WHY DO INVESTORS LIKE SHORT-LEG SECURITIES?

EVIDENCE FROM A TEXTUAL ANALYSIS OF BUY RECOMMENDATIONS

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

This Draft: June 2025

Our paper examines 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 ("short-leg securities"). We conduct textual analysis and test whether the justifications provided in these buy recommendations primarily (1) emphasize a stock's safe-haven quality, (2) indicate general investor exuberance, or (3) highlight a stock's lottery-like features. We find that the buy recommendations mostly emphasize stocks' lottery-like characteristics. We subsequently validate our text-based inferences through a one-time survey of institutional investors and retail investors with long positions in short-leg securities. Overall, our results suggest that perceived upside potential plays a material role in explaining why investors invest in stocks that reside in short legs of anomalies.

JEL Classification: G11, G12, G14, G40, G50.

Keywords: Textual Analysis, Survey Analysis, Anomalies, Risk, Behavioral Finance.

---

* Chen is affiliated with the Faculty of Business and Economics, The University of Hong Kong, Hong Kong. Hwang is with NTU-Singapore. Peng is with the Central University of Finance and Economics, Beijing. E-mail: chen19@hku.hk, bh.hwang@ntu.edu.sg and zz.peng@cufe.edu.cn. The authors would like to thank the Editor and the Reviewers, Nick Barberis, Andrew Chen, Tony Cookson, Zhi Da, Griffin Jiang, Lawrence Jin, Kelvin Tan, Chishen Wei, seminar participants at the Chinese University of Hong Kong, Hong Kong Polytechnic University, Hong Kong University of Science and Technology, Korea University, National University of Singapore, KAEA, the 2021 GSU Virtual Seminar Series, and audiences at the 2022 AFBC, the 2022 USC Social & Behavioral Finance Conference, the 2023 SFFS and the 2024 ABFER Annual Meeting for their 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 developed a wide range of theories to explain the behavior of financial markets (e.g., Lucas 1978; Breeden 1979; Barberis, Shleifer, and Vishny 1998; Daniel, Hirshleifer, and Subrahmanyam 1998; Barberis and Shleifer 2003). These theories differ markedly in the assumed investor preferences and belief formation processes, raising the natural question of which framework best captures the real world. This question is difficult to resolve, however, as these models, despite their stark differences in assumed investor behavior, make similar predictions. It is thus difficult to differentiate between the various theoretical frameworks based on observational data alone (e.g., Liu et al. 2022).

In response to this challenge, researchers increasingly turn to investor surveys. Surveys allow for the direct elicitation of investors' subjective beliefs, which, in turn, can shed light on which theory best captures investors' decision-making processes. For example, surveys can reveal whether investors consider rare disaster risk when selecting stocks (e.g., Choi and Robertson 2000; Bender et al. 2022; Chinco, Hartzmark, and Sussman 2022) or whether they extrapolate from past returns (e.g., Fisher and Statman 2000; Vissing-Jorgensen 2003; Greenwood and Shleifer 2014; Liu et al. 2022).

While surveys can provide valuable insights, they are also costly and logistically challenging to administer at scale. In this paper, we propose a complementary method to capture investors' subjective beliefs that is both cheaper and scalable. Financial market participants frequently express their opinions in writing. Some of these written expressions, such as analyst reports and social media posts, are publicly accessible. We posit that by parsing these textual data and extracting the key considerations that drive the authors' opinions, we can make progress in better understanding which framework best captures investors' thoughts and considerations and, consequently, the behavior of financial markets.

To test our proposition, we focus on one of the most prominent empirical regularities: the cross-section of average stock returns. As we sort stocks each month based on certain firm characteristics, form decile portfolios, and compute the ensuing average raw returns for each portfolio, we observe a monotonically increasing (or decreasing) pattern in portfolio returns. Most of the predictability comes from the stocks in the outer decile earning very low returns (e.g., Stambaugh, Yu, and Yuan 2012). In the literature, these stocks are referred to as "short-leg securities" or as stocks residing "in the short leg" of an "anomaly."

There are at least three theoretical frameworks that explain the low average returns of short-leg securities. First, investors may view short-leg securities as "safe-haven assets," that is, stocks that perform relatively well during bad states of the world. These investors rationally pay higher prices and accept lower average returns for the insurance-like properties these stocks provide (e.g., Lucas 1978; Breeden 1979). Alternatively, investors may exhibit a tendency to become exuberant about short-leg securities because of extrapolative expectations (e.g., Barberis, Shleifer, and Vishny 1998; Barberis and Shleifer 2003) or because of overconfidence in private signals (Daniel, Hirshleifer, and Subrahmanyam 1998). These tendencies, combined with short-sale constraints, can push prices above fundamentals and lead to poor future performances (e.g., Barberis and Thaler 2003). Finally, investors may perceive short-leg securities to exhibit lottery-like features. These investors, knowingly or not, end up paying a premium for the small chance of earning extreme payoffs (e.g., Barberis 2018).[1]

To assess which of these theories best explains investor behavior, we analyze two large corpora of stock opinion articles. The first corpus includes 1,718,353 analyst reports covering 7,009 U.S. stocks. The second contains 141,746 Seeking Alpha (SA) articles covering 5,793 U.S. stocks. We conjecture that analyst reports primarily reflect institutional investors' opinions, while SA articles mostly capture retail investors' perspectives. Our sample spans the period from January 2006 through December 2021.

We build on Chen and Zimmermann (2022), who survey and replicate the anomalies literature. For each of 186 anomalies, we compile all analyst- and SA buy recommendations for the stocks that – as of the recommendation issuance month – reside in the short leg. For example, each month, we identify all stocks that reside in the top asset growth decile and compile all corresponding analyst and SA buy

[1] There are other possible theories. Our study focuses on these three because they are among the most prominent explanations in the literature. We discuss additional theories in Sections 4.2 and 6.2.4. In this paper, we use the terms "lottery" and "upside potential" interchangeably.

recommendations (Cooper, Gulen, and Schill 2008). We then examine whether, in explaining why they like high-asset-growth stocks, analysts and SA contributors unusually fixate on the stocks' perceived safe-haven quality, perceived supremacy, or perceived lottery-like features.

We measure these unusual fixations using a dictionary-based approach. We construct three wordlists by collaborating with CoreData Research, a market research firm that conducts investor surveys for large financial institutions. In particular, we ask one hundred US-based institutional investors to list five words that they would use to describe (1) "*a stock that, to you, is a 'safe-haven asset': a stock that does relatively well when times are bad*"; (2) "*a stock that has been doing well and that you expect will continue to do very well or, in general, a stock that you are very confident will earn above-normal returns*"; and (3) "*a stock that offers somewhat of a gamble: the stock will most likely not produce above-normal returns, but, if it does, the payoff will be enormous*."

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

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. We later confirm the robustness of our results using alternative wordlists and a more advanced natural language processing (NLP) model in place of the dictionary-based approach.

Our main results are as follows: We find that analysts' rationales for liking short-leg securities most strongly point to perceived safety for 15 out of the 186 firm characteristics, or 8% of the time. For instance, we find that the buy recommendations for stocks with low operating leverage unusually heavily rely on safety words. We observe no abnormal use of supremacy or lottery words. These patterns suggest the key reason investors like low-leverage stocks is their perceived safety, and that this may explain why these stocks trade at comparatively high prices and earn low returns on average (Novy-Marx 2011).

In 17% of all cases, analysts like short-leg securities mostly for their perceived supremacy. For instance, we find that the buy recommendations for stocks with high returns over the past three years most notably stand out for their unusually heavy reliance on supremacy words.

55% of the time, analysts primarily like the upside potential they see in the corresponding short-leg securities. For instance, we find that analysts substantially more frequently use lottery words for high-asset growth stocks than for stocks that do not have high asset growth. We observe no reliable differences in the use of safety or supremacy words.

19% of the time, the results are inconclusive as we detect no abnormal use of either safety, supremacy, or lottery words in the buy recommendations for short-leg securities. For SA articles, the corresponding fractions are 6% (perceived safety), 11% (perceived supremacy), and 54% (perceived lottery). The results are inconclusive 30% of the time.

Overall, a comparison of the fractions across analyst reports and SA articles suggests that while perceived safety, supremacy, and lottery all contribute to investors' attraction to short-leg securities, perceived upside potential emerges as the dominant factor.

Our method and inferences hinge on the assumption that the beliefs expressed in analyst reports and SA articles at least partially reflect those of the investor population. To assess the validity of this assumption, we conduct a one-time survey of investors with actual long positions in short-leg securities. Working with the same market research firm, CoreData Research, we run a separate survey on 450 institutional investors with substantial assets under management (AUM). We also recruit 314 US retail investors.

We ask respondents to list up to eight stocks they bought in the past year and explain the main reason for each purchase. We analyze responses separately for stocks that resided in the short leg of a specific anomaly (as of the time of the survey) and those that did not. We then test for each anomaly whether our investors disproportionately frequently report having bought the corresponding short-leg securities for their perceived safety, supremacy, or lottery.

Two key findings emerge. First, for 64% of anomalies, institutional investors disproportionately frequently report having bought the corresponding short-leg securities for their perceived lottery-like features. Retail investors do so in 81% of cases. These figures closely align with our text-based results. Second, we observe a strong positive correlation between our text-based and survey-based inferences of the primary reason investors like stocks that reside in the short leg of a particular anomaly.

On the whole, we acknowledge that our proposed method of extracting investors' subjective beliefs from analysts' and bloggers' written expressions can be less precise than directly asking investors about their beliefs. However, the results from our validation exercise suggest that our text-based approach is not entirely futile and can provide meaningful insights into investor behavior and stock market dynamics.

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

2. Literature Review and Contribution

2.1 Investor Surveys and Textual Analyses

Our paper closely relates to recent studies using surveys to better understand investors' decision-making. Choi and Robertson (2020) survey US households and ask what factors determine their equity allocations. Relatedly, Chinco, Hartzmark, and Sussman (2022) survey US investors to test the finance textbook prediction that the correlation between stock returns and consumption growth plays a material role in investors' decision-making. One of their key findings is that "*only 11% reported thinking about consumption-growth correlations in a manner consistent with textbook theory*" (page 2186). Bender et al. (2022) survey US individuals with at least $1 million in investable assets to examine how well leading academic theories explain the beliefs and investment decisions of "millionaires." The authors find that their beliefs about financial markets are similar to those of the average US household.

Liu et al. (2022) conduct a survey of Chinese retail investors, which they link to transaction data from the Shenzhen Stock Exchange, to compare competing behavioral explanations for excessive trading. The authors identify perceived information advantage and gambling preference as the dominant drivers of excessive trading.

The key contribution of our paper is the proposition of a complementary method to uncover what causes investors to make certain decisions. Financial market participants constantly express their views and reasoning in various forms of text, many of which are publicly available. By parsing these texts, we posit that we can obtain a good perspective into investors' minds without having to directly ask them.

The primary limitation of our approach is that even advanced NLP models, currently, cannot match surveys in their accuracy. The key strength of the text-based approach lies in its cost efficiency and scalability, as it can be applied continuously to an almost unlimited volume of textual data. Unlike surveys, our method can also be applied retroactively to historical data to understand what may have driven investors' decisions in the past. Ultimately, it appears to us that both surveys and textual analysis offer valuable and complementary tools for future research on investor behavior.

By parsing written stock recommendations, our paper also relates to the textual analysis literature in finance. The textual analysis literature, so far, primarily 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 goes a level deeper and examines *why* investors like a particular stock and what the *why* tells us about investors' decision-making processes.

2.2 The Cross-Section of Expected Stock Returns and Behavioral Finance

By exploring why investors are drawn to short-leg securities, our paper also adds to the debate of what determines the cross-section of expected stock returns (Fama and French 1992, 1996). The empirical asset-pricing literature has documented many cross-sectional stock-return predictabilities. Recently, there has been increasing debate about whether these predictabilities are real or economically important (e.g., Harvey, Liu, and Zhu 2016; Linnainmaa and Roberts 2018; Hou, Xue, and Zhang 2020).

Our study builds on McLean and Pontiff (2016), Jacobs and Müller (2020), Chen (2021), and Jensen, Kelly, and Pedersen (2023) and assumes that the cross-sectional differences in average returns are real; short-leg securities are thus distinct not only in the minds of some econometricians but also in the minds of investors. Our finding that there are systematic differences in how sell-side analysts and SA contributors describe short-leg securities is consistent with this conjecture.

Assuming that the cross-sectional differences are real, we contribute to the empirical asset-pricing and behavioral finance literatures by estimating the extent to which the low returns of short-leg securities, a key feature of the cross-section of expected stock returns, stem from differences in perceived safety, supremacy, or lottery.

3. Data and Variables

This section describes our key data sources and our wordlists.

3.1 The Three Frameworks

There are at least three possible explanations for what perceptions draw investors to short-leg securities. To motivate our data and the construction of our wordlists, we begin with a brief outline of these three frameworks.

3.1.1 The Traditional Risk Framework

In the traditional finance paradigm, investors willingly pay a high price and accept low average returns for stocks they perceive as providing insurance against adverse future economic conditions. Under this framework, investors must therefore view short-leg securities as providing especially strong insurance against unfavorable states. The finance and economics literature identifies several forms these bad states can take, including low consumption growth (Lucas 1978; Breeden 1979), economic disasters (Barro 2006), and heightened uncertainty about the standard of living (Bansal and Yaron 2004), among others.

3.1.2 The Irrational Beliefs Framework

In contrast, behavioral finance contends that many of the observed stock market patterns reflect systematically mistaken beliefs coupled with limits to arbitrage.

There are two reasons investors may systematically form incorrect beliefs. The first reason is that people extrapolate, in the sense that "*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 and that the average belief negatively predicts future returns, suggesting that investors *over*-extrapolate. Theoretical models show that the extrapolation of past stock returns or past growth in fundamentals, coupled with frictions, can generate many cross-sectional return patterns such as momentum and long-run reversal (e.g., Barberis, Shleifer, and Vishny 1998; Barberis and Shleifer 2003; Barberis et al. 2015, 2018).

The second reason investors may systematically form incorrect beliefs is that they 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 part of the cross-section of expected stock returns.

To the extent that short-leg securities more frequently experience strong past growth or incite greater overconfidence in positive private signals, irrational beliefs, coupled with short-sale constraints, offer another explanation for the low average returns of short-leg securities.

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 (Kahneman and Tversky 1979; Tversky and Kahneman 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 assign a greater "decision weight" to low-probability events at the tails of a distribution ("probability weighting"). These greater decision weights do not represent mistaken beliefs. Instead, they reflect how people psychologically react to the likelihood of small-probability outcomes: A 1-in-1,000 chance of gaining $10,000 feels more attractive than earning a certain $10. Moreover, a 1-in-1,000 chance of losing $10,000 feels scarier than losing a certain $10. Probability weighting can explain why people invest in both lotteries and insurance policies, a behavior that is difficult to capture under the traditional Expected Utility paradigm.

To the degree that investors believe the return distributions of short-leg securities resemble those of lotteries, probability weighting offers another explanation for why investors may be willing to pay a premium for these stocks.

While the irrational beliefs framework mostly underscores broad investor enthusiasm, probability weighting emphasizes the allure of extreme, positive, small-probability events. Probability weighting can therefore explain why certain stocks, such as those with high failure probability, can become overpriced even when there is little ground for general exuberance.[2]

3.2 Survey-Based Wordlists

To examine whether investors are primarily drawn to short-leg securities for their perceived safety, supremacy, or lottery, we parse written buy recommendations with a dictionary-based approach.

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 in October 2022. To reach institutional investors, we collaborate with CoreData Research

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

(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; 41% report having more than $1 billion in AUM. 97% of the institutional investors in our sample have more than ten years of work experience; 69% have more than twenty years of work experience.

Our key questions are as follows:

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

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

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

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

Our first question asks what words our survey participants would use to describe a stock that provides insurance against bad states of the world. Our second question elicits words of general excitement tied to extrapolation and overconfidence. Our third question asks what words our survey participants would use to describe the lottery-like features of a stock that will most likely not produce above-normal returns.[3]

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

Despite our instructions, some investors provide company names and industries. We delete these terms. We 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*.[4]

We consider all possible word forms of our base terms, including plural forms, noun forms, verb forms, adjective forms, adverb forms, and verb conjugations.[5] 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 supremacy 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, supremacy, and lottery words in Panel A of Table 1.

Our wordlists and the resulting inferences are subject to an important caveat. The objective of question Q3 in our survey is to elicit words people use when describing stocks they find appealing because of their cumulative prospect theory preferences. However, the use of lottery-related terms is not a sufficient condition for cumulative prospect theory preferences.

Suppose we observe that investors unusually heavily rely on lottery words as they explain why they like a particular stock. Such language would certainly be congruent with cumulative prospect theory: Investors may psychologically react strongly and assign high decision weights to the possibility of extreme positive outcomes and, as a result, extensively discuss the lottery-like features of a stock.

[4] When comparing the supremacy with the lottery words, it appears that the former are more backward-looking, while the latter are more forward-looking. One implication is that investor beliefs about future excellence are rooted in events and trends that have already occurred. But when thinking about lottery features, investors are mostly guided by the imagination of possible events in the future.

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

But such language may also reflect overly optimistic beliefs about the likelihood of these extreme outcomes materializing. That is, while the actual probability of a large positive return becoming realized is low, investors may erroneously believe that it is relatively high and thus make frequent references to it.[6]

It is difficult to discern with our method whether an unusually heavy reliance on lottery terms stems from prospect-theory-driven overweighting of small probabilities or from an unrealistic belief about the right tail of the return distribution. In reality, people may exhibit both cumulative prospect theory preferences and irrational beliefs about the tail of a distribution, which only exacerbates their preference for lottery stocks. All of our ensuing results should be interpreted with this caveat in mind.

3.2.2 Retail Investors' Survey-Based Wordlists

One could argue that the language institutions use to describe safe, superior, or lottery-like stocks differs from that of retail investors. Consequently, while the institutional investor wordlists may be well-suited for parsing analyst reports, they may be less appropriate for SA articles, which are primarily written by retail investors for retail investors. To gauge the sensitivity of our results, we therefore also create wordlists based on an online survey sent to 303 US-based retail investors in November 2022. We recruit retail investors through Prolific (https://www.prolific.co), a platform that enables researchers to recruit prescreened participants for online surveys and experiments. Online Appendix Table A2 provides some descriptive statistics for the participants in the retail investor survey.

As displayed in Online Appendix Figure A2, the survey we send to retail investors is very similar to the one sent to institutional investors. The five most frequently listed safety terms by retail investors are *reliable*, *safe*, *secure*, *stable*, and *steady*; the five supremacy terms are *consistent*, *excellent*, *growth*, *innovative*, and *winner*; and the five lottery terms are *exciting*, *gamble*, *potential*, *speculative*, and *volatile*.

[6] Unlike most theoretical models on biased beliefs, which emphasize general investor exuberance and the center of the belief distribution, here, the mistaken belief concentrates on the tail of the distribution.

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

3.3 Stock Opinion Articles

We apply our lists of safety, supremacy, and lottery words to two types of written stock opinions: analyst reports and SA articles. We focus on reports and articles written about common shares that trade on the NYSE, AMEX, or NASDAQ. Our sample period starts in January 2006 and ends in December 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 in portable document format (PDF). 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 account for 98% of all analyst reports. The reason we focus on the top 100 brokerages is that extracting the text in the analyst reports 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.

We retain only analyst reports that are issued for a single company.[7] Our final sample comprises 1,718,353 analyst reports. Of these, 1,032,917 represent buy recommendations, 625,232 represent hold

[7] Investext tags each report by ticker(s), brokerage firm, and date, so that Investext users may search for and retrieve their desired reports. However, we caution future researchers that Investext's tagging is both inaccurate and systematically biased. First, Investext removes ticker tags when a company delists, resulting in a substantial number of reports without ticker tags and introducing survivorship bias. Second, Investext may assign multiple tickers to a report, even when the report focuses on a single company and mentions other firms only as industry peers. In our study, we directly extract the primary ticker(s) listed in each analyst report PDF.

recommendations, and 60,204 represent sell recommendations. 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,[8] and number of pages.

For each article, we calculate the number of safety words scaled by the total number of words, *Safety [%]*. Similarly, we calculate the number of supremacy words scaled by the total number of words, *Supremacy [%]*, 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[9] (*no*, *not*, *none*, *neither*, *never*, *nobody*, *few*, *little*, *less*,[10] *low*, *hardly*, *rarely*) occurring within three words preceding a word from our wordlists.

In the end, we find that the average *Safety [%]*, *Supremacy [%]*, and *Lottery [%]* across all analyst buy recommendations based on the institutional investors wordlists are 0.061%, 0.372%, and 0.255%, respectively; the average report length is 1,146 words.[11]

To provide readers with a sense of what analyst reports with high fractions of safety, supremacy, or lottery words look like, we show in Online Appendix Figure A5 the first pages of analyst reports with a high *Safety [%]*, *Supremacy [%]*, and *Lottery [%]*, respectively. The red boxes in Online Appendix Figure A5 also illustrate which sections of the analyst reports we extract for parsing.

The first report in Online Appendix Figure A5 contains 1,184 words, of which six are safety words. The report has no supremacy 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 perceived 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 … we*

[8] 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.
[9] We take the first six words from Loughran and McDonald (2011) and add to them the ensuing six words.
[10] 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.
[11] Online Appendix Figures A3 and A4 present word clouds for our safety, supremacy, and lottery words across analyst reports and SA articles.

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 supremacy 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 supremacy words. The analysts recommend that investors buy Lone Pine Resources (LPR), an oil and gas company. Unlike in the second report, the analysts are not uniformly positive and emphasize the uncertainty that accompanies any investment in LPR. At the same time, the analysts highlight the upside potential and conclude: "*We see near-term headwinds from investor perceptions of high exposure to natural gas as well as from reduced production and increased cost guidance, but we believe the shares will find appeal among more risk-seeking investors as well as investors comfortable with a management team that is still establishing a track record with a newly-independent company*."

3.3.2 SA Articles

Our second source of stock opinions is SA. Any user can submit a stock opinion article for possible publication on the SA website. A team of editors curates these submissions. If articles are deemed of adequate quality and published on the SA website, 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 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,[12] whether the article is tagged as a "long idea" or a "short idea," the title, and the main text.[13] Our sample comprises 141,746 articles from 2006 through 2021. Of these, 73,061 are tagged as long ideas ("buy recommendations"), 9,530 are tagged as short ideas ("sell recommendations"), and 59,155 are untagged.

The average *Safety [%]*, *Supremacy [%]*, and *Lottery [%]* across SA long ideas based on the institutional investors wordlists after accounting for simple negation are 0.057%, 0.313%, and 0.205%, respectively; the average article length is 1,217 words.

Online Appendix Figure A6 displays the beginning paragraphs of SA articles with a high *Safety [%]*, *Supremacy [%]*, and *Lottery [%]*, respectively. The first article in Online Appendix Figure A6 contains 621 words, of which four are safety words. The article has no supremacy 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 supremacy 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*."

[12] We caution future researchers that the stock ticker tagged to an article by SA sometimes does not match the focal firm's ticker as of the article publication time and thus needs to be corrected.
[13] As with analyst reports, we expand contractions (e.g., *couldn't* → *could not*) and remove digits, punctuations, special characters, tables, and figures.

The third article contains 769 words, of which six are lottery words. The article has no safety words and no supremacy 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 perceived safety, supremacy, and lottery 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.

4.1 Text-Based Evidence of Why Investors Like Short-Leg Securities

Chen and Zimmermann (2022) survey the literature and identify 205 "*clear and likely* [cross-sectional] *predictors*" of raw stock returns in the US. For each predictor, the authors construct long-short portfolios and test whether the mean monthly long-short portfolio return is zero. The authors reject this 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 (https://www.openassetpricing.com/data).[14] 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. We then reinsert price, size, and past one-month returns and restrict our analysis to common shares that trade on the NYSE, AMEX, or NASDAQ from January 2006 through December 2021.

[14] This information pertains to their "August 2024 Data Release."

We adopt the following procedure separately for each of the 205 firm-level characteristics: Each month *t*, we rank stocks based on firm characteristic *i*. The stocks in the bottom decile as of month *t* represent the short-leg securities. We refer to reports and articles published in the same month *t* that recommend that investors buy these short-leg securities as "short-leg recommendations." We refer to all other buy recommendations published in month *t* as "other recommendations." In additional analyses, we repeat this procedure but focus on buy recommendations published in month *t* on stocks that have consistently been in the short leg from months *t*-3 through *t* or from months *t*-6 through *t*.

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 like the respective stocks. To understand the sources of these positive views, we parse all buy recommendations for short-leg securities and test whether these short-leg recommendations unusually fixate on the stocks' safe-haven quality, their seeming superiority, or their perceived lottery-like characteristics.

To this end, we compute the average *Safety [%]* across analysts' short-leg recommendations. To determine whether the use of safety words in short-leg recommendations is abnormally high, we need a benchmark for what constitutes a normal frequency of safety words in a typical analyst buy recommendation. In our main analysis, we take the average *Safety [%]* across analysts' all other buy recommendations as our benchmark. We then calculate the relative difference between the former and the latter and test whether the difference is positive and statistically significant at the 5% level:

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

In additional analyses, tabulated in Online Appendix Table A3, we compute the relative difference when considering all other buy recommendations written by the *same* analyst in the same month on non-short-leg securities. The results based on this alternate measure echo our main findings.

We analogously compute Δ*Supremacy [%]* and Δ*Lottery [%]* and test whether they are positive and statistically significant at the 5% level. We repeat this process for SA articles.

For some of the 205 firm characteristics we consider, there are no analyst reports or SA articles for the stocks in the short leg, either because there are very few stocks in the short leg or because the short-leg securities represent microcap stocks, which sell-side analysts and SA contributors rarely cover. Our final analysis thus comprises 186 firm-level characteristics.

Suppose that, for firm characteristic *i*, the difference in the fraction of safety words was positive and significant. In other words, suppose that in explaining why investors like stocks in the short leg of anomaly *i*, they unusually frequently use safety words. In that case, we would conclude that one important reason investors are drawn to these stocks is their perceived safety. Similarly, suppose that the average *Supremacy [%]* or the average *Lottery [%]* across the short-leg recommendations were abnormally high. In those cases, we would infer that one reason investors like these short-leg securities is their perceived supremacy or lottery, respectively. For some firm characteristics *i*, we observe abnormally high uses of terms from more than one wordlist. In such cases, we label the perception associated with the largest abnormal difference as the one that "primarily" explains investors' liking of the short-leg securities.[15] Finally, for some firm characteristics *i*, we detect no statistically significant difference in the occurrence of either safety, supremacy, or lottery words. We label such results "inconclusive."

In Table 2, we report our findings aggregated across all 186 firm characteristics for the institutional investor wordlists (Panel A) and the retail investor wordlists (Panel B). We first describe our Panel A results. When we apply our institutional investor wordlists and parse analysts' explanations of why they like stocks that reside in the short leg of a particular anomaly, we find that in 22 out of the 186 cases, or 12% of the time, short-leg recommendations exhibit an abnormally high occurrence of safety words. That is, the short-leg recommendations unusually frequently tell a safe-haven story. In 54 out of the 186 cases, or 29% of the time, short-leg recommendations are marked by an abnormally high use of supremacy words. That is, analysts unusually frequently tell stories of continuous growth or positive developments that they are certain will come to fruition. In 126 out of the 186 cases, or 68% of the time, the short-leg

[15] If there is only one difference that is positive and significant, it naturally serves as the primary explanation for investors' liking of short-leg securities.

recommendations exhibit an unusually high occurrence of lottery words. That is, the short-leg recommendations stand out in their unusually heavy emphasis on the stocks' upside potential.

The fractions of times analysts *primarily* explain their liking of short-leg securities through perceived safety, supremacy, and lottery considerations are 8%, 17%, and 55%, respectively. The remaining 20% of the time, the results are inconclusive.

The patterns are similar for SA articles. In their justifications for why investors should buy short-leg securities, SA contributors point to perceived safety 7% of the time, perceived supremacy 16%, and perceived lottery 57%. The fractions of times SA contributors *primarily* explain their buy recommendations through perceived safety, supremacy, and lottery are 6%, 11%, and 54%, respectively. The remaining 30% of the time, the results are inconclusive.

Panel B of Table 2 presents the results based on the retail investor wordlists. The results are similar to those in Panel A. The fractions of times analysts primarily point to perceived safety, supremacy, and lottery are 5%, 18%, and 65%, respectively. For SA articles, the corresponding fractions are 8%, 13%, and 57%, respectively.

Overall, our results in Table 2 indicate that no single perception fully explains why investors are drawn to short-leg securities. For some firm characteristics and their corresponding short-leg securities, perceptions of safety matter most, while for others, perceived supremacy or lottery are the key. However, a comparison of the fractions also shows that – irrespective of whether we consider the institutional investor wordlists, the retail investor wordlists, analyst reports, or SA articles – perceived upside potential is the dominant consideration.

That even institutional investors heavily emphasize the upside potential of short-leg securities may surprise. One possible explanation is that institutions themselves harbor strong non-traditional preferences or exhibit frequent exuberance about the right tail of return distributions. Alternatively, fund managers may not personally prioritize such upside potential but feel compelled to do so because their retail clients do. If the ultimate asset owners value the possibility of extreme positive outcomes, fund managers may, in turn, place greater weight on these features when selecting stocks, and sell-side analysts will assist accordingly (Akbas and Genc 2020; Agarwal, Jiang, and Wen 2022).

4.2 Text-Based Evidence by Economic Categories

To expound on our main result, we take advantage of Chen and Zimmerman's (2022) classification of all 205 firm characteristics into 32 distinct economic categories.[16] These categories range from "Accruals" to "Volume." Some categories, such as "External Financing," comprise many firm characteristics.[17] Other categories, such as "Default Risk," include only one firm characteristic.[18]

Table 3 reports the results separately for each of the 32 economic categories.[19] For brevity, going forward, we report only results based on the institutional investor wordlists. Results using the retail investor wordlists are similar and available upon request.

The results from both analyst reports and SA articles indicate that short-leg securities in the "Information Proxy" category are primarily favored for their perceived safety, while short-leg securities in the "R&D" and "Long-Term Reversal" categories are mostly liked for their perceived supremacy. Perceived lottery emerges as the primary explanation for short-leg securities in the following economic categories: "Cash Flow Risk," "Default Risk," "Earnings Forecasts," "Earnings Growth," "External Financing," "Investment," "Investment Growth," "Profitability," "Short-Term Reversal," "Valuation," and "Volatility."

For the remaining economic categories, such as "Accruals" and "Lead Lag," the results are less conclusive either because there is no uniform pattern or because the results for most firm characteristics are inconclusive. This ambiguity may reflect the inherent noise in our text-based approach. It may also indicate that forces other than perceived safety, supremacy, or lottery matter in explaining investors' liking of the

[16] Chen and Zimmermann (2022) identify 34 categories; however, we merge "Investment" with "Investment (Alternative)" and "Profitability" with "Profitability (Alternative)," reducing the total number of categories in our paper to 32.

[17] Change in Current Operating Liabilities, Change in Financial Liabilities, Composite Debt Issuance, Composite Equity Issuance, Convertible Debt Indicator, Initial Public Offerings, Net Debt Financing, Net Equity Financing, Net External Financing, Share Issuance (1 Year), and Share Issuance (5 Year).

[18] O-Score.

[19] Online Appendix Table A4 reports the results separately for each of the 186 characteristics.

corresponding short-leg securities. For instance, the poor performance of short-leg securities in the "Lead Lag" category may arise because boundedly rational investors are not fully aware of recent negative events and developments, leading them to temporarily hold overly positive perceptions of these stocks ("underreaction to bad news").

On the whole, our economic category results appear tenable. It is conceivable that investors would view high information proxy stocks as safer and, thus, be willing to accept lower returns on average. Similarly, it is theoretically defensible that stocks with high past long-run returns and stocks with high levels of R&D would be particularly susceptible to investor exuberance (Barberis 2018). Finally, it is plausible that investors would view stocks with high volatility, high risk, and high growth as more lottery-like.

To more systematically gauge the sensibility of our results, we conduct the following test. Investors likely see more upside potential in small stocks than in large stocks. In contrast, extrapolative tendencies are likely stronger among larger companies (Barberis 2018). We should thus expect that investors more frequently like short-leg securities for their perceived lottery when short-leg securities have comparatively small market capitalizations, while investors more frequently like short-leg securities for their perceived supremacy when they are comparatively large.

To test these conjectures, we compute, separately for each of the 186 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 Online Appendix Table A5, we report our results aggregated across the 93 firm characteristics whose average market capitalization is above the median ("Less Likely to be Lottery Based"). In Panel B, we report our results aggregated across the 93 firm characteristics whose average market capitalization is below the median ("More Likely to be Lottery Based").

In line with expectations, perceived lottery plays a larger role when the short-leg securities are comparatively small. Consider the analyst-report results. When the short-leg securities are comparatively small, the writing points to perceived lottery 78% of the time compared with 31% when the short-leg securities are large. We observe the opposite pattern for the relevance of perceived supremacy. When the short-leg securities are comparatively small, the writing points to perceived supremacy 6% of the time compared with 28% when short-leg securities are large. The results are similar when considering SA articles.

4.3 Validation of Text-Based Evidence Through One-Time Investor Survey

Our text-based inferences rely on the assumption that the beliefs expressed in analyst reports and SA articles at least partially reflect those of the broader investor population. This assumption is not unreasonable: When sell-side analysts and SA contributors issue buy recommendations and explain their reasoning, they must choose what story to tell. For example, when recommending a high-volatility stock, they must choose whether to pitch the stock as a safe-haven asset, an opportunity to invest in a supreme stock, or a chance to possibly reap tremendous profits in a relatively short period of time.

An extensive literature argues that readers have confirmation bias and prefer consuming news that matches their belief system (Nickerson 1998; Park et al. 2013; Cookson, Engelberg, and Mullins 2023).

If analysts and SA contributors want their reports to be read and if investors exhibit confirmation bias, it is sensible for analysts and contributors to cater to their readers' belief systems. Thus, when analysts and SA contributors choose to tell an upside-potential story for high-volatility stocks, it is conceivable that they do so because they feel that an upside-potential story most strongly resonates with investors' belief system regarding high-volatility stocks. In that regard, analysts' and SA contributors' story choices *should* mirror investors' thoughts and considerations.

To directly test the assumption that the beliefs expressed in analyst reports and SA articles at least partially reflect those of the broader investor population, we conduct a one-time survey of investors with actual long positions in short-leg securities and ask them directly about their motivations. We then compare their responses to our text-based inferences.

We collaborate again with CoreData Research. In March and April 2024, we survey 450 institutional investors. These investors manage even larger portfolios than those used to generate our institutional investor-based wordlists. 69% have AUMs of at least $1 billion. 51% manage more than $5 billion, and 21% manage over $50 billion. All our survey participants actively invest in US stocks. In June 2024, we also survey 314 US retail investors, whom we recruit through Prolific using the same filters detailed in Section 3.2.2.

As illustrated in Online Appendix Figures A7 and A8, we ask investors to reflect on the stocks they purchased over the past year. If they bought more than eight stocks, we instruct them to consider the first eight that come to mind. For each stock, we ask whether their primary motivation for the purchase was the stock's perceived safety, supremacy, or lottery.

For each of the 186 firm characteristics, we then conduct the following analysis. Similar to Table 2, we define "short-leg securities" as stocks that our survey participants report having bought and that, at any point between April 2023 and March 2024 for institutional investors and between July 2023 and June 2024 for retail investors, appeared in the short leg based on firm characteristic *i* Stocks that participants purchased but that never appeared in the short leg based on characteristic *i* during these periods are labeled as "other stocks." The "August 2024 Data Release" of Chen and Zimmerman does not provide data through June 2024 for all firm characteristics. Furthermore, for some firm characteristics, none of our participants reports having purchased any stocks in the corresponding short leg. As a result, we can perform our validation exercise only across 53 firm characteristics for institutional investors and 48 for retail investors. For each of the 53 (48) firm characteristics, we compute *ΔSafety [%]*, which measures the fraction of times institutional investors report having bought the short-leg securities for their perceived safety relative to how often the same investors cite safety as the key reason for purchasing other stocks. We then test whether this difference is positive and statistically significant at the 5% level. We conduct similar tests for *ΔSupremacy [%]* and *ΔLottery [%]*. If *ΔSafety [%]* is positive and significant, we conclude that our survey participants disproportionately purchased short-leg securities for their perceived safety. Similarly, if *ΔSupremacy [%]* or *ΔLottery [%]* is positive and significant, we conclude that our participants disproportionately purchased short-leg securities for their perceived supremacy or lottery-like features, respectively. If none of the differences are statistically significant at the 5% level, we classify the result as inconclusive. If more than one difference is statistically significant, we assign the perception with the largest relative difference as the *primary* explanation for why our survey participants disproportionately bought the short-leg securities.

We perform this analysis for each of the 53 (48) firm characteristics and, in the end, calculate the fractions of times our investors primarily explain their actual purchases of short-leg securities through perceptions of safety, supremacy, or lottery. We also compute how frequently our results are inconclusive.

Table 4 reports the findings separately for institutional investors and retail investors. Institutional investors primarily explain their actual purchases of short-leg securities through perceived safety in 9 out of 53 cases, or 17% of the time. Supremacy and lottery are the key reason 13% and 64% of the time, respectively. For retail investors, perceived safety, supremacy and lottery are the primary reason 6%, 10%, and 81% of the time, respectively.

These figures closely align with our text-based analysis in Panel A of Table 2, where the corresponding fractions are 8%, 17%, and 55% based on analyst reports and 6%, 11%, and 54% based on SA articles. The prevalence of inconclusive results is noticeably lower in Table 4 than in Table 2 (6% and 2% versus 20% and 30%). One possible interpretation is that survey responses yield more precise results than our text-based method.

To further compare our text-based and survey-based inferences, we compute the Proportion Agreement between our text-based inferences from Table 2 and our survey-based inferences from Table 4. Specifically, we ask, for each anomaly, whether the key argument analysts and bloggers cite for recommending short-leg securities (safety, supremacy, or lottery) matches the primary explanation our surveyed investors provide for their actual purchases of those securities.

We find that the Proportion Agreement between analyst reports and institutional investor survey responses is 65%. That is, in 65% of anomalies, the key rationale analysts use in justifying their buy recommendations for short-leg securirities is exactly the same reason institutional investors give for having bought those stocks. The Proportion Agreement between SA articles and retail investor responses is 91%. These values indicate a strong alignment between our text-based inferences and investors' self-reported motivations.

Overall, the results from our one-time survey of institutional and retail investors who actually invested in short-leg securities strongly support the findings derived from our text-based methodology.[20]

5. Additional Analyses

We conduct a series of robustness tests and additional analyses.

5.1 Alternative Scalars

Our main analysis draws inferences based on equation (1), which computes the differences in the fractions of safety, supremacy, 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, supremacy, 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 repeat our main analyses with two alternative measures, both of which tilt the balance towards firm characteristic *j*:

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

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

[20] Another tangential, yet related way to gauge whether the beliefs captured in reports at least partially reflect those of the investor population is to correlate the tone of analyst reports and SA articles with order imbalances and abnormal stock returns. To assess whether the use of supremacy words reflects overly optimistic beliefs, we can also examine whether the use of supremacy words positively correlates with analysts' long-term earnings growth forecasts (Bordalo et al. 2019, 2024). We conduct these analyses and detail our method and corresponding results in Online Appendix Tables A6 and A7.

We report the corresponding results in Panels A and B of Table 5. The results are very similar to those reported in Panel A of Table 2 and suggest that perceived lottery plays a dominant role irrespective of whether we consider differences on a relative basis or an absolute basis.

5.2 Wordlist Iterations

In our second exercise, we evaluate whether the removal of individual words from our wordlists alters our conclusions. Specifically, 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 supremacy 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 A9. We focus on the top five words as they are most likely to alter our findings. Overall, the results presented in Online Appendix Figure A9 closely align with those reported in Table 2.

Additionally, only two of our five lottery terms, *potential* and *upside*, carry positive connotations. We find that our results are similar when we exclude these two terms, suggesting that investors are disproportionately drawn to short-leg securities for speculative reasons rather than general exuberance (Online Appendix Table A8).

5.3 Self-Defined Wordlists

In our third exercise, we follow Loughran and McDonald (2011), who self-define positive and negative words, and create our own lists of safety, supremacy, and lottery words.

Traditional finance theory suggests that it is the covariance of a stock's payoff with bad states of the world that matters. However, it seems 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 view, we find that the term *covariance* appears only 123 times across our 1.72 million analyst reports comprising roughly 1.88 billion words. Across our 141,746 SA articles comprising more than 152 million words, the term covariance appears only 17 times. Our safety words are therefore *safe*, *low risk*, *certain*, *predictable*, and *stable*. We consider all possible word forms of the five safety terms and account for negation.

Behavioral finance theory proposes that perceptions of supremacy are rooted in (1) investors' extrapolative tendencies and (2) overconfidence in the precision of their private signals. To capture investors' extrapolative tendencies, we proceed as follows: We extract all sequences of four words in a text.[21] We then examine for each 4-gram whether any of five "continuation terms" appear jointly with any of five "growth terms." The continuation terms are *carry on*, *continue*, *extend*, *go on*, and *keep on*. The growth terms are *excel*, *expand*, *grow*, *outperform*, and *rise*. If we observe a continuation word appearing jointly with a growth word in a 4-gram (e.g., *continue – grow*), we mark the two words as supremacy words. As before, we consider all possible word forms that are meaningfully tied to the business realm and account for negation.

To capture overconfidence in the precision of private signals, 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 supremacy 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*, *take a chance*, *potential*, and *upside*. Again, we consider all possible word forms tied to the business realm and account for negation.

The key advantage of our self-defined wordlists is that the construction of the supremacy words is more theoretically grounded. The disadvantage of our self-defined wordlists is that we insert ourselves into the data-generating process.

21 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.

We report the results in Panel C of Table 5. Our analyst report results suggest that investors like short-leg securities primarily for their perceived safety in 12% of cases, supremacy in 18%, and lottery in 54%. For SA articles, the corresponding fractions are 10%, 14%, and 48%, respectively. These patterns are very similar to our main results reported in Panel A of Table 2.

5.4 *Bidirectional Encoder Representations from Transformers (BERT)*

Our primary analysis employs a dictionary-based approach. While this method is straightforward and transparent, it is limited in its ability to capture contextual nuances and semantic relationships, and it does not represent the most advanced NLP technique available. To assess the robustness of our findings, we repeat our analysis using BERT, a more advanced NLP method.

BERT is a transformer-based language model. It is pre-trained using a masked language modeling objective, which enables the model to capture bidirectional context by jointly conditioning on both left and right tokens surrounding a masked word. This capacity allows BERT to develop a nuanced semantic understanding, making it particularly effective for tasks such as sentiment analysis and text classification, and adaptable for topic categorization with appropriate fine-tuning (Devlin et al. 2018). Given that our objective is to classify whether a writer's rationale for favoring a stock falls in a particular category, BERT should be well-suited for our analysis.

We construct three binary classification models, one for each perception category: perceived safety, perceived supremacy, and perceived lottery.[22] Each model classifies a given sentence as either relevant or irrelevant to its respective category. We iteratively train these models by fine-tuning FinBERT (Huang, Wang, and Yang 2023), which is a BERT variant pre-trained on a large corpus of financial texts, including 10-K filings, conference call transcripts, and analyst reports. We provide technical details in Online Appendix Figure A10.

[22] We make our three models publicly available on Hugging Face: perceived safety (https://huggingface.co/ZhZhPeng/3f_safe_final), perceived supremacy (https://huggingface.co/ZhZhPeng/3f_Supremacy_final), and perceived lottery (https://huggingface.co/ZhZhPeng/3f_Lottery_final).

In short, our results presented in Panel D of Table 5 point to similar conclusions as the findings from our dictionary-based analysis. Specifically, our analyst report results indicate that investors primarily favor short-leg securities for their perceived safety in 16% of cases, supremacy in 10%, and lottery-like features in 62%. For SA articles, the corresponding fractions are 10%, 10%, and 59%, respectively.

Compared to our dictionary-based approach, BERT yields fewer inconclusive results. Under BERT, analyst reports and SA articles yield inconclusive results 12% and 21% of the time. The corresponding figures under the dictionary-based method are 20% and 30%. This difference likely stems from how each method processes text. While the dictionary-based approach offers the highest level of transparency and interpretability, it lacks the ability to capture contextual semantics and word dependencies. In comparison, transformer-based language models, such as BERT, are less transparent but consider contextual embeddings, allowing for a more nuanced understanding of textual data. Each approach provides complementary perspectives and considering them jointly should reduce the risk that conclusions are sensitive to a particular method.

5.5 Subsample Analyses and Out-of-Sample Considerations

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 perceived lottery offers 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 reported in Panel E of Table 5 echo our primary results.

In another test, we restrict our analysis to analyst reports and SA articles published only after the corresponding cross-sectional stock-return predictability has been documented in an academic study. For instance, to understand investors' liking of stocks with high asset growth, we consider only analyst reports and SA articles published since January 1, 2009, after Cooper, Gulen, and Schill's (2008) documentation that stocks with high asset growth earn unusually low future returns. The corresponding results reported in Panel F of Table 5 again show similar patterns.

Finally, we repeat our analysis for the firm characteristics for which the corresponding paper's publication year is in the bottom quartile of the distribution, specifically papers published in the year 2001 or earlier. With the growing attention paid to prospect theory, researchers may have been data-mining for predictors that are consistent with prospect theory. If that is the case, our results should be weaker for "older" predictors. We find the opposite to be the case. The explanatory power of lottery is even stronger among the older cross-sectional stock-return predictors. The results reported in Panel G of Table 5 suggest that, when considering analyst reports, investors' rationales for liking short-leg securities are most congruent with perceived safety 11% of the time, supremacy 13%, and lottery 60%. The corresponding percentages for SA articles are 7%, 11%, and 62%, respectively.[23]

5.6 Sell Recommendations of Long-Leg Securities

Our primary analysis focuses on the buy recommendations for short-leg securities. We can also apply our method to sell recommendations for "long-leg securities," that is, stocks in the outer decile earning comparatively high returns, to better understand why investors may not (sufficiently) like them.

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

More importantly, the unusually low returns of short-leg securities constitute a more important component of the cross-section of expected stock returns (e.g., Stambaugh, Yu, and Yuan 2012). This imbalance accords with theory. Probability weighting in prospect theory predicts overpricing and unusually low returns for stocks with positively skewed returns. Probability weighting does *not* predict underpricing

[23] We conduct a series of additional robustness and subsample analyses, all of which we discuss and present in Online Appendix Table A9.

and abnormally high returns unless we set the model parameters at unrealistic values.[24] Similarly, as long as there are investors with overly optimistic beliefs about short-leg securities and frictions, specifically, short-sale constraints, which keep arbitrageurs out of the market, we will observe overpricing and unusually low future returns. There is no equivalent friction that would keep arbitrageurs out of the market when investors with overly pessimistic beliefs temporarily depress market prices. It is thus theoretically less clear how irrational beliefs would generate underpricing and abnormally high future returns among long-leg securities.

Although the asset pricing implications are limited, understanding the most pervasive reason analysts and SA contributors do not like long-leg securities still appears of interest. Is the most important perceived shortcoming the high level of risk, the poor state and trajectory of the company, or the lack of upside? These explanations need not mirror our findings for short-leg securities. For instance, while investors may favor short-leg securities primarily for their perceived upside, they may dislike long-leg securities primarily for their perceived lack of safety.

To find the most important shortcoming, we first compute (1) the fraction of negated safety words (e.g., *little – protection*), (2) the fraction of negated supremacy words (e.g., *not – competitive*), and (3) the fraction of negated lottery words (e.g., *no – potential*) in the sell recommendations for long-leg securities ("long-leg recommendations"). 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.

As shown in Panel A of Online Appendix Table A10, the explanatory power of our dictionary-based approach is much weaker for sell recommendations than for buy recommendations. For analyst reports, the results are inconclusive 87% of the time. That is, in 87% of cases, we find no statistically

[24] Individual stocks rarely have highly negatively skewed return distributions.

significant differences in the occurrences of negated safety words, negated supremacy words, or negated lottery words. For SA articles, the corresponding figure is 98%.

The key reason for the mostly inconclusive results is the extremely low overall usage of negated safety, supremacy, and lottery terms, which limits our dictionary-based method's ability to distill what key consideration drove the writer's sell recommendation.

To explore this point further, we repeat our analysis using BERT. We develop three additional BERT-based binary classification models for sell recommendations, following procedures analogous to those described in Section 5.5. As can be seen in Panel B of Online Appendix Table A10, there are substantially fewer inconclusive results when we apply BERT (38% for analyst reports and 55% for SA articles). The BERT results indicate that analysts primarily dislike long-leg securities for their perceived lack of safety in 15% of cases, supremacy in 17%, and lottery in 30%. For SA articles, the corresponding fractions are 7%, 14%, and 23%, respectively.

The BERT-based results suggest that investors often avoid long-leg securities for their perceived lack of lottery. However, we also observe that, in a non-negligible number of cases, negative perceptions of a company's state or trajectory also help explain why certain investors stay away from long-leg securities.[25]

6. Limitations and Possible Directions for Future Research

Before concluding, we discuss limitations of our method and outline possible directions for future research.

6.1 Noise and Ambiguity Regarding Low-Level Constructs

The key limitation of our text-based approach is the presence of noise. Inferring an individual's subjective beliefs from written expressions is inherently indirect and less precise than directly asking the individual

[25] Relatedly, we find that, for a non-negligible number of anomalies, the occurrence of negated supremacy terms is not only unusually high in the sell recommendations for long-leg securities; the frequency of supremacy terms is also unusually low in the buy recommendations (Online Appendix Table A11). For reasons that are beyond the scope of this study, the fraction of inconclusive results remains higher for the BERT analysis on sell recommendations than the BERT analysis on buy recommendations. One possible reason could be that even sophisticated NLP models encounter greater difficulty in distilling aversion against perceived negatives than enthusiasm for perceived positives. This is a methodological challenge that, interestingly, surveys also encounter (e.g., Scheibehenne, Mata, and Richter 2019).

about their beliefs. The challenge becomes more pronounced the more nuanced the subjective belief in question becomes.

Our findings suggest that text-based methods remain effective in distilling why an investor favors a stock at a high level, in particular, whether the primary appeal is perceived safety, supremacy, or lottery. Text-based methods, in their current form, would struggle when attempting to capture lower layers of reasoning.

To illustrate, the analyst report displayed in Online Appendix Figure A5 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*." Our text-based approach can successfully assign this sentence to the perceived safety category. However, it would struggle to specify, for instance, whether, within the perceived safety category, this statement is a manifestation of the belief that the stock offers good protection against economic disaster or whether the author wanted to express that this stock is a good hedge against the risk of aggregate consumption uncertainty. Ideally, one would construct a follow-up survey and directly ask the authors.

Similarly, consider the sample SA article displayed in Online Appendix Figure A6, which recommends that investors buy Magnum Hunter because at "*$0.50, Magnum Hunter appears to offer some potential as a purely speculative play for monetizing its assets.*" As already noted in Section 3.2, it is difficult to discern with current textual methods whether the emphasis on the "*speculative play*" stems from the contributor's prospect-theory-driven overweighting of small probabilities or from an unrealistic belief about the right tail of the return distribution. Again, a follow-up survey could clarify the author's thinking.

6.2 Possible Directions for Future Research

6.2.1 Taking Advantage of Advances in NLP

The noise and ambiguity regarding low-level constructs represent a challenge, but also an opportunity for future research. Textual analysis is still relatively young and evolving. BERT, for example, was only introduced to the public in 2018 (Devlin et al. 2018). As NLP techniques advance and an increasing volume of written expressions becomes publicly available, textual analysis may evolve into a more precise tool for examining investor perceptions at lower levels.

Relatedly, our study adopts a top-down approach. First, we pre-define three categories, each linked to a canonical framework in finance. We then measure the prevalence of statements tied to these categories in analyst reports and SA articles. A complementary strategy would be to pursue more of a bottom-up approach, using large language models (LLMs) to inductively identify the most salient topics in investor discourse. Three recent studies have started conducting such analyses. Ke (2025) applies LLMs to analyst reports and shows that references to topics, including profitability, corporate management, financial conditions, and macroeconomic trends, vary systematically over time and across firms; these topics differentially impact analysts' earnings forecasts. Bastinaella, Décaire, and Guenzel (2025) analyze how attention to topics, such as sales, costs, and margins, varies over time and influences analysts' price targets. Décaire, Sosyura, and Wittry (2025) apply a related methodology to study how analysts determine discount rates when estimating the present value of future cash flows.

Our top-down approach enables a direct evaluation of how well leading academic theories explain investor reasoning. However, our method misses out on considerations that matter to investors but fall outside existing theoretical frameworks. A promising avenue for future research would be to pursue more of a bottom-up approach and take advantage of advances in LLMs to identify themes that are frequent and important to investors, but not yet formally captured by finance theories. Such insights could guide the development of new or improved theoretical frameworks that better reflect how real-world investors think and make decisions.

6.2.2 Considering Other Sources of Text

Our study examines stock opinion reports and articles. Our method can be easily extended to other types of text. For example, social media platforms such as SA publish not only individual stock opinion articles but also financial advice articles. The household finance literature provides distinct explanations regarding the factors that determine people's equity allocations, including human capital risk (e.g., Pratt and Zeckhauser

1987), house price risk (e.g., Flavin and Yamashita 2002), inflation concerns (e.g., Boudoukh and Richardson 1993), trust (e.g., Guiso, Sapienza, and Zingales 2008), and personal experience (e.g., Malmendier and Nagel 2011), among others. Future research could gauge the relevance of each of these factors by parsing financial advice articles and identifying which of them are most frequently discussed.

Future research in household finance could also augment written expressions with author characteristics to better understand what types of agents are more likely to form certain beliefs and under what conditions.[26]

6.2.3 Contrasting Perceptions with Facts

Our analysis suggests that investors' perceptions do not always align with actual patterns in the data. For instance, Online Appendix Table A4 shows that the fraction of lottery words is reliably higher for growth stocks, indicating that investors perceive them as having more positively skewed returns. However, in the data, it is value stocks that have more positively skewed returns.

Future research could systematically examine when and why investor perceptions diverge from reality, explore the origins of these misperceptions, and discuss possible implications. For instance, one important implication of the results in our paper is that the rejection of a particular theory in an empirical test based on observational data does not necessarily imply the failure of the theory *per se*. In particular, prospect theory predicts that stocks with more positively skewed returns have lower average returns. Standard empirical tests based on observational data would examine whether growth stocks have both higher return skewness and lower average returns. Since, in observational data, growth stocks have lower average returns but less positively skewed returns, these tests would reject prospect theory (Barberis, Jin, and Wang 2021).[27]

[26] Online Appendix Table A12 provides some suggestive survey evidence on which investors more frequently invest for lottery considerations.
[27] In the past decade, growth stocks outperformed value stocks (Arnott et al. 2021).

However, the fact that growth stocks earn relatively low returns need not be inconsistent with prospect theory. It could just be that investors have the wrong perception of which stocks offer greater upside potential. If investors (incorrectly) perceive growth stocks to have greater skewness, as suggested by our textual analysis, then the fact that growth stocks have lower average returns perfectly accords with prospect theory.

The same point applies to 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 greater skewness. Yet, while stocks with high accruals, asset growth, market capitalization, and short-term performance earn comparatively low returns, they do not have more positively skewed returns.

The results reported in Online Appendix Table A4 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.

6.2.4 Specifying Social Transmission Biases

Our method can also play an important role in the rapidly growing social finance literature. The finance field increasingly recognizes that investors turn to each other for investment advice and that these social interactions can influence investment decisions and, in turn, affect asset prices (Hirshleifer 2020; Han, Hirshleifer, and Walden 2022; Hwang 2023).

To illustrate, suppose that investors exhibit systematic preferences for discussing stocks with certain features. Suppose further 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 can generate overpricing and unusually low future returns among certain types of stocks.

Textual methods can be a useful tool for examining which story types investors rely on primarily as they converse about stocks. Textual analysis can also offer insights into the channels through which stocks can become viral and ultimately overpriced.

7. Conclusion

The question of what drives the cross-section of expected stock returns lies at the heart of asset pricing and has been the focus of extensive research (e.g., Fama and French 1992; Davis, Fama, and French 2000; Chen and Zimmermann 2022). Our paper revisits this classic question. We parse investors' buy recommendations and record their primary reasoning for liking short-leg securities. The results from our analysis indicate that sell-side analysts and SA contributors predominantly favor short-leg securities for their perceived upside. Our finding suggests that any theory of the cross-section of expected stock returns must account for investors' preoccupation with the possibility of extreme positive outcomes, stemming either from non-traditional investor preferences or overly optimistic beliefs about the right tail of return distributions.

More broadly, our paper advocates for text-based methods as a complementary tool for extracting investor perceptions. With our growing reliance on modern information technologies, an increasing share of our thoughts and conversations are written out and publicly expressed. With continued advancements in text-based methods, we believe this approach offers valuable opportunities to gain deeper insights into investor decision-making and the dynamics of financial markets.

Code Availability

The replication code is available in the Harvard Dataverse at https://doi.org/10.7910/DVN/SGL1RM.

References

Agarwal, V., L. Jiang, and Q. Wen. 2022. Why do mutual funds hold lottery stocks? *Journal of Financial and Quantitative Analysis* 57:825–856.

Akbas, F., and E. Genc. 2020. Do mutual fund investors overweight the probability of extreme payoffs in the return distribution? *Journal of Financial and Quantitative Analysis* 55:223–261.

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

Antweiler, W., and M. Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59:1259–1294.

Arnott, R. D., C. R. Harvey, V. Kalesnik, and J. T. Linnainmaa. 2021. Reports of value's death may be greatly exaggerated. *Financial Analysts Journal* 77:44–67.

Bansal, R., and A. Yaron. 2004. Risks for the long run: a potential resolution of asset pricing puzzles. *Journal of Finance* 59:1481–1509.

Barber, B. M., and T. Odean. 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Barberis, N. 2018. Psychology-based models of asset prices and trading volume. In *Handbook of Behavioral Economics: Applications and Foundations 1*, eds. B. Bernheim, S. DellaVigna, and D. Laibson, 79–175. Amsterdam: North-Holland.

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

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer. 2018. Extrapolation and bubbles. *Journal of Financial Economics* 129:203–227.

Barberis, N., L. J. Jin, and B. Wang. 2021. Prospect theory and stock market anomalies. *Journal of Finance* 76:2639–2687.

Barberis, N., and A. Shleifer. 2003. Style investing. *Journal of Financial Economics* 68:161–199.

Barberis, N., A. Shleifer, and R. Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 49:307–343.

Barberis, N., and R. Thaler. 2003. A survey of behavioral finance. In *Handbook of the Economics of Finance*, eds. G. M. Constantinides, M. Harris, and R. M. Stulz, 1053–1128. Amsterdam: Elsevier.

Barro, R. J. 2006. Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics* 121:823–866.

Bastianello, F., P. H. Décaire, and M. Guenzel. 2025. Mental models and financial forecasts. Working Paper, University of Chicago Booth School of Business.

Bender, S., J. J. Choi, D. Dyson, and A. Z. Robertson. 2022. Millionaires speak: What drives their personal investment decisions? *Journal of Financial Economics* 146:305–330.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer. 2019. Diagnostic expectations and stock returns. *Journal of Finance* 74:2839–2874.

Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer. 2024. Belief overreaction and stock market puzzles. *Journal of Political Economy* 132:1450–1484.

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

Boudoukh, J., and M. Richardson. 1993. Stock Returns and Inflation: A Long-Horizon Perspective. The *American Economic Review* 83:1346–1355.

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

Chen, A. Y., and T. Zimmermann. 2022. Open source cross-sectional asset pricing. *Critical Finance Review* 27:207–264.

Chinco, A., S. M. Hartzmark, and A. B. Sussman. 2022. A new test of risk factor relevance. *Journal of Finance* 77:2183–2238.

Choi, J. J., and A. Z. Robertson. 2020. What matters to individual investors? Evidence from the horse's mouth. *Journal of Finance* 75:1965–2020.

Cookson, J. A., J. E. Engelberg, and W. Mullins. 2023. Echo chambers. *Review of Financial Studies* 36:450–500.

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

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

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

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

Décaire, P. H., D. Sosyura, and M. Wittry. 2025. Resolving estimation ambiguity. Working Paper, Arizona State University.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv, preprint, https://arxiv.org/abs/1810.04805.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47:427–465.

Fama, E. F., and K. R. French. 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51:55–84.

Fisher, K. L., and M. Statman. 2000. Investor Sentiment and Stock Returns. *Financial Analysts Journal* 56:16–23.

Flavin, M., and T. Yamashita. 2002. Owner-Occupied Housing and the Composition of the Household Portfolio. *American Economic Review* 92:345–362.

Garcia, D. 2013. Sentiment during recessions. *Journal of Finance* 68:1267–1300.

Greenwood, R., and A. Shleifer. 2014. Expectations of returns and expected returns. *Review of Financial Studies* 27:714–746.

Guiso, L., P. Sapienza, and L. Zingales. 2008. Trusting the stock market. *Journal of Finance* 63:2557–2600.

Han, B., D. Hirshleifer, and J. Walden. 2022. Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis* 57:390–412.

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

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

Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–2133.

Huang, A. H., H. Wang, and Y. Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research* 40:806–841.

Hwang, B.-H. 2023. The impact of word-of-mouth communication on investors' decisions and asset prices. In *Handbook of Financial Decision Making*, eds. G. Hilary and D. McLean, 171–191. Cheltenham, UK: Edward Elgar Publishing.

Jacobs, H., and S. Müller. 2020. Anomalies across the globe: once public, no longer existent? *Journal of Financial Economics* 135:213–230.

Jensen, T. I., B. Kelly, and L. H. Pedersen. 2023. Is there a replication crisis in finance? *Journal of Finance* 78:2465–2518.

Kahneman, D., and A. Tversky. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291.

Ke, S. 2025. Analysts' belief formation in their own words. Working Paper, Yale School of Management.

Linnainmaa, J. T., and M. R. Roberts. 2018. The history of the cross-section of stock returns. *Review of Financial Studies* 31:2606–2649.

Liu, H., C. Peng, W. A. Xiong, and W. Xiong. 2022. Taming the bias zoo. *Journal of Financial Economics* 143:716–741.

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

Lucas Jr, R. E. 1978. Asset prices in an exchange economy. *Econometrica* 46:1429–1445.

Malmendier, U., and S. Nagel. 2011. Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics* 126:373–416.

McLean, R.D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–32.

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

Novy-Marx, R. 2011. Operating leverage. *Review of Finance* 15:103–134.

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

Pratt, J. W., and R. J. Zeckhauser. 1987. Proper Risk Aversion. *Econometrica 55*:143–154.

Scheibehenne, B., J. Mata, and D. Richter. 2019. Accuracy of food preference predictions in couples. *Appetite 133*:344–352.

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

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

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

Vissing-Jorgensen, A. 2003. Perspectives on behavioral finance: Does "irrationality" disappear with wealth? Evidence from expectations and actions. *NBER Macroeconomics Annual* 18:139–194.

Table 1
Survey-Based Wordlists

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

Table 2
Why Do Investors Like Short-Leg Securities? – Text-Based Evidence

This table reports the frequency with which a particular reason is used to explain the buy recommendation of a stock that resides in the short leg of an anomaly. For each of 186 firm characteristics, we conduct the following analysis: Each month, we rank stocks based on the firm characteristic. Prior literature indicates that stocks in the lowest decile ("short-leg securities") trade at relatively high prices and earn unusually low future returns. We term analyst reports and Seeking Alpha (SA) articles that recommend buying these short-leg securities as "short-leg recommendations," and all other buy recommendations as "other recommendations." Analysts and SA contributors may recommend buying a short-leg security for at least three reasons: (1) They believe that the stock adds safety to the overall portfolio as it exhibits stability, consistency, and resilience during economic downturns or adverse conditions ("Perceived Safety"). (2) They are very confident that something good will happen to the company, or the stock has been doing well and is expected to continue to do well ("Perceived Supremacy"). (3) While risky and uncertain, the stock has the potential for quick, substantial gains ("Perceived Lottery"). To examine whether short-leg recommendations emphasize a stock's perceived safety, supremacy, or lottery-like feature, we compute the average fraction of safety words, *Safety [%]*, in analysts' (or SA contributors') short-leg recommendations. To determine whether the use of safety words is abnormally high, we also compute the average *Safety [%]* across analysts' (or SA contributors') other recommendations. We then calculate the difference between the former and the latter (on a relative basis), *ΔSafety [%]*, and test whether this difference is positive and statistically significant at the 5% level. We conduct similar tests for *ΔSupremacy [%]* and *ΔLottery [%]*. If *ΔSafety [%]* is positive and statistically significant at the 5% level, we conclude that one reason investors like the short-leg securities is the perceived safety. Similarly, if *ΔSupremacy [%]* or *ΔLottery [%]* are positive and significant, we conclude that investors like the short-leg securities for their perceived supremacy or lottery-like feature, respectively. If none of the differences are statistically significant at the 5% level, we deem our result "inconclusive." If more than one of the differences is positive and significant, we label the rationale associated with the largest relative difference as the one that "primarily explains" investors' liking of the short-leg securities. We perform this analysis for each of the 186 firm characteristics and, in the end, calculate the fraction of times investors explain [primarily explain] their liking of short-leg securities through the stocks' perceived safety, supremacy, or lottery-like feature, respectively. We also compute how many times our results are inconclusive. Panel A reports the fractions based on the institutional investor wordlist shown in Table 1, while Panel B reports the fractions based on the retail investor wordlist, also listed in Table 1.

	Fraction of Times Investors Explain [Primarily Explain] Their Liking of Short-Leg Securities Through			
	Perceived Safety (1)	Perceived Supremacy (2)	Perceived Lottery (3)	Inconclusive
Panel A: Institutional Investors' Survey-Based Wordlists				
Sell-Side Analyst Reports	12% [8%]	29% [17%]	68% [55%]	20%
Seeking Alpha Articles	7% [6%]	16% [11%]	57% [54%]	30%
Panel B: Retail Investors' Survey-Based Wordlists				
Sell-Side Analyst Reports	24% [5%]	27% [18%]	69% [65%]	12%
Seeking Alpha Articles	10% [8%]	23% [13%]	63% [57%]	22%

Table 3
Why Do Investors Like Short-Leg Securities? – Evidence by Economic Categories

Chen and Zimmerman (2022) assign all "anomalies" into 32 distinct "economic categories." This table reports the results from Panel A in Table 2, broken down into these 32 economic categories. To give a sense of which anomalies reside in a particular economic category, we list the top one or top two anomalies within each category based on the Google Scholar citations of the corresponding academic papers as of June 2022. Columns (1) through (3) display the fraction of anomalies where sell-side analysts primarily explain their liking of short-leg securities based on perceived safety, supremacy, and lottery. Column (4) reports the number of instances where our findings are inconclusive. Columns (5) through (8) display the corresponding numbers for Seeking Alpha contributors. We bold all cases where both the fraction based on sell-side analysts and that based on Seeking Alpha contributors are greater than 66%.

	Sell-Side Analyst Reports				Seeking Alpha Articles			
	Fraction of Times Investors Primarily Explain Their Liking of Short-Leg Securities Through				Fraction of Times Investors Primarily Explain Their Liking of Short-Leg Securities Through			
Economic Category [Number of Anomalies in Category] (*Top One or Top Two Anomalies in Category*)	Perceived Safety (1)	Perceived Supremacy (2)	Perceived Lottery (3)	Inconclusive (4)	Perceived Safety (5)	Perceived Supremacy (6)	Perceived Lottery (7)	Inconclusive (8)
Accruals [5] *(e.g., Accruals, Abnormal Accruals)*	0%	20%	40%	40%	0%	0%	60%	40%
Asset Composition [4] *(e.g., Net Op. Assets, Cash to Assets)*	0%	25%	50%	25%	50%	0%	25%	25%
Cash Flow Risk [1] *(Cash Flow to Price Variance)*	0%	0%	**100%**	0%	0%	0%	**100%**	0%
Composite Account [5] *(e.g., Piotroski F-score, Excluded Expenses)*	20%	20%	40%	20%	0%	20%	0%	80%
Default Risk [1] *(O-Score)*	0%	0%	**100%**	0%	0%	0%	**100%**	0%
Earnings Event [2] *(Ret. Over Earn. Announcement Period, Decline in Coverage)*	0%	0%	50%	50%	0%	0%	50%	50%
Earnings Forecast [5] *(e.g., Earn. Forecast to Stock Price; Predicted Analyst Forecast Errors)*	0%	0%	**80%**	20%	0%	0%	**100%**	0%
Earnings Growth [3] *(e.g., Average Earn. Growth, Streak in Earn. Surprise)*	0%	0%	**67%**	33%	0%	0%	**67%**	33%
External Financing [11] *(e.g., Indicator of Initial Public Offering, ΔCurrent Op. Liabilities)*	0%	18%	**82%**	0%	0%	0%	**82%**	18%

Table 3. Continued.

	Sell-Side Analyst Reports				Seeking Alpha Articles			
	Fraction of Times Investors Primarily Explain Their Liking of Short-Leg Securities Through				Fraction of Times Investors Primarily Explain Their Liking of Short-Leg Securities Through			
Economic Category [Number of Anomalies in Category] (*Top One or Top Two Anomalies in Category*)	Perceived Safety (1)	Perceived Supremacy (2)	Perceived Lottery (3)	Inconclusive (4)	Perceived Safety (5)	Perceived Supremacy (6)	Perceived Lottery (7)	Inconclusive (8)
Information Proxy [1] *(Firm Age)*	**100%**	0%	0%	0%	**100%**	0%	0%	0%
Investment [18] *(e.g., CAPX to Revenue, ΔEquity to Assets)*	6%	17%	**72%**	6%	6%	11%	**67%**	17%
Investment Growth [3] *(e.g., CAPEX Growth Over Ind., CAPEX Growth Over Past Three Yrs.)*	0%	0%	**100%**	0%	0%	0%	**100%**	0%
Lead Lag [9] *(e.g., Economic Link Momentum, Price Delay Coefficient)*	22%	11%	11%	56%	11%	0%	22%	67%
Leverage [4] *(e.g., Book Leverage, Mkt. Leverage)*	25%	0%	25%	50%	0%	0%	25%	75%
Liquidity [9] *(e.g., Illiquidity, Bid-Ask Spread)*	22%	0%	67%	11%	11%	11%	56%	22%
Long-Term Reversal [6] *(e.g., Long-Run Reversal, Medium-Run Reversal)*	0%	83%	17%	0%	0%	50%	33%	17%
Momentum [9] *(e.g., Momentum (12-Month), Momentum (6-Month))*	0%	0%	67%	33%	0%	0%	56%	44%
Option Risk [2] *(Volatility Smirk, Put Volatility Minus Call Volatility)*	0%	0%	50%	50%	0%	0%	100%	0%
Other [25] *(e.g., GIM Governance Index, IPO Age)*	12%	28%	44%	16%	8%	20%	40%	32%
Ownership [3] *(e.g., Active Shareholders, Number of Inst. Owners)*	0%	0%	33%	67%	0%	0%	0%	100%
Payout Indicator [1] *(First Month When Consistent Div. Payers Fail to Pay Div.)*	0%	0%	0%	100%	0%	0%	0%	100%

Table 3. Continued.

Economic Category [Number of Anomalies in Category] (*Top One or Top Two Anomalies in Category*)	Sell-Side Analyst Reports: Fraction of Times Investors Primarily Explain Their Liking of Short-Leg Securities Through Perceived Safety (1)	Perceived Supremacy (2)	Perceived Lottery (3)	Inconclusive (4)	Seeking Alpha Articles: Fraction of Times Investors Primarily Explain Their Liking of Short-Leg Securities Through Perceived Safety (5)	Perceived Supremacy (6)	Perceived Lottery (7)	Inconclusive (8)
Profitability [9] *(e.g., Gross Profits to Total Assets, Net Inc. to Book Equity)*	0%	11%	**89%**	0%	0%	11%	**89%**	0%
R&D [4] *(e.g., Advertising Expense, R&D Expense to Mkt. Value)*	0%	**75%**	0%	25%	0%	**75%**	25%	0%
Recommendation [2] *(Analysts Consensus Rating)*	0%	0%	0%	100%	0%	0%	0%	100%
Risk [6] *(e.g., CAPM Beta, Coskewness Using Monthly Ret.)*	17%	33%	33%	17%	33%	0%	67%	0%
Sales Growth [6] *(e.g., Revenue Growth Rank, Sales Growth Over Inventory Growth)*	0%	17%	50%	33%	0%	0%	17%	83%
Short Sale Constraints [5] *(e.g., SIR, IO and Forecast Dispersion; SIR, IO and Idio. Volatility)*	0%	0%	20%	80%	0%	0%	20%	80%
Short-Term Reversal [1] *(Short-Term Reversal)*	0%	0%	**100%**	0%	0%	0%	**100%**	0%
Size [1] *(Size)*	100%	0%	0%	0%	0%	100%	0%	0%
Valuation [15] *(e.g., Book-to-Mkt. Ratio, Total Assets to Mkt. Value)*	7%	27%	**67%**	0%	7%	13%	**73%**	7%
Volatility [5] *(e.g., Idio. Risk Based on CAPM, Idio. Risk Based on FF3)*	0%	0%	**100%**	0%	0%	0%	**100%**	0%
Volume [5] *(e.g., Lagged 2-Month Trading Volume, Monthly Trading Volume Trend)*	20%	0%	80%	0%	0%	20%	60%	20%

Table 4
Why Do Investors Like Short-Leg Securities? – Survey-Based Evidence

This table presents responses from institutional- and retail investors regarding their primary motivation for purchasing short-leg securities. In March/April 2024, we survey 450 institutional investors, asking them to reflect on the stocks they purchased over the past year. If an investor bought more than eight stocks, we instruct them to consider the first eight that come to mind. For each stock, we ask whether their key motivation for the purchase was the stock's perceived safety, supremacy, or lottery. In June 2024, we ask the same questions to 314 U.S. retail investors. For each of the 186 firm characteristics *i*, we then conduct the following analysis. Similar to Table 2, we define "short-leg securities" as stocks that our survey participants report having bought and that reside in the short leg based on firm characteristic *i* at any point between April 2023 and March 2024 for institutional investors and between July 2023 and June 2024 for retail investors. We label stocks that our participants purchased but that never resided in the short leg as "other stocks." For each firm characteristic with the necessary data, we compute *ΔSafety [%]*, which measures the fraction of times institutional investors report having bought the short-leg securities for their perceived safety relative to how often the same investors cite safety as the key reason for purchasing other stocks. We then test whether this difference is positive and statistically significant at the 5% level. We conduct similar tests for *ΔSupremacy [%]* and *ΔLottery [%]*. If *ΔSafety [%]* is positive and significant, we conclude that our survey participants disproportionately purchased short-leg securities for their perceived safety. Similarly, if *ΔSupremacy [%]* or *ΔLottery [%]* is positive and significant, we conclude that our participants disproportionately purchased short-leg securities for their perceived supremacy or lottery-like features, respectively. If none of the differences are statistically significant at the 5% level, we classify the result as inconclusive. If more than one difference is statistically significant, we assign the perception with the largest relative difference as the primary explanation for why our survey participants bought the short-leg securities. We perform this analysis for each firm characteristic and, in the end, calculate the fraction of times investors primarily explain their actual purchases of short-leg securities through perceptions of safety, supremacy, or lottery. We also compute how frequently our results are inconclusive. Below, we report the fractions separately for institutional investors and retail investors. Additionally, we compute the percentage of cases where our text-based inferences of why investors primarily favor short-leg securities (from Table 2) and our survey-based inferences (from this table) are exactly the same.

	Fraction of Times Investors [Primarily Explain] Their Actual Purchase of Short-Leg Securities Through			
	Perceived Safety (1)	Perceived Supremacy (2)	Perceived Lottery (3)	Inconclusive
Institutional Investors	[17%]	[13%]	[64%]	6%
Proportion Agreement (Sell-Side Analyst Reports; Institutional Investor Survey) = 65.2%				
Retail Investors	[6%]	[10%]	[81%]	2%
Proportion Agreement (Seeking Alpha Articles; Retail Investor Survey) = 91.2%				

Table 5
Why Do Investors Like Short-Leg Securities? – Sensitivity Analyses

This table reports the frequency with which investors use a particular reason to explain their buy recommendations of stocks that reside in the short leg of an anomaly. The analyses are identical to those in Panel A of Table 2, except that we now report results based on a different denominator and based on the numerator only as detailed in Section 5.1 (Panel A and B), apply our self-defined wordlists as detailed in Section 5.3 (Panel C), apply BERT rather than compute the fractions of safety, supremacy or lottery words as detailed in Section 5.4 (Panel D), consider only the firm characteristics whose corresponding academic paper's Google Scholar citation as of June 2022 is in the top quartile (Panel E), consider only data since the corresponding academic paper has been published (Panel F), or consider only the firm characteristics whose corresponding academic paper was published in the first quartile of our sample period (Panel G).

	Fraction of Times Investors Explain [Primarily Explain] Their Liking of Short-Leg Securities Through			
	Perceived Safety (1)	Perceived Supremacy (2)	Perceived Lottery (3)	Inconclusive
Panel A: Alternate Denominator				
Sell-Side Analyst Reports	12% [8%]	29% [17%]	68% [55%]	20%
Seeking Alpha Articles	7% [6%]	16% [11%]	57% [54%]	30%
Panel B: Numerator Only				
Sell-Side Analyst Reports	14% [9%]	31% [18%]	65% [49%]	24%
Seeking Alpha Articles	6% [4%]	21% [10%]	59% [54%]	31%
Panel C: "Theory-Driven" Self-Defined Wordlists				
Sell-Side Analyst Reports	17% [12%]	27% [13%]	66% [58%]	16%
Seeking Alpha Articles	12% [10%]	18% [11%]	56% [53%]	25%
Panel D: BERT				
Sell-Side Analyst Reports	18% [16%]	26% [10%]	68% [62%]	12%
Seeking Alpha Articles	13% [10%]	16% [10%]	61% [59%]	21%

Table 5. Continued.

	Fraction of Times Investors Explain [Primarily Explain] Their Liking of Short-Leg Securities Through			
	Perceived Safety (1)	Perceived Supremacy (2)	Perceived Lottery (3)	Inconclusive
Panel E: High Google Scholar Citations Anomalies Only				
Sell-Side Analyst Reports	13% [13%]	34% [17%]	70% [57%]	13%
Seeking Alpha Articles	6% [6%]	21% [13%]	64% [60%]	21%
Panel F: Data Since Publication Only				
Sell-Side Analyst Reports	10% [8%]	28% [18%]	68% [54%]	19%
Seeking Alpha Articles	8% [6%]	16% [11%]	56% [53%]	30%
Panel G: Older Anomalies Only				
Sell-Side Analyst Reports	13% [11%]	33% [13%]	73% [60%]	16%
Seeking Alpha Articles	9% [7%]	20% [11%]	64% [62%]	20%