INSIDE THE MINDS OF EXPECTED STOCK RETURNS

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

This Draft: January 2023

Our paper conducts textual analysis on 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 test whether the justifications provided in these buy recommendations mostly (1) emphasize a stock's safe-haven quality, (2) indicate investor exuberance, or (3) point to a 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 dominant 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.

* Chen is affiliated with the Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong. Hwang and Peng are affiliated with Nanyang Business School, NTU, Singapore. E-mail: chen19@hku.hk, bh.hwang@ntu.edu.sg and zhuozhen001@e.ntu.edu.sg. The authors would like to thank Nick Barberis, Andrew Chen, Tony Cookson, Zhi Da, Lawrence Jin, Kelvin Tan, and seminar participants at Georgia State University, Korea University, the National University of Singapore, KAEA, the 2022 AFBC, and the 2022 USC Social & Behavioral Finance Conference for valuable suggestions and comments. The authors would also like to thank Andrew Chen and Tom Zimmermann for generously providing their cross-sectional return predictability data on https://www.openassetpricing.com.

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. 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 returns in the cross-section.

There are at least three prominent perspectives regarding the sources of this predictability. The traditional perspective is that stocks with certain characteristics tend to do better when the marginal utility of money is high. These stocks are safe-haven assets as they provide insurance against bad states of the world. It is rational behavior for risk-averse investors to pay high prices and accept low average returns for stocks with these characteristics (e.g., Lucas, 1978; Breeden, 1979; Fama, 1998). An alternative perspective is that certain firm characteristics correlate with or induce overly optimistic beliefs. Overly optimistic beliefs can arise from investors' extrapolation of 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 subsequently becomes corrected, these characteristics predict 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. It may also explain why investors overpay for stocks with lottery-like features (e.g., Barberis, 2018).

Gauging which of these three perspectives best explains the cross-section of expected stock returns is challenging, as the patterns we observe in the stock market 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 low average returns. One possibility is that investors perceive stocks with high asset growth as less risky 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 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 for differentiating between these 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 gauge which of the three frameworks best explains the prices investors are willing to pay for certain stocks. 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.[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' opinions.

We then proceed as follows: We separately consider each of the 205 firm-level characteristics that Chen and Zimmermann (2022) study in their main tests. Consider the asset-growth anomaly again. Each month, we examine stocks in the top-asset-growth decile. These are stocks that earn relatively low future returns. We hereafter refer to these stocks as "short-leg securities" or as stocks residing "in the short leg of the asset-growth anomaly." We parse all analyst- and SA buy recommendations written about the short-leg securities and examine whether, in explaining why they like high-asset-growth stocks, analysts and SA contributors fixate on the stocks' safe-haven quality, the stocks' 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 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"), (2) a list

[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.

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. The five safety terms most frequently mentioned by our survey participants are *conservative*, *defensive*, *protection*, *reliable*, and *stable*. The five exuberance terms are *competitive*, *expanding*, *leader*, *outperformer*, and *strong*. The five lottery terms are *gamble*, *potential*, *speculative*, *upside*, and *volatile*.

Our analysis considers all possible word forms of the above terms that are meaningfully tied to the business realm, including plural forms, noun forms, verb forms, adjective forms, adverb forms, and verb conjugations. We account for negation. In additional analyses, we conduct a series of tests to gauge the sensitivity of the results to variations of our wordlists.

To illustrate our method and discuss some of our findings at the firm-characteristic level, Ang, Hodrick, Xing, and Zhang (2006) show that stocks with high idiosyncratic risk subsequently earn unusually low returns. We parse analysts' and SA contributors' buy recommendations for stocks with high idiosyncratic risk and study what draws investors to these stocks. We find that analyst reports substantially more frequently use lottery words when discussing high idiosyncratic risk stocks than when discussing stocks that do not have high idiosyncratic risk: the fraction of lottery words is 19% higher (t-statistic = 35.68). SA articles also more frequently use lottery words when discussing stocks with high idiosyncratic risk (+26%, t-statistic = 16.44). We observe no reliable differences in the use of safety or exuberance words. The abnormal use of lottery words in the buy recommendations for high idiosyncratic risk 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 idiosyncratic risk 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 are marked with an abnormally heavy use of exuberance words. We observe this pattern in both analyst reports and SA articles. We observe no abnormal use of safety words. We observe some abnormal use of lottery words in analyst reports but not in SA articles. Overall, it appears that the poor future performance of stocks with high returns over the past three years (DeBondt and Thaler, 1985) is most congruent with overpricing due to overly optimistic beliefs.

For some of the firm characteristics we consider, there are no analyst reports or SA articles for 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 (such as in the above low-leverage illustration) for 12 out of the 181 firm characteristics, or 7%

of the time. That is, for 12 “anomalies,” 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 (such as in the above high-past-returns illustration) 17% of the time. Analysts’ rationales are most consistent with the non-traditional preferences framework (such as in the above high-idiosyncratic-risk illustration) 58% of the time. 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 suggests that while all three frameworks explain components of the cross-section of expected stock returns, non-traditional investor preferences offer the most comprehensive explanation.

In additional analyses, we send an almost identical version of our institutional investor survey to 303 US retail investors and construct retail investor-based lists of safety, exuberance and lottery words. Our results rooted in the retail investor wordlists point even more strongly to the relevance of non-traditional preferences. When parsing analyst reports, our results suggest that 6% of the cross-sectional stock return predictabilities arise because of differences in the level of risk. 14% are rooted in irrational beliefs, while 66% are rooted in non-traditional preferences. For SA articles, the corresponding fractions are 8%, 13%, and 57%, respectively.

Our results are similar when we consider only the “most important” predictabilities, that is, firm characteristics for which the citation number of the paper documenting the corresponding predictability is in the top quartile. Our results are also similar 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.

As we discuss in the main body of the text, we conduct a series of other tests and consider possible confounding factors, such as analyst incentives. The results from all these tests point to the same conclusion:

non-traditional investor preferences play a dominant role in explaining the cross-section of expected stock returns.

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. In Sections 4 and 5, we present our main results and results from additional analyses, respectively. Section 6 discusses caveats and important limitations of our study, and Section 7 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 cross-sectional stock-return 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.

Relatedly, 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 (2022) arrive at similar conclusions.

Our study builds on McLean and Pontiff (2016), Jacobs and Müller (2020), Chen (2021), and Jensen, Kelly, and Pedersen (2022) and assumes that the cross-sectional stock-return predictabilities 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 stock-return predictabilities are real, we contribute to the empirical asset-pricing literature by providing estimates of the extent to which the patterns in average raw 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). While both departures likely play a role, to guide future research in behavioral finance, it is important that we know which of the two departures more comprehensively explains the cross-section of expected stock returns.

In a recent study, Moskowitz and Vasudevan (2022) try to answer this question. Moskowitz and Vasudevan 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 cross-section of expected stock returns.

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 view into investors' minds without continuously surveying them. Surveys have the advantage of allowing 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. Before describing our data and wordlists, we briefly outline the three frameworks that our wordlists intend to capture.

3.1 The Three Frameworks

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 possible 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 mistaken investor beliefs, coupled with limits to arbitrage, or non-traditional preferences. One reason investors may systematically form incorrect beliefs 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 stock-return predictabilities 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.

Another 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 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.

3.2 Survey-Based Wordlists

To examine which of these frameworks best explains investors' decision-making, we study investor buy recommendations and the words they use to explain why they like these stocks.

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.

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 investors what words they would use to describe a stock that provides insurance against bad states of the world. Our second question elicits words of excitement tied to extrapolation and overconfidence. Our third question asks what words investors would use to describe the lottery-like features of a stock.

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 cross-section of expected stock returns.[2]

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.[3] 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

---

[2] 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.

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

platform that allows 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 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.

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 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.[4]

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 (*no*, *not*, *none*, *neither*, *never*, *nobody*, *few*, *little*, *less*,[5] *low*, *hardly*, *rarely*) occurring within three words preceding a word from our wordlists.

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.

[4] 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 Investexts misassignments in our sample.
[5] 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.

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. The key feature driving the analysts' recommendation is the perceived upside potential: "*We rate shares of LPR Overweight, with our $8/share December 2012 price target implying 134% potential upside. . . . 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,[6] 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.

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*.

[6] 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.

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 *raw* 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.[7] 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 download the dataset, add back 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.

[7] 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:

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

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 they like stocks on the "short-spectrum" of a firm characteristic, investors 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 particular 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.[8]

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 institutional investors' survey-based results presented in Panel A. When we apply our institutional investor wordlists to analyst reports 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 analysts 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.

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

The fractions of times investors' 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.

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

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 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 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 present findings separately for each of the 181 firm characteristics. To keep the table size manageable, we present only the findings based on the institutional investors' survey-based wordlists.[9] The results are similar for the retail investors' survey-based wordlists and are available upon request.

The results reported in Table 3 indicate 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 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

[9] In Online Appendix Figure A5 we also plot our findings separately for each firm characteristic.

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 in analysts' and SA contributors' explanations of why they like 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 beta (Fama and MacBeth, 1973).

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 (DeBondt 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 for the above types of stocks 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

Our results suggest that the non-traditional preferences framework explains the largest number of anomalies. Among others, 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.

4.3.1 Non-Traditional Preferences versus Exuberance

We think that two features of our results on non-traditional preferences are worth discussing further. In particular, one possible concern with our analysis is that investors' discussion of a stock's upside potential is just another reflection of optimism and extrapolation of recent, positive events.

While we cannot rule out this possibility, 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 as analysts and SA contributors explain 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.

Relatedly, 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 A3 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. When considering analyst reports, the fraction of times the cross-sectional return predictabilities are consistent with non-traditional preferences ranges from 67% to 68%. When considering SA articles, the corresponding fractions range from 55% to 56%.

4.3.2 Relation to Existing Evidence

The second feature of our results 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.[10]

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 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

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

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 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.

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. Additional 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; investors' rationales are congruent with the irrational beliefs framework 35% of the time, and investors' rationales are congruent with 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 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, which were documented before researchers started more seriously considering prospect theory's relevance to finance.

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 indicate 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 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. To further gauge the sensitivity of our results to the wordlist composition, we conduct the following two exercises.

5.2.1 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.

5.2.2 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.

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.

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 investor exuberance is rooted in (1) extrapolation and (2) investor 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 ("4 grams").[11] 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* – *rise*), 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 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.

[11] 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.

The advantage of our self-defined wordlists over our survey-based wordlists is that the construction of the exuberance words is more rooted in theory. The disadvantage is that our self-defined wordlists are subjective. By creating our own wordlists, we also insert ourselves into the data-generating process.

We report the results in Panel D of Table 4. 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.

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 dominant role.

5.2.3 Other Sensitivity Analyses

We conduct several additional sensitivity analyses. 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% (→ ΔSafety[%] = 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*:

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

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

We report the corresponding results in Panels E and F 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.

5.2.4 Sell Recommendations of Long-Leg Securities

Our primary analysis focuses on buy recommendations for short-leg securities. The reason is that there are many more buy recommendations than sell recommendations, which increases the power of our analysis. Sell recommendations are particularly rare on SA.

More importantly, the vast majority of the seemingly anomalous returns come from the short leg rather than the long leg (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.[12] 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 optimistic investors out of the market and that would allow a subset of investors with overly pessimistic views to generate underpricing and abnormally high future returns.

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 recommend buying stocks that, on average, deliver moderately higher returns. Is the most commonly noted shortcoming the high level of risk, adverse recent events or signals, or the perceived lack of upside potential?

To find the most commonly noted shortcoming, we consider the bearish recommendations for stocks that reside in the long leg, and we compute the fraction of negated safety words (e.g., *little – protection*), the fraction of negated exuberance words (e.g., *not – competitive*) and the fraction

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

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 A4. 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 appear to be 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.

5.3 Moderating Factors

In Table 5, we return our focus to the buy recommendations for short-leg securities and consider a moderating factor for the relevance of irrational beliefs over that of non-traditional preferences. As alluded to in Section 4.3, 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 5, 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, analysts' rationales point 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 analysts' emphasis on a stock's upside potential. When short-leg securities are comparatively large, analysts' rationales 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 and find that their prospect-theory-based model can match the long-short returns of most of the 23 anomalies. The authors' model does better for some anomalies than for others. 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 consider 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.

6. Caveats and Limitations

Before concluding, we discuss several factors that could limit the generalizability of our findings. We also discuss important shortcomings of our analysis.

6.1 Sell-Side Analysts, SA Contributors, and the General Investor Population

In this paper, we assume that sell-side analysts' and SA contributors' rationales for liking certain stocks either directly impact or mirror the views of a meaningful portion of the investor population. That is, as analysts and SA contributors like certain stocks for a particular reason, so does a meaningful portion of the investor population, thereby pushing up the prices of these stocks. To gauge the plausibility of this assumption, we estimate regressions of abnormal stock returns on the tones of analyst reports and the tones of SA articles.

We compute the abnormal return on stock *i* on day *t* as the difference between stock *i*'s raw return on day *t* and the return on a value-weighted portfolio of firms of similar size, book-to-market ratios, and past returns (Daniel, Grinblatt, Titman, and Wermers, 1997).

The tone of a report or an article is the fraction of positive words minus the fraction of negative words scaled by the total number of words. $Tone_{\text{Sell-Side Analysts } i,t}$ is the average tone across all analyst reports published about stock *i* on day *t*. $Tone_{\text{Seeking Alpha } i,t}$ is the average tone across all SA articles published about stock *i* on day *t*. We also construct $I_{\text{Sell-Side Analysts } i,t}$ and $I_{\text{Seeking Alpha } i,t}$, which equal one if there are analyst reports and SA articles published about stock *i* on day *t*, respectively. $Tone_{\text{Sell-Side Analysts } i,t}$ and $Tone_{\text{Seeking}}$

$_{Alpha\ i,t}$ are set to zero when there are no analyst reports and no SA articles published about stock *i* on day *t*, respectively. We adjust our standard errors for heteroscedasticity, and we cluster by day.

It is unlikely that analyst reports and SA articles are both 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 any positive association between stock returns and tone reflects company news announced on day *t*, which moves stock prices on day *t* and, at the same time, is reported in analyst reports and SA articles published on day *t*. Still, to account for the possibility that analyst reports and SA articles published on day *t* reflect company news announced on day *t*, we obtain composite sentiment scores ("CSS") from Ravenpack for all Dow Jones Newswires. We then include in all our regression equations $Sentiment_{\ Dow\ Jones\ Newswires}$, which is the average CSS on stock *i* on day *t*, and $I_{\ Dow\ Jones\ Newswires}$, which equals one if there are Dow Jones Newswires published about stock *i* on day *t*; we set $Sentiment_{\ Dow\ Jones\ Newswires}$ to zero when there are no Dow Jones Newswires.

The results reported in Table 6 indicate that both the tones of analyst reports and the tones of articles published in SA strongly positively correlate with contemporaneous abnormal returns. When including both $Tone_{\ Sell\text{-}Side\ Analysts}$ and $Tone_{\ Seeking\ Alpha}$ as independent variables, the estimate of $Tone_{\ Sell\text{-}Side\ Analysts}$ is 0.441 (*t*-statistic = 57.49) and the estimate of $Tone_{\ Seeking\ Alpha}$ is 0.212 (*t*-statistic = 15.35). These results are consistent with our assumption that the views expressed in analyst reports and SA provide a representative glimpse of the views of a meaningful portion of the investor population and, therefore, could relate to stock prices.

6.2 Analyst Incentives versus True Beliefs

Another related concern is that analysts might not express their true beliefs in their reports. Analysts have incentives to issue overly favorable reports (Barber, Lehavy, Trueman, 2007). As analysts decide how to best support their favorable views, they must choose what story to tell. For instance, in explaining their buy recommendation for a high-volatility stock, analysts must choose whether to tell a safe-haven story, a continued-growth story, or an upside-potential story. Analysts may choose an upside-potential story not because they truly believe in it but because they think it will make their justification of a buy recommendation of a high-volatility stock appear the most plausible to their investor clients.

Bradley (2022) provides an overview of the academic literature on analysts and argues that two forces incentivize analysts to issue overly favorable reports: (1) the desire to generate investment banking business and (2) the aspiration to trigger stock purchases and earn trading commissions.

Kadan, Madureira, Wang, and Zach (2009), among others, find evidence that the desire to generate investment banking business no longer materially impacts analyst recommendations since the Global Settlement in April 2003. The Global Settlement separated brokerages' research activities from their investment banking pursuits. Our sample period begins in January 2006 and thus falls entirely in the post-Global Settlement period.

The aspiration to trigger stock purchases and earn trading commissions comes from the practice in virtue of which analyst research is bundled and paid for with soft dollar arrangements through trading commissions. Bradley (2022) provides a detailed explanation of this practice. On January 3, 2018, the European Union ended this practice in Europe by enacting the Markets in Financial Instruments Directive II (MiFID II). Fang, Hope, Huang, Moldovan (2020), Guo and Mota (2021), and Lang, Pinto, and Sul (2022) find that the enactment of MiFID II significantly altered analyst incentives.

While MiFID II applies to European financial markets only, some global asset managers follow MiFID II both in the US and Europe, either in anticipation of a similar directive by the US Securities and Exchange Commission or for the sake of parsimony (Bradley, 2022). If analyst incentives generate parts of our findings, we should thus observe a change in our main findings from the pre-MiFID II period to the post-MiFID II period. We report the corresponding results in Online Appendix Table A5. In short, we observe no meaningful changes in our main results from the pre-MiFID II period to the post-MiFID II period.

In general, we do not think our conclusions would change materially even if analyst incentives continued to be a factor. The reason is that even if analysts cater their reports to investors' preconceptions, their reports ultimately still mirror investors' views and the presumptions on which these investors trade.

6.3 Investor Underreaction

Our paper suffers from two important shortcomings.[13] 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.

6.4 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, 2022; Hwang, 2022).

[13] There may be more, but these are the two that are apparent to us.

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 and, consequently, testing 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 impact asset pricing. The findings in our paper provide some guidance for future research on this question. Our study examines which story types investors rely on primarily as they converse about stocks that appear overpriced. Our study, therefore, offers insights into the content of investor conversations and the possible channels through which stocks can become viral and overpriced.

7. 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 studies this classic question with a new approach. We parse investors' buy recommendations and record their primary reasoning for liking stocks. 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 determinant 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, low-cost, and scalable method for capturing real-world investors' perceptions.

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., Lehavy, R., & Trueman, B. (2007). Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics*, *85*(2), 490-517.

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*, *1*, 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.

Bradley, D. (2022). Financial analysts. Handbook of Financial Decision Making, forthcoming.

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.

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., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance*, *52*(3), 1035-1058.

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.

Fang, B., Hope, O.-K., Huang, Z., & Moldovan, R. (2020). The effects of MiFID II on sell-side analysts, buy-side analysts, and firms. *Review of Accounting Studies*, *25*(3), 855-902.

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.

Guo, Y., & Mota, L. (2021). Should information be sold separately? Evidence from MiFID II. *Journal of Financial Economics*, *142*(1), 97-126.

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. (2022). News diffusion in social networks and stock market reactions. Working Paper.

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. *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. (2022). Is there a replication crisis in finance? *Journal of Finance*, forthcoming.

Johnson, T. L., & So, E. C. (2012). The option to stock volume ratio and future returns. *Journal of Financial Economics*, *106*(2), 262-286.

Kadan, O., Madureira, L., Wang, R., & Zach, T. (2009). Conflicts of interest and stock recommendations: the effects of the Global Settlement and related regulations. *Review of Financial Studies*, *22*(10), 4189-4217.

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

Lang, M. H., Pinto, J., & Sul, E. (2022). MiFID II unbundling and sell side analyst research. Working Paper.

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.

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.

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? *Accounting Review*, *71*(3), 289-315.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *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. *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 *defensively, protected, protections, reliability, reliabilities, reliableness, reliablenesses, reliably, stability, stabilities, stableness, stablenesses, stably*	*competitive, expanding, leader, outperformer, strong* *competitively, expand, expanded, expands, expansion, expansions, lead,*[i] *leaders, leadership, leaderships, leading, leads, led, strongly*	*gamble, potential, speculative, upside, volatile* *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 *reliability, reliabilities, reliableness, reliablenesses, reliably, safeness, safenesses, secured, securely, secureness, securenesses, secures, securing, stability, stabilities, stableness, stablenesses, stably, steadied, steadies, steadily, steadiness, steadinesses, steadying*	*consistent, excellent, growth, innovative, winner* *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*	*exciting, gamble, potential, speculative, volatile* *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 Δ*Exuberance [%]* and Δ*Lottery [%]* 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% [7%]	29% [17%]	69% [58%]	19%
Seeking Alpha Articles	6% [6%]	15% [10%]	58% [54%]	29%
Panel B: Retail Investors' Survey-Based Wordlists				
Sell-Side Analyst Reports	21% [6%]	24% [14%]	70% [66%]	14%
Seeking Alpha Articles	9% [8%]	22% [13%]	64% [57%]	22%

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 each of the 181 firm characteristics. For each of 181 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. We bold the difference that is the largest economically speaking. An empty cell implies that the difference is not positive and statistically significant at the 5% level.

	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		
Firm Characteristic	Safety Words?	Exuberance Words?	Lottery Words?	Safety Words?	Exuberance Words?	Lottery Words?
5-Yr. Growth in Number of Shares		2% (5.16)	**3% (7.22)**			
52-Week High			**13% (20.71)**			**18% (10.43)**
Δ Asset Turnover		**8% (16.21)**				
Δ Current Op. Assets		**13% (30.17)**	6% (11.41)		**6% (4.45)**	
Δ Current Op. Liabilities		**13% (32.38)**	7% (15.50)			**4% (2.53)**
Δ Deferred Revenue to Assets			**17% (17.62)**			**24% (7.75)**
Δ Equity to Assets			**18% (38.88)**			**12% (8.33)**
Δ Financial Liabilities			**9% (18.91)**			**4% (2.60)**
Δ Long-term Investment	**9% (10.48)**					
Δ Net Financial Assets			**14% (28.28)**			**10% (5.94)**
Δ Net Noncurrent Op. Assets		4% (9.38)	**8% (15.02)**			**8% (4.80)**
Δ Net Op. Assets		2% (4.66)	**8% (16.42)**			
Δ Net Working Capital			**9% (15.36)**			**13% (7.11)**
Δ PPE & Inventory to Total Assets			**7% (12.26)**			
Δ Tax to Total Assets			**4% (8.68)**			**3% (2.28)**
Abnormal Accruals			**10% (14.08)**			**13% (6.59)**
Accruals			**7% (10.79)**			**12% (6.16)**
Active Shareholders						

Table 3. Continued.

	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					
Firm Characteristic	Safety Words?		Exuberance Words?		Lottery Words?		Safety Words?		Exuberance Words?		Lottery Words?	
Advertising Expense			**16%**	**(28.61)**	13%	(17.45)					**9%**	**(4.17)**
Advertising Expense to Brand Capital							**15%**	**(4.11)**				
Analyst Consensus Forecast					**40%**	**(67.83)**					**33%**	**(16.51)**
Analyst EPS Forecast Dispersion					**12%**	**(16.49)**					**17%**	**(7.16)**
Analyst EPS Forecast Revision											**6%**	**(3.38)**
Analyst Forcast on EPS												
Analyst Optimism			**6%**	**(12.95)**	6%	(10.31)					**6%**	**(2.90)**
Analysts Consensus Ratings of the Mkt.												
Annual Sales to Mkt. Value					**42%**	**(81.25)**					**41%**	**(22.09)**
Asset Growth					**15%**	**(32.72)**					**9%**	**(6.65)**
Asset Tangibility	3%	(3.05)	**8%**	**(14.85)**			**7%**	**(2.02)**				
Average Earn. Growth												
Average Earn. Growth					**8%**	**(8.44)**					**19%**	**(5.93)**
Average Earn. Surprise of Big Firms												
Bid-Ask Spread			**6%**	**(3.73)**			**34%**	**(2.29)**				
Book Leverage	**36%**	**(32.59)**										
Book-to-Mkt. Ratio			5%	(14.36)	**14%**	**(31.65)**					**7%**	**(5.51)**
Book-to-Mkt. Ratio (most recent Mkt.)			17%	(50.50)	**17%**	**(41.65)**			**6%**	**(5.26)**	3%	(2.63)
CAPM Beta	**7%**	**(5.06)**					**40%**	**(6.54)**			9%	(2.99)
CAPX to Revenue					**2%**	**(2.87)**					**10%**	**(4.80)**
Cash Flow to Price Variance					**8%**	**(9.33)**					**11%**	**(5.10)**
Cash to Assets					**1%**	**(2.48)**	**13%**	**(3.60)**				
Cash-based Op. Profitability					**25%**	**(32.80)**					**36%**	**(15.00)**
Convertible Debt Indicator					**9%**	**(19.61)**						
Coskewness Using Daily Ret.					**3%**	**(4.36)**					**8%**	**(3.57)**

Table 3. Continued.

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?	
Coskewness Using Monthly Ret.					**6%**	**(10.50)**					**10%**	**(5.30)**
Credit Rating Downgrade												
Customer/Supplier Momentum			**3%**	**(2.51)**								
Earn. Forcast to Stock Price					**7%**	**(4.32)**					**9%**	**(2.06)**
Earn.-to-Price Ratio			**13%**	**(27.48)**	10%	(16.32)			6%	(3.38)	**10%**	**(4.81)**
Economic Link Momentum												
Efficient Frontier Index			**15%**	**(36.05)**	7%	(13.95)						
Enterprise Component of Book-to-Mkt.	**4%**	**(5.13)**										
Enterprise Multiple			**20%**	**(45.41)**	11%	(20.72)			7%	(5.05)	**8%**	**(4.70)**
Equity Duration					**31%**	**(43.80)**					**31%**	**(15.28)**
Exchange Switch												
Excluded Expenses					**4%**	**(7.86)**						
Firm Age					**6%**	**(10.28)**	**10%**	**(3.41)**				
First Month When Consistent Div. Payers Fail to Pay Div.												
Frazzini-Pedersen Beta							**51%**	**(4.59)**			17%	(2.94)
Fundamental Value-to-Mkt. Value					**34%**	**(43.74)**					**32%**	**(12.02)**
GIM Governance Index												
Gross Profits to Total Assets					**43%**	**(68.73)**					**49%**	**(23.03)**
Gross Profits with R&D Adjusted					**29%**	**(33.98)**					**31%**	**(12.55)**
Growth CAPEX Over Ind. CAPEX Growth					**3%**	**(5.30)**					**9%**	**(5.05)**
Growth CAPEX Over Past Three Yrs.					**8%**	**(14.65)**					**8%**	**(4.94)**
Growth CAPEX Over Past Two Yrs.					**13%**	**(21.98)**					**17%**	**(8.80)**
Growth in Advertising Expenses			**8%**	**(12.54)**	7%	(8.66)					**9%**	**(3.89)**
Growth in Book Equity					**13%**	**(27.94)**					**7%**	**(4.97)**
Growth in Long-term Op. Assets					**12%**	**(21.26)**					**15%**	**(8.00)**

Table 3. Continued.

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?	
Growth in Number of Employees			2%	(4.40)	**9%**	**(21.21)**					**6%**	**(4.68)**
Growth in Number of Shares					**11%**	**(23.05)**					**19%**	**(10.66)**
Growth in Order Backlog to Total Assets											**10%**	**(2.04)**
Growth of Liabilities			5%	(11.29)	**10%**	**(17.05)**					**6%**	**(3.09)**
Growth of Mkt. Value Over Stock Ret.					**17%**	**(26.88)**					**25%**	**(11.53)**
IPO Age					**8%**	**(2.50)**						
IPO and No R&D Expenses			**19%**	**(11.46)**					**41%**	**(6.18)**		
Idio. Risk Based on CAPM					**19%**	**(35.68)**					**26%**	**(16.44)**
Idio. Risk Based on CAPM Past Daily Data					**32%**	**(46.51)**					**37%**	**(19.18)**
Idio. Risk Based on FF3					**20%**	**(36.98)**					**28%**	**(17.42)**
Idio. Skewness Based on FF3			**12%**	**(36.80)**	6%	(14.51)					**5%**	**(3.83)**
Illiquidity	**6%**	**(12.51)**	4%	(15.24)					**3%**	**(3.95)**		
Implied Volatility Smirk											**18%**	**(5.44)**
Inc. Taxes to Tax Share of Net Inc.												
Ind. Concentration Based on Assets			**9%**	**(19.66)**					**9%**	**(5.92)**		
Ind. Concentration Based on Book Value			**10%**	**(21.70)**					**5%**	**(3.41)**		
Ind. Concentration Based on Revenues			**9%**	**(21.59)**					**8%**	**(5.81)**		
Ind. Momentum												
Ind. Ret. of Big Firms												
Indicator of Initial Public Offerings			9%	(21.80)	**13%**	**(26.27)**			9%	(5.85)	**18%**	**(9.25)**
Inst. Ownership Among Highly Shorted Stocks												
Inst. Ownership and Forecast Dispersion												
Inst. Ownership and Idio. Volatility												
Inst. Ownership and Mkt. to Book												
Inst. Ownership and Turnover												

Table 3. Continued.

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?	
Intangible Ret. Using Book-to-Mkt.			**19%**	**(46.88)**	17%	(34.07)			**11%**	**(8.52)**	4%	(2.51)
Intangible Ret. Using Earn. to Mkt. Value			**18%**	**(40.41)**	14%	(26.30)					**5%**	**(2.98)**
Intangible Ret. Using Net Inc. to Mkt. Value			**19%**	**(42.27)**	13%	(23.89)						
Intangible Ret. Using Sales to Mkt. Value			**24%**	**(60.3)**	12%	(25.30)			**14%**	**(11.05)**		
Intermediate Momentum					**15%**	**(20.13)**					**24%**	**(11.45)**
Inventory Growth			5%	(16.67)	**6%**	**(17.29)**						
Inverse Stock Price	3%	(4.81)	**16%**	**(56.34)**	4%	(11.71)			**9%**	**(10.01)**		
Junk Stock Momentum					**2%**	**(2.19)**					**11%**	**(4.13)**
Lagged 2-month Trading Volume	**3%**	**(6.86)**	3%	(13.96)					**3%**	**(3.83)**		
Leverage Component of Book-to-Mkt.	**5%**	**(4.88)**										
Long-run Reversal			**10%**	**(27.98)**	9%	(19.61)			**4%**	**(3.48)**		
Long-term EPS Forecast					**9%**	**(13.91)**					**10%**	**(4.88)**
Maximum Ret. over Previous Month					**18%**	**(37.66)**					**25%**	**(15.93)**
Median EPS Forcast to Stock Price					**49%**	**(37.05)**					**28%**	**(7.55)**
Medium-run Reversal			4%	(9.37)	**12%**	**(23.60)**					**8%**	**(5.11)**
Mkt. Leverage			12%	(32.30)	**21%**	**(45.87)**					**18%**	**(12.03)**
Mkt. Value to Liquid Assets			**23%**	**(66.37)**	8%	(21.25)			**16%**	**(13.16)**		
Mohanram G-score					**3%**	**(3.13)**			**16%**	**(5.09)**	11%	(2.95)
Momentum (12-Month)					**12%**	**(14.95)**					**16%**	**(8.11)**
Momentum (6-Month)					**9%**	**(13.64)**					**13%**	**(7.17)**
Momentum Based on FF3 Residuals												
Momentum Filtered by Firm Age					**5%**	**(3.43)**						
Momentum in High Volume Stocks					**6%**	**(6.90)**					**11%**	**(4.91)**
Monthly Trading Volume Trend					**10%**	**(17.43)**					**23%**	**(11.70)**
Net Debt Financing					**8%**	**(15.17)**			**4%**	**(2.62)**	3%	(1.98)

Table 3. Continued.

	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					
Firm Characteristic	Safety Words?		Exuberance Words?		Lottery Words?		Safety Words?		Exuberance Words?		Lottery Words?	
Net Debt to Mkt. Value												
Net Equity Financing					**32%**	**(56.76)**					**32%**	**(17.22)**
Net External Financing					**32%**	**(60.38)**					**39%**	**(20.29)**
Net Inc. to Book Equity					**29%**	**(48.98)**					**30%**	**(16.10)**
Net Mkt. Cash Inflow					**27%**	**(30.79)**					**30%**	**(12.15)**
Net Op. Assets					**7%**	**(12.76)**						
Net Payout Yield					**26%**	**(36.22)**					**41%**	**(15.48)**
Normalized Δ Revenue Per Share												
Number of No Trade Days (Monthly)					**7%**	**(28.55)**					**8%**	**(10.92)**
Numer of Inst. Owners												
O Score					**35%**	**(39.17)**					**59%**	**(19.71)**
Off-season Momentum					**7%**	**(10.48)**					**12%**	**(6.69)**
Off-season Reversal Yrs. 11-15			5%	(10.60)	**8%**	**(13.29)**					**4%**	**(2.06)**
Off-season Reversal Yrs. 16-20			**9%**	**(14.33)**	8%	(11.95)						
Off-season Reversal Yrs. 2-5			4%	(10.67)	**15%**	**(33.37)**					**10%**	**(6.66)**
Off-season Reversal Yrs. 6-10					**14%**	**(25.87)**						
Op. Accruals to |Inc. Before Extraord. Items|												
Op. Cash Flow to Mkt. Value					**27%**	**(33.38)**					**33%**	**(13.74)**
Op. Leverage	**53%**	**(39.76)**					**27%**	**(5.90)**				
Op. Profits to Book Equity					**18%**	**(29.63)**					**21%**	**(9.36)**
Option Volume to Average			**3%**	**(5.44)**			**12%**	**(2.88)**	5%	(2.33)	11%	(4.58)
Option Volume to Stock Trading Volume			2%	(5.31)	**2%**	**(6.29)**						
Order Backlog to Average Total Assets					**7%**	**(6.30)**						
Organizational Capital									**8%**	**(4.15)**		
Pastor-Stambaugh Liquidity Beta					**11%**	**(20.36)**					**19%**	**(11.23)**

Table 3. Continued.

	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					
Firm Characteristic	Safety Words?		Exuberance Words?		Lottery Words?		Safety Words?		Exuberance Words?		Lottery Words?	
Payout Yield					**11%**	**(16.14)**					**9%**	**(4.08)**
Pension Funding Status												
Piotroski F-score					**11%**	**(3.39)**						
Predicted Analyst Earn.' Forecast Errors			11%	(16.71)	**13%**	**(15.23)**					**15%**	**(5.07)**
Price delay Coefficient			**4%**	**(9.92)**	4%	(8.23)			5%	(3.42)	**5%**	**(3.15)**
Price delay R-Square	**11%**	**(12.89)**	2%	(5.89)			**16%**	**(4.51)**	7%	(4.32)		
Price delay with Standard Error Adjusted					**3%**	**(6.12)**			3%	(2.41)	**4%**	**(2.52)**
Put Volatility Minus Call Volatility					**6%**	**(7.94)**					**21%**	**(9.37)**
R&D Ability	**37%**	**(11.51)**	25%	(15.53)					**17%**	**(3.33)**		
R&D Expense to Mkt. Value	3%	(3.13)	**5%**	**(11.74)**					**17%**	**(11.73)**		
Real Dirty Surplus	2%	(2.53)	**2%**	**(7.39)**								
Real Estate Holdings					**8%**	**(10.53)**					**20%**	**(7.27)**
Ret. On Assets					**42%**	**(66.17)**					**51%**	**(24.04)**
Ret. Over Earn. Announcement Period											**7%**	**(4.64)**
Ret. Seasonality Over Last Yr.					**16%**	**(24.50)**					**19%**	**(9.47)**
Ret. Seasonality Over Past 11-15 Yrs.					**10%**	**(14.30)**					**12%**	**(5.55)**
Ret. Seasonality Over Past 16-20 Yrs.			4%	(5.95)	**11%**	**(13.79)**					**9%**	**(3.88)**
Ret. Seasonality Over Past 2-5 Yrs.					**14%**	**(21.04)**					**22%**	**(10.29)**
Ret. Seasonality Over Past 6-10 Yrs.					**8%**	**(11.86)**					**14%**	**(6.32)**
Ret. Skewness			**10%**	**(29.64)**	8%	(19.04)					**8%**	**(5.84)**
Ret. of Non-Conglomerates to Total Assets of Conglomerates												
Revenue Growth Rank			6%	(15.24)	**11%**	**(21.45)**						
Rolling Std. Dev. of Trading Volume												
Sales Growth Over Inventory Growth					**3%**	**(5.02)**						
Sales Growth Over Overhead Growth					**16%**	**(21.43)**					**12%**	**(5.79)**

Table 3. Continued.

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?	
Share Turnover Volatility					**28%**	**(18.54)**					**13%**	**(3.78)**
Short Interest to Shares Outstanding					**11%**	**(24.10)**					**10%**	**(7.51)**
Short-Term Reversal					**14%**	**(29.32)**					**13%**	**(8.95)**
Size	**7%**	**(14.65)**	5%	(19.34)					**3%**	**(3.97)**		
Streaks in Earn. Surprise					**10%**	**(11.29)**					**24%**	**(8.72)**
Suppliers Momentum					**4%**	**(3.21)**						
Systematic Risk Volatility					**7%**	**(15.61)**					**10%**	**(6.88)**
Tail Risk Beta					**6%**	**(10.09)**	**29%**	**(8.37)**				
Takeover Vulnerability	**28%**	**(2.61)**										
Total Accruals					**15%**	**(31.00)**					**8%**	**(5.38)**
Total Accruals to Absolute Net Inc.											**6%**	**(3.12)**
Total Assets to Mkt. Value			17%	(52.72)	**20%**	**(49.78)**			7%	(6.19)	**11%**	**(8.72)**
Volume to Mkt. Equity					**5%**	**(8.97)**					**9%**	**(6.40)**

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). Panel D reports results based on our self-defined wordlists as detailed in Section 5.2.2. Panels E and F report results based on a different denominator and based on the numerator as detailed in Section 5.2.3, 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: Self-Defined Wordlists				
Sell-Side Analyst Reports	16% [10%]	17% [12%]	66% [55%]	23%
Seeking Alpha Articles	13% [9%]	15% [15%]	55% [48%]	28%
Panel E: Alternate Denominator				
Sell-Side Analyst Reports	9% [7%]	29% [17%]	69% [58%]	19%
Seeking Alpha Articles	6% [6%]	15% [10%]	58% [54%]	29%
Panel F: 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? – 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%	47%	55%	23%
	[12%]	[27%]	[37%]	
Seeking Alpha Articles	10%	24%	32%	45%
	[10%]	[19%]	[26%]	
Panel B: Less Likely to be Irrational Beliefs Based / More Likely to be Non-Traditional Preferences Based				
Sell-Side Analyst Reports	1%	11%	82%	14%
	[1%]	[6%]	[79%]	
Seeking Alpha Articles	2%	6%	84%	13%
	[2%]	[2%]	[82%]	

Table 6
Do Analyst and SA Views Reflect Views of the General Investor Population?

This table reports results from regressions of DGTW-adjusted stock returns on the tones of sell-side analyst reports and articles published on Seeking Alpha. The sample includes 12,044,943 stock/day observations from January 2006 through October 2021. To construct $Tone_{\text{Sell-Side Analysts}}$, we compute for each stock, at the end of each day, the average tone across all sell-side analyst reports published on the corresponding stock on the corresponding day. Tone is the number of positive words in the report minus the number of negative words divided by the total number of words in the report. We account for negation. To construct $Tone_{\text{Seeking Alpha}}$, we compute for each stock, at the end of each day, the average tone across all Seeking Alpha articles published on the corresponding stock on the corresponding day. $Sentiment_{\text{Dow Jones Newswires}}$ is the average composite sentiment score ("CSS") in Ravenpack across Dow Jones Newswires on the corresponding stock on the corresponding day. We also construct $I_{\text{Sell-side Analysts}}$, $I_{\text{Seeking Alpha}}$, and $I_{\text{Dow Jones Newswires}}$, which equal one if there are sell-side analyst reports, Seeking Alpha articles and Dow Jones Newswires published on the corresponding stock on the corresponding day, respectively. $Tone_{\text{Sell-Side Analysts}}$, $Tone_{\text{Seeking Alpha}}$, and $Sentiment_{\text{Dow Jones Newswires}}$ are set to zero when there are no sell-side analyst reports, no Seeking Alpha articles, and no Dow Jones Newswire published on the corresponding stock on the corresponding day, respectively. *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)
$Tone_{\text{Sell-Side Analysts}}$	0.443***		0.441**
	(57.63)		(57.49)
$Tone_{\text{Seeking Alpha}}$		0.234***	0.212***
		(16.86)	(15.35)
$Sentiment_{\text{Dow Jones Newswires}}$	0.043***	0.044***	0.043**
	(55.65)	(55.95)	(55.63)
$I_{\text{Sell-Side Analysts}}$	0.001***		0.001***
	(6.71)		(6.76)
$I_{\text{Seeking Alpha}}$		0.000	0.000
		(1.13)	(0.84)
$I_{\text{Dow Jones Newswires}}$	0.001***	0.001***	0.001***
	(14.48)	(15.01)	(14.51)
# Obs.	12,044,943	12,044,943	12,044,943
Adj. R^2	0.002	0.001	0.002