analyst reports and SA articles is 0.74.

Before discussing the sensitivity of our results, we point to two features of our results on non-traditional preferences that we believe are particularly noteworthy. First, our results suggest that non-traditional preferences can explain not only anomalies in which the short-leg securities have been performing well but also anomalies in which the short-leg securities have been performing poorly. Specifically, we detect abnormally high uses of lottery words in analysts' and SA contributors' justifications for why they like stocks with low forms of profitability, stocks with high failure probability, stocks with high short interest, and stocks with poor recent returns.[13]

---

[13] On a related note, in additional analyses, we repeat our main test but exclude recommendations if the corresponding stock's performance over the past month is in the top 10% of its cross-sectional distribution as of the recommendation issuance date. We also experiment with excluding stocks that are in the top 20% and the top 50% of their distributions. The results reported in Online Appendix Table A4 show that we arrive at the same conclusion even as we remove high performers and restrict our analysis to stocks for which there is little ground for exuberance.

These results are consistent with our proposition that our lottery words capture one specific feature about a stock that appeals to investors rather than mirror general investor exuberance and overly optimistic beliefs.

The second feature we think is worth expounding upon is that our findings can help explain a major blemish of prospect theory (Barberis, Jin, and Wang, 2021). In the data, value stocks have more positively skewed returns. Prospect theory, therefore, predicts lower average returns on value stocks. In reality, the opposite holds: value stocks earn higher returns on average.[14]

Table 3 shows that the fraction of lottery words is reliably higher for growth stocks. In other words, it appears that investors (incorrectly) believe that growth stocks have more positively skewed returns. The fact that growth stocks earn relatively low returns and that, correspondingly, value stocks earn relatively high returns thus need not be inconsistent with prospect theory. It could just be that investors have the wrong perception of which stocks offer greater upside potential. Unlike in the irrational beliefs framework, which emphasizes generally overly optimistic beliefs due to investor extrapolation or overconfidence, here, the mistaken belief comes from investors' misunderstanding of the tails of a stock's return distribution.

We make the same observation for other seeming failures of prospect theory. Based on historical data, Barberis, Jin, and Wang (2021) note that prospect theory does a poor job of explaining the accruals, asset growth, size, and short-term reversal anomalies. Under prospect theory, stocks earning lower average returns should have more positively skewed return distributions. Yet, while stocks with high accruals, asset growth, market capitalization, and short-term performance earn comparatively low returns, they do not have more positively skewed returns.

The results reported in Table 3 and Online Appendix Table A3 indicate that the fraction of lottery words is reliably higher for stocks with high accounting accruals, asset growth, and past-one-month stock returns. Again, it appears that investors (incorrectly) believe that these stocks have more positively skewed returns, potentially explaining why these stocks earn such low returns on average.

---

[14] In the past decade, growth stocks outperformed value stocks (Arnott, Harvey, Kalesnika, and Linnainmaa, 2021).

The exception is the size anomaly. The fraction of lottery words is higher for small stocks, suggesting that investors (correctly) perceive small stocks to have greater upside potential. Like Barberis, Jing, and Wang (2021), we thus arrive at the conclusion that prospect theory cannot explain why small stocks earn higher returns on average.

5. Sensitivity Analyses

We conduct a series of tests to gauge the robustness of our results.

5.1 Subsample Analyses

We first assess whether our results vary across subsets of our data. Again, we report only the results based on the institutional investors' survey-based wordlists to keep the table sizes manageable. The results based on retail investors' survey-based wordlists are very similar and available upon request.

Some cross-sectional stock-return predictabilities have received more attention than others. For instance, as of June 2022, Alwathainani's (2009) study, which documents the predictability of "earnings consistency," has received 39 Google Scholar citations. In comparison, Fama and French (1992), which discusses the predictability of the book-to-market ratio, has received 24,946 Google Scholar citations.

To gauge whether non-traditional preferences offer the dominant explanation among the most widely studied anomalies, we repeat our analysis for firm characteristics for which the corresponding papers' Google Scholar citations are in the top quartile of the distribution as of June 2022 (> 1,750 citations).

Our results echo our earlier results that non-traditional preferences offer the most comprehensive explanation. The results reported in Panel A of Table 4 indicate that, when considering analyst reports, investors' rationales for liking short-leg securities are congruent with the risk framework 17% of the time, the irrational beliefs framework 35% of the time, and the non-traditional preferences framework 70% of the time. The corresponding percentages for SA articles are 7%, 20%, and 65%, respectively.

In another test, we restrict our analysis to analyst reports and SA articles published only after the corresponding cross-sectional stock-return predictability has been documented in an academic study. For

instance, to understand investors' liking of stocks with high asset growth, we consider only analyst reports and SA articles published since January 1, 2009, after Cooper, Gulen, and Schill's (2008) documentation that stocks with high asset growth earn unusually low future returns.

Non-traditional preferences continue to offer the most comprehensive explanation. The results reported in Panel B of Table 4 indicate that, when considering analyst reports, investors' rationales for liking short-leg securities are congruent with the risk framework 10% of the time, the irrational beliefs framework 29% of the time, and the non-traditional preferences framework 67% of the time. The corresponding percentages for SA articles are 6%, 15%, and 56%, respectively.

Finally, we repeat our analysis for the firm characteristics for which the corresponding paper's publication year is in the bottom quartile of the distribution, specifically papers published in the year 2001 or earlier. With the growing attention paid to prospect theory, researchers may have been data-mining for predictors that are consistent with prospect theory. If that is the case, our results should be weaker for "older" predictors. We find the opposite to be the case. The explanatory power of prospect theory is even stronger among the older cross-sectional stock-return predictors. The results reported in Panel C of Table 4 show that, when considering analyst reports, investors' rationales for liking short-leg securities are congruent with the risk framework 11% of the time, the irrational beliefs framework 32% of the time, and the non-traditional preferences framework 74% of the time. The corresponding percentages for SA articles are 6%, 17%, and 66%, respectively.

5.2 Analyses Based on Alternative Denominator and Numerator Only

Our main analysis draws inferences based on equation (1), which computes the differences in the fractions of safety, exuberance, and lottery words on a relative basis. One concern with our relative measure is that any strong positive difference may be driven by an unusually small denominator rather than an unusually heavy reliance on safety, exuberance, or lottery words.

For instance, suppose that, for firm characteristic i, *Safety[%]*$_{short\text{-}leg\ recommendations}$ was 1% and *Safety[%]*$_{other\ recommendations}$ was 0.1% ($\rightarrow \Delta$Safety[%] = 900%), while for firm characteristic j, *Safety[%]*$_{short\text{-}}$

leg recommendations was 4% and *Safety[%]*_{other recommendations} was 1% (→ ΔSafety[%] = 300%). Based on equation (1), we would infer that the use of safety words is more unusual for firm characteristic *i* than for firm characteristic *j* while one could argue for the reverse.

To account for this possibility, we conduct our main analyses with two alternative measures, both of which tilt the balance towards firm characteristic *j*:

$$\Delta Safety[\%] = \frac{\overline{Safety[\%]_{short-leg\ recommendations}} - \overline{Safety[\%]_{other\ recommendations}}}{\overline{Safety[\%]_{short-leg\ recommendations}} + \overline{Safety[\%]_{other\ recommendations}}} \qquad (2)$$

$$\Delta Safety[\%] = \overline{Safety[\%]_{short-leg\ recommendations}} - \overline{Safety[\%]_{other\ recommendations}} \qquad (3)$$

We report the corresponding results in Panels D and E of Table 4. The results are very similar to those reported in Panel A of Table 2 and suggest that non-traditional preferences play a dominant role irrespective of whether we consider differences on a relative basis or an absolute basis.

5.3 Sell Recommendations of Long-Leg Securities

Our primary analysis focuses on the buy recommendations for short-leg securities. One possible alternative is to also consider the sell recommendations for long-leg securities to better understand why long-leg securities appear comparatively underpriced and earn somewhat higher average returns.

We focus on the buy recommendations for short-leg securities because the unusually low returns of short-leg securities constitute a more important component of the cross-section of expected stock returns (e.g., Stambaugh, Yu, and Yuan, 2012). This imbalance accords with theory. Probability weighting in prospect theory predicts overpricing and unusually low returns for stocks with positively skewed returns. Probability weighting does not predict underpricing and abnormally high returns unless we set the model parameters at unrealistic values.[15] Similarly, as long as there are investors with overly optimistic beliefs and frictions, specifically short-sale constraints, which keep the pessimistic investors out of the market, we will observe overpricing and unusually low future returns. There is no equivalent friction that would keep

---

[15] Stocks rarely have highly negatively skewed return distributions.

optimistic investors out of the market, and that would enable a subset of investors with overly pessimistic views to generate underpricing and abnormally high future returns.

Another reason for our emphasis on the buy recommendations is that there are many more buy recommendations than sell recommendations, which greatly increases the power of our analysis. Sell recommendations are particularly rare in SA.

Still, in additional tests, we consider analyst sell- and hold recommendations as well as SA short ideas written on long-leg securities, and we search for the primary reason investors do not like these stocks even though they deliver moderately higher returns on average. Is the most commonly noted shortcoming the high level of risk, adverse recent events, or the perceived lack of upside potential?

To find the most commonly noted shortcoming, we compute (1) the fraction of negated safety words (e.g., *little – protection*), (2) the fraction of negated exuberance words (e.g., *not – competitive*), and (3) the fraction of negated lottery words (e.g., *no – potential*). To assess whether the occurrences of negated words in the "long-leg recommendations" are abnormally high, we also compute the fractions of negated words in the "non-long-leg recommendations" and test whether the relative difference between the former and the latter is positive and statistically significant at the 5% level.

We report the results in Online Appendix Table A5. For analyst reports, the fractions of times investors' rationales are consistent with the risk framework, the irrational beliefs framework, and the non-traditional preferences framework are 3%, 7%, and 10%, respectively. Thus, while it appears that non-traditional preferences play a more critical role than risk considerations and irrational beliefs, the explanatory power is much weaker for the long side than for the short side: 88% of the time, we detect no statistically significant differences in the occurrences of negated safety words, negated exuberance words, or negated lottery words. Our tests for the long side are even less powerful for SA articles. The low power may not surprise, given that there are only a total of 9,459 SA short ideas in our sample. The corresponding fractions are 1% (risk framework), 0% (irrational beliefs framework), 0% (non-traditional preferences framework), and 99% (inconclusive), respectively.

5.4 Alternate Wordlists

Our findings thus far show that we make similar observations irrespective of whether we apply the institutional investor wordlists or the retail investor wordlists. We conduct the following two final sensitivity analyses.

5.4.1 Wordlist Iterations

In our first exercise, we examine, for each of the five most frequently used safety words, whether its removal changes our conclusion. We repeat this exercise for each of the five most frequently used exuberance words and each of the five most frequently used lottery words. We thus consider 15 variations of our wordlists, and we plot the fractions pertinent to each of the 15 variations in Online Appendix Figure A6. We focus on the top five words as they are most likely to alter our findings.

In short, the results presented in Online Appendix Figure A6 are similar to those reported in Table 2. Across all variations, investors' rationales are congruent with the non-traditional preferences framework the vast majority of the time. On a related note, only two of our five lottery terms (*gamble*, *potential*, *speculative*, *upside*, and *volatile*) are "positive:" *potential* and *upside*. Online Appendix Figure A6 shows that our results are similar when we do not consider *potential* or *upside*,[16] suggesting that the use of lottery words is related more to non-traditional preferences than to general investor exuberance tied to overly optimistic beliefs.

5.4.2 Self-Defined Wordlists

In our second exercise, we follow Loughran and McDonald (2011), who self-define what constitutes a positive and a negative word, and we create our own lists of safety, exuberance, and lottery words.

---

[16] Our results are also robust to the removal of both terms, *potential* and *upside*. The results are available upon request.

In particular, we conjecture that investors describe stocks providing insurance against bad states of the world in the following five safety terms: *certain*, *low risk*, *predictable*, *safe*, and *stable*. We consider all possible word forms of the five safety terms and account for negation.

While traditional finance theory suggests that it is the *covariance* of a stock's payoff with bad states of the world that matters, we believe that it is unlikely that sell-side analysts and SA contributors would describe stocks in covariance terms even if they behaved according to the traditional finance paradigm. Consistent with this belief, we find that the terms *covariance* and *economic disaster* appear only 107 and 14 times, respectively, across our 1.17 million analyst reports comprising roughly 1.4 billion words. Across our 140,420 SA articles comprising more than 150 million words, the two terms appear only 49 and 64 times, respectively.

Behavioral finance theory proposes that general investor exuberance is rooted in (1) investors' extrapolative tendencies and (2) overconfidence in the precision of their private signals. To capture investors' extrapolative tendencies regarding positive events, we proceed as follows: We extract all sequences of four words in a text.[17] We then examine for each 4-gram whether any of five continuation terms appear jointly with any of five growth terms. The continuation terms are *carry on*, *continue*, *extend*, *go on*, and *keep on*. The growth terms are *excel*, *expand*, *grow*, *outperform*, and *rise*. If we observe a continuation word appearing jointly with a growth word in a 4-gram (e.g., *continue – grow*), we mark the two words as exuberance words. As before, we consider all possible word forms that are meaningfully tied to the business realm and account for negation.

To capture possibly overconfident bullish beliefs, we consider all 4-grams in a text and examine whether investors use a strong modal word in conjunction with a positive word. We use the lists of strong modal words and positive words of Loughran and McDonald (2011). If we observe a strong modal word

---

[17] In this task, we remove stop words (excluding any of our eight negation words) in order to retain more meaningful words within 4-grams. For instance, "*is likely to continue the pattern of strong growth*" is analyzed as "*continue pattern strong growth*." We do not exclude stop words in our main tests or when calculating the word length of each article.

jointly appearing with a positive word in a 4-gram (e.g., *definitely – achieve*), we tag the two words as exuberance words. We account for negation.

Finally, to see whether investors emphasize the lottery-like aspects of a stock, we search for the following five lottery terms: *bet*, *gamble*, *potential*, *take a chance*, and *upside*. We again consider all possible word forms tied to the business realm and account for negation.

The advantage of our self-defined wordlists over our survey-based wordlists is that the construction of the exuberance words is rooted more in theory. The disadvantage is that self-defined wordlists insert the researcher more closely into the data-generating process.

We report the results in Table 5. For analyst reports, the fractions of times investors' rationales are consistent with the risk framework, the irrational beliefs framework, and the non-traditional preferences framework are 16%, 17%, and 66%, respectively. For SA articles, the corresponding fractions are 13%, 15%, and 55%, respectively. The results are very similar when we combine the self-defined exuberance words with the survey-based safety and lottery words (results available upon request).

Again, our results suggest that all three frameworks can explain components of the cross-section of expected stock returns. But, again, it appears that non-traditional preferences play a material role.

6.   Methodological Considerations

As alluded to in the introduction, our conclusion that non-traditional investor preferences play a dominant role in explaining the low returns of short-leg securities hinges on two assumptions: (1) Our wordlists and dictionary-based approach capture beliefs and preferences. (2) In addition, the beliefs and preferences that we capture in analyst reports and SA articles at least partially reflect those of the investor population. In this section, we discuss the validity of these two assumptions.

6.1 Can Our Method Capture Beliefs and Preferences?

Our ability to say anything regarding the low returns of short-leg securities depends on the use of safety words being tied to risk aversion, the use of exuberance words being tied to overly optimistic beliefs, and the use of lottery words being tied to non-traditional preferences.

A prominent view in the irrational-beliefs-based models is that biased cash-flow forecasts – in particular, long-term earnings growth forecasts – drive stock price fluctuations (Barberis, 2018). Bordalo, Gennaioli, LaPorta, and Shleifer (2019, 2022) find that analysts' long-term earnings growth forecasts are strongly positively related to past earnings growth. The authors also find that when the long-term earnings growth forecasts are particularly high, these forecasts tend to be much too high in hindsight.

To evaluate whether the use of exuberance words associates with overly optimistic beliefs, we consider whether the use of exuberance words positively correlates with analysts' long-term earnings growth forecasts. We estimate a regression of analysts' long-term earnings growth forecasts on the tone and the fractions of safety, exuberance, and lottery words in the corresponding written reports. We include combinations of firm-, year- and analyst fixed effects and cluster our standard errors by time.

As reported in Table 6, we find that the analysts with the highest long-term earnings growth forecasts use substantially more exuberance words in their reports, suggesting that the use of exuberance words is indeed tied to overly optimistic beliefs.

We observe no associations between analysts' long-term earnings growth forecasts and the use of safety or lottery words. The lack of associations indicates that lottery words are tied more to non-traditional preferences than to biased beliefs.

6.2 Do Beliefs and Preferences in Analyst Reports and SA Articles Reflect Those of the Investor Population?

The evidence in the previous subsection points to the possibility that our wordlists and dictionary-based approach can capture differences in beliefs and preferences. In this subsection, we gauge whether the beliefs and preferences captured in analyst reports and SA articles may also tell us something about the beliefs and preferences of the investor population.

6.2.1 Analyst and SA Contributor Incentives

Sell-side analysts and SA contributors must choose what story to tell as they explain their buy recommendations. For instance, in explaining their buy recommendation for a high-volatility stock, analysts and SA contributors must choose whether to tell a safe-haven story, a continued-growth story, or an upside-potential story.

An extensive literature documents that readers have confirmation bias and prefer consuming news that matches their belief system (Nickerson, 1998). Park, Konana, Gu, Kumar, and Raghunathan (2013) and Cookson, Engelberg, and Mullins (2023) show that confirmation bias' reach extends to investors.

If analysts and SA contributors want their reports to be read and if investors exhibit confirmation bias, it is sensible for analysts and contributors to cater to their readers' belief systems. Thus, when analysts and SA contributors choose to tell an upside-potential story for high-volatility stocks, it is conceivable that they do so because they feel that an upside-potential story most strongly agrees with investors' belief system regarding high-volatility stocks. In that regard, analysts' and SA contributors' story choices should mirror investors' beliefs and preferences even in, or precisely because of the presence of analysts' and SA contributors' incentives.

6.2.2 Sensibility of Results

One way to empirically gauge whether the beliefs captured in analyst reports and SA articles at least partially reflect those of the investor population is to gauge the sensibility of our results.[18]

As discussed in Sections 4.1 through 4.3, our method points to (1) the risk framework as the most likely explanation for the unusually low returns of stocks with low leverage, high debt capacity, and low

---

[18] Another tangential, yet related way to gauge whether the beliefs captured in reports at least partially reflect those of the investor population is to correlate the tone of analyst reports and SA articles with order imbalances and abnormal stock returns. We detail our method and tabulate our results in Online Appendix Table A6. In short, we find that the tone in analyst reports and SA articles strongly positively correlates with order imbalances and abnormal stock returns for the corresponding stock on the publication day. We note that analyst reports and SA articles are generally not written and published on the same day: analyst reports need to be vetted and formatted; SA articles generally need to be submitted for review by an editorial team. This time gap mitigates concerns that the positive association reflects company news announced on day t, which moves trading and stock prices on day t and, at the same time, is reported in analyst reports and SA articles published on day t.

market beta; (2) the irrational beliefs framework as the most likely explanation for the unusually low returns of stocks with high returns over the past three years, high valuation ratios, high price levels, and large market capitalizations, among others; and (3) the non-traditional preferences framework as the most likely explanation for the unusually low returns of stocks with high forms of volatility, high skewness, recently listed stocks, stocks with high failure probability and stocks with high short interest, among others.

These findings appear sensible as there are strong theoretical arguments for why stocks with low leverage, high debt capacity, and low market beta should be considered safer and, thus, earn lower returns on average. Similarly, it is theoretically defensible that stocks with high past returns, high valuation ratios, and high price levels are particularly susceptible to investor exuberance (Barberis, 2018). Finally, stocks with high volatility, high skewness, recently listed stocks, stocks with high failure probability, and stocks with high short interest all exhibit positively skewed return distributions.

Our findings also agree with prior empirical studies. For instance, Bordalo, Gennaioli, LaPorta, and Shleifer (2019) provide strong evidence that stocks with high returns over the past three years subsequently earn low returns because investors form overly optimistic beliefs and create overpricing.

To systematically gauge the sensibility of our results, we conduct two final tests. Investors likely see more upside potential in small stocks than in large stocks. In contrast, extrapolative tendencies are likely stronger among larger companies that have been growing steadily (Barberis, 2018). We should thus expect the explanatory power of the irrational beliefs framework to be stronger when short-leg securities have comparatively *large* market capitalizations, while the explanatory power of the non-traditional preferences framework should be stronger when the short-leg securities have comparatively *small* market capitalizations.

To test these conjectures, we compute, separately for each of the 181 firm characteristics, the average market capitalization of the short-leg securities for each month. We then compute the time-series means. In Panel A of Table 7, we report our results aggregated across the 91 firm characteristics whose average market capitalization is above the median ("More Likely to be Irrational Beliefs Based/Less Likely to be Non-Traditional Preferences Based"). In Panel B, we report our results aggregated across the 90 firm

characteristics whose average market capitalization is below the median ("Less Likely to be Irrational Beliefs Based/More Likely to be Non-Traditional Preferences Based").

In line with expectations, the irrational beliefs framework has greater explanatory power when the short-leg securities are comparatively large. Consider the analyst-report results. When the short-leg securities are comparatively large, the writing points to exuberance 47% of the time compared with 11% of the time when short-leg securities are comparatively small. We observe the opposite pattern for a stock's upside potential. When the short-leg securities are comparatively large, the buying propositions are congruent with non-traditional preferences 55% of the time compared with 82% of the time when short-leg securities are comparatively small. The differences are even larger when considering SA articles.

We also conduct the following test. Barberis, Jin, and Wang (2021) construct a model with prospect theory investors and test whether their model-implied returns can match the returns observed in the real world. The authors consider 23 anomalies. The authors' model does better for some anomalies than for others. In particular, the authors compute for each anomaly what portion of its long-short performance accrues during earnings announcements. The authors argue that a high portion suggests that the corresponding anomaly is driven by biased beliefs, which become partially corrected around the earnings announcement. Consistent with this argument, the authors find that their model-implied returns do not align closely with real-world data for anomalies with high portions.

Barberis, Jin, and Wang (2021) report their portions for each of the 23 anomalies in their Online Appendix A.III. We borrow their portions and test whether they correlate positively with whether our method classifies an anomaly as consistent with the irrational beliefs framework. We also test whether the portions correlate negatively with whether our method classifies an anomaly as consistent with the non-traditional preferences framework.

We find that the correlations are 0.34 and -0.28, respectively. In other words, our method is more likely to assign anomalies, which accrue more of their long-short performance during earnings announcements, to the irrational beliefs framework. Our method is less likely to assign them to the non-

traditional preferences framework. When extending the analysis to all 181 anomalies, the correlations turn to 0.46 and -0.39, respectively.

7. Limitations

7.1 Investor Underreaction

While we believe that our method and results can help us better understand an important component of the cross-section of expected stock returns, we also acknowledge that, in addition to the above methodological caveats, our paper suffers from two important shortcomings. The first shortcoming is that we do not test the relevance of a fourth prominent framework, the underreaction framework. The underreaction framework builds either on conservatism or on the finite processing capacities of our brains. Conservatism reflects the phenomenon whereby people cling to their prior beliefs and are overly skeptical of new information, causing them to insufficiently update their prior beliefs when new signals arrive (Barberis, Shleifer, and Vishny, 1998). The finite processing capacities view suggests that the daily volume of new information is too large for investors to handle. Value-relevant information, therefore, does not become fully impounded into stock prices (Hong and Stein, 1999; Hong, Lim, and Stein, 2000).

Our research design is not well suited to capturing underreaction due to investors' finite processing capacities, and we cannot think of a wordlist that would cleanly capture underreaction due to conservatism.

Our conclusion that almost two-thirds of cross-sectional stock-return predictabilities are consistent with non-traditional preferences would not change if we considered the relevance of the investor underreaction framework. However, whether non-traditional preferences would remain the *most* consistent framework the majority of the time once we include and compare with the underreaction framework is unclear.

7.2 Social Finance

A second and related shortcoming is that we do not test the relevance of an emerging fifth framework, the social finance framework.

The finance field increasingly recognizes that investors turn to each other for investment advice and that these social interactions can impact investment decisions and potentially alter asset prices (Hirshleifer, 2020; Han, Hirshleifer, and Walden, 2022; Hirshleifer, Peng, and Wang, 2023; Hwang, 2022).

To illustrate this phenomenon, suppose that investors exhibit a systematic preference for discussing stocks with specific features, such as high volatility, because they make for more interesting conversations. Further, suppose that investors tend to purchase stocks that enter their radar (Barber and Odean, 2008) and that there are short-sale constraints. Under these conditions, social interactions and investors' synchronous purchases of high-volatility stocks will generate overpricing and unusually low returns in the long run among these stocks.

The above line of argument can be extended to other cross-sectional determinants of average stock returns. Can systematic sharing preferences explain why stocks with extreme returns earn unusually low returns (Bali, Cakici, and Whitelaw, 2011)? Can systematic sharing preferences help explain why growth stocks earn lower returns than value stocks (Fama and French, 1992; 2015)?

Our research design is not well suited to capturing investors' sharing preferences. We are thus unable to test whether investor conversations can help explain the cross-section of expected stock returns. As before, our conclusion that most predictabilities are consistent with non-traditional preferences would not change if we considered social interaction. However, whether non-traditional preferences would remain the *most* consistent framework the majority of the time is unclear.

There is much current work examining the degree to which investor social interactions can impact asset pricing. The findings in our paper provide some guidance for future research on this question. Our study sheds light on which story types investors rely on primarily as they converse about stocks that appear overpriced. Our study, therefore, offers insights into the channels through which stocks can become viral and overpriced.

8. Conclusion

The question of what drives the cross-section of expected stock returns lies at the heart of asset pricing and has motivated a significant body of research (e.g., Fama and French, 1992; Davis, Fama, and French, 2000; Chen and Zimmermann, 2022). Our paper revisits this classic question with a new approach. We parse investors' buy recommendations and record their primary reasoning for liking short-leg securities. We then test whether investors' reasoning is most congruent with a particular theory. We find that sell-side analysts and SA contributors like short-leg securities primarily for their upside potential, suggesting that prospect theory is an important component of the cross-section of expected stock returns.

Our analysis also shows that the stocks that investors perceive as lottery-like are often not lottery-like in the data. The rejection of a particular theory in an empirical test, therefore, does not imply the failure of the theory per se. Instead, it could reflect inaccuracies in how we measure the sometimes incorrect perceptions of real-world investors. Our paper introduces a new method for capturing real-world investors' perceptions. On the one hand, our method is not as precise as investor surveys. On the other hand, it is cheap and scalable and may thus represent an interesting alternative for future research. In this paper, we apply a dictionary-based textual analysis approach. We made this choice for transparency. Future research should consider less transparent but likely more powerful machine-learning-based approaches such as Latent Dirichlet Allocation or Bidirectional Encoder Representations from Transformers.

References

Ali, A., Hwang, L. S., & Trombley, M. A. (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, *69*(2), 355-373.

Alwathainani, A. M. (2009). Consistency of firms' past financial performance measures and future returns. *British Accounting Review*, *41*(3), 184-196.

Ang, A., Chen, J., & Xing, Y. (2006). Downside risk. *Review of Financial Studies*, *19*(4), 1191-1239.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, *61*(1), 259-299.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*, *59*(3), 1259-1294.

Arnott, R. D., Harvey, C. R., Kalesnik, V., & Linnainmaa, J. T. (2021). Reports of value's death may be greatly exaggerated. *Financial Analysts Journal*, *77*(1), 44-67.

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, *99*(2), 427-446.

Bali, T. G., Engle, R. F., & Murray, S. (2016). *Empirical asset pricing: the cross-section of stock returns*. John Wiley & Sons.

Bansal, R., & Yaron, A. (2004). Risks for the long run: a potential resolution of asset pricing puzzles. *Journal of Finance*, *59*(4), 1481-1509.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3-18.

Barber, B. M., & 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), 785-818.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of Behavioral Economics: Applications and Foundations 1* (Vol. 1, pp. 79-175). North-Holland.

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2015). X-CAPM: an extrapolative capital asset pricing model. *Journal of Financial Economics*, *115*(1), 1-24.

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, *129*(2), 203-227.

Barberis, N., Jin, L. J., & Wang, B. (2021). Prospect theory and stock market anomalies. *Journal of Finance*, *76*(5), 2639-2687.

Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, *68*(2), 161-199.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*, 307-343.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. Handbook of the Economics of Finance, 1053-1128.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, *121*(3), 823-866.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis. *Journal of Finance*, *32*(3), 663-682.

Blume, M. E., & Husic, F. (1973). Price, beta, and exchange listing. *Journal of Finance*, *28*(2), 283-299.

Bordalo, P., Gennaioli, N., Porta, R. L., & Shleifer, A. (2019). Diagnostic expectations and stock returns. *Journal of Finance*, *74*(6), 2839-2874.

Bordalo, P., Gennaioli, N., La Porta, R., & Shleifer, A. (2022). Belief overreaction and stock market puzzles. (No. w27283). National Bureau of Economic Research.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, *7*(3), 265-296.

Brennan, M., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, *49*(3), 345-373.

Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *Journal of Finance*, *56*(6), 2431-2456.

Chen, A. Y. (2021). The limits of p-hacking: some thought experiments. *Journal of Finance*, *76*(5), 2447-2480.

Chen, A. Y., & Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review, 27*(2), 207-264.

Chinco, A., Hartzmark, S. M., & Sussman, A. B. (2022). A new test of risk factor relevance. *Journal of Finance*, *77*(4), 2183-2238.

Choi, J. J., & Robertson, A. Z. (2020). What matters to individual investors? Evidence from the horse's mouth. *Journal of Finance*, *75*(4), 1965-2020.

Cookson, J. A., Engelberg, J. E., & Mullins, W. (2023). Echo chambers. *Review of Financial Studies*, *36*(2), 450-500.

Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *Journal of Finance*, *63*(4), 1609-1651.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, *53*(6), 1839-1885.

Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *Journal of Finance*, *61*(4), 1605-1643.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: sentiment extraction from small talk on the web. *Management Science*, *53*(9), 1375-1388.

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance*, *55*(1), 389-406.

De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, *40*(3), 793-805.

De Franco, G., Hope, O.-K., Vyas, D., & Zhou, Y. (2015). Analyst report readability. *Contemporary Accounting Research*, *32*(1), 76-104.

Dechow, P. M., Hutton, A. P., Meulbroek, L., & Sloan, R. G. (2001). Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics*, *61*(1), 77-106.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *Journal of Finance*, *53*(3), 1131-1147.

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross-section of stock returns. *Journal of Finance*, *57*(5), 2113-2141.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*, 283-306.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427-465.

Fama, E. F., & French, K. R. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, *82*(3), 491-518.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, *81*(3), 607-636.

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance*, *68*(3), 1267-1300.

Greenwood, R., & Shleifer, A. (2014). Expectations of returns and expected returns. *Review of Financial Studies*, *27*(3), 714-746.

Hahn, J., & Lee, H. (2009). Financial constraints, debt capacity, and the cross-section of stock returns. *Journal of Finance*, *64*(2), 891-921.

Han, B., Hirshleifer, D., & Walden, J. (2022). Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, *57*(1), 390-412.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *Review of Financial Studies*, *29*(1), 5-68.

Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance*, *55*(3), 1263-1295.

Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3), 401-439.

Hirshleifer, D. (2020). Presidential address: social transmission bias in economics and finance. *Journal of Finance*, *75*(4), 1779-1831.

Hirshleifer, D., Peng, L., & Wang, Q. (2023). News diffusion in social networks and stock market reactions (No. w30860). National Bureau of Economic Research.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, *54*(6), 2143-2184.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, *55*(1), 265-295.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, *33*(5), 2019-2133.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: a topic modeling approach. *Management Science*, *64*(6), 2833-2855.

Huang, A. H., Zang, A. Y., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, *89*(6), 2151-2180.

Hwang, B.-H. (2022). The impact of word-of-mouth communication on investors' decisions and asset prices. Handbook of Financial Decision Making, forthcoming.

Jacobs, H., & Müller, S. (2020). Anomalies across the globe: once public, no longer existent? *Journal of Financial Economics*, *135*(1), 213-230.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, *45*(3), 881-898.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, *48*(1), 65-91.

Jensen, T. I., Kelly, B., & Pedersen, L. H. (2023). Is there a replication crisis in finance? *Journal of Finance*, forthcoming.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *Review of Financial Studies*, *31*(7), 2606-2649.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, *66*(1), 35-65.

Loughran, T., & Wellman, J. W. (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis*, *46*(6), 1629-1650.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica*, 1429-1445.

McLean, R.D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, *71*(1), 5-32.

Moskowitz, T., & Vasudevan, K. (2022). Betting without beta. Yale University Working Paper.

Nagel, S., & Xu, Z. (2022). Asset pricing with fading memory. *Review of Financial Studies*, *35*(5), 2190-2245.

Nickerson, R. S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220.

Novy-Marx, R. (2011). Operating leverage. *Review of Finance*, *15*(1), 103-134.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, *103*(3), 429-453.

Novy-Marx, R. (2013). The other side of value: the gross profitability premium. *Journal of Financial Economics*, *108*(1), 1-28.

Park, J., Konana, P., Gu, B., Kumar, A., & Raghunathan, R. (2013). Information valuation and confirmation bias in virtual communities: evidence from stock message boards. *Information Systems Research*, 24(4), 1050-1067.

Pontiff, J., & Woodgate, A. (2008). Share issuance and cross-sectional returns. *Journal of Finance*, *63*(2), 921-945.

Richardson, S. A., Sloan, R. G., Soliman, M. T., & Tuna, I. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, *39*(3), 437-485.

Ritter, J. R. (1991). The long-run performance of initial public offerings. *Journal of Finance*, *46*(1), 3-27.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, *71*(3), 289-315.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The Accounting Review*, *83*(3), 823-853.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288-302.

Tetlock, P. C. (2007). Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance*, *62*(3), 1139-1168.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297-323.

Xie, H. (2001). The mispricing of abnormal accruals. *The Accounting Review*, *76*(3), 357-373.

Figure 1

This figure displays word clouds for the safety, exuberance and lottery words analysts and SA contributors use in their buy recommendations. The safety, exuberance and lottery words are rooted in a survey sent to 100 institutional investors (Section 3.2).

Analyst Reports:

SA Articles:

Figure 2

This figure displays word clouds for the safety, exuberance and lottery words analysts and SA contributors use in their buy recommendations. The safety, exuberance and lottery words are rooted in a survey sent to 303 retail investors (Section 3.2).

Analyst Reports:

SA Articles:

Table 1
Survey-Based Wordlists

This table reports our survey-based safety words, exuberance words, and lottery words. We first report our five base terms, followed by all the word forms of the base terms. We describe how we arrive at the base terms and construct the corresponding word forms in Section 3.2.

Safety Words	Exuberance Words	Lottery Words
Panel A: Institutional Investors' Survey-Based Wordlists		
conservative, defensive, protection, reliable, stable	*competitive, expanding, leader, outperformer, strong*	*gamble, potential, speculative, upside, volatile*
defensively, protected, protections, reliability, reliabilities, reliableness, reliablenesses, reliably, stability, stabilities, stableness, stablenesses, stably	*competitively, expand, expanded, expands, expansion, expansions, lead,[i] leaders, leadership, leaderships, leading, leads, led, strongly*	*gambled, gambler, gamblers, gambles, gambling, gamblings, potentialities, potentiality, potentially, potentials, speculate, speculated, speculates, speculating, speculation, speculations, speculatively, speculativeness, speculativenesses, speculator, speculators, upsides, volatiles, volatilities, volatility, volatilizable, volatilize, volatilized, volatilizes, volatilizing*
Panel B: Retail Investors' Survey-Based Wordlists		
reliable, safe, secure, stable, steady	*consistent, excellent, growth, innovative, winner*	*exciting, gamble, potential, speculative, volatile*
reliability, reliabilities, reliableness, reliablenesses, reliably, safeness, safenesses, secured, securely, secureness, securenesses, secures, securing, stability, stabilities, stableness, stablenesses, stably, steadied, steadies, steadily, steadiness, steadinesses, steadying	*consistence, consistences, consistency, consistencies, consistently, excel, excelled, excellence, excellences, excellently, excelling, excels, grew, grow, grower, growers, growing, growings, grown, grows, growths, innovate, innovated, innovates, innovating, innovation, innovations, innovational, innovativeness, innovativenesses, innovator, innovators, win, wins, winners, winning, winnings, won*	*excite, excited, excitement, excitements, excites, excitingly, gambled, gambler, gamblers, gambles, gambling, gamblings, potentiality, potentialities, potentially, potentials, speculate, speculated, speculates, speculating, speculation, speculations, speculatively, speculativeness, speculativenesses, speculator, speculators, volatiles, volatility, volatilities, volatilizable, volatilize, volatilized, volatilizes, volatilizing*

---

[i] We exclude "*lead to, leads to, leading to*" as these words do not carry the same meaning as "*leader*."

Table 2
Which Framework Best Explains the Cross-Section of Expected Stock Returns? – Aggregated Evidence

This table reports the fraction of times a particular framework best explains the cross-section of expected stock returns. For each of 181 firm characteristics, we proceed as follows: For each month t, we rank stocks based on firm characteristic i. The stocks in the utmost decile that, based on prior literature, trade at comparatively high prices and earn unusually low future returns represent the "short-leg securities." We refer to analyst reports and SA articles recommending that investors buy these short-leg securities as "short-leg recommendations." We refer to all other analyst and SA buy recommendations as "other recommendations." To understand the source of investors' bullish views of short-leg securities, we parse all the bullish views of the short-leg securities and test whether these short-leg recommendations unusually fixate on the stocks' safe-haven quality, their seeming superiority, or their upside potential. To this end, we compute the average *Safety [%]* across analysts' (SA contributors') short-leg recommendations. To assess whether the use of safety words in the short-leg recommendations is abnormally high, we also compute the average *Safety [%]* across analysts' (SA contributors') other recommendations. We then calculate the difference between the former and the latter (on a relative basis), *ΔSafety [%]*, and test whether the difference is positive and statistically significant at the 5% level. Similarly, we compute the average *Exuberance [%]* and *Lottery [%]* across analysts' (SA contributors') short-leg recommendations and the average *Exuberance [%]* and *Lottery [%]* across analysts' (SA contributors') other recommendations. We then test whether *ΔExuberance [%]* and *ΔLottery [%]* are positive and statistically significant at the 5% level. If *ΔSafety [%]* is positive and statistically significant at the 5% level, then we label investors' rationale for liking the short-leg securities as "consistent with the risk framework." If *ΔExuberance [%]* or *ΔLottery [%]* is positive and statistically significant at the 5% level, then we label investors' rationale as "consistent with the irrational beliefs framework" or as "consistent with the non-traditional preferences framework." If neither *ΔSafety [%]*, *ΔExuberance [%]* nor *ΔLottery [%]* is positive and statistically significant at the 5% level, then we label our result as "inconclusive." If more than one of the fractions is statistically significant at the 5% level, we label the framework associated with the largest relative difference as the one that is "most consistent." We report the fraction of times investors' rationale is consistent with [most consistent with] a particular framework. In Panel A, *Safety [%]*, *Exuberance [%]* and *Lottery [%]* are rooted in the institutional investor wordlist listed in Table 1; in Panel B, *Safety [%]*, *Exuberance [%]* and *Lottery [%]* are rooted in the retail investor wordlist also listed in Table 1.

	Fraction of Times Investors' Rationale for Buying Short-Leg Securities Consistent With [Most Consistent With]			
	Risk Framework	Irrational Beliefs Framework	Non-Traditional Preferences Framework	Inconclusive
Panel A: Institutional Investors' Survey-Based Wordlists				
Sell-Side Analyst Reports	9%	29%	69%	19%
	[7%]	[17%]	[58%]	
Seeking Alpha Articles	6%	15%	58%	29%
	[6%]	[10%]	[54%]	
Panel B: Retail Investors' Survey-Based Wordlists				
Sell-Side Analyst Reports	21%	24%	70%	14%
	[6%]	[14%]	[66%]	
Seeking Alpha Articles	9%	22%	64%	22%
	[8%]	[13%]	[57%]	

Table 3

Which Framework Best Explains the Cross-Section of Expected Stock Returns? – Evidence by Firm Characteristic

This table reports the results from Panel A in Table 2 separately for 22 out of the 181 firm characteristics that Barberis, Jin and Wang (2021) examine. For each of the 22 firm characteristics, we compute the average *Safety [%]*, *Exuberance [%]*, and *Lottery [%]* across the buy recommendations of short-leg securities; we also compute the average fractions across the buy recommendations written on all other stocks. We then compute the difference between the former and the latter (on a relative basis) and test whether the difference is positive and statistically significant at the 5% level. We report the relative difference and the corresponding *t*-statistic in parentheses if the difference is positive and statistically significant at the 5% level. An empty cell indicates that the difference is not positive and statistically significant at the 5% level.

Firm Characteristic	Do Sell-Side Analyst Buy Recommendations of Short-Leg Securities Disproportionately Frequently Use			Do Seeking Alpha Buy Recommendations of Short-Leg Securities Disproportionately Frequently Use		
	Safety Words?	Exuberance Words?	Lottery Words?	Safety Words?	Exuberance Words?	Lottery Words?
Accrual			7% (10.79)			12% (6.16)
Difference of Opinion			12% (16.49)			17% (7.16)
Asset Growth			15% (32.72)			9% (6.65)
Value		5% (14.36)	14% (31.65)			7% (5.51)
Gross Profitability			43% (68.73)			49% (23.03)
Investment			8% (14.65)			8% (4.94)
Composite Equity Issuance			11% (23.05)			19% (10.66)
Idiosyncratic Volatility			20% (36.98)			28% (17.42)
Expected Idiosyncratic Skewness		12% (36.80)	6% (14.51)			5% (3.83)
Long-term Reversal		10% (27.98)	9% (19.61)	4% (3.48)		
Maximum Daily Return			18% (37.66)			25% (15.93)
Momentum			9% (13.64)			13% (7.17)
Net Stock Issuance			32% (56.76)			32% (17.22)
External Finance			32% (60.38)			39% (20.29)
Net Operating Assets			7% (12.76)			
O-Score			35% (39.17)			59% (19.71)
Organizational Capital				8% (4.15)		
Failure Probability			11% (3.39)			
Return On Assets			42% (66.17)			51% (24.04)
Post-Earnings Announcement Drift						7% (4.64)
Short-Term Reversal			14% (29.32)			13% (8.95)
Market Capitalization	7% (14.65)	5% (19.34)			3% (3.97)	

Table 4

Which Framework Best Explains the Cross-Section of Expected Stock Returns? – Sensitivity Analyses

This table reports the fraction of times a particular framework best explains the cross-section of expected stock returns. The analyses are identical to those in Panel A of Table 2, except that we now consider only the firm characteristics whose corresponding academic paper's Google Scholar citation as of June 2022 is in the top quartile (Panel A), only data since the corresponding academic paper has been published (Panel B), only the firm characteristics whose corresponding academic paper was published in the first quartile (Panel C). Panels D and E report results based on a different denominator and based on the numerator only as detailed in Section 5.2, respectively.

	Fraction of Times Investors' Rationale for Buying Short-Leg Securities Consistent With [Most Consistent With]			
	Risk Framework	Irrational Beliefs Framework	Non-Traditional Preferences Framework	Inconclusive
Panel A: "High Google Scholar Citations Anomalies" Only				
Sell-Side Analyst Reports	17% [15%]	35% [15%]	70% [59%]	11%
Seeking Alpha Articles	7% [7%]	20% [13%]	65% [61%]	20%
Panel B: "Data Since Publication" Only				
Sell-Side Analyst Reports	10% [8%]	29% [17%]	67% [57%]	19%
Seeking Alpha Articles	6% [5%]	15% [10%]	56% [54%]	31%
Panel C: "Older Anomalies" Only				
Sell-Side Analyst Reports	11% [8%]	32% [13%]	74% [64%]	15%
Seeking Alpha Articles	6% [6%]	17% [11%]	66% [62%]	21%
Panel D: Alternate Denominator				
Sell-Side Analyst Reports	9% [7%]	29% [17%]	69% [58%]	19%
Seeking Alpha Articles	6% [6%]	15% [10%]	58% [54%]	29%
Panel E: Numerator Only				
Sell-Side Analyst Reports	14% [9%]	30% [18%]	66% [51%]	22%
Seeking Alpha Articles	6% [4%]	21% [10%]	59% [55%]	31%

Table 5

Which Framework Best Explains the Cross-Section of Expected Stock Returns?

– Exuberance Terms Rooted in Theory

This table reports the fraction of times a particular framework best explains the cross-section of expected stock returns. The analyses are identical to those in Panel A of Table 2, except that we now apply our self-defined wordlists as detailed in Section 5.4.2.

	Fraction of Times Investors' Rationale for Buying Short-Leg Securities Consistent With [Most Consistent With]			
	Risk Framework	Irrational Beliefs Framework	Non-Traditional Preferences Framework	Inconclusive
Sell-Side Analyst Reports	16% [10%]	17% [12%]	66% [55%]	23%
Seeking Alpha Articles	13% [9%]	15% [15%]	55% [48%]	28%

Table 6

Analysts' Long-Term Earnings Growth Forecasts and Their Use of Exuberance Words

This table reports results from regressions of analysts' long-term earnings growth forecasts on the tone and the fractions of safety, exuberance and lottery words in the corresponding written reports. *Tone [%]* is the number of positive words minus the number of negative words scaled by the total number of words. *Safety [%]* is the number of safety words scaled by the total number of words. *Exuberance [%]* is the number of exuberance words scaled by the total number of words. *Lottery [%]* is the number of lottery words scaled by the total number of words. We winsorize variables at the top and bottom 1%. *T*-statistics are reported in parentheses and are based on standard errors adjusted for heteroscedasticity and clustered by day. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Tone [%]	0.699***		0.619***	0.713***		0.633***	0.711***		0.629***
	(7.05)		(6.36)	(7.90)		(7.12)	(7.93)		(7.12)
Safety [%]		0.535	0.224		0.285	-0.058		0.345	-0.007
		(0.75)	(0.32)		(0.46)	(-0.09)		(0.57)	(-0.01)
Exuberance [%]		1.533***	0.835***		1.554***	0.855***		1.581***	0.889***
		(5.18)	(2.94)		(6.48)	(3.78)		(6.63)	(3.96)
Lottery [%]		0.082	-0.028		0.394	0.287		0.285	0.180
		(0.21)	(-0.07)		(1.30)	(0.96)		(0.96)	(0.61)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Analyst FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	38,780	38,780	38,780	38,663	38,663	38,663	38,643	38,643	38,643
Adj. R^2	0.475	0.474	0.476	0.576	0.575	0.576	0.581	0.580	0.582

Table 7

Which Framework Best Explains the Cross-Section of Expected Stock Returns? – Moderating Factor

This table reports the fraction of times a particular framework best explains the cross-section of expected stock returns. The analyses are identical to those in Panel A of Table 2, except that we now report results for different subsets of firm characteristics. We compute for each of the 181 firm characteristics the average market capitalization of the stocks in the short leg. Extrapolative tendencies are likely stronger for larger stocks; non-traditional preferences are likely more relevant when evaluating smaller stocks. Panel A reports the results for the firm characteristics whose average market capitalization of the short-leg securities is above the median ("more likely to be irrational beliefs based / less likely to be non-traditional preferences based"). Panel B reports the results for the firm characteristics whose average market capitalization of the short-leg securities is below the median ("less likely to be irrational beliefs based / more likely to be non-traditional preferences based").

	Fraction of Times Investors' Rationale for Buying Short-Leg Securities Consistent With [Most Consistent With]			
	Risk Framework	Irrational Beliefs Framework	Non-Traditional Preferences Framework	Inconclusive
Panel A: More Likely to be Irrational Beliefs Based / Less Likely to be Non-Traditional Preferences Based				
Sell-Side Analyst Reports	16% [12%]	47% [27%]	55% [37%]	23%
Seeking Alpha Articles	10% [10%]	24% [19%]	32% [26%]	45%
Panel B: Less Likely to be Irrational Beliefs Based / More Likely to be Non-Traditional Preferences Based				
Sell-Side Analyst Reports	1% [1%]	11% [6%]	82% [79%]	14%
Seeking Alpha Articles	2% [2%]	6% [2%]	84% [82%]	13%