Soothing Investors: The Impact of Manager Communication on Mutual Fund Flows

Ahmed Guecioueur*

January 9, 2024

Latest: `ahmedgc.com/files/latest.pdf`

Abstract

I show that communication by fund managers to their investor clients fosters trust and encourages these investors to bear risk. Using an institutional setting that enables causal identification, I find that more detailed communication about risk encourages investors to *increase* their holdings in the market portfolio, driving flows into the stock market. I rule out learning about risk, returns or manager skill, and other potential explanations. Instead, my analysis shows this communication soothes investors' anxiety and alleviates their effective risk aversion, consistent with the money doctors framework of Gennaioli, Shleifer, and Vishny (2015).

Keywords: Fund Flows, Anxiety, Trust, Persuasion, Text Data

JEL Codes: G11, G23, G41, G50, D01, D83

---

*INSEAD, Boulevard de Constance, 77300 Fontainebleau, France, email: ahmed.guecioueur@insead.edu
I am grateful to my advisor, Joël Peress, for his invaluable guidance. I would like to thank Pat Akey, Bronson Argyle, Alexandru Barbu, Frederico Belo, Pedro Bordalo, Gavin Cassar, Aditya Chaudhry, Olivier Dessaint, Tilman Fries (discussant), Andreas Fuster, Nicola Gennaioli, Sergei Glebkin, Naveen Gondhi, Richard Grice, Markus Ibert, Philipp Illeditsch, Pascal Maenhout, Massimo Massa, Cristina Ortiz (discussant), Alessandro Previtero, Rodolfo Prieto, Dmitrii Pugachev, Vesa Pursiainen, Paola Sapienza (discussant), Lucie Teplá, Boris Vallée, Rüdiger Weber, Scott Yonker (discussant), Bart Zhou Yueshen, and many others, as well as participants at the Fall 2023 NBER Behavioral Finance Working Group Meeting, the 2023 Financial Intermediation Research Society (FIRS) Conference PhD Sessions, the 7th SAFE Household Finance Workshop, the 16th Behavioural Finance Working Group Conference, the LSE-Oxford-Imperial Behavioural Finance Group, the 6th Monash-Warwick-Zürich Text-as-Data Workshop, the 16th International Risk Management Conference, the 2023 Washington University in St. Louis Economics Graduate Student Conference, the 7th HEC Paris Finance PhD Workshop, the 11th Wharton-INSEAD Doctoral Consortium, and the PhD Economics Virtual Seminar for their helpful comments. I am grateful to Yehuda Izhakian, Gerald S. Martin, and Melissa Porras Prado for sharing data, and to Google for granting me access to its Trends API. A previous version of this paper was circulated under the title "Money Doctors and Their Prognoses."

1 Introduction

Investing is a risky endeavor. Low equity holding rates indicate that many individuals are reluctant to bear these risks.[1] To the extent that certain behavioral biases are at fault, these might be overcome by governments nudging savers to take more risks by regulating their choice architecture (Thaler and Sunstein, 2008), or by asset managers designing investment products that foster risk-taking (Calvet, Célérier, Sodini, and Vallée, 2023). However, less attention has been paid to the role of communication, which is ubiquitous in the asset management industry. Can investors be encouraged to shoulder risk through information provision?

This paper shows that communication by asset managers significantly enhances investors' willingness to assume risks. I identify a mechanism through which communication builds up trust and reduces investors' anxiety. Gennaioli, Shleifer, and Vishny (2015, henceforth **GSV**) argue that trusted financial intermediaries can give their clients the confidence to overcome their anxieties and take risks, including through "communication" and "persuasive advertising." Indeed, the medical literature has long recognized that providing information to patients reduces their anxiety (Hayward, 1975).[2] Motivated by this evidence and by GSV's model of the role of anxiety in financial risk-taking, I investigate how investors adjust their equity holdings in response to communication about the very risks they fear. To do so, I use a setting that facilitates identification and allows me to evaluate a number of distinct mechanisms, including updates to investors' beliefs and reductions in their anxiety.

Mutual funds regularly write and disseminate free-form letters to their investors (Hillert, Niessen-Ruenzi, and Ruenzi, 2021). I isolate the effect of this communication by focusing on managers tracking the exact same market portfolio; namely, S&P 500 index funds. I find that fund managers who provide more detailed information about risks in their fund letters – in the sense of dedicating a larger proportion of the letter to discussing this topic – enjoy *higher* net flows, even though the actual risk and return of the underlying portfolio is the same. Furthermore, the reaction of investors does not depend on whether the letter conveys a high or low level of risk. The effect on flows is economically significant: a 1 standard deviation (SD) growth in the amount of detail communicated about risk predicts a 0.23 SD growth in assets

[1] The majority of the US population does not participate at all in the stock market (Mankiw and Zeldes, 1991; Haliassos and Bertaut, 1995; Melcangi and Sterk, 2020). In addition, many households that *do* hold stocks hold a low share of their financial wealth in them (Calvet, Campbell, and Sodini, 2007; Calvet, Célérier, Sodini, and Vallée, 2023). This paper focuses on investments in diversified large-cap equity portfolios such as the S&P 500, which are generally considered beneficial for households' welfare (in contrast to trading in individual stocks).

[2] Individual risk aversion is not a stable trait (Schildberg-Hörisch, 2018), and increases with anxiety (Raghunathan and Pham, 1999; Lerner and Keltner, 2001; Maner and Schmidt, 2006; Kuhnen and Knutson, 2011; Giorgetta et al., 2012; Guiso, Sapienza, and Zingales, 2018). In the field of medicine, Hayward (1975) shows in an early field experiment that communication alleviates both anxiety and pain among patients. In a meta-analysis of 41 medical studies, Hall, Roter, and Katz (1988) compare different aspects of communication and find the *amount* of communication is the strongest predictor of patient satisfaction measures.

under management during the subsequent month – equivalent to an additional 0.67 percentage points of assets, on average. The relationship is present both over time within a fund, and also in the cross-section of S&P 500 index funds. Decomposing net flows, I find that the effect is driven by inflows; given that the letters are disseminated to existing investors, these results indicate that communication about risk increases *existing* investors' willingness to bear risk. These findings are consistent with GSV's hypothesis that "money doctors" can decrease investors' effective risk aversion. Furthermore, I uncover communication as an antidote that soothes investors' anxieties.

Focusing on quasi-identical S&P 500 index funds allows me to isolate the effect of communication on flows. To begin with, any differences in the funds' portfolios, risk and returns are minimal, eliminating these as drivers of flows. I also explicitly control for observable drivers of fund flows that have been identified by the literature, most notably past performance, fees and other fund and family characteristics. Importantly, I incorporate time effects, which control for investors' (common) prior beliefs, the growth in passive investing, and any news about or shocks to the common portfolio that might simultaneously (and independently) drive both flows and communication. I also include fund fixed effects to control for unobservable but persistent differences between funds, the families they belong to, and their investor bases. Given the implicit and explicit controls that are present in this setting, I attribute differences in fund flows to differences in funds' communication to investors.

To mitigate the concern that results are driven by other variables that are potentially unobservable – such as additional content in fund letters, or other attributes of the funds, their families or their investors – I employ three distinct and complementary identification strategies. First, I conduct a sensitivity analysis to reassure that unobservables cannot reasonably explain the main effect (Altonji, Elder, and Taber, 2005; Oster, 2019). Second, I construct an instrument for the amount of communication by the index funds in my sample that captures within-fund family information spillovers from active to passive funds. This instrumental variables strategy produces economically similar coefficient estimates to the baseline panel regression, supporting a causal interpretation. And third, I exploit the presence of corner bunching in the empirical distribution of the treatment variable (defined as the fraction of a letter discussing risk) to detect and correct for potential selection bias (Caetano, Caetano, and Nielsen, forthcoming). These diverse empirical strategies consistently confirm that the positive effect of communicated risk detail on fund flows is robust to unobservables.

In order to identify the mechanism underlying the effect of communication, I consider and ultimately rule out a number of potential explanations, of which the first three involve learning. *First*, I show that investors are not simply learning about risk; if they were, I should observe a lower flow when the amount of detail communicated is indicative of higher risk, and yet I find the opposite.[3] Furthermore, the level of risk discussed by fund letters and the sentiment of these risk discussions are each informative about implied volatility, but investor flows do not react to these features of the text. *Second*, investors do not seem to be learning about returns more broadly from these letters, as they do not respond to communication that is specifically about returns and performance. In addition, returns are not predictable based on the information conveyed about risk. *Third*, investors are not learning about fund manager skill: more detailed communication is not predictive of index funds' tracking error. *Finally*, I conduct a battery of tests that rule out a rebalancing motive by investors from other equity holdings (such as a flight-to-quality motive), an educational motive (Lusardi and Mitchell, 2014), a shrouding motive (Célérier and Vallée, 2017), and other potential channels.

Instead of these possible explanations, the evidence supports an anxiety-alleviating mechanism: I find that low-anxiety readers barely respond to more detailed communication, whereas high-anxiety readers respond strongly and positively through increased inflows. To arrive at this conclusion, and to circumvent the necessity of measuring individual investors' anxiety attitudes at scale, I employ a novel and comprehensive geographic measure of anxiety attitudes. I geographically locate the readers of fund letters, and infer their anxiety attitudes in aggregate from the frequency of local Google searches for terms relating to anxiety and worry (using categories defined by Google, and which comprise a multitude of phrases, such as "stressed," "anxiety" and "stop worrying"). I then produce a fund letter readership measure of investor anxiety by taking an asset-weighted average of fund letter readers' local anxiety attitudes, and find the positive effect of communicated detail on flows is concentrated among funds with high-anxiety readers. This finding indicates that high prior anxiety levels had been holding back investors from taking risks – until being soothed by communication. The finding is thus consistent with the "money doctors" hypothesis that financial intermediaries can alleviate investors' anxiety and give them the confidence to bear risk. By alleviating investors' effective risk aversion, this communication acts as a form of preference-based persuasion (DellaVigna and Gentzkow, 2010). Furthermore, to the extent that anxiety alleviation increases investors' trust in the stock market (GSV), this finding sheds light on the origins of trust.

While trust in the stock market can be fostered by communication from fund managers, trust in the fund managers themselves is arguably necessary for the communication to be effective. I investigate this potential mechanism. To avoid the challenge of directly measuring perceived trustworthiness, I follow Gurun, Stoffman, and Yonker's (2018) approach of using investors' indirect exposure to fraud as a negative shock to financial intermediaries' trustworthiness. Specifically, I measure the exposure of fund letter readers to the contemporaneous revelation of fraud by local firms (Giannetti and Wang, 2016; Karpoff, Koester, Lee, and Mar-

[3] As a general test, I compare the strength of potential belief updating with the strength of the prior about risk; results are not consistent with Bayesian updating or belief-based persuasion (DellaVigna and Gentzkow, 2010).

tin, 2017), and find that the effect of communication is attenuated by recipients' exposure to fraud. Therefore, a higher level of trust in a fund manager strengthens the impact of her communication, which in turn builds the trust of her investors in the stock market.

Finally, I quantify the aggregate asset pricing implications of communication, and find them to be economically important. My focus on funds that track the equity market portfolio allows me to measure the extent to which communication-driven inflows into the stock market affect its value. To translate fund flow effects into stock market returns, I use estimates of the macro multiplier of stock market inflows by Gabaix and Koijen (2021) and Hartzmark and Solomon (2023). A back-of-the-envelope estimate is that the communication-driven net flows in my sample are responsible for an increase of between 27 and 67 basis points a year in the value of the S&P 500 (depending on which macro multiplier is used). For comparison, the S&P 500 index returned around 8% per annum over this period. This simple analysis suggests that the asset pricing consequences of anxiety alleviation are meaningful in aggregate.

This paper contributes to five main streams of literature. First and foremost, it contributes to the literature on investor decision-making by highlighting the role played by investors' anxiety. I provide evidence that financial intermediaries can reduce investor anxiety through communication, as hypothesized by GSV. To the best of my knowledge, my paper is the first to uncover a channel through which asset managers can actively alleviate their clients' anxieties.[4] I find that effective communication stimulates investing in financial markets and, to the extent that a greater share of wealth held in the stock market is beneficial (Calvet, Campbell, and Sodini, 2007; Calvet, Célérier, Sodini, and Vallée, 2023), my findings therefore shed light on the benefits of (non-instrumental)[5] information provision and social influence for households (Gomes, Haliassos, and Ramadorai, 2021; Kuchler and Stroebel, 2021).

Second, this study contributes new insights on both the determinants and effects of trust in financial markets. Trust is known to encourage stock market participation (Guiso, Sapienza, and Zingales, 2008; Giannetti and Wang, 2016) and economic development (Arrow, 1972;

[4]Loewenstein, Weber, Hsee, and Welch (2001) argue via their "risk-as-feelings" hypothesis that emotions play a crucial part in individuals' risk-related behavior, over and above cognitive-consequentialist processes. In the field, Lo and Repin (2002) find that even experienced professional traders undergo emotional responses when taking risks. In a relevant setting, Hartzmark and Sussman (2019) and Wang and Young (2020) find that emotions guide the decisions of US mutual fund investors. More broadly, Engelberg and Parsons (2016) find that stock market-induced anxiety increases hospital admissions; Sergeyev, Lian, and Gorodnichenko (2023) show that individual psychic costs aggregate up to macroeconomic outcomes; Kaur, Mullainathan, Oh, and Schilbach (2021) show that financial anxiety hurts labor productivity; Lin and Pursiainen (2023) find that financial stress increases domestic violence; and Haushofer and Fehr (2014) argue that poverty-induced stress and anxiety can make it more difficult for individuals to escape poverty because of the increased risk aversion that is induced by negative affect.

[5]I define the non-instrumental usage of information by an economic agent as a use for some purpose besides forming more accurate beliefs. The literature on this topic tends to focus on the potential drawbacks of information aversion (Andries and Haddad, 2020) and selective attention (Sicherman, Loewenstein, Seppi, and Utkus, 2016), or analyzes the use of information as entertainment (Ely, Frankel, and Kamenica, 2015). I provide evidence on the bright side of non-instrumental information provision, to the extent that anxiety alleviation is beneficial.

Fukuyama, 1995). While various studies have shown that investors' trust can be broken (Kostovetsky, 2016; Giannetti and Wang, 2016; Gurun, Stoffman, and Yonker, 2018), there is less understanding of how to build trust in the first place. My findings address this question by showing that communication by financial intermediaries can build trust in the stock market. Furthermore, the effectiveness of communication increases with the trustworthiness of the financial intermediaries themselves, highlighting the complementary effects of trust.

Third, this paper contributes to a broad literature on the role of communication in economic interactions. Persuasion and communication are widespread phenomena (McCloskey and Klamer, 1995), and this paper shows that investors are persuaded to bear risk by shifting their effective risk aversion. DellaVigna and Gentzkow (2010) distinguish such a preference-based channel from a belief-based form of persuasion.[6]

Fourth, this paper advances the literature on information transmission in the asset management industry. Prior work tends to concentrate on the topics of disclosure accuracy and manager skill;[7] for example, Cassar et al. (2018) analyze the choice of hedge funds to disclose past information. Some studies examine the contents of mandatory prospectus disclosures: Abis, Buffa, Javadekar, and Lines (2022), Sheng, Xu, and Zheng (2022) and Mitali (2022) analyze strategy descriptions, risk factor exposures, and confidence in writing style, respectively. Others analyze voluntary communication by mutual funds: Hillert, Niessen-Ruenzi, and Ruenzi (2021), Du, Jiao, Ye, and Fan (2019) and Cao, Yang, and Zhang (2021) study whether shareholder letters are informative about manager skill or future performance, and whether investors can thus learn from this communication. I find that communication plays a distinct, anxiety-alleviating role in the asset management industry. Methodologically, I contribute multiple identification strategies for text data that can also be employed to study classic topics in disclosure and learning about skill.

Last but not least, this study contributes to the literature on equity flows. At a macro level, I answer Gabaix and Koijen's (2021) call to investigate the drivers of flows to the aggregate stock market. I also add to a rich fund-level literature that documents the drivers of mutual fund investors' capital allocation decisions, including Sirri and Tufano (1998), Chevalier and Ellison (1999), Jain and Wu (2000), Berk and Green (2004), Huang, Wei, and Yan (2007),

[6]In preference-based persuasion (for example, Galbraith (1958) and Becker and Murphy (1993)), the receiver's *preferences* are influenced by communication; in the present paper, investors' effective risk aversion is a preference. In belief-based persuasion models (for example, Stigler (1961), Crawford and Sobel (1982), Milgrom and Roberts (1986), and Kamenica and Gentzkow (2011)), the receiver's *beliefs* are shifted by communication; this class includes models where receivers are not fully Bayesian (for example, DeMarzo, Vayanos, and Zwiebel (2003) and Mullainathan, Schwartzstein, and Shleifer (2008)). DellaVigna and Gentzkow (2010) comprehensively survey these two streams of the literature. Recent work studies the use of models (Schwartzstein and Sunderam, 2021), rhetoric (Paugam, Stolowy, and Gendron, 2021) and message delivery (Hu and Ma, 2021) to persuade.

[7]A smaller literature studies advertising by fund families (Jain and Wu, 2000; Cronqvist, 2006; Gallaher, Kaniel, and Starks, 2015). Advertising may attract new investors by raising awareness or attracting attention; focusing on communication to existing investors rules out these channels to expose a persuasion effect.

Del Guercio and Tkac (2008), Bailey, Kumar, and Ng (2011), and many others.

2 Money Doctors

This section briefly reviews the core of GSV's money doctors theory, which provides a concise empirical framework for how communication can encourage investor risk-taking.

2.1 Classic Portfolio Choice

Consider the simple case of an investor who can access a riskless asset that returns RF_{t+1} and a risky asset with excess return R_{t+1}. I interpret the risky asset as the market portfolio; note that the present paper studies investors in S&P 500 index funds. The investor will allocate fraction x_t of her portfolio to the risky asset and the remainder $(1-x_t)$ to the riskless asset. For brevity, I omit investor subscripts to focus on a single investor. Assume the investor has mean-variance preferences and γ is her coefficient of risk aversion.[8,9] She therefore seeks to maximize the objective function

$$U(x_t) = RF_{t+1} + x_t \mathbb{E}_t[R_{t+1}] - \frac{\gamma}{2} x_t^2 \, \mathbb{V}\mathrm{ar}_t(R_{t+1}), \tag{1}$$

and as a result will allocate the fraction

$$x_t = \frac{\mathbb{E}_t[R_{t+1}]}{\gamma \mathbb{V}\mathrm{ar}_t(R_{t+1})} \tag{2}$$

of her portfolio to the risky asset, based upon her risk aversion and beliefs about the payoff.

2.2 Incorporating Anxiety

GSV include a parameter in the investor's objective that specifically captures the "anxiety suffered by [the investor] for bearing risk." Likewise, I augment Eqn. (1) with the parameter $a_t \geq 1$; the presence of a subscript t highlights that investor anxiety may vary over time:

$$U(x_t) = RF_{t+1} + x_t \mathbb{E}_t[R_{t+1}] - \frac{a_t \gamma}{2} x_t^2 \, \mathbb{V}\mathrm{ar}_t(R_{t+1}), \tag{3}$$

[8]This classic baseline makes the standard assumption that risk tolerance γ is fixed. Incorporating anxiety (next) will allow the effective risk aversion to vary, consistent with empirical evidence (Schildberg-Hörisch, 2018). One experiment attributes a fifth of investors' risk aversion changes as due to emotion (Huang and Xu, 2023).

[9]The assumptions of mean-variance preferences and a static horizon are not crucial. Under widely used CARA-Normal assumptions, the behavior is identical; a log-utility investor behaves similarly (Koijen and Yogo, 2019, pp. 1481). In dynamic settings, a continuous-time investor behaves identically under various sets of assumptions on the utility function, investment opportunity set, and so on; in any case, the mean-variance optimal portfolio (2) is an important component of dynamic portfolio choice.

This now leads the investor to allocate a (weakly) smaller portfolio share to the risky asset:

$$x_t = \frac{1}{a_t \gamma} \cdot \frac{\mathbb{E}_t[R_{t+1}]}{\mathbb{V}\mathrm{ar}_t(R_{t+1})}. \tag{4}$$

The limit $a_t \to \infty$ captures the behavior of an investor who experiences so much anxiety that she does not participate in the stock market at time t at all. Conversely, the lower the anxiety a_t that the investor experiences, the higher the fraction x_t of wealth that she invests in the risky asset (instead of the riskless one). Thus, the term $a_t\gamma$ parametrizes the investor's effective risk aversion, which varies with her anxiety levels.

Money Doctors Can Reduce Investor Anxiety and Build Trust GSV argue that financial intermediaries play a key role in giving investors the confidence to take risks. In their model, a money manager (or "money doctor") can reduce an investor's anxiety a_t. Money doctors thus "provide investors peace of mind." Furthermore, GSV define "trust" as "reducing investor anxiety about taking risk;" under this view, trust and anxiety are inextricably linked. Finally, GSV suggest some possible sources of trust: "personal relationships, familiarity, persuasive advertising, connections to friends and colleagues, communication, and schmoozing."

Consider an investor in a mutual fund. Assuming that $\mathbb{E}_t[R_{t+1}] > 0$, Eqn. (4) predicts that

$$\frac{\partial x_t}{\partial a_t} < 0. \tag{5}$$

That is, any decreases in her anxiety over a time period $\Delta a_{t+1} < 0$ (or, equivalently, increases in her trust) encourage her to increase the fraction of her wealth x that she places in the fund, *ceteris paribus*. These increases in her portfolio allocation $\Delta x_{t+1} > 0$ empirically manifest as inflows into the fund. Holding the investor's beliefs fixed in Eqn. (4) and assuming $x_t \neq 0$, the investor's predicted net flows into the index mutual fund simplify to

$$\text{Net Flow} = \frac{x_{t+1} - x_t}{x_t} = \frac{a_t}{a_{t+1}} - 1. \tag{6}$$

Existing Evidence Two streams of literature support the arguments made by GSV. The first examines exogenous trust-breaking events and finds they lead to outflows, consistent with Eqn. (5). This includes observational studies by Kostovetsky (2016) examining mutual fund clients' flows, and by Gurun, Stoffman, and Yonker (2018) analyzing investment advisor clients' flows. The second stream directly measures the effective risk aversion of investors, and finds it does indeed increase with their anxiety a_t. This includes experimental studies by Raghunathan and Pham (1999), Lerner and Keltner (2001), Maner and Schmidt (2006), Kuhnen and Knutson (2011), Giorgetta et al. (2012), and Guiso, Sapienza, and Zingales (2018).

Roles of Communication Equation (4) makes clear that communication can influence investor risk-taking along two separate paths. First, under a classic view, communication might alter the beliefs $\mathbb{E}_t[R_{t+1}]$ and $\mathbb{V}\mathrm{ar}_t(R_{t+1})$ that investors hold about their future payoff. In the case of active management, these would be dependent on active fund manager skill, while this paper focuses on the passive market portfolio. Nevertheless, information of all kinds could shift these beliefs, as might belief-based persuasion (DellaVigna and Gentzkow, 2010).

The second path for communication to foster risk-taking is by reducing the investor's anxiety a_t. As per Eqns. (5) & (6), a mutual fund investor whose anxiety is soothed by this communication would react with inflows into the fund. A number of experimental studies show that providing information decreases anxiety. In a medical context, Hayward (1975) finds that surgical patients report lower anxiety (and even less physical pain) when provided with information. In a more closely related setting, Eliaz and Schotter (2010) find that subjects who must bear a risky prospect exhibit a demand for information even when it cannot be used instrumentally. Furthermore, the subjects reported that possessing information gave them "confidence in [their] decision" and "a sense of security" (Eliaz and Schotter, 2010, pp. 316). Since $a_t\gamma$ is interpretable as a preference parameter (that governs the investor's effective risk aversion), communication of this type can be categorized as preference-based persuasion (DellaVigna and Gentzkow, 2010).

2.3 Empirical Approach

The GSV empirical framework informs the current paper's approach to studying the impact of communication on investor flows. I examine investors in the S&P 500 market portfolio, intermediated by mutual fund managers. I study the effect of communications about unrealized risks transmitted by mutual fund managers to these investors, including their interplay with investors' prior anxiety levels and trust shocks. Importantly, I control for and rule out various belief-updating channels to home in on the effect of anxiety (Eqn. (5)).

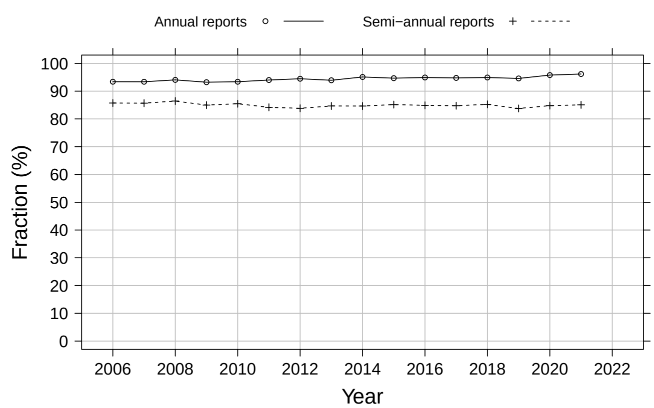
3 Institutional Setting and Data

3.1 Letters to Investors Contained in Funds' Annual and Semi-Annual Reports

The SEC requires funds to disseminate annual reports and semi-annual reports to their investor clients, and these must also be filed within 10 days on the SEC's EDGAR system as Form N-CSR and N-CSRS, respectively. These reports contain mandatory information about historical performance, fees and holdings. They also often contain free-form text communications to shareholders. Such communications do not follow a standardized format, are completely optional, and may contain unaudited statements. I call these statements "fund letters."[10,11]

Fund letters are often formatted as just that, beginning with phrases such as "Dear investors" and ending with a signoff. However, they may also appear in other guises; for example, under a section titled "Management discussion and analysis" or "Interview with portfolio manager." Extracting them requires casting a net wide enough to cover the many forms that these letters can take. My solution to this challenge is to define and then iteratively refine a set of regular expressions that identify the start and end locations of letters based on common phrases and section headings. I repeatedly sample from reports in which no letters were detected and manually read through the reports to select new delimiting phrases, then update the regular expressions and iterate once more. I also sample among extracted letters and cross-check them against the reports from which they were extracted to verify the accuracy of the results. Figure 1 indicates that my letter extraction methodology identifies letters in over nine out of ten annual reports and eight out of ten semi-annual reports in each year of my sample. Therefore, such free-form text is a very common form of communication from mutual fund managers to their investor clients.

Figure 1: Fraction of reports with letters. The vast majority of annual (N-CSR) and semi-annual (N-CSRS) reports published by fund families since 2006 contain at least one letter intended for investor clients. Fund letters within reports were detected using my text extraction methodology.



Both annual and semi-annual reports are published at the level of a legal entity, as identified by a Central Index Key (CIK) rather than at the level of an individual fund.[12] This

[10]According to an Investment Company Institute (2018) survey, 81% of mutual fund investors recall receiving shareholder reports; of those investors, 37% read at least some portion of it. Note that fund letters appear near the beginning of shareholder reports, and are thus likely to be read. Furthermore, 83% of respondents to the survey agree that including a textual discussion of market conditions and performance is "important."

[11]The literature also uses the more general phrase "shareholder letter" (Hillert, Niessen-Ruenzi, and Ruenzi, 2021) to connote that letter contents are not matched to individual funds. I use "fund letter" to emphasize that the letters I extract are targeted at the investors in an individual fund.

[12]For example, "Schwab Equity Index Funds" rather than "Schwab S&P 500 Index Fund."

presents a further challenge for matching letters to individual funds, but also the opportunity to use contemporaneous sibling funds' letter contents as part of an identification strategy.[13] To match letters in the fund family-level report to individual funds, I search for fund names or tickers in the contents of each letter. Some fund families only contain a single fund; in that case, matching is trivial but my identification strategy cannot be employed.

My sample begins in 2006, which is the first year individual fund letters can be identified; prior to that, the SEC EDGAR database did not record the mappings from funds to CIKs that are needed for my matching procedure. The sample ends in August 2021.

3.2 Isolating and Quantifying Risk Discussions in Fund Letters

I contribute a transparent methodology to precisely measure the contents of fund letters that are devoted to risk. I find that discussions about risk are prevalent in fund letters: 8.47% of sentences in the entire corpus discuss this topic.

3.2.1 Screening for Sentences that Discuss Risk

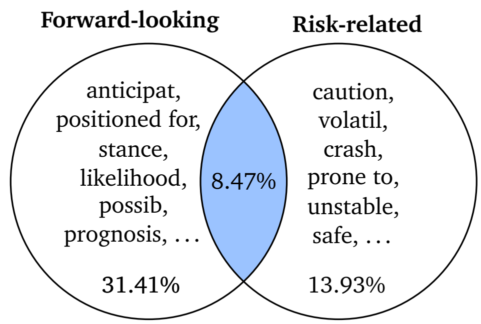
To prepare the text corpus, I first extract grammatically valid sentences from fund letters using Stanford CoreNLP tools (Manning et al., 2014). I then use a set of regular expressions (defined by manually sampling letters) to recognize sentences that contain formulaic legal boilerplate, and discard those.

I also use regular expressions (defined by manually sampling letters) to screen for sentences that discuss the topic of risk. One challenge in isolating information provision about risks is to exclude discussions that simply recap past events: Fund letters frequently discuss past portfolio holdings, returns, and economic and world events that occurred over the past few months. As Equation (4) emphasizes, investor portfolio choices are forward-looking, and this paper focuses on communication about the unknown risks that investors fear. I therefore define two sets of regular expressions to screen separately for sentences that contain risk-related words and word stems, as well as forward-looking words and word stems, and extract sentences that jointly meet these two criteria; Figure 2 graphically illustrates these joint criteria. All regular expressions used in this paper are detailed in Internet Appendix I for ease of extension and interpretation.

Validation In Internet Appendix B.1, I compare my methodology to other approaches that are adapted to different contexts, or that use off-the-shelf software. I find that my customized approach detects a comprehensive set of forward-looking and risk-related statements, while

[13] The identification strategies used in this paper are discussed in detail in Section 4.2.

Figure 2: Screening for statements about risk. Statements about risk are defined as sentences that are matched by both forward-looking and risk-related regular expressions. The percentage figures denote the fraction of all sentences in the fund letter corpus that are matched by each type of regular expression.



also being transparent and adapted to the context of fund letters. Quantitatively, this methodology meets or exceeds the fraction of forward-looking or risk-related statements detected by state-of-the-art software used in the Psychology literature.

3.2.2 Quantifying the Level of Risk Conveyed by Text

I measure the level of risk conveyed by a fund letter (fixing fund i and year-month t) using two approaches.

Sentiment-based Approach The first approach measures the level of the risk conveyed using the net sentiment of the text that discusses risk, specifically. Individual words are classified as having positive or negative (or neutral) sentiment using Loughran and McDonald's (2011) financial sentiment dictionary, with an adjustment to correct for negations in the 3 words preceding positive terms. The net sentiment is then calculated as the difference of positive-sentiment and negative-sentiment word counts, normalized by the total word count of the discussion about risk:

$$\text{Net sentiment of risk text} = \frac{\text{Positive word count} - \text{Negative word count}}{\text{Total word count}} \tag{7}$$

Finally, the level of risk conveyed by this text is simply the negative of the above net sentiment measure, standardized within each fund to normalize for any fund-specific mean sentiment (and thus facilitate comparisons across funds):

$$\text{Level of risk conveyed} = -z(\text{Net sentiment of risk text}). \tag{8}$$

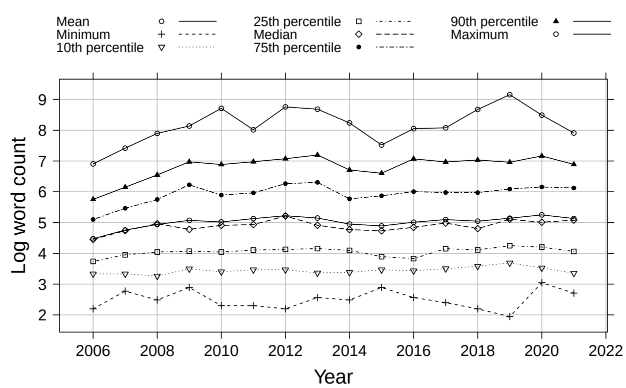
Contextual Approach The second approach focuses on words in the neighborhood of each occurrence of the word "risk" (or its synonyms such as "volatility" or "uncertainty") and counts the number of words denoting a high level of risk (such as "high," "increased," "elevated," "severe," and so on) and the number of words denoting a low level of risk (such as "low," "decreased," "reduced," "moderate," and so on). The window around "risk" or its synonym consists of the 3 words before and the 3 words after. (The results are not affected by widening the window to 5 words.) The contextual measure of the level of risk conveyed is simply the difference of the high-risk and low-risk descriptor word counts in the fund letter:

$$\text{Level of risk conveyed} = \text{High risk word count} - \text{Low risk word count} \tag{9}$$

3.2.3 Variation in Communication About Risk

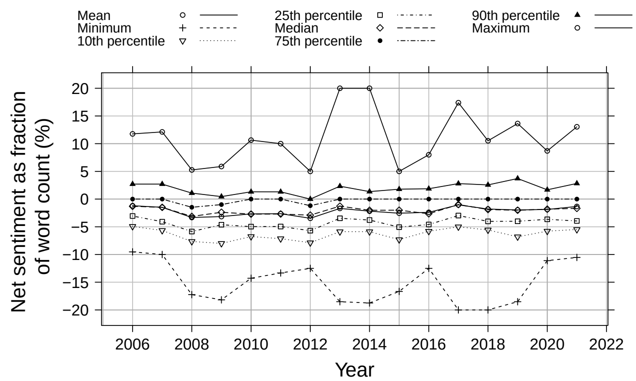
Cross-Sectional Dispersion Figures 3 and 4 summarize the cross-sectional dispersion in the amount of communication about risk (by the log word count) and the net sentiment of that text, respectively. Both variables exhibit considerable dispersion in the cross-section. Sentiment is on average negative (as might be expected when discussing risks) but both net-positive and net-negative sentiment letters coexist at any given point in time.

Figure 3: Dispersion in detail. Quantiles and the mean for the (log) word count of the risk discussions contained in S&P 500 index fund letters.



Timeseries Variation & Determinants Even within a fund, there is considerable variation in both the amount of communication about risk, as well as the level of the risk conveyed by the textual discussion. The mean (median) within-fund autocorrelation of the fraction of the letter about risk is only 0.12 (0.09). The mean (median) within-fund autocorrelation of the number of words devoted to risk is similar, at 0.13 (0.19). The mean (median) within-fund autocorrelation of the level of risk conveyed by the sentiment is only 0.02 (0.08). The mean (median) within-fund autocorrelation of the level of risk conveyed by contextual words is also

Figure 4: Dispersion in sentiment. Quantiles and the mean for the net sentiment conveyed by the risk discussions contained in S&P 500 index fund letters. The Loughran and McDonald (2011) sentiment dictionary is used to calculate net sentiment.



low, at -0.05 (-0.05). As I show next, market conditions are an important determinant of this timeseries variation.

Table 1 analyzes the timeseries means across funds of two detail measures, as well as two measures of the level of risk conveyed. All measures are an equally weighted cross-sectional year-month t mean of the within-fund-standardized (i.e. z-scores) equivalents; each measure is thus a simple index of the features of funds' communication about risks over time. The primary driver of all four timeseries is the contemporaneous level of the VIX. Within-fund variation is thus informative, in the sense that managers provide more detail about risk at precisely the moments when the VIX is higher. The communication is also informative about the level of risk conveyed by letters: managers write more pessimistically about risk when the VIX is high – showing again that they are reflecting an objective measure of risk in their communications.

3.3 Mutual Fund Classifications, Characteristics and Fees

I use the CRSP US Survivor-Bias-Free Mutual Fund Database (MFDB) to compute the mutual fund and fund family characteristics used in this study. These characteristics include classifications of funds as domestic equity funds, or otherwise.

I clean fee data and calculate total fund fees by closely following the procedure of Roussanov, Ruan, and Wei (2021, Internet Appendix). Total fees are therefore defined as the sum of expense ratios (which include management fees, and other expenses) and front loads (annualized by dividing by 7).

Funds are defined based on EDGAR identifiers; namely, CIKs and Series IDs. CRSP fund characteristics are matched to SEC identifiers using the "CRSP_cik_map" table from the CRSP

Table 1: Timeseries indices of communicated detail and sentiment. The dependent variables are equally-weighted indices of the amount of detail in fund letters devoted to risk using the number of words (column 1) or fraction of the total number of words (column 2), and of the level of the risk conveyed using the negative of the Loughran and McDonald (2011) net sentiment of the risk discussion (column 3) or the net score of contextual risk words defined by Equation (9) (column 4). Each dependent variable y_t is calculated as the cross sectional mean over individual fund i variables, $y_t = \frac{1}{N_t}\sum_i z_{i,t}$. These individual fund variables $z_{i,t}$ are pre-standardized to focus purely on within-fund variation, $z_{i,t} = (x_{i,t} - \overline{x_i})/\sigma(x_i)$. As well as the (log) VIX and prior month's return of the S&P 500 index, the independent variables comprise a measure of aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)), and a measure of aggregate perceived ambiguity of the S&P 500 market index (measured by Brenner and Izhakian (2018)). The sample of funds used to construct the dependent variables consists of S&P 500 index funds only.

Dependent Variables:	Risk Detail$_t$		Risk Level$_t$	
Methodology:	Number of Words	Fraction of Letter	Using Sentiment	Using Context
	(1)	(2)	(3)	(4)
Log VIX$_t$	0.2719**	0.2332**	0.3205***	0.2439**
	(0.1191)	(0.1151)	(0.1205)	(0.1205)
AAII Expected Return$_t$ (%)	-0.0002	0.0003	0.0060	-0.0075**
	(0.0022)	(0.0024)	(0.0040)	(0.0033)
Log Ambiguity$_t$	0.0387	0.0115	-0.1327**	0.1521**
	(0.0556)	(0.0596)	(0.0603)	(0.0635)
S&P 500 Return$_{t-1}$ (%)	0.0087	0.0040	0.0029	0.0123*
	(0.0060)	(0.0063)	(0.0079)	(0.0071)
N	183	183	183	183
R^2	0.05	0.04	0.14	0.05

*Newey-West standard errors (in parentheses) use automatically selected lags. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Mutual Fund Database. Share class-level characteristics are aggregated to the fund level by total net assets (TNA).

I obtain fund-level monthly net returns from the CRSP MFDB using a similar procedure. These are then winsorized at the 5% and 95% levels over the entire sample to avoid the undue influence of outliers.

One of my identification strategies relies on measuring communication by active funds; to identify these, I first classify index funds (in general) based on their names, following the list of keywords proposed by Ben-David, Li, Rossi, and Song (2022). Active funds are then defined as non-index funds.

3.4 Fund Flows

Following the extant literature, mutual funds' monthly net flows (from the end of month t to the end of month $t+1$) are calculated based on the fund returns and Total Net Assets (TNA) recorded in the CRSP US Survivor-Bias-Free Mutual Fund Database:

$$\text{Net Flow}_{i,t\to t+1}(\$) \;=\; \text{TNA}_{i,t+1} \;-\; (1+R_{i,t\to t+1}) \times \text{TNA}_{i,t}. \tag{10}$$

In addition to the above measure of net flows, I also make use of data on gross monthly inflows and outflows reported to the SEC at a fund level. These gross dollar flows are extracted from mandatory SEC filing documents (N-SAR, N-PORT, and any amendments), with the latest available observation retained per month.

All dollar flows are normalized to a percentage of lagged assets:

$$\text{Net Flow}_{i,t\to t+1}(\%) \;=\; \frac{\text{Net Flow}_{i,t\to t+1}(\$)}{\text{TNA}_{i,t}} \times 100\%. \tag{11}$$

A similar normalization to Eqn. (11) is applied to gross inflows and gross outflows.

To avoid the undue influence of outliers, including those due to corporate actions, the net flow percentage variable is winsorized at the 5% and 95% levels over the full sample. Similarly, each gross flow percentage variable is winsorized at the 95% level.

3.5 S&P 500 Index Funds

S&P 500 index funds may be either mutual funds or exchange traded funds (ETFs). This study focuses on index mutual funds, which are more prevalent. To select S&P 500 index mutual funds, I source monthly fund return information from the CRSP US Survivor-Bias-Free Mutual Fund Database: I aggregate returns from the share class level to the fund level, then correlate these returns with the returns on the S&P 500 index and discard funds with a correlation of less than 98%. Finally, I filter out funds whose names indicate they are not following a pure passive indexing strategy (such as quant funds, funds tilted to a particular sector, or funds that track a different index). This procedure selects a total of 211 S&P 500 index mutual funds over the full sample, of which 197 can be assigned SEC EDGAR identifiers.[14]

3.6 SEC EDGAR Usage

I measure the geographic locations of fund letter readers using SEC EDGAR usage logs, beginning in 2006 and ending in June 2017. I follow Lee, Ma, and Wang (2015) in identifying and

[14]By way of comparison, I repeated the procedure on the ETFs recorded in the CRSP monthly stock database, and found only 12 S&P 500 index ETFs. Focussing on mutual funds thus captures the vast majority of S&P 500 tracker funds.

removing robots from the usage logs, and de-anonymize IP addresses using the lookup table provided by Chen et al. (2020). Similarly to Grice and Gueciioueur (2023), I then geolocate IP addresses using historical mapping tables provided by the MaxMind geolocation service. I recalculate the geographic distribution of a fund's investor base (as measured through their readership of fund letters) at an annual frequency. Internet Appendix B.2 shows that, overall, fund letter readers are geographically distributed in a very similar manner to the overall population of the United States. I later exploit fund-level cross-sectional variation in the locations of their letter readers.

4 Fund Manager Communication and Investor Risk-taking

This section shows that more detailed fund manager communication encourages investor risk-taking. I first quantify the effects of detail and sentiment on individual S&P 500 tracker fund flows. I show that sentiment does not, in fact, predict fund flows, which is initial evidence against a belief-updating channel; in Section 6 I rule out such a channel more strongly. Section 4.2 formulates three distinct (and hence complementary) identification strategies that each confirm the presence of a causal effect.

4.1 Communication and Fund Flows

I begin by analyzing the relationship between net flows into S&P 500 index funds and the amount of information about risk transmitted by each fund to existing investors. I find that they are positively associated, irrespective of whether the conveyed sentiment is positive or negative.

Amount of Detail The panel that I analyze consists of fund i and year-month t pairs for each disseminated fund letter. Letters are published at a semi-annual frequency (approximately) and I focus on months t during which a letter was disseminated. I capture the amount of detail provided about risk by the (log) word count devoted to this topic. My baseline panel regression takes the following form:

$$\text{Flow}_{i,t\to t+1}(\%) = \beta \times \text{Risk Detail}_{i,t} + \Gamma \times \text{Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}. \tag{12}$$

That is, I measure the strength of association β between the amount of communication about risk during a month t and the normalized percentage flow into the fund from the end of month t to the end of month $t+1$. I separately study net flows, gross inflows and gross outflows as the dependent variable. The richest specifications include fund fixed effects π_i and year-month time effects ϕ_t. I always include controls for time-varying fund level characteristics that have been previously found to drive fund flows (Sirri and Tufano, 1998; Chevalier and Ellison, 1999) including the fund's past returns (and its square, to allow for a nonlinear relationship), fees and size, and the fund family's age and size. The controls also include the (log of the) total number of words in a fund letter, which effectively normalizes the length of communication about risks by the overall length of the letter. Γ is a vector of coefficients on the set of controls.

Table 2 columns 1-4 present the results of panel regressions of the net flows of S&P 500 funds regressed against the amount of communication about risk contained in fund letters disseminated the previous month. Importantly, controlling for year-month time effects ϕ_t in columns (1) and (4) implicitly controls for *any* time-varying potential confounders; this includes past or future market returns and measures of risk or risk premia, market-wide sentiment, any (common) prior held by investors, the growth of passive investing, and any other market-wide or economy-wide events or shocks. All estimated β coefficients in columns 1-4 are positive and significantly different to zero, with the coefficient magnitudes increasing slightly as more controls are incorporated. Furthermore, the positive and significant association survives controls for other known drivers of fund flows such as past returns, as well as the fund fixed effects π_i and time effects ϕ_t.

To illustrate that the effect is not only cross-sectional (i.e. present when time effects are included) but also occurs over time within a given fund, the specifications in columns (2) and (3) do not include time effects at all. In column (3) I explicitly control for investors' beliefs about the distribution of the S&P 500's future return: I use survey expectations (Greenwood and Shleifer, 2014) and the (log) level of the VIX for the first and second moment, respectively. In column (4), I include both fund fixed effects and time effects and find that the magnitude slightly strengthens.

Alternative Specifications Internet Appendix D presents a number of alternative specifications: (i) a simple linear specification, (ii) measuring detail using the fraction of a letter devoted to discussing risk, and (iii) using log-log panel specifications. For each, the conclusion remains that an increase in the amount of detail attracts more inflows. The log-log specification also admits a structural interpretation.

Economic Significance The flow-detail effect is economically significant. This is most clear when regressing standardized (i.e. z-scored) net flows against standardized detail measures in Table 2 columns 5-8. In column (8), the richest specification, a growth of 1 standard deviation (SD) in the amount of detail above its average predicts a next-month increase in the fund's assets of 0.23 SDs in magnitude – this in turn translates to a 0.67 percentage point increase of assets under management, on average. To put this number in context, a 1 SD increase in

Table 2: Detail communicated about risk vs. net flows into S&P 500 index funds. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). Columns 1-4 present results with no variable transformations. Columns 5-8 present results where both the dependent and the main independent variables have been standardized (z-scores). *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only.

Dependent Variables:	Net Flow$_{i,t\to t+1}$ (%)				z(Net Flow$_{i,t\to t+1}$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Detail$_{i,t}$	0.3893***	0.4747***	0.4754***	0.5044***				
	(0.1238)	(0.1529)	(0.1528)	(0.1808)				
z(Risk Detail$_{i,t}$)					0.1762***	0.2149***	0.2152***	0.2283***
					(0.0560)	(0.0692)	(0.0692)	(0.0819)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls$_t$			✓				✓	
Year-month FEs	✓			✓	✓			✓
Fund FEs		✓	✓	✓		✓	✓	✓
N	1,155	1,155	1,155	1,155	1,155	1,155	1,155	1,155
R^2	0.21	0.31	0.31	0.44	0.21	0.31	0.31	0.44

*Clustered (Fund & Year-month) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 3: Detail communicated about risk vs. gross flows to/from S&P 500 index funds. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Gross dollar flows are extracted from mandatory SEC filing documents (N-SAR, N-PORT), and then normalized as a percentage of lagged total net assets. Columns 1-4 focus on the association with gross inflows. Columns 5-8 focus on the association with gross outflows. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only.

Dependent Variables:	Inflow$_{i,t\rightarrow t+1}$ (%)				Outflow$_{i,t\rightarrow t+1}$ (%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Detail$_{i,t}$	0.2272	0.4901***	0.4932***	0.4709**	-0.0772	0.0194	0.0203	0.0503
	(0.2277)	(0.1731)	(0.1742)	(0.2021)	(0.1676)	(0.0820)	(0.0823)	(0.0971)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls$_t$			✓				✓	
Year-month FEs	✓			✓	✓			✓
Fund FEs		✓	✓	✓		✓	✓	✓
N	1,155	1,155	1,155	1,155	1,155	1,155	1,155	1,155
R^2	0.26	0.74	0.74	0.79	0.25	0.83	0.83	0.85

*Clustered (Fund & Year-month) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

fees is associated with a 0.60 SD decrease in net flows; the effect of communicated risk detail on flows is therefore approximately a third of the effect of fees on flows.

Un-standardized estimates also convey the magnitude of the effect. In column (4), the richest specification, the estimated coefficient value of $\hat{\beta} \approx 1/2$ indicates that a fund that doubles the detail it communicates about risk attracts an additional $1/2 \times \log(1+1) \approx 0.35\%$ worth of assets under management in the month following its dissemination of a fund letter. The economic magnitude also remains notable when assessing the aggregate effect in Section 7.

Unobservables The stability of the estimated coefficients in Table 2 is informative about what effect any unobservables might have. The baseline results in this table show that including additional controls, especially fund-level fixed effects, does not change the main coefficient estimates in any qualitative or economically meaningful way, but does increase the R^2: comparing columns (1) to (4) and columns (5) to (8), we see that including fund fixed effects more than doubles the R^2. Unobservables are hence unlikely to explain the main effect, reducing potential concerns about selection bias and omitted variable bias (Altonji, Elder, and Taber, 2005; Oster, 2019). This finding further justifies the present paper's focus on S&P 500 index funds to aid in identification. Section 4.2.1 elaborates further by formally applying the Oster (2019) test.

Decomposition into Inflows & Outflows I now decompose net flows into gross inflows and gross outflows. Table 3 repeats the baseline panel regression (12) on each gross flow variable, including the same fixed effects, time effects and controls as previously. These estimates show that the effect acts through increasing inflows, rather than decreasing outflows: the estimates for the effect on inflows in columns 1-4 are positive and significant; by contrast, the estimates for the effect on outflows in columns 5-8 are not significantly different to zero. Since fund letters are targeted at existing investors rather than potential clients (who are attracted by fund prospectuses or advertising), this positive coefficient can be interpreted as existing investors increasing their share of the risky asset. Furthermore, the absence of a negative effect on outflows suggests investors do not reallocate capital among S&P 500 index funds according to how much they communicate about risk; therefore, this is also consistent with additional risk-taking by investors.

Role of Directional Information (or Lack Thereof) Given that the level of risk conveyed in fund letters is informative, in the sense that it is positively associated with the contemporaneous level of the VIX (Table 1), I now examine whether the level of risk conveyed alters the baseline detail effect; I find it does not.

Table 4: Adding the level of risk communicated to the amount of detail, and their joint relation with S&P 500 index fund flows. This table incorporates the level of risk conveyed by the discussion about risk into the baseline flow-detail panel regression. The amount of detail is measured as the log of the number of words devoted to discussion of risk. The level of risk is measured as the negative of the within-fund-standardized net sentiment of the text devoted to discussing risk, using the Loughran and McDonald (2011) sentiment dictionary (columns 2-3), or the net score of contextual risk words defined by Equation (9) (column 4), or a modified score that counts only high-risk words (columns 5-6). Column 1 repeats the baseline flow-detail panel regression. Columns 2, 4 & 5 augment it with the level of risk conveyed by the letter, using three different measures. Columns 3 & 6 include an interaction between the effect of detail and whether the letter conveys a high level of risk (i.e. when $\mathbb{1}\{\text{High Risk Level}\}_{i,t}$ is 1); a high level of risk is defined as a below-median sentiment in column 3, and the presence of one or more high-risk words in column 6. Fund controls (not shown) are for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. A control for the total (log) word count is displayed separately. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Risk Detail$_{i,t\,i,t}$	0.5044***	0.4976***	0.5258***	0.5022***	0.4950***	0.4795**
	(0.1808)	(0.1768)	(0.1974)	(0.1842)	(0.1877)	(0.2002)
Risk Level$_{i,t}$		0.1043		0.0058	0.0144	
		(0.1123)		(0.0169)	(0.0201)	
$\mathbb{1}\{\text{High Risk Level}\}_{i,t}$			0.3339			0.5278
			(0.8204)			(1.242)
Risk Detail$_{i,t}$			-0.0457			-0.0404
$\times\ \mathbb{1}\{\text{High Risk Level}\}_{i,t}$			(0.1380)			(0.2232)
Total Detail$_{i,t}$	-0.3843	-0.3792	-0.3879	-0.3829	-0.3791	-0.3653
	(0.2368)	(0.2355)	(0.2399)	(0.2391)	(0.2406)	(0.2417)
Risk Level measure		Sentiment	Sentiment	High-Low Words	High Words	High Words
$\mathbb{1}\{\text{High Risk Level}\}$ threshold			Median			0
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓	✓
N	1,155	1,137	1,137	1,154	1,154	1,154
R^2	0.44	0.43	0.43	0.44	0.44	0.44

*Clustered (Fund & Year-month) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

In Table 4, columns 2, 4 and 5, I show the results of augmenting the baseline panel regression (of Eqn. (12)) to incorporate multiple measures of the level of risk conveyed:

$$\begin{aligned}\text{Net Flow}_{i,t\to t+1}(\%) = {} & \beta_1 \times \text{Risk Detail}_{i,t} + \beta_2 \times \text{Risk Level}_{i,t} \\ & + \Gamma \times \text{Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}. \end{aligned} \tag{13}$$

In each case, the estimates $\hat{\beta}_1$ are positive, significant and of the same order of economic magnitude as the baseline effect (which is repeated in column 1 for reference). These results hold no matter whether the level of risk is measured based on the Loughran and McDonald (2011) sentiment (column 2), or by counting high-risk and low-risk words (column 4), or by counting high-risk words only (column 5). Furthermore, insignificant estimates for $\hat{\beta}_2$ indicate the level of risk conveyed is not associated with any variation in fund flows. Investors thus appear to ignore the level of risk conveyed by the text, even though an earlier analysis found it covaries with the level of implied volatility (Table 1).

Columns 3 and 6 of Table 4 display panel regression results that examine whether the response of flows to the amount of detail provided about risk is attenuated by the level of risk conveyed in the letters, using variants of the following specification:

$$\begin{aligned}\text{Net Flow}_{i,t\to t+1}(\%) = {} & \beta_1 \times \text{Risk Detail}_{i,t} + \beta_3 \times \mathbb{1}\{\text{High Risk Level}\}_{i,t} \\ & + \beta_4 \times \mathbb{1}\{\text{High Risk Level}\}_{i,t} \times \text{Risk Detail}_{i,t} \\ & + \Gamma \times \text{Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}. \end{aligned} \tag{14}$$

Once again, the positive and significant $\hat{\beta}_1$ estimates remain similar. The non-significance of the $\hat{\beta}_4$ estimates indicates that more detailed communication about risk remains associated with higher net flows even when the level of risk communicated is high. This holds whether the letter conveys a high level of risk by expressing a below-median sentiment (column 3), or explicitly describes risk levels as high (column 6). As a further test, Internet Appendix H.5 confirms the main flow-detail effect is robust to focusing only on letters that convey a high level of risk.

Taken together, these results highlight that the primary driver of fund flows is the amount of detail communicated *irrespective* of the information conveyed about the actual level of risk. Further analyses in Section 6.2 provide evidence that investors do not update their beliefs about the level of risk based on the information conveyed by these fund letters.

4.2 Identification

The fundamental pillar of this paper's ability to identify the effect of communication is its focus on quasi-identical S&P 500 index funds — in such a setting, important confounding channels such as learning about or signaling/disclosing manager skill are not relevant. This section seeks to confirm that the estimates in the previous section are not driven by omitted variables or alternative forms of selection bias. To achieve this, I employ three distinct and complementary methodologies to assess potential endogeneity.

The first relies on partial identification results (Altonji, Elder, and Taber, 2005; Oster, 2019). The second relies on an exclusion restriction, by employing an instrumental variables approach. The third relies on a shape restriction, by exploiting a discontinuity induced by corner bunching (Caetano, Caetano, and Nielsen, forthcoming). Despite the very different assumptions underlying these methods, the three analyses consistently confirm the presence of a causal effect.

4.2.1 Ruling Out Unobservable Drivers

Building on the seminal work of Altonji, Elder, and Taber (2005), Oster (2019) shows how to test for the presence of unobservable selection using empirical measures of coefficient stability. I now apply her test to confirm my baseline results are not driven by further unobservable omitted variables. Reassuringly, the test implies unobservables do not drive the findings.

The test is based on calculating a coefficient of proportionality $\tilde{\delta}$ that measures the proportional degree of selection on unobservables relative to observables that would be needed to nullify the estimated treatment effect. The calculation of $\tilde{\delta}$ itself involves a sensitivity analysis of the baseline regression, and is described in detail by Oster (2019). The main parameterization required is for the hypothetical share of variance that could be explained; I adopt her suggested assumption of $\tilde{R}^2 = \min\{1.3\hat{R}^2, 1\}$. I set $\hat{R}^2$ to that of Table 2 column (4), which is my richest specification, and which has the most explanatory power. Given the above parameterization, Oster (2019) argues a coefficient of proportionality with value $\tilde{\delta} > 1$ signifies the effect is robust to the influence of unobservables.

The resulting degree of proportionality is calculated to be $\tilde{\delta} = 1.21$. Since this is strictly greater than 1, it provides reassuring evidence that unobservable drivers are not responsible for the baseline effect.

4.2.2 Instrumental Variables Identification Strategy

As a further check, I now formulate an instrumental variables (IV)-based identification strategy that shifts the explanatory variable (the amount of detail communicated). The strategy is facilitated by this paper's focus on passive funds.

The key quantity of interest in my analyses is the (log of the) number of words within an S&P 500 fund letter devoted to risk. To construct an instrument for this quantity, I calculate similar measures for the contemporaneous letters of other funds within the same fund family,

Table 5: Instrumented effect of the amount of detail communicated about risk on the net flows of S&P 500 index funds. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. The instrument for the main covariate used in the 2SLS specifications is constructed from the log word counts (about the same topic of risk) of the letters written by active domestic equity sibling funds within the focal index fund's overall family. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only. Columns 1-3 show un-instrumented panel specifications. Columns 4-6 show 2SLS specifications, with F-stats corresponding to the first stage; the first-stage regressions are shown in Table IA.8.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)							
Regression:	OLS				2SLS IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Detail$_{i,t}$	0.3893***	0.4747***	0.4754***	0.5044***	0.4896***	0.6087***	0.6102***	0.6503***
	(0.1238)	(0.1529)	(0.1528)	(0.1808)	(0.1298)	(0.1777)	(0.1785)	(0.2044)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls$_t$			✓				✓	
Year-month FEs	✓			✓	✓			✓
Fund FEs		✓	✓	✓		✓	✓	✓
N	1,155	1,155	1,155	1,155	1,155	1,155	1,155	1,155
R^2	0.21	0.31	0.31	0.44	0.21	0.31	0.31	0.44
F-stat (IV 1st-stage)					253.4	294.8	236.2	361.3

Clustered (Fund & Year-month) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

then take the family-level mean over all active domestic equity funds. This excludes any passive index funds, including the focal S&P 500 fund of interest.

To understand the information spillover mechanism captured by this instrument, recall first that the market portfolio return, in which S&P 500 index fund investors are invested in, is the aggregated return of all constituent firms or industries i:

$$R_{t+1} = \frac{1}{\sum_i \text{Size}_{i,t}} \sum_i \text{Size}_{i,t} R_{i,t+1} \tag{15}$$

In the ordinary course of business, an actively-managed fund uncovers a series of signals ε_i about some subset of individual returns $R_{i,t+1}$ as part of its portfolio choice process. These signals ε_i about individual firms or industries i can be aggregated up to a signal ε about the market portfolio return R_{t+1}. I do not observe the internal research of active funds within the family (i.e. the ε_i realizations), but can observe the amount of detail about risk that these active funds communicate to their own clients, via their own fund letters. $\bar{w}_A = \frac{1}{N_k} \sum_{k \in \text{Active}} w_k$, the mean over all these variables w_k, captures the amount of detail about risk that emerges from active funds' internal research. This instrument captures information diffusion among funds in the same family, which drives variation in the content of the focal S&P 500 index fund's letter.[15]

Two assumptions must hold for this instrument to be valid. First, there must be no reverse information spillovers from passive funds to active funds. Since passive funds do not conduct fundamental research of their own, a reverse spillover is unlikely.[16] Second, any information about risks contained in active funds' letters should not affect flows to passive funds (other than via information spillovers to passive funds' letters). Analyses to follow in Section 6.1 show that more detailed communication about risk in sibling active funds' letters does *not* trigger outflows from these active funds themselves, which is the most likely source of rebalancing inflows to the passive funds. Hence, such a potential violation is also unlikely.

My IV-based identification strategy produces similar results to the paper's baseline regressions. Table 5 columns 1-4 repeats these baseline results (from Table 2), and then columns 5-8 show two-stage least squares (2SLS) estimates that have been instrumented using the sibling domestic active equity fund communication detail instrument. Note that the inclusion of

[15] I do *not* assume that funds share fundamental research or trade ideas with one another, although there is evidence of this in prior work (Cici, Jaspersen, and Kempf, 2017; Auh and Bai, 2020). I assume only that the market-wide views expressed in fund letters by passive funds incorporate information produced by active funds within the same fund family.

[16] It might be argued that passive funds still conduct fundamental research as part of their shareholder duties (voting and monitoring), but the literature finds this is not the case. Iliev, Kalodimos, and Lowry (2021) measure research activity at the fund family level using SEC EDGAR usage, and find that fund families with a larger share of passive assets conduct less research. Heath, Macciocchi, Michaely, and Ringgenberg (2022) find that index funds are poor monitors as a result. Moreover, while some index funds are closet active funds, these tend to use custom indices (Akey, Robertson, and Simutin, 2021); the present paper focuses on S&P 500 index funds.

fund fixed effects in columns 6-8 also implicitly controls for any unobserved and persistent differences that are due to membership in a particular fund family, since fund fixed effects also subsume family fixed effects. The positive and significant effect remains, and can now be interpreted as the causal effect of increased communication about risk: more detailed communication about risk thus causes a comparative increase in flows into the risky portfolio managed by the fund. Further details of the IV identification strategy – including evidence of information spillovers within fund families – are provided in Internet Appendix E.1.

4.2.3 Corner Bunching Identification Strategy

My third and final analysis exploits corner bunching in the detail treatment variable for identification. Recall that the baseline specifications in the paper so far normalize for the total (log) word count of the entire fund letter by including that variable as a control. I now construct a related measure of the amount of detail about risk: the fraction of a fund letter devoted to this topic, without taking logs of either the numerator or denominator:

$$\text{Detail Fraction}_{i,t} = \frac{\text{Number of words about risk}_{i,t}}{\text{Total number of words in letter}_{i,t}}. \tag{16}$$

Figure 5: Corner bunching used in identification. These figures visualize the distribution of the fractions of S&P 500 index fund letters that are devoted to risk. The left chart shows the empirical CDF of this variable, and the right chart shows its histogram. The histogram highlights corner bunching at the value 0%: this is shown by the single bar colored red, while all other bars are colored blue. A corresponding sharp discontinuity is visible in the empirical CDF at that value. The presence of this feature of the empirical distribution is used in the corner bunching identification strategy described in Section 4.2.3.

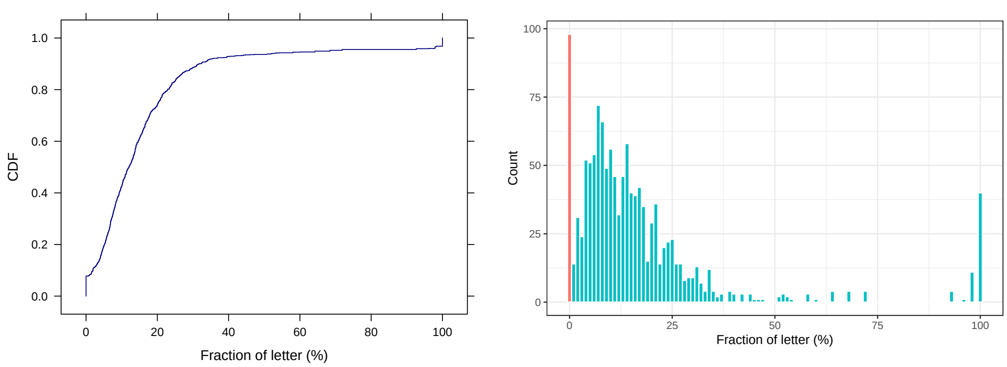


Figure 5 highlights the presence of bunching at the 0% corner value. Recent methodological contributions by Caetano, Caetano, and Nielsen (forthcoming, henceforth **CCN**) have been developed to exploit precisely such a feature of the distribution for identification purposes. CCN's contribution is to show that, using the samples whose treatment variable is bunched at the lower support, a combination of the discontinuity and shape restrictions on the controls can be leveraged to recover (and correct for) unobserved endogeneity in the treatment effect at large. The key intuition is that the treatment variable is constrained at the corner solution, while the unobservables that potentially drive endogeneity are not, and hence can be modeled. Further details of the methodology are described in Internet Appendix E.2. In the current section, I employ CCN's approach to detect and correct potential endogeneity.

Table 6: Corner bunching-based identification of the effect of the amount of detail communicated about risk on the net flows of S&P 500 funds. This table shows estimates of Eqn. (17), for various correction function specifications (columns 2-4). β is the coefficient capturing the corrected effect of the amount of detail contained in the letter written by fund i at time t about risk, using the fraction of the letter devoted to that topic. δ is the coefficient on the endogeneity correction term. The procedures for estimating the endogeneity correction terms used in specifications 2-4 are described in Internet Appendix E.2. Fund controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. Standard errors are based on 10,000 bootstrap replicates. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)			
Correction:	Uncorrected	Tobit	Semiparam. Tobit	Symmetric
	(1)	(2)	(3)	(4)
β	0.026***	0.067*	0.091*	0.057**
	(0.008)	(0.04)	(0.051)	(0.029)
δ		-0.038	-0.062	-0.029
		(0.036)	(0.048)	(0.026)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	1,253	1,253	1,253	1,253

Bootstrapped standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

I now apply two modifications to this paper's main baseline regression (12) of the amount of detail about risk in a fund letter on the next-month net flows, with the main coefficient of interest once more named β. The usual additive year-month and fund fixed effects are included, together with fund controls. The first modification is to use the fraction of a letter devoted to risk (16), rather than the log word count; I thus remove the total (log) word count from the included controls. The second modification is to estimate CCN's endogeneity correction term, and include it in the regression with coefficient δ:

$$\begin{aligned}\text{Flow}_{i,t\to t+1}(\%) = \; & \beta \times \text{Risk Detail Fraction}_{i,t} + \delta \times \text{Correction}_{i,t} \\ & + \Gamma \times \text{Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}. \end{aligned} \tag{17}$$

Table 6 column (1) shows the result of estimating Eqn. (17) while omitting the correction term. Here, β is positive and significantly different to zero. However, β remains potentially subject to endogeneity concerns, and therefore the remaining specifications exploit corner bunching to estimate the endogeneity correction term according to the methodology detailed in Internet Appendix E.2.

Table 6 column (2) shows the result of estimating Eqn. (17) while including a CCN correction term that is parameterized using a Tobit shape restriction. Column (3) likewise shows estimates when specifying the CCN correction term using a semiparameteric Tobit. Finally, column (4) shows estimates due to a weaker, nonparametric distribution assumption: that the distribution's tails should be symmetric. Importantly, the $\widehat{\delta}$ estimates are not significantly different to zero in either specification, suggesting there is no discernible endogeneity: the non-significant $\widehat{\delta}$ estimates directly map to the absence of unobservable confounders, as explained in Internet Appendix E.2. Accordingly, the treatment effect estimates $\widehat{\beta}$ remain positive and significantly different to zero. Therefore, according to CCN's framework, the positive effect of the amount of detail on flows can be interpreted as a causal treatment effect.

The absence of detectable endogeneity provides further support for a causal interpretation of the effect of communication. The three identification strategies employed in this section are therefore consistent in their conclusion.

5 Inspecting the Mechanism

This section provides evidence that communication-induced inflows are consistent with an anxiety-alleviating mechanism.

5.1 Detailed Communication Alleviates Anxiety

The money doctors hypothesis by GSV posits that financial intermediaries can alleviate investors' anxiety. This section tests whether the anxiety attitudes of fund letter readers impact their response to the provision of information about risk by fund managers.

To test for investors' anxiety playing a role, I examine the differential effect of communication on high-anxiety versus low-anxiety readerships. My measure for the *ex ante* level of anxiety experienced by a letters' readers is constructed by combining the geolocations of readers with local search activity on Google. The Google Trends engine compiles Search Volume

Indexes (SVIs) for umbrella "topics," which group together related search terms, and defines the following two anxiety-related topics:

1. The first anxiety topic, which Google labels "Psychological Stress", includes the search terms "stressed", "what is stress", "anxiety" & "relieve stress."

2. The second anxiety topic, which Google labels "Worry", includes the search terms "worry about", "anxiety", "stop worrying" & "bible worry."

Figure IA.3 in Internet Appendix C charts geographic variation in the SVIs of the above anxiety topics, highlighting their similarity. I employ panels of these SVI measures (varying at a geographic and monthly level) as measures of the aggregate anxiety levels of the population in each US state. Analyses in Internet Appendix B.4 demonstrate that, as expected, these Google search-based anxiety variables have predictive power for households' financial risk-taking in aggregate.[17]

I take the mean of both anxiety indices. Then, for this geographic state s-month t anxiety measure, I aggregate across fund letter readers to construct a panel of the prior anxiety attitudes of the readers of the letter issued by fund i at month t:

$$\text{Readership Anxiety}_{i,t} = \frac{\sum_j \text{Investment}_{j,t} \times \text{Geographic Anxiety}_{j,t}}{\sum_j \text{Investment}_{j,t}}. \tag{18}$$

I measure the geographic distribution of a fund's letter readership base by geolocating the IP addresses of readers who have downloaded annual or semi-annual reports through the SEC EDGAR website, as described in Section 3.6. In constructing the weighted average (18), I proxy for the amount invested by an investor in a geographic region by the mean dollar investment in a mutual fund, according to the Census Bureau's latest Surveys of Income and Program Participation (SIPP).

Finally, I repeat the flow-risk detail panel regressions of Tables 2 & 4, with a cross-sectional sample split based on letter readers' *ex ante* anxiety attitudes: at each year-month t, I split fund letter readerships into those with below-median and above-median anxiety attitudes (measured by Eqn. (18)). I also control for each fund letter readership's (mean investment-weighted) exposure to local economic growth and to inflation, as measured by the state-level Philadelphia Fed Coincident Index and by Hazell, Herreno, Nakamura, and Steinsson's (2022) state-level inflation rates, respectively. Table 7 presents the results of these panel regressions.

[17]Note that, while I acquired SVI data using Google's restricted Trends API, the publicly accessible Google Trends website also allows SVI data to be downloaded. At the time of writing (August 2023), the "Psychological Stress" topic SVI was accessible at `https://trends.google.com/trends/explore?q=/m/0glpkn8&date=all&geo=US`. Similarly, the "Worry" topic SVI was accessible at `https://trends.google.com/trends/explore?q=/m/0bg_x1&date=all&geo=US`.

Table 7: Readership anxiety levels, and the differential effects of the amount of detail & level of risk communicated on S&P 500 index fund flows. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. The level of risk conveyed is calculated according to Eqn. (8). The sample consists of S&P 500 index funds only, and is split cross-sectionally (i.e. at every year-month t) into fund letter readerships with below-median anxiety levels (columns 1-2), and above median anxiety levels (columns 3-4). In measuring readership anxiety levels, the sum of search volume indices for the Google Trends topics of "Psychological Stress" and "Worry" is used. *Local economy controls* (not shown) are for each fund letter readership's contemporaneous asset-weighted exposure to local economic activity & employment (measured by the Philadelphia Fed's State Coincident Index) and to local inflation rates (measured at a state level by Hazell, Herreno, Nakamura, and Steinsson (2022)). *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)			
Sub-sample:	Low Anxiety		High Anxiety	
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	-0.1019	-0.1170	0.5850***	0.5795***
	(0.1386)	(0.1365)	(0.2064)	(0.2065)
Risk Level$_{i,t}$		0.1137		0.2221**
		(0.1060)		(0.1109)
Local Economy Controls$_{i,t}$	✓	✓	✓	✓
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	553	552	686	685
R^2	0.61	0.61	0.52	0.52

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The estimates in Table 7 show readers' anxiety attitudes are crucial determinants of their response to communication about risk. Low-anxiety readers do not respond to more detailed communication in a statistically significant manner (columns 1-2). By contrast, high-anxiety readers respond strongly to more detailed communication about risks (columns 3-4). Given that anxious investors are predisposed to take less risk to begin with (e.g. Guiso, Sapienza, and Zingales, 2018), these results support a mechanism in which communication alleviates investor anxiety.

Robustness Checks Internet Appendix G.4 confirms the robustness of my findings on the influence of investors' anxiety attitudes. I first recast Table 7 to pool both sub-samples and use interaction terms to verify the statistical significance of the finding that more anxious investors respond more strongly to communication about risk. I then include a number of demographic controls for the readership base of fund letters, including variables such as wealth and age (which may enter into investors' risk aversion) and the likelihood of living in an urban area (which may enter into anxiety attitudes). The regression estimates remain similar when demographic controls are included.

Further Results on the Effectiveness of Reassurance I conduct additional analyses in Internet Appendix G.2 to understand when this anxiety-alleviating communication is most effective. I find that, within a fund, the same amount of communication is more effective in times of high market stress, which is precisely when the communication is most needed by an anxious investor. I also find that, across funds, the same amount of communication is more effective when transmitted to a less sophisticated base of investor clients. Both results are consistent with a mechanism in which communication reassures anxious investors.

5.2 Distrust in Financial Intermediaries Attenuates Investor Reassurance

To the extent that alleviating investors' anxiety about risk-taking increases their trust in the stock market (GSV), the findings of Section 5.1 shed light on the origins of investor trust – namely, that trust in the stock market can be established through communication about the very risk that investors fear. I now investigate whether trust in the fund managers who transmit this communication plays an additional, complementary role. Is trust in the source of communication necessary for it to be effective?

Given the difficulty of measuring the perceived trustworthiness of fund managers in the eyes of the investors, I follow the literature in exploiting the occurrence of fraud as a negative shock to financial intermediaries' trustworthiness (Kostovetsky, 2016; Gurun, Stoffman, and Yonker, 2018). Specifically, I employ a geographic measure of negative shocks to investors' trust, which therefore allows me to measure the exposure of different groups of fund letter readerships to the incidence of this trust-breaking fraud.

Like Giannetti and Wang (2016), I measure local negative trust shocks using the public revelation of securities fraud investigations into local companies, as compiled by Karpoff, Koester, Lee, and Martin (2017). Using the state in which each company was headquartered in, I construct a state-level measure of the local investor population's exposure to securities fraud incidents. Specifically, for an individual j, her exposure to securities fraud during a period t is measured based on the number of fraud incidents by companies headquartered in the

Table 8: Communication-driven flows and readership exposure to securities fraud. The amount of detail contained in the letter written by fund i at time t about risk is measured using the log of the number of words devoted to that topic. The indicator variable $\mathbb{1}\{\text{High Fraud Exposure}\}_{i,t}$ is 1 when the readership of the fund's letter currently has an above-median Fraud Exposure Index, and zero otherwise; a baseline effect is estimated in addition to the interaction shown in the table. The variable Fraud $\text{Exposure}_{i,t}$ is rank-standardized, and thus runs from 0 to 1; a baseline effect is estimated in addition to the interaction shown in the table. The Fraud Exposure Index is a weighted average (by unique report downloads) of revelations of securities fraud (recorded by Karpoff, Koester, Lee, and Martin (2017)) committed by publicly listed companies headquartered in the local areas of each fund i's annual/semi-annual report readers, varying annually. The overall sample period begins in 2006 (due to EDGAR data availability) and ends in 2012 (with the securities fraud database). *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net $\text{Flow}_{i,t\to t+1}$ (%)	
	(1)	(2)
Risk $\text{Detail}_{i,t}$	0.6919^{*}	0.9146^{*}
	(0.4142)	(0.4828)
Risk $\text{Detail}_{i,t}$ × $\mathbb{1}\{\text{High Fraud Exposure}\}_{i,t}$	-0.4601^{*}	
	(0.2621)	
Risk $\text{Detail}_{i,t}$ × Fraud $\text{Exposure}_{i,t}$		-0.8466^{*}
		(0.4361)
Fund $\text{Controls}_{i,t}$	✓	✓
Year-month FEs	✓	✓
Fund FEs	✓	✓
N	440	440
R^2	0.54	0.54

Clustered (Fund & Year-month) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

state in which she is located, divided by the total number of public companies headquartered in the same state:

$$\text{Geographic Fraud Exposure}_{j,t} = \frac{\text{\# Fraud incidents by in-state public firms}}{\text{\# Total in-state public firms}} \times 100\%. \tag{19}$$

Given the geographic locations of fund letter readers that I measured from the SEC EDGAR logs, I aggregate from individuals' geographic fraud exposures to fund-level fraud exposures:

$$\text{Readership Fraud Exposure}_{i,t} = \frac{1}{\sum_j \text{Investment}_{j,t}} \sum_j \text{Investment}_{j,t} \times \text{Geographic Fraud Exposure}_{j,t}. \tag{20}$$

This aggregation is performed similarly to that in the previous section; i.e. proxying for investment weights using SIPP survey values. The fund level readership fraud exposure captured by Equation (20) proxies the extent to which the readers of the letter issued by fund i during month t are simultaneously exposed to local trust shocks.

Table 8 shows the results of repeating this paper's baseline regression specification, augmented with an additional interaction for local securities fraud exposure by the readership of each fund letter. The specification in column (1) employs an indicator variable for above-median exposure, and the specification in column (2) instead uses the rank-scaled exposure. Baseline effects are also calculated, though not shown. In both specifications, there is a negative and statistically significant interaction effect between the amount of detail communicated about risk in a fund letter, and the contemporaneous exposure of the letter's readers to local securities fraud. Nevertheless, a fund letter readership that is maximally exposed to negative fraud shocks (i.e. setting Fraud Exposure$_{i,t} = 1$, in column 2) can still be soothed by providing more detailed communication about risks, as the net effect of risk detail (+0.07) remains positive. Note that the smaller sample size (and lower statistical power of the test) is due to the shorter sample, as the current version of the Karpoff, Koester, Lee, and Martin (2017) database ends in 2012.

Both across funds and within a fund, fund letter readers with a greater exposure to trust-breaking fraud shocks respond less to the same amount of communication about risk than fund letter readers with a lower exposure to fraud shocks. I conclude from these results that trust in fund managers (or financial intermediaries more broadly) strengthens and complements their ability to foster trust in the stock market using communication.

6 Evaluating Potential Alternative Mechanisms

I now consider a number of potential alternative explanations for my baseline empirical findings. Going through the proposed mechanisms in turn, I conduct additional empirical tests that rule them out as potential explanations. Notably, the evidence is not consistent with mechanisms involving learning, such as belief-based persuasion models.

In addition to this section, Internet Appendix F further examines and rules out additional potential channels, including shrouding, education, behavioral pandering, and the possibility that more communication about risk attracts new investors.

6.1 Can Inflows Be Explained by Rebalancing Into Passive Trackers?

One potential explanation for the positive effect of communication about risk on S&P 500 index fund inflows is that consuming communication about risk shifts investors' perceptions about risk higher in general; similarly, investors' risk aversion may in fact be increased. If there is indeed such a risk perception or risk aversion spillover, investors should withdraw their holdings from funds with a riskier mandate and increase their allocations to lower-risk S&P 500 index funds.

An ideal test of this candidate explanation would exploit individual investor-level data, instead of relying on aggregate flow data; yet this is not an important limitation due to the structure of investors' switching costs. Mutual fund investors are likely to keep their investments within the same fund family due to high across-family switching costs (Massa, 2003), as compared to within-family switching costs. In other words, studying rebalancing flows within fund families focuses precisely on the most frictionless setting, in which such flows are most likely to appear – if the rebalancing hypothesis is indeed true.

I test the hypothesis according to the following logic: a necessary condition for this explanation to hold is that communication about risks induces investors to withdraw their holdings from riskier funds. Therefore, communication about risks in S&P 500 index fund letters should predict an increase in outflows from riskier active domestic equity funds.

The regressions in Table 9 columns 1-2 test this hypothesis: instead of being positive, the predictive coefficient is negative in magnitude and not statistically significantly different to zero. If investors are indeed rebalancing from riskier holdings after reading S&P 500 fund letters, these outflows do not originate from their holdings in active domestic equity funds within the same fund family.

I next conduct a further test of the rebalancing hypothesis: this time, I test whether communication about risks contained in the letters of the higher-risk active domestic equity funds drives flows to lower-risk S&P 500 index funds. Columns 3-4 in Table 9 conduct this analysis, and the results are similar since there is no statistically significant effect.

6.2 Are Investors Learning About Risk From Letters?

Instead of communication decreasing investors' effective risk aversion, an obvious alternative is for investors to simply update their beliefs about risk based on the communication they

Table 9: Testing for communication-driven rebalancing flows out of active funds. This table shows the relationship between outflows from active domestic equity funds and the amount of detail about risk contained in the letters of their sibling S&P index tracker funds (columns 1-2) or in their own letters (columns 3-4). The amount of detail is measured as the log of the number of words devoted to a discussion of risk. Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of active domestic equity funds that belong to fund families that also contain an S&P 500 index fund.

Dependent Variable:	Sibling Outflow$_{i,t\to t+1}$ (%)			
	(1)	(2)	(3)	(4)
S&P 500 Index Fund Risk Detail$_{i,t}$	-0.0878	-0.1215		
	(0.2137)	(0.0956)		
Sibling Active Fund Risk Detail$_{i,t}$			-0.0421	-0.1524
			(0.2221)	(0.1004)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓		✓	
Fund FEs		✓		✓
N	9,123	9,123	9,123	9,123
R^2	0.36	0.79	0.36	0.79

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

receive (which is also about risk). In this section, I investigate the potential presence of this channel and do not find any evidence that it operates in the present paper's setting.

Bayesian Investor Who Learns About Risk To fix ideas in this section, consider a Bayesian investor who learns about the variance parameter σ^2 of the return distribution, or indeed some other distributional parameter that captures the notion of "risk." The Bayesian investor starts with a Gaussian prior on this parameter, $\sigma_p^2 \sim \mathcal{N}(m_p, {}^1/_{\rho_p})$, in which ρ_p denotes the prior precision. The investor receives an independent signal $s = m_s + \varepsilon$ about the variance, with $\varepsilon \sim \mathcal{N}(0, {}^1/_{\rho_s})$. Upon receipt of the signal, the investor updates her beliefs and so now perceives the posterior distribution of the variance to be the following:

$$\sigma^2|s \sim \mathcal{N}\left(\alpha s + (1-\alpha)m_p \ , \ \frac{1}{\rho_s + \rho_p}\right), \tag{21}$$

where $\alpha = \frac{\rho_s}{\rho_s+\rho_p}$. Upon receipt of the signal, therefore, the shift in the investor's expected variance from her prior mean m_p to the posterior is:

$$\Delta(\sigma^2) = [\alpha s + (1-\alpha)m_p] - m_p = (s-m_p)\frac{\rho_s}{\rho_s+\rho_p}. \tag{22}$$

As Equation (22) makes clear, the difference between the signal value and the investor's prior, $(s-m_p)$ is a first-order determinant of the change in her perceived risk. If the investor chooses her portfolio as in Section 2, the difference $(s-m_p)$ should be proportional to her *reduction* of her allocation to the risky asset. This paper takes the risky asset to be an S&P 500 index mutual fund, and therefore $(s-m_p)$ should be negatively associated with the fund flow, holding all else equal.

Extension to Lognormal-Lognormal Updating The assumption that the return variance is Normally distributed is not necessarily ideal, since the variance should have positive support. However, the conclusions from the previous example apply more generally. In fact, the previous argument using Normal-Normal updating carries through exactly if the *log* of the variance is assumed to be Normally distributed, which guarantees a positive variance.

6.2.1 Detail as a Signal of Risk

I first discuss whether investors respond to the amount of detail conveyed for instrumental reasons – that is, whether they respond to it *because* it signals the level of risk σ^2. Table 1 indicates that the amount of detail is positively related to the VIX, which is an objective measure of S&P 500 risk. Perhaps the amount of detail thus corresponds to the signal s in Eqn. (22)? If so, when the investor's prior perceived level of risk (m_p) is controlled for, any increases in the amount of detail should predict a *reduction* by the investor of her allocation to the risky asset; i.e. a decreased inflow or increased outflow (or both) from an S&P 500 index fund.

However, the effect of the amount of detail about risk acts in the opposite direction. As displayed in Table 3, panel regressions of the amount of detail against gross flows show that increases in s actually predict increased inflows (columns 1-4) rather than decreased inflows, and there is no discernible increase in outflows (columns 5-8). This fact holds no matter whether the prior perceived level of risk m_p is controlled for using time effects ($m_p := \phi_t$ in columns 1 & 5), or controlled for using fund fixed effects ($m_p := \pi_i$ in columns 2 & 4), or even controlled for with additive fund fixed effects and time fixed effects ($m_p := \pi_i + \phi_t$ in columns 4 & 8). In each case, the effect of the update to the investor's perceived level of risk (the difference $s-m_p$ in Eqn. (22)) on flows is in the opposite direction to that expected.[18]

[18]Another potential mechanism that conflicts with this fact is salience theory (Bordalo, Gennaioli, and Shleifer, 2013). If increased detail about risk makes this attribute of the product more salient, then investors should be

6.2.2 Level of Risk Conveyed

Investors ignore the signal conveyed by the letter about the level of risk. This fact is also not consistent with a belief-updating mechanism.

With Additive Fund + Month Measures of Prior Risk Perception Table 1 column (3) indicates that the level of risk conveyed by the text in fund letters is a proxy for s in Eqn. (22): the direction is high when the contemporaneous level of the VIX is high. However, Table 4 shows that the sentiment conveyed has no predictive power for fund flows. This conclusion holds no matter how the prior m_p is controlled for: using time effects, fund fixed effects, or an additive combination of the two. Once again, the empirical facts are inconsistent with investors treating fund letters as a source of information to update their beliefs about the level of risk.

With Interactive Fund × Month Measures of Prior Risk Perception The above discussion has assumed that an additive combination of fund fixed effects and time effects is sufficient to control for the prior held by the readership of each fund i's letter at time t. However, it is possible that differences in investor opinion could be richer, with investor priors varying at the fund i and year-month t level. I therefore produce a measure of prior risk perception at that unit of analysis, and repeat my test for any effect of the level of risk conveyed. I find that belief-updating still does not explain the baseline empirical findings.

The measure of investors' prior perceived risk that I use is a fund i- and year-month t-varying asset-weighted average of local Google search activity. Previous work by Da, Engelberg, and Gao (2015) uses Google search activity to measure aggregate sentiment; I show how to measure the prior risk perceptions of mutual fund investors.

I begin by creating a measure of who reads each letter. To meet this challenge, I focus on the readers of fund letters who download them in the form of shareholder reports from the SEC's EDGAR website. The SEC makes EDGAR website usage data up to June 2017 public (Lee, Ma, and Wang, 2015), and a number of studies have used these logs to measure information consumption. Further details on the SEC EDGAR usage logs are provided in Section 3.6 and Internet Appendix B.2.

Employing a methodology similar to Grice and Guecioueur's (2023), I measure which states the readers of fund shareholder reports are located in; these fund shareholder reports contain the fund letters that the present paper focusses on. I then use the geo-locations of

less willing to purchase the product in this case: Recall from Eqn. (1) that investors are typically modeled as disliking risk as an attribute of an investment product. All else being equal, an increase in the salience of risk should therefore predict decreased inflows, rather than increased inflows.

these readers to construct an aggregate index for the prior held about risk by the readership of each fund i at each year-month t:

$$\text{Readership Prior}_{i,t} = z\left(\frac{1}{\sum_j \text{Investment}_{j,t}} \sum_j \text{Investment}_{j,t} \times \text{Geographic Prior}_{j,t}\right). \tag{23}$$

The above index is standardized within-fund to ensure comparability. Investment weights are proxied using SIPP survey values. A Geographic Prior index, which varies at a state-month level, is used to compute the aggregate Readership Prior. The state-to-fund mappings for each fund's investor base are re-computed annually.

To define the Geographic Prior index, I obtain state s- and year-month t-varying Search Volume Indices (SVI) from the Google Trends API, for search terms relating to the general topic of stock market crashes. This topic is defined by Google's search engine itself, and nests individual searches for "stock market crash" and related terms such as "market crash" and "stock crash." I obtain a monthly timeseries of the aggregated SVI across the entire United States, as well as within-month SVIs across the cross-section of states. For each term, I separately merge and rescale individual SVIs to produce a term-specific month-state panel $\text{SVI}_{s,t}$ of comparable values. The Geographic Prior is defined simply as the local Google search volume for the stock market crash topic:

$$\text{Geographic Prior}_{s,t} = \text{SVI}_{s,t}^{\text{stock market crash topic}} \tag{24}$$

The United States-wide aggregate version of the index $\text{SVI}_t^{\text{stock market crash topic}}$ covaries strongly with the VIX: in Internet Appendix B.3, I compare the two using an array of regression specifications including various time trends (to capture growth in search engine usage) and varying functional forms, and show this relationship is robust. In the current section, I make use of state s-month t cross-sectional variation.[19]

It remains to measure the strength of belief updating about investors' perceived level of risk that is induced by the receipt of a fund letter. In close analogy to the term $s - m_p$ in Eqn. (22), the update is defined as the difference between the letter's signal about risk and investors' priors,

$$\text{Risk Level Update}_{i,t} = \text{Letter Signal}_{i,t} - \text{Readership Prior}_{i,t}, \tag{25}$$

with the first term is defined by either Eqn. (8) or (9) (depending on the measure employed), and the second term is defined by Eqn. (23).

[19]This part of the analysis would be unnecessary if it were possible to compute an index such as the VIX that is localized for specific geographic subgroups of investors — however no such index exists. I therefore compute my own index based on Google search traffic. The validation checks in Internet Appendix B.3 confirm that the aggregate version of the index and the VIX both capture investors' perceptions of risk about the stock market. At the time of writing (August 2023), the publicly-accessible location to download the "Stock Market Crash" topic SVI was at `https://trends.google.com/trends/explore?q=/m/0gz_4&date=all&geo=US`.

Table 10: Joint relation of the amount of detail & the update to the level of risk conveyed with S&P 500 index fund flows. This table shows the relationship between net flows and both communication detail and the update to perceived risk due to the communication. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. The letter's signal about risk is measured in columns (2) and (4) based on the net sentiment of the text discussing risk, as defined by Eqn. (8); in columns (3) and (5) it is measured as the net score of contextual risk words, as defined by Eqn. (9). The prior perceived level of risk is measured based on local Google searches, as defined by Eqn. (23) The update term is the difference between the two, as defined by Eqn. (25). Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)				
	(1)	(2)	(3)	(4)	(5)
Risk Detail$_{i,t}$	0.5207***	0.4988***	0.4824**	0.5019***	0.4812**
	(0.1854)	(0.1783)	(0.1958)	(0.1766)	(0.1957)
Risk Level Update$_{i,t}$		0.1574	0.1339		
		(0.1378)	(0.1164)		
Letter Signal$_{i,t}$				0.1745	0.1621
				(0.1398)	(0.1246)
Readership Prior$_{i,t}$				0.0621	0.0967
				(0.3251)	(0.3169)
Risk Level measure		Sentiment	High-Low Words	Sentiment	High-Low Words
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓
N	901	882	890	882	890
R^2	0.45	0.43	0.45	0.43	0.45

Clustered (Fund & Year-month) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 10 shows a set of panel regression estimates that build upon the previous regressions of Table 4, with a further modification: as well including the level of the risk conveyed by the communication, the specifications also incorporate fund i-year-month-t-varying priors about risk held by the letter's readership. This prior is incorporated implicitly into the Risk Update$_{i,t}$ term in columns 2-3, or explicitly as a standalone covariate in the bottom row of columns 4-5.

The coefficients on Risk Update$_{i,t}$ in columns 2-3 of Table 10 are positive in sign and not significantly different to zero. If fund letter readers update their prior about risk using the signal conveyed in the letter, one would expect *negative* and significant estimates for these coefficients — that is, a revision upwards in the level of perceived risk should predict a smaller net flow. In columns 4-5, the Risk Update$_{i,t}$ term is split apart into its constituent signal and prior terms; the estimated coefficients on both are not significant. By contrast, the main coefficient on the amount of detail in the first row always remains positive and significant.

Once again, these results are not consistent with a classic belief-updating mechanism in which the investor is a Bayesian who learns from the content of the fund letter about about some risk parameter (such as the variance parameter of a conditional distribution of market returns).

6.2.3 Does Communication Invoke Belief-Updating at All?

Having ruled out a number of belief-related hypotheses, I now consider more broadly whether communication about risk acts to change investor beliefs about risk, at all. A wide class of what DellaVigna and Gentzkow (2010) term belief-based persuasion models involve the sender inducing a shift in the receiver's posterior belief through communication. DellaVigna and Gentzkow (2010) suggest a single test for this class of models: "persuasion will tend to be more effective when receivers are less certain about the truth. In Bayesian models, the weaker receivers' priors are, the more their beliefs are affected by a given piece of new information (holding everything else constant)."[20] Classic strategic models of communication and persuasion include a core Bayesian updating element, including those of Crawford and Sobel (1982) and Kamenica and Gentzkow (2011).

Effect of the Strength of the Investor's Prior As a concise illustration of how the strength of the prior affects the strength of belief updating, consider again the Bayesian investor who learns about the variance parameter σ^2 of the return distribution, as described at the beginning of Section 6.2. Define the strength of her update from the prior mean m_p to the posterior

[20] A key advantage of this test is that it also applies to belief-based models in which agents deviate from rational Bayesian updating, as long as their behavior varies less as they are (*ex ante*) more certain about the state of the world before receiving the communication. Furthermore, by testing another relationship between flows and belief-based parameters, the current test is complementary to the previous ones, including that in Section 6.2.2.

as the magnitude of the signed update (Eqn. 22):

$$\left|\Delta(\sigma^2)\right| \;=\; \left|s - m_p\right| \frac{\rho_s}{\rho_s + \rho_p}. \tag{26}$$

It is then clear that the strength of the update is decreasing in the strength of the prior:

$$\frac{\partial}{\partial \rho_p}\left|\Delta(\sigma^2)\right| \;=\; -\left|s - m_p\right| \frac{\rho_s}{(\rho_s + \rho_p)^2} \;<\; 0. \tag{27}$$

As before, this conclusion is not overly dependent on the assumption of Normal-Normal updating; for instance, Lognormal-Lognormal updating about the variance leads to a similar implication.

Table 11: Communication-driven flows and *ex ante* strength of the prior about risk. The amount of detail contained in the letter written by fund i at time t about risks is measured using the log of the number of words devoted to that topic. The indicator variable $\mathbb{1}\{\text{Low Prior Confidence}\}_t$ is 1 when the VVIX is above the median, and 0 otherwise; a baseline effect is estimated in addition to the interaction shown in the table. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only.

Dependent Variables:	Net Flow$_{i,t\to t+1}$ (%)		Inflow$_{i,t\to t+1}$ (%)		Outflow$_{i,t\to t+1}$ (%)	
	(1)	(2)	(3)	(4)	(5)	(6)
Risk Detail$_{i,t}$	0.5248***	0.5195***	0.5569***	0.5160**	0.0350	0.0958
	(0.1633)	(0.1917)	(0.1886)	(0.2005)	(0.0667)	(0.0804)
Risk Detail$_{i,t}$	-0.0778	-0.0346	-0.1485	-0.1028	-0.0546	-0.1038
× $\mathbb{1}\{\text{Low Prior Strength}\}_t$	(0.1141)	(0.1328)	(0.1271)	(0.1536)	(0.0873)	(0.0948)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓	✓
Market Controls$_t$	✓		✓		✓	
Year-month FEs		✓		✓		✓
N	1,155	1,155	1,155	1,155	1,155	1,155
R^2	0.31	0.44	0.74	0.79	0.83	0.85

*Clustered (Fund & Year-month) standard-errors in parentheses: ***: 0.01, **: 0.05, *: 0.1*

Empirical Test I now test whether more detailed communication about risk interacts with the strength of the representative investor's prior about the level of risk. I measure the prior by the VVIX, the so-called "VIX of the VIX." This publicly-available index measures the expected volatility of the 30-day forward price of the VIX. The VVIX is therefore informative of "the degree of confidence the market has in its forecast of future values of the VIX," according to its publisher, the CBOE.[21]

I interpret a higher level of the VVIX as corresponding to a weaker, more diffuse prior about the level of risk held by an investor before receiving communication via a fund letter. A high value of the VVIX thus indicates the investor is less confident about her assessment of risk. Under a belief-based persuasion model, two identical signals (with the same level of precision) should invoke different responses according to whether the investor has strong priors or weak priors: if she holds a strong prior, the investor should respond only weakly and take little action; if her prior is weak, the investor should update her beliefs more strongly, thus manifesting as higher inflows.

Table 11 presents the results of just such an analysis. To be precise, I repeat this paper's main flow-versus-detail panel regressions, and also include an interaction term for low prior strength/precision in the level of risk. This corresponds to market states with above-median levels of the VVIX. The coefficient on this interaction between having low prior strength about risk and the amount of detail about risk conveyed in fund letters is insignificant. Moreover, the direction is opposite to that expected: instead of being positive, the estimate is in fact negative, suggesting the representative investor updates *less* strongly when she is less confident *ex ante*.

These results suggest the underlying persuasive mechanism is unlikely to be belief-based.

6.3 Are Investors Learning About the Return From Fund Letters?

I now examine two possible explanations involving investors learning about the return from fund letters.

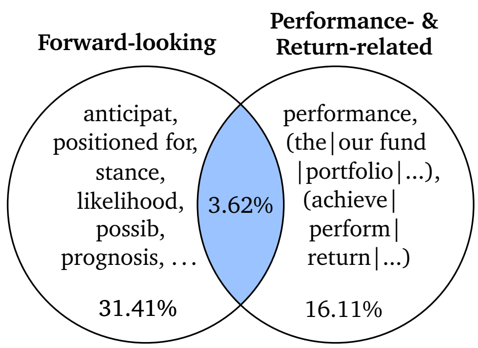
6.3.1 Are Investors Learning About Returns, Increasing Their Posterior Precision?

Instead of updating their beliefs about risk directly, could the results be explained by investors learning about future *returns* from the content of the fund letters, and *then* forming beliefs about the second moment of the return/payoff? Put differently: is an uncertainty-reduction channel at play?

Example with Normal-Normal Bayesian Updating About the Return To illustrate such a mechanism, consider an investor who learns about the return r (or payoff, or fundamental value) of an asset in general — instead of learning about the variance or some risk-related parameter, specifically. The investor starts with a Gaussian prior on the return, $r_p \sim \mathcal{N}(m_p, 1/\rho_p)$,

[21] Source: https://www.cboe.com/us/indices/dashboard/vvix/.

Figure 6: Screening for forward-looking performance- and return-related statements. Forward-looking statements within fund letters relating to performance & returns are defined as sentences that are matched by both forward-looking and performance or return-related regular expressions. The percentage figures denote the fraction of all sentences in the fund letter corpus that are matched by each type.



in which ρ_p denotes the prior precision. The investor receives an independent signal $s = m_s + \varepsilon$ about the overall return, with $\varepsilon \sim \mathcal{N}(0, 1/\rho_s)$. Upon receipt of the signal, the investor updates her beliefs and so now perceives the posterior distribution of the return to be the following:

$$r|s \sim \mathcal{N}\left(\alpha s + (1-\alpha)m_p\ ,\ \frac{1}{\rho_s + \rho_p}\right), \tag{28}$$

where $\alpha = \frac{\rho_s}{\rho_s + \rho_p}$. Importantly, the precision has increased from ρ_p to $(\rho_s + \rho_p)$; equivalently the posterior variance $\mathbb{V}\text{ar}[r|s] = 1/(\rho_s + \rho_p)$ has decreased in comparison to the prior variance $1/\rho_p$. Uncertainty is thus reduced, irrespective of the actual value (realization) of the signal s that the investor receives. Consider now that the investor chooses between a risky asset (with return r) and a riskless asset, and the quantity $\mathbb{V}\text{ar}[r|s]$ is defined to be her perceived risk. Then the receipt of *any* signal about the return will mechanically reduce the investors' perceived uncertainty, and hence lead her to increase her allocation to the risky asset.

I now conduct an explicit test for the presence of such an uncertainty-reduction channel, precisely where it is most likely to be apparent: I focus on statements in fund letters that are forward-looking and devoted to performance or returns specifically (rather than risk). I do so by modifying the set of regular expressions used, as illustrated by Figure 6. While statements devoted to performance and returns occur frequently in the fund letter corpus, only 23% of these are also forward-looking (compared to 61% of risk-related statements).

I next examine how fund flows respond to the amount of detail in fund letters concerning future performance or returns, explicitly. The logic behind this test is the following: if investors are learning about the return from statements about risk, they should do the same from statements that are explicitly about the return itself — if anything, the effect should be

Table 12: Placebo test for the amount of detail communicated about future performance & returns on the flows of S&P 500 index funds. The main covariate measures the amount of detail contained in the letter written by fund i at time t about forward-looking performance & returns, using the log of the number of words devoted to that topic. Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only. Each flow variable is the next month's flow, in percentage points of TNA.

Dependent Variables:	Net Flow	Inflow	Outflow
	(1)	(2)	(3)
Detail about Performance & Returns$_{i,t}$	-0.0725	0.0447	0.0991
	(0.0983)	(0.1272)	(0.1124)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	882	882	882
R^2	0.50	0.81	0.87

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

even stronger when using the measure. Conversely, if a different channel specifically related to risk-related information is at play, the analysis will be a placebo with no effect detected.

Table 12 presents the results of repeating the paper's main flow-detail regressions, but where the amount of detail in fund letters is measured from forward-looking performance- and return-related statements. The estimated coefficients in Table 12 exhibit no significant relation between the amount of detail conveyed and fund flows. The absence of such an effect suggests that detailed textual communication about performance or returns (explicitly) does not act to reduce investors' perceived risks via a mechanical reduction in perceived uncertainty. It is thus even less likely that communication specifically about risk will have such an effect.

6.3.2 Do Investors Expect a Higher Return With More Communication About Risk?

Perhaps investors are inferring a higher expected return from more communication about risks? If investors either completely ignore information about risk, or increase their expected return more sharply than the increase in their perceived risk, this may lead the investors to increase their allocations to the risky asset, on net. Reviewing the evidence, I find this channel is unlikely no matter whether investors hold subjective expectations of returns or attempt to rationally form expected returns according to the communication that they receive.

Subjective Expected Returns Prior work surveys retail investors and shows they do not perceive a positive subjective relationship between risk and return. In fact, the opposite holds: surveyed investors perceive that a higher level of risk jointly occurs with a *lower* subjective expected return (Amromin and Sharpe, 2014; Giglio, Maggiori, Stroebel, and Utkus, 2021; Jo, Lin, and You, 2022). These subjective expected returns then impact flows into the stock market (Greenwood and Shleifer, 2014). This subjective relationship holds over the current paper's sample period.[22] Since unsophisticated investors' subjective expected returns do not increase simultaneously with their perceived risk, changes to their expected returns should not attenuate any outflows that are driven by higher risk perceptions, but exacerbate them.

Rational Expected Returns On the other hand, investors may act rationally: if communication about risk predicts returns, then investors may rationally use this information when forming their beliefs about returns. The timeseries regressions in Table 13 columns 1-2 show that investors during month t can successfully predict the current level of market-implied *risk* over the next month ($t+1$) for the S&P 500 market portfolio. However, the coefficient estimates in columns 3-4 show that this information does not predict the actual level of the *return* over the same period: as well as not being significantly different to zero, the coefficients' signs have contradictory signs. When extending the time horizon to the next quarter (columns 5-6) or the full 5-month period before the dissemination of the next letter (columns 7-8), the R^2 and coefficient on the number of words grow, but remains mostly insignificant. The coefficient on sentiment is never significantly different to zero, and continues to have a contradictory sign even over those longer horizons.

Furthermore, an investor who attempts to predict returns in this manner would still perceive a higher growth in risk than in expected return. The estimate on column (1) is 4.48 times as large as the estimate in column (3), meaning that even if an investor can predict an increased return based on more detailed communication about risk, she would simultaneously infer a far greater increase in risk, and hence *reduce* her risk-taking overall (see Eqn. (2)).

Altogether, there is only weak evidence for any kind of return predictability based on information about risks provided in the fund letters of S&P 500 mutual funds – and not enough to explain observed investor behavior.

[22] I compare aggregate subjective expected returns to the level of the VIX (which covaries positively with the amount of detail and negatively with its sentiment, as shown in Table 1). Consistent with the literature on subjective expected returns, I find a contemporaneous correlation between the AAII survey index of subjective expected returns (studied by Greenwood and Shleifer (2014)) and the VIX of -0.43 over my sample period. A one-tailed hypothesis test rejects the null of a non-negative correlation with a t-stat of 6.5 and an approximately zero p-value.

Table 13: Forming beliefs about the S&P 500 from timeseries indices of communicated detail and sentiment. The dependent variables (in percentage points) are the square of the level of the VIX at the end of month t (which connotes the market-implied conditional variance of the next-month S&P 500 return), and the return of the S&P 500 index from the end of month t to the end of month $t+h$. The independent variables are the equally-weighted across funds and standardized within fund number of words devoted to discussing risk, and a measure of the level of risk implied by the sentiment of that communication; i.e. columns (1) and (3) from Table 1. The sample of funds used to construct the independent variables consists of S&P 500 index funds only.

			S&P 500 Return					
Dependent Variables:	Implied $\mathbb{V}\text{ar}_t(R_{t+1})$		Period $t \to t+1$		Period $t \to t+3$		Period $t \to t+5$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Detail$_t$	2.348**		0.5246		1.840		3.689*	
	(0.9910)		(0.7618)		(1.282)		(1.973)	
Risk Level$_t$		2.353***		-0.4570		-0.1984		-1.335
		(0.7588)		(0.7359)		(1.269)		(2.638)
N	187	187	187	187	187	187	187	187
R^2	0.02	0.05	0.002	0.003	0.007	0.0001	0.01	0.004

Newey-West standard errors (in parentheses) use automatically selected lags.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

6.4 Are Investors Learning About Manager Skill?

Studying homogeneous S&P 500 index-tracking funds is useful for identification, as it implicitly controls for unobservable differences across funds that may confound the empirical analyses (Hortaçsu and Syverson, 2004). In particular, it is unlikely that investors are learning about unobserved active management skill, in the style of Berk and Green (2004), since no active management is involved. To the extent that manager skill is involved, though, it is in the minimization of a fund's tracking error with respect to the S&P 500 index. I therefore investigate whether investors are learning about future tracking errors from the amount of detail provided about risk.

In the literature and in practice, the tracking error is defined as the standard deviation of the difference between monthly fund returns and the S&P 500 index returns (e.g. Chevalier and Ellison, 1997; Grinold and Kahn, 1999; Cremers and Petajisto, 2009). Likewise, I define the future tracking error from the end of the current month t over some future horizon of h months as follows:

$$\text{Tracking Error}_{i,t\to t+h} = \text{StdDev}\left\{ r^{Fund}_{i,t+k} - r^{S\&P\,500}_{t+k} \;\middle|\; 1 \le k \le h \right\}. \tag{29}$$

Table 14: Placebo test for tracking error prediction. The main independent variable is the (log) number of words devoted to discussing risks. In addition, the total (log) number of words in a fund letter is included as a control. The dependent variable is the fund's tracking error, as defined in Eqn. (29) for different future horizons h. The sample consists of S&P 500 index funds only.

Dependent Variables:	Tracking Error$_{i,t\to t+h}$			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.0001	-0.0011	0.0006	0.0005
	(0.0005)	(0.0008)	(0.0008)	(0.0010)
Total Detail$_{i,t}$	0.0005	0.0016*	-0.0009	-0.0028**
	(0.0008)	(0.0010)	(0.0014)	(0.0014)
Horizon h (months)	6	12	24	36
Fund FEs	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
N	1,159	1,153	964	899
R^2	0.83	0.75	0.59	0.58

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The assumption underlying the test is as follows: if investors respond to the amount of detail because it signals a low tracking error, this detail should have predictive power for the realized tracking error. The placebo test in Table 14 confirms that this is *not* actually the case: the amount of detail conveyed about risk is not a signal for how well an S&P 500 index fund will track the S&P 500 index return going forward.

7 Effect of Aggregate Flows on Prices

This section quantifies the aggregate asset pricing implications of this paper's fund-level results. Given my focus on funds that track the market portfolio, it is simple to calculate the effect of such flows on the market value. While reliant on macro multiplier estimates from other studies, this analysis nevertheless is a useful confirmation that the effects identified in this paper are economically meaningful and have important asset pricing implications.

I begin by aggregating the fund-level flow effects of Section 4.1; specifically, the estimates in Table 5 columns (4) and (8). Figure 7 charts these cumulative net flows into the S&P 500 (via index funds) as a percentage of index fund AuM. The flows are distributed evenly over this paper's sample period of January 2006 to August 2021.

Aggregate Effect Based on Gabaix and Koijen's (2021) Multiplier Estimate Gabaix and Koijen (2021) show that flows into the stock market have a large price impact: their estimate of the macro multiplier is that a \$1 inflow translates into an increase in value of \$5. For the purposes of this paper, a 1% communication-driven flow would therefore produce a 5% return in the S&P 500 index. I use this measure of the stock market's (in)elasticity to translate communication-driven market flows into price impacts, according to a simple back-of-the-envelope calculation: I recalculate the communication-driven flows as a percentage of the total market capitalization of the S&P 500, and then multiply by 5 to estimate the causal effect of communication on market returns. To put the numbers in context, I adjust the historical levels of the S&P 500 by subtracting communication-driven returns induced from January 2006 onwards. Figure 8 plots the counterfactual levels of the S&P 500 that would have obtained in the absence of communication about risks to the market portfolio.

The level of the S&P 500 index stood at 4,523 at the end of this paper's sample in August 2021. In the absence of communication-driven net flows, the counterfactual level of the index would have stood at only 4,015 (or a slightly lower 3,879 using 2SLS estimates). The counterfactual gap is therefore around 11% of the capitalization of the aggregate stock market. This translates to an average annual S&P 500 index return of approximately 67 basis points. As a benchmark, the S&P 500 index returned 8% annually on average over this period, excluding dividends. Although sensitive to the macro flow multiplier estimate, this aggregation exercise

Figure 7: Aggregate communication-driven flows into S&P 500 index funds. Cumulative total of the net flows since January 2006 that were driven by increased communication about risk, as a percentage of the assets under management of the S&P 500 index funds disseminating fund letters. The effect is aggregated from fund-level effects estimated using the specifications of Table 5 columns (2) and (6), and plotted as the orange and red series, respectively.

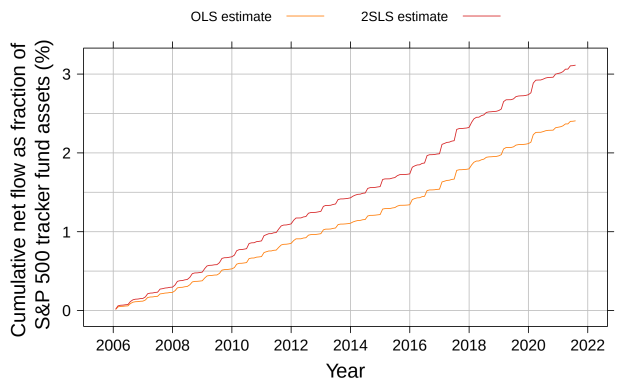
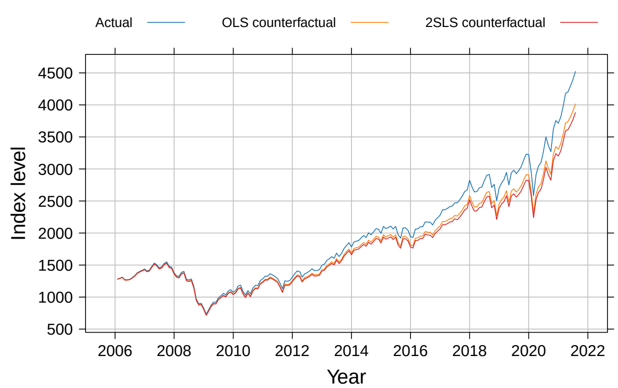


Figure 8: Illustration of price effect of communication-driven flows on the S&P 500 index. The blue series plots the actual historical level of the S&P 500 over this paper's sample period of January 2006 to August 2021. The orange and red series plot counterfactual historical levels of the index in the absence of risk-related communication disseminated the letters of S&P 500 index funds: First, the fund-level effect on net flows are aggregated from the fund-level specifications estimated in specifications of Table 5 columns (4) and (8). Next, the aggregate net flows are multiplied by 5 to compute the causal effect of market flows on the market index, based on the measured price (in)elasticity estimated by Gabaix and Koijen (2021). Finally, S&P 500 returns are adjusted (downward) to exclude these (positive) monthly return effects to compute counterfactual levels of the index.



puts the magnitude of the anxiety-alleviation mechanism into context, and confirms that it is economically meaningful. Communication-driven flows thus lead to substantial changes in asset prices.

Aggregate Effect Based on Hartzmark and Solomon's (2023) Multiplier Estimate As Gabaix and Koijen (2021) discuss, the market multiplier may be lower if the flows are considered to be anticipated. This would decrease the magnitude of the above estimates of the marketwide impact of anxiety-alleviating communication. It is simple to assess the robustness of the previous result to a different flow multiplier estimate.

Using anticipated divided payments, Hartzmark and Solomon (2023) find the market remains inelastic (albeit less so), with an estimated flow multiplier of 1.9. Combining this lower multiplier with my own estimates, the counterfactual level of the S&P 500 would have stood at 4,323 (or a slightly lower 4,267 using 2SLS estimates) at the end of my sample period in the absence of anxiety-alleviating communication. The counterfactual gap thus shrinks to around 4% of the stock market's capitalization, which translates to an average annual S&P 500 index return of 27 basis points. The effect thus remains economically meaningful.

8 Conclusion

In this paper, I find that money doctors who communicate in more detail about the riskiness of the fund's portfolio induce their investors to *increase* their share allocated to the risky asset. I contribute to the literature by illuminating a behavioral channel consistent with the notion of fund managers as money doctors (GSV). This has implications for the origins of trust in other economic settings. My findings also highlight that effective communication about risk can alleviate investors' anxiety about taking risk. Finally, I show that communication plays a persuasive role in the asset management industry (DellaVigna and Gentzkow, 2010).

Recent work by Gorodnichenko, Pham, and Talavera (2023) and Laudenbach and Siegel (forthcoming) highlights that vocal cues can impact investor and borrow behavior. Likewise, further examination of the content of fund letters may shed additional light on the role of affect in driving investor responses.

References

Abis, S., Buffa, A. M., Javadekar, A., & Lines, A. (2022). Learning from prospectuses. *Available at SSRN, 3865753*.

Akey, P., Robertson, A., & Simutin, M. (2021). Closet active management of passive funds. *Rotman School of Management Working Paper, 3874582*.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151–184.

Amromin, G., & Sharpe, S. A. (2014). From the horse's mouth: Economic conditions and investor expectations of risk and return. *Management Science*, *60*(4), 845–866.

Andries, M., & Haddad, V. (2020). Information aversion. *Journal of Political Economy*, *128*(5), 1901–1939.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, 343–362.

Auh, J. K., & Bai, J. (2020). Cross-asset information synergy in mutual fund families. *National Bureau of Economic Research Working Paper*, *26626*.

Bailey, W., Kumar, A., & Ng, D. (2011). Behavioral biases of mutual fund investors. *Journal of Financial Economics*, *102*(1), 1–27.

Becker, G. S., & Murphy, K. M. (1993). A simple theory of advertising as a good or bad. *Quarterly Journal of Economics*, *108*(4), 941–964.

Ben-David, I., Li, J., Rossi, A., & Song, Y. (2022). What do mutual fund investors really care about? *Review of Financial Studies*, *35*(4), 1723–1774.

Berk, J. B., & Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, *112*(6), 1269–1295.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, *121*(5), 803–843.

Brenner, M., & Izhakian, Y. (2018). Asset pricing and ambiguity: Empirical evidence. *Journal of Financial Economics*, *130*(3), 503–531.

Caetano, C., Caetano, G., & Nielsen, E. R. (forthcoming). Correcting for endogeneity in models with bunching. *Journal of Business & Economic Statistics*. https://doi.org/10.1080/07350015.2023.2252471

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, *115*(5), 707–747.

Calvet, L. E., Célérier, C., Sodini, P., & Vallée, B. (2023). Can security design foster household risk-taking? *Journal of Finance*, *78*(4), 1917–1966.

Cao, S., Yang, B., & Zhang, A. L. (2021). Deep learning mutual fund disclosure: Risk sentiment, risk taking, and performance. *Available at SSRN*, *4060307*.

Cassar, G. J., Gerakos, J. J., Green, J. R., Hand, J. R., & Neal, M. (2018). Hedge fund voluntary disclosure. *The Accounting Review*, *93*(2), 117–135.

Célérier, C., & Vallée, B. (2017). Catering to investors through security design: Headline rate and complexity. *Quarterly Journal of Economics*, *132*(3), 1469–1508.

Chen, H., Cohen, L., Gurun, U., Lou, D., & Malloy, C. (2020). IQ from IP: Simplifying search in portfolio choice. *Journal of Financial Economics*, *138*(1), 118–137.

Chevalier, J., & Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, *105*(6), 1167–1200.

Chevalier, J., & Ellison, G. (1999). Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance. *Journal of Finance*, *54*(3), 875–899.

Cici, G., Jaspersen, S., & Kempf, A. (2017). Speed of information diffusion within fund families. *Review of Asset Pricing Studies*, *7*(1), 144–170.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, 1431–1451.

Cremers, K. M., & Petajisto, A. (2009). How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies*, *22*(9), 3329–3365.

Cronqvist, H. (2006). Advertising and portfolio choice. *Available at SSRN*, *920693*.

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. *Review of Financial Studies*, *28*(1), 1–32.

Del Guercio, D., & Tkac, P. A. (2008). Star power: The effect of Morningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, *43*(4), 907–936.

DellaVigna, S., & Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, *2*(1), 643–669.

DeMarzo, P. M., Vayanos, D., & Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, *118*(3), 909–968.

Du, Q., Jiao, Y., Ye, P., & Fan, W. (2019). When mutual fund managers write confidently. *Available at SSRN*, *3513288*.

Eliaz, K., & Schotter, A. (2010). Paying for confidence: An experimental study of the demand for non-instrumental information. *Games and Economic Behavior*, *70*(2), 304–324.

Ely, J., Frankel, A., & Kamenica, E. (2015). Suspense and surprise. *Journal of Political Economy*, *123*(1), 215–260.

Engelberg, J., & Parsons, C. A. (2016). Worrying about the stock market: Evidence from hospital admissions. *Journal of Finance*, *71*(3), 1227–1250.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. Free Press.

Gabaix, X., & Koijen, R. S. (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis. *National Bureau of Economic Research Working Paper*, *28967*.

Galbraith, J. K. (1958). *The Affluent Society*. Houghton Mifflin Harcourt.

Gallaher, S., Kaniel, R., & Starks, L. T. (2015). Advertising and mutual funds: From families to individual funds. *CEPR Discussion Paper*, (10329).

Gennaioli, N., Shleifer, A., & Vishny, R. (2015). Money doctors. *Journal of Finance*, *70*(1), 91–114.

Giannetti, M., & Wang, T. Y. (2016). Corporate scandals and household stock market participation. *Journal of Finance*, *71*(6), 2591–2636.

Giglio, S., Maggiori, M., Stroebel, J., & Utkus, S. (2021). Five facts about beliefs and portfolios. *American Economic Review*, *111*(5), 1481–1522.

Giorgetta, C., Grecucci, A., Zuanon, S., Perini, L., Balestrieri, M., Bonini, N., Sanfey, A. G., & Brambilla, P. (2012). Reduced risk-taking behavior as a trait feature of anxiety. *Emotion*, *12*(6), 1373.

Gomes, F., Haliassos, M., & Ramadorai, T. (2021). Household finance. *Journal of Economic Literature*, *59*(3), 919–1000.

Gorodnichenko, Y., Pham, T., & Talavera, O. (2023). The voice of monetary policy. *American Economic Review*, *113*(2), 548–584.

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

Grice, R., & Gueciouer, A. (2023). Mutual fund market structure and company fee competition: Theory and evidence. *Available at SSRN*, *4449026*.

Grinold, R. C., & Kahn, R. N. (1999). *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk* (2nd ed.). McGraw Hill.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *Journal of Finance*, *63*(6), 2557–2600.

Guiso, L., Sapienza, P., & Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, *128*(3), 403–421.

Gurun, U. G., Stoffman, N., & Yonker, S. E. (2018). Trust busting: The effect of fraud on investor behavior. *Review of Financial Studies*, *31*(4), 1341–1376.

Haliassos, M., & Bertaut, C. C. (1995). Why do so few hold stocks? *The Economic Journal*, *105*(432), 1110–1129.

Hall, J. A., Roter, D. L., & Katz, N. R. (1988). Meta-analysis of correlates of provider behavior in medical encounters. *Medical Care*, *26*(7), 657–675.

Hartzmark, S. M., & Solomon, D. H. (2023). Marketwide predictable price pressure. *Available at SSRN*, *3853096*.

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *Journal of Finance*, *74*(6), 2789–2837.

Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *Science*, *344*(6186), 862–867.

Hayward, J. (1975). Information – a prescription against pain. *The Study of Nursing Care Project Reports*, (5). Royal College of Nursing, London.

Hazell, J., Herreno, J., Nakamura, E., & Steinsson, J. (2022). The slope of the Phillips Curve: Evidence from US states. *Quarterly Journal of Economics*, *137*(3), 1299–1344.

Heath, D., Macciocchi, D., Michaely, R., & Ringgenberg, M. C. (2022). Do index funds monitor? *Review of Financial Studies*, *35*(1), 91–131.

Hillert, A., Niessen-Ruenzi, A., & Ruenzi, S. (2021). Mutual fund shareholder letters: Flows, performance, and managerial behavior. *Available at SSRN*, *2524610*.

Hortaçsu, A., & Syverson, C. (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics*, *119*(2), 403–456.

Hu, A., & Ma, S. (2021). Persuading investors: A video-based study. *National Bureau of Economic Research Working Paper*, *29048*.

Huang, J., Wei, K. D., & Yan, H. (2007). Participation costs and the sensitivity of fund flows to past performance. *Journal of Finance*, *62*(3), 1273–1311.

Huang, X., & Xu, N. R. (2023). Risk aversion spillover: Evidence from financial markets and controlled experiments. *Available at SSRN*, *3852463*.

Iliev, P., Kalodimos, J., & Lowry, M. (2021). Investors' attention to corporate governance. *Review of Financial Studies*, *34*(12), 5581–5628.

Investment Company Institute. (2018). Mutual fund investors' views on shareholder reports: Reactions to a summary shareholder report prototype.

Jain, P. C., & Wu, J. S. (2000). Truth in mutual fund advertising: Evidence on future performance and fund flows. *Journal of Finance*, *55*(2), 937–958.

Jo, C., Lin, C., & You, Y. (2022). Negative subjective risk-return trade-offs. *Available at SSRN*, *4096443*.

Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*(6), 2590–2615.

Karpoff, J. M., Koester, A., Lee, D. S., & Martin, G. S. (2017). Proxies and databases in financial misconduct research. *The Accounting Review*, *92*(6), 129–163.

Kaur, S., Mullainathan, S., Oh, S., & Schilbach, F. (2021). Do financial concerns make workers less productive? *National Bureau of Economic Research Working Paper*, *28338*.

Koijen, R. S., & Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, *127*(4), 1475–1515.

Kostovetsky, L. (2016). Whom do you trust?: Investor-advisor relationships and mutual fund flows. *Review of Financial Studies*, *29*(4), 898–936.

Kuchler, T., & Stroebel, J. (2021). Social finance. *Annual Review of Financial Economics*, *13*, 37–55.

Kuhnen, C. M., & Knutson, B. (2011). The influence of affect on beliefs, preferences, and financial decisions. *Journal of Financial and Quantitative Analysis*, *46*(3), 605–626.

Laudenbach, C., & Siegel, S. (forthcoming). Personal communication in an automated world: Evidence from loan repayments. *Journal of Finance*.

Lee, C. M., Ma, P., & Wang, C. C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, *116*(2), 410–431.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), 146.

Lin, T.-C., & Pursiainen, V. (2023). The disutility of stock market losses: Evidence from domestic violence. *Review of Financial Studies*, *36*(4), 1703–1736.

Lo, A. W., & Repin, D. V. (2002). The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience*, *14*(3), 323–339.

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

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

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, *52*(1), 5–44.

Maner, J. K., & Schmidt, N. B. (2006). The role of risk avoidance in anxiety. *Behavior Therapy*, *37*(2), 181–189.

Mankiw, N. G., & Zeldes, S. P. (1991). The consumption of stockholders and nonstockholders. *Journal of Financial Economics*, *29*(1), 97–112.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Association for Computational Linguistics (ACL) System Demonstrations*, 55–60.

Massa, M. (2003). How do family strategies affect fund performance? When performance-maximization is not the only game in town. *Journal of Financial Economics*, *67*(2), 249–304.

McCloskey, D., & Klamer, A. (1995). One quarter of GDP is persuasion. *American Economic Review (Papers and Proceedings)*, *85*(2), 191–195.

Melcangi, D., & Sterk, V. (2020). Stock market participation, inequality, and monetary policy. *FRB of New York Staff Report*, (932).

Milgrom, P., & Roberts, J. (1986). Relying on the information of interested parties. *RAND Journal of Economics*, 18–32.

Mitali, S. F. (2022). How do mutual fund investors react to text-based uncertainty? *Available at SSRN*, *4259075*.

Mullainathan, S., Schwartzstein, J., & Shleifer, A. (2008). Coarse thinking and persuasion. *Quarterly Journal of Economics*, *123*(2), 577–619.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.

Paugam, L., Stolowy, H., & Gendron, Y. (2021). Deploying narrative economics to understand financial market dynamics: An analysis of activist short sellers' rhetoric. *Contemporary Accounting Research*, *38*(3), 1809–1848.

Raghunathan, R., & Pham, M. T. (1999). All negative moods are not equal: Motivational influences of anxiety and sadness on decision making. *Organizational Behavior and Human Decision Processes*, *79*(1), 56–77.

Roussanov, N., Ruan, H., & Wei, Y. (2021). Marketing mutual funds. *Review of Financial Studies*, *34*(6), 3045–3094.

Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives*, *32*(2), 135–154.

Schwartzstein, J., & Sunderam, A. (2021). Using models to persuade. *American Economic Review*, *111*(1), 276–323.

Sergeyev, D., Lian, C., & Gorodnichenko, Y. (2023). The economics of financial stress. *National Bureau of Economic Research Working Paper*, *31285*.

Sheng, J., Xu, N., & Zheng, L. (2022). Do mutual funds walk the talk? A textual analysis of risk disclosure by mutual funds. *Available on SSRN*, *3757077*.

Sicherman, N., Loewenstein, G., Seppi, D. J., & Utkus, S. P. (2016). Financial attention. *Review of Financial Studies*, *29*(4), 863–897.

Sirri, E. R., & Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance*, *53*(5), 1589–1622.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, *69*(3), 213–225.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Wang, A. Y., & Young, M. (2020). Terrorist attacks and investor risk preference: Evidence from mutual fund flows. *Journal of Financial Economics*, *137*(2), 491–514.

Internet Appendix

A Example Risk Discussions from Fund Letters

Figure IA.1 shows an example fund letter addressed to the investors in an S&P 500 index fund. The remainder of this appendix shows example statements about risk extracted from a wide sample of fund letters using this paper's methodology – first from the sample of S&P 500 index fund letters that this paper focusses on, and then from the overall corpus of fund letters.

Figure IA.1: Example Fund Letter. This letter opens the section on the Schwab S&P 500 Index Fund, within the Schwab Equity Index Funds 2020 semi-annual report (filed with the SEC on 2nd July 2020).

From the President

Jonathan de St. Paer
President of Charles Schwab Investment Management, Inc. and the funds covered in this report.

Dear Shareholder,

The six-month reporting period ended April 30, 2020 represented one of the most volatile investing environments on record, as the COVID-19 pandemic continued to impact nearly every aspect of daily life. Up until the last week of February 2020, U.S. stock market performance was strong, with the longest bull market cycle in history continuing and major equity indices hitting new highs. But the rapid spread of COVID-19 around the world in late February and throughout March prompted increasingly strict government social distancing policies and travel restrictions that brought many economies to an abrupt halt and sent stock markets reeling. Against this unprecedented backdrop, a new record was set for the fastest U.S. market decline, with the S&P 500® Index, a bellwether for the overall U.S. stock market, dropping 30% in just 22 trading days. Global markets saw similar sell-offs. Stocks in the U.S. and abroad regained some lost ground in April, but investors continued to grapple with historic market volatility. For the six-month reporting period ended April 30, 2020, the S&P 500® Index returned -3.2%, while the Russell 2000® Index, a measure of U.S. small-cap stocks, returned -15.5% and the MSCI EAFE® Index (Net)*, a broad measure of developed international equity performance, returned -14.2%.

We don't know yet what the full impact of the COVID-19 pandemic will be or how long it will last. The sudden health crisis, sharp increase in unemployment, and market declines have understandably rattled investors, leading many to seek the perceived safety of asset classes such as cash. Market volatility and declines can be difficult to withstand, and can engender strong emotional reactions, such as selling out of fear or staying on the sidelines. In the longer-term, this can often make investing outcomes worse, as investors who missed out on the market's rebound in April may now realize. At times like these it's important to remember that all market cycles, no matter how long, ultimately come to an end.

At Charles Schwab Investment Management, we believe that maintaining a long-term investing plan and a portfolio with exposure to a mix of asset classes that perform differently over time is one way to weather the ups and downs that come with investing. We designed the Schwab Equity Index Funds with this long-term perspective in mind. The funds provide access to broad segments of the equities markets with different risk and return profiles, including small-cap, mid-cap, and large-cap stocks, those oriented toward value or growth, and equities from both U.S. and international markets. In addition, the Schwab Equity Index Funds can help investors achieve their

3

Schwab Equity Index Funds | Semiannual Report

Example Statements About Risk From S&P 500 Index Fund Letters

- *"As such, we continue to expect to see higher interest rates and increased equity market volatility."*
- *"Periods of investment uncertainty can present challenges, but experience has taught us that maintaining long-term investment goals can be an effective way to plan for the future."*
- *"While equity market volatility and political uncertainties are likely to remain with us for the rest of 2018, we remain fairly constructive on equities and credit markets, but are carefully monitoring economic and political developments."*
- *"Risks in the new year include the possible end of the boom in the housing market, where we believe prices are more likely to stall than plunge."*
- *"Looking ahead, we expect to see positive returns from equity markets, though the potential for a market correction remains high."*

Example Statements About Risk From Entire Fund Letter Corpus

- *"Market volatility readings have been remarkably low of late, but conditions can change quickly."*
- *"While we optimistically await to see what next year will bring, we note that many of the major cap-weighted and equal-weighted indices still carry significant and diversifiable related business risk."*
- *"For that reason, the portfolio manager remains vigilant, and stands ready to move the portfolio back into a "Risk Off" position should his quantitative model dictate such a move."*
- *"However, we do recognize that tighter U.S. labor markets and continued very easy monetary policy could give rise to an inflation scare later in 2016, with a temporary sell-off in risky assets and a more sustained sell-off in government bonds due to market concerns about the Fed being behind the curve a possibility."*
- *"Meanwhile, expectations for coordinated central bank easing have replaced fears of an impending recession in the U.S., and many important economic indicators appear to have troughed and now seem to be on the rise."*

B Validation of Data Sources

B.1 Validation of Textual Measures

I now validate this study's methodology for detecting sentence types in the fund letter corpus. To do this, I compare my methodology – which has been manually adapted to the fund letter setting – with methodologies from other studies that are not intentionally adapted to my study's context. This includes well-accepted methods in the Psychology literature that unfortunately suffer from a lack of transparency in how they operate to detect sentence types because they are implemented as "black-box" software packages.

I compare and contrast the power of each of these methodologies to detect forward-looking and risk-related statements in the context of fund letters. I find that the present paper's methodology is either comparable to, or exceeds, the ability of other methodologies employed in other parts of the Accounting and Psychology literatures to detect such statements. This paper's methodology thus succeeds in its goal, while also having the advantages of being (i) transparent, and (ii) intentionally designed for the fund letter corpus studied in this paper.

The first two alternative methodologies that I examine in this section (Li, 2010; Hassan, Hollander, van Lent, and Tahoun, 2019) were designed for detecting either forward-looking (only) or risk-related (only) statements in listed company filings.

Although neither methodology was designed with fund letters in mind, a study by Hillert, Niessen-Ruenzi, and Ruenzi (2021) used the Li (2010) methodology as part of an analysis of mutual fund shareholder letters, suggesting that this comparison will be informative. A statement is classified as forward-looking if it contains at least one forward-looking word, as defined by Hillert, Niessen-Ruenzi, and Ruenzi (2021) & Li (2010).

Similarly, using the methodology of Hassan, Hollander, van Lent, and Tahoun (2019), a statement is classified as risk-related if it contains at least one risk-related word, as defined by Hassan, Hollander, van Lent, and Tahoun (2019) & the Oxford English Dictionary.

The next alternative methodology I consider is the Linguistic Inquiry and Word Count (LIWC) software, of which I use the latest version, LIWC-22 (Boyd, Ashokkumar, Seraj, and Pennebaker, 2022). This software contains a number of proprietary dictionaries that are kept confidential by its vendor; it is likely these dictionaries comprise both individual word stems and regular expressions. It is operated by inputting a block of text, and returns the number of words in the input text that match a given dictionary. While the software is well accepted in the field of Psychology, its lack of transparency is a barrier to adopting this software for analyzing fund letters. I run the software to count the number of words matching two dictionaries – future focus in time orientation, and risk motives – in order to detect forward-looking and risk-related statements (respectively) in my fund letter corpus. To be precise, a statement is defined as being forward-looking if it contains at least one forward-looking word, as determined by

Table IA.1: Forward-looking sentence detection. Number & fraction of forward-looking statements detected in my fund letter corpus using three different methodologies, and (in the final column) the correlations of statement classifications of the two alternative methodologies with my own.

Methodology	**Description**	**Sentences detected**	**Fraction of corpus**	**Correlation to baseline**
This paper	Set of custom regular expressions designed for the fund letter corpus	1,529,823	31.41%	—
Hillert, Niessen-Ruenzi, and Ruenzi (2021), using the dictionary of Li (2010, Appendix B)	17 words used by Li (2010) to detect forward-looking statements in listed company 10-K & 10-Q filings	877,338	18.01%	0.3103
LIWC-22 (Boyd, Ashokkumar, Seraj, and Pennebaker, 2022)	Proprietary set of 138 words & word stems related to future focus in time orientation, designed for general-purpose psychology research	854,105	17.53%	0.4023

the LIWC software; similarly, a statement is defined as being risk-related if the LIWC software detects the presence of a risk-related word.

Table IA.1 compares the results of this paper's baseline methodology for detecting forward-looking statements to the two alternative measures. The baseline methodology detects the highest fraction of forward-looking sentences, and its classifications are positively correlated with both alternative methodologies'.

Table IA.2 compares the results of this paper's baseline methodology for detecting risk-related statements to the two alternative measures. The baseline methodology detects a higher fraction of risk-related statements than that of Hassan, Hollander, van Lent, and Tahoun (2019), and a comparable rate to the LIWC-22 software. The baseline method's classifications are positively correlated with both alternative methodologies'.

Table IA.2: Risk-related sentence detection. Number & fraction of risk-related statements detected in my fund letter corpus using three different methodologies, and (in the final column) the correlations of statement classifications of the two alternative methodologies with my own.

Methodology	**Description**	**Sentences detected**	**Fraction of corpus**	**Correlation to baseline**
This paper	Set of custom regular expressions designed for the fund letter corpus	678,338	13.93%	—
Hassan, Hollander, van Lent, and Tahoun (2019, Internet Appendix Table 3)	111 synonyms for the word "risk" according to the Oxford Dictionary, used to analyze listed company filings by Hassan, Hollander, van Lent, and Tahoun (2019)	380,268	7.81%	0.6341
LIWC-22 (Boyd, Ashokkumar, Seraj, and Pennebaker, 2022)	Proprietary set of 128 words & word stems related to risk motives, designed for general-purpose psychology research	689,979	14.17%	0.4684

B.2 Validation of Fund Letter Readership Measure

This section presents evidence that the individuals who downloaded annual & semi-annual reports from the SEC EDGAR system are broadly representative of the overall US population, based on their observable geographic dispersion. The locations of these readers are used to measure their characteristics, such as their anxiety attitudes, by matching fund reader locations with geographic local trust data.

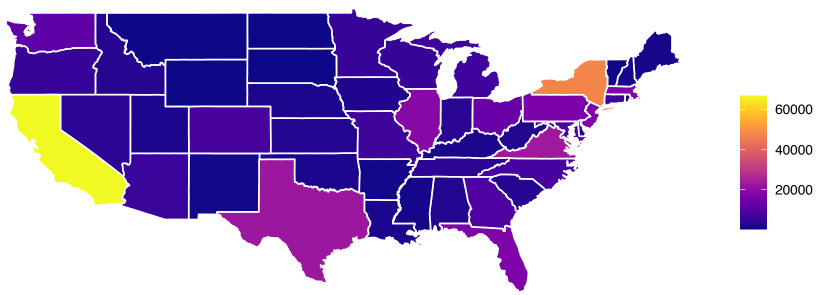
Figure IA.2(a) charts the state-level counts of annual and semi-annual report downloads from the SEC EDGAR system. Readers of funds' annual & semi-annual reports – which contain the fund letters used in this study – are located throughout the United States, rather than being concentrated in states that are home to financial institutions and professional investors, such as New York and Massachusetts (Kim, Wang, and Wang, 2022).

In fact, the correlation between state-level reader counts and the total state populations is 0.90 over the full sample used in this study, which runs from 2006 to the end of the SEC EDGAR log files in June 2017. At a more fine-grained unit of analysis, the correlation between county-level reader counts and the total county populations is also high, at 0.72. Figure IA.2(b) visualizes how county-level correlations between unique readers and the total population vary from state to state. All such correlations are positive, and for the most part uniformly high.

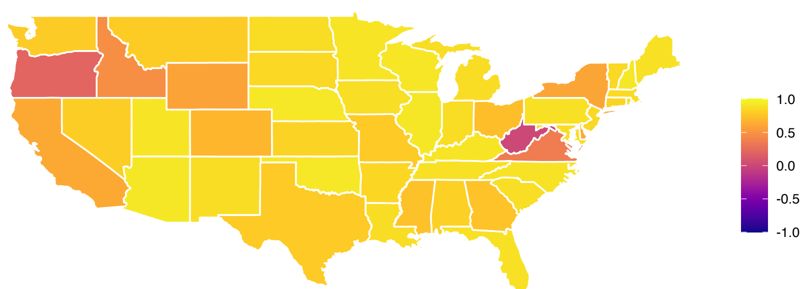
It is worth highlighting that similarly high downloader-to-population correlations were found by Grice and Guecioueur (2023), who used a related (but separate) sample of fund *prospectus* readerships from SEC EDGAR, and found that model-derived competition measures based on this prospectus readership dataset are predictive of mutual fund fee dispersion. Their study's findings validate SEC EDGAR as a source of information on investor behavior.

Figure IA.2: Annual & semi-annual report readership counts are broadly representative. Both charts visualize statistics derived from unique EDGAR user counts over the sample period 2006-2017. Users are geolocated based on their IP addresses. Sub-figure (a) displays the number of unique EDGAR annual & semi-annual report downloaders per state; lighter states have more unique readers, colored according to the legend. These reader counts are highly correlated with total population counts. Sub-figure (b) displays the correlations of unique reader counts with county population counts, for each state; lighter states exhibit higher correlations, colored according to the legend.

(a) State-level unique reader counts.



(b) Within-state reader-to-population correlations.



B.3 Validation of Prior Geographic Risk Perception Measure

In Equation (24), state s and year-month t Google search volume indices $\text{SVI}_{s,t}$ were used to define a geographic measure of local investors' prior perceptions of the level of stock market risk. This appendix examines the aggregate United States-wide version of the index, SVI_t, and compares it to the VIX index of aggregate implied stock market volatility.

Columns 1-3 of Table IA.3 regress the aggregate SVI_t against the VIX, together with a time trend (t) and a quadratic trend (t^2) to account for the growth in the use of Google searches over the entire sample period. In each specification, the SVI_t and VIX are positively and significantly associated. This validates Google search interest in this topic as a measure of perceived stock market risk at an aggregate, country-wide level. Therefore geographic variation in Google search interest should also be a valid measure of local perceptions about stock market risk.

Columns 4-5 of Table IA.3 regress the log of the SVI_t against the log of the VIX. Interpreting the log-log coefficient in column (4), a 10% increase in the level of the VIX is associated with an approximately 3% increase in Google search volume about the topic of stock market crashes. The positive association between the growths of the two variables thus parallels the positive association between their levels.

Table IA.3: Comparing aggregate Google search interest to the VIX. SVI_t is the monthly Google Search Volume Index across the United States about the topic of a stock market crash, as defined by Google themselves. The controls t and t^2 denote the presence of time trends and a quadratic growth term, respectively. The sample period of this analysis runs from 2004-2022.

Dependent Variables:	SVI_t			Log SVI_t	
	(1)	(2)	(3)	(4)	(5)
VIX_t	0.5228***	0.4938***	0.4896***		
	(0.1723)	(0.1669)	(0.1643)		
Log VIX_t				0.3076**	0.2638**
				(0.1291)	(0.1180)
Constant	10.25***	5.148	11.65***	2.020***	1.884***
	(3.022)	(3.265)	(3.165)	(0.3597)	(0.3464)
t		0.0501***	-0.1197***		0.0023***
		(0.0109)	(0.0342)		(0.0007)
t^2			0.0008***		
			(0.0002)		
N	225	225	225	225	225
R^2	0.15	0.23	0.29	0.07	0.19

Newey-West standard errors (in parentheses) use automatically selected lags.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

B.4 Validation of Anxiety Measures

Section 5.1 uses two Google search volume indices, for the topics of Worry and Psychological Stress, as proxies for fund letter readers' anxiety attitudes. I now compare these aggregate measures of anxiety to two measures of households' financial risk-taking, and find the anxiety measures do indeed predict risk-taking in aggregate.

I employ two measures of household risk-taking that are compiled by the Census Bureau Survey of Income and Program Participation (SIPP), and made publicly available as part of the SIPP Wealth Tables for each Census region.[23] The first measure, Participation Rate, is simply the percentage of households owning stocks or mutual fund shares among the population, taken from SIPP Wealth Tables 2 or 3. The second measure is the ratio of the mean dollar value of stocks and mutual funds held by a household to the mean dollar net worth of a household (excluding home equity):

$$\text{Average Risky Share} = \frac{\text{Mean Value of Stocks and Mutual Funds Owned}}{\text{Mean Net Worth}}. \tag{30}$$

This serves as a proxy for an average household's financial wealth share held in risky assets, and is computed from SIPP Wealth Table 5. Both measures are computed at an annual frequency, for the years 2005, 2009-2011, and 2013-2017; i.e. for all the data releases that fall within the sample period used in Section 5.1.

In order to compare Google SVI values, which are collected at a state s and year-month m value, I first compute the mean state s year y value, and then take a population-weighted mean to aggregate up to the region r level:

$$\text{SVI}_{r,y} = \sum_{s\in r}\left(\frac{\text{Population}_s}{\text{Population}_r} \times \frac{1}{12}\sum_{m\in y}\text{SVI}_{s,m}\right), \tag{31}$$

with a slight abuse of the $\in$ notation. I then cross-sectionally standardize SVI values within each year y in order to eliminate the influence of any timeseries growth in search engine usage over the sample period.

Table IA.4 presents the results of cross-sectional regressions that compare the two measures of household risk-taking to the two measures of anxiety. Column (1) shows that regions with a greater search intensity for Worry-related terms have a lower household participation rate, and column (4) shows that a lower average risky share held by households is associated with a greater search intensity for terms relating to Psychological Stress.

[23]The four Census regions are the Northeast, Midwest, South and West. Note that publicly-released microdata for major household surveys, including the Fed's Survey of Consumer Finances (SCF), the Panel Study of Income Dynamics (PSID) and the Michigan Surveys of Consumers, also report households' locations at the level of the Census region, with the goal of protecting respondents' privacy. Confidential access to survey microdata may allow access to households' locations at a finer level of geographic resolution (e.g. Giannetti and Wang (2016)).

While subject to the limitations of a coarse-grained unit of analysis, these negative and significant associations are consistent with the known causal effect of heightened anxiety on financial risk-taking (e.g. Guiso, Sapienza, and Zingales (2018)).

Table IA.4: Comparing aggregate household risk-taking to search-based anxiety. This table shows the results of cross-sectionally comparing two measures of aggregate household risk-taking (dependent variables) to two measures of aggregate anxiety (independent variables). Risk-taking is measured by the household participation rate and the average risky share held in stocks or mutual fund shares. The aggregate measures of Worry and Psychological stress are Google search volume indices (SVI) for these two topics; refer to Section 5.1 for a more detailed description. The unit of analysis is at the Census region and year level, for all available survey results from 2005 to 2017.

Dependent Variables:	Participation Rate (%)		Average Risky Share	
	(1)	(2)	(3)	(4)
Worry	-2.144***		0.1394	
	(0.2558)		(0.0951)	
Psychological Stress		0.4559		-0.1553**
		(0.2640)		(0.0655)
Year FEs	✓	✓	✓	✓
N	32	32	32	32
R^2	0.08	0.42	0.10	0.44

Clustered (year) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

C The Geography of Anxiety and Stock Market Risk Perceptions

This appendix charts the geographic variation in Google search-based measures of anxiety (Figure IA.3) and stock market risk perceptions (Figure IA.4). These geographic measures are used in the main paper to construct fund letter readership-level measures of anxiety and perceived risk.

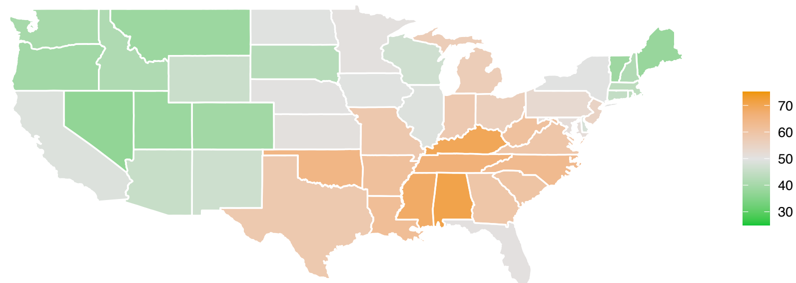
Recall that the Google search volume index (SVI) panel consists of state s and year-month t search volume indices for a particular topic k: $\text{SVI}_{s,t}^{k}$. To produce visualizations of search activity for a particular topic k, I first compute the rank-percentiles for each state s within a given month t, and then take the mean of these rank-percentiles over the sample period. These mean rank-percentiles are then displayed on a map of the continental United States, in order to visualize time-invariant geographic dispersion in Google-search based measures of anxiety and stock market risk perceptions (with the topic k defined accordingly).

Figure IA.3 charts geographic dispersion in the population's anxiety attitudes. Panel (a) measures anxiety attitudes by the SVI for Google's "Psychological Stress" topic. Panel (b) measures anxiety attitudes by the SVI for Google's "Worry" topic. Finally, panel (c) uses the average of both topic SVIs. Figure IA.3 reveals that anxiety is not persistently high in any one geographic region: the East Coast, the South and the Midwest all contain a mixture of high-anxiety and low-anxiety investors. This suggests that anxiety attitudes are not driven by persistent demographic attributes, such as the level of urbanization or cultural values.

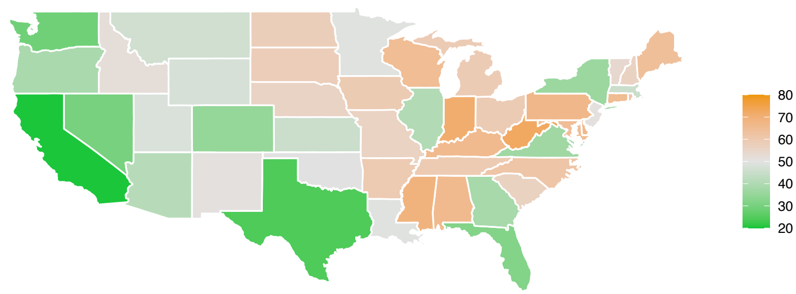
Figure IA.4 charts geographic dispersion in the population's stock market risk perceptions. This is measured by the SVI for Google's "stock market crash" topic. Internet Appendix B.3 confirms that this is correlated with the level of the VIX in the timeseries, and thus serves as a proxy for perceived risk. The lack of overlap with anxiety attitudes (which influence risk aversion) confirms that this measure (which captures the perceived level of risk) is distinct.

Figure IA.3: Geography of anxiety. Means of cross-sectional rank-percentiles of state-level Google Search Volume indices, over the period 2006 to 2017, for various search topics. Sub-figure (a) shows the distribution of the "Psychological Stress" topic. Sub-figure (b) shows the distribution of the "Worry" topic. Sub-figure (c) shows the distribution of the mean of both topics. Orange states denote a cross-sectionally higher level of anxiety; green states denote a lower level of anxiety.

(a) "Psychological Stress" topic.



(b) "Worry" topic.



(c) Mean of both topics.

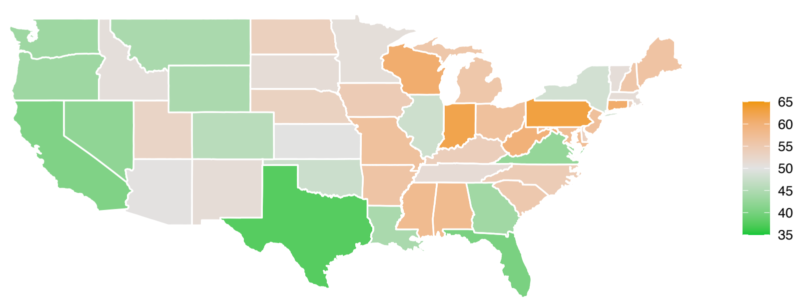
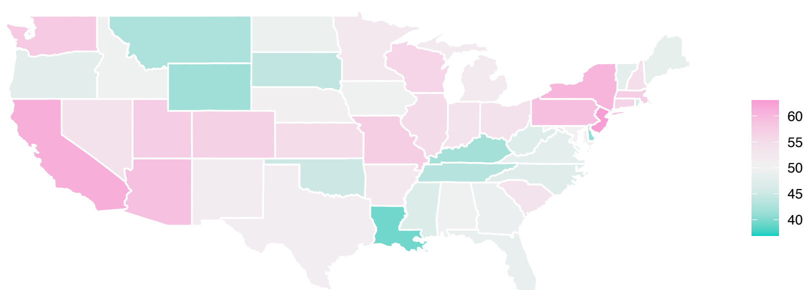


Figure IA.4: Geography of stock market risk perceptions. Means of cross-sectional rank-percentiles of state-level Google Search Volume indices, over the period 2006 to 2017, for the "stock market crash" topic. Purple states denote a cross-sectionally higher level of perceived stock market risk; teal states denote a lower level of perceived risk.



D Alternative Specifications for the Effect of Communication

This appendix conducts panel regressions using alternative measures for the amount of detail communicated about risk.

D.1 Linear Specifications

The first alternative measure is a simple linear regression of the word count devoted to discussing risk during month t. The coefficient estimates for this specification, presented in Table IA.5, show that this measure is significantly and positively associated with the net flows into the fund from the end of month t to the end of month $t+1$.

Table IA.5: Linear estimates of effect of communication about risk on S&P 500 index fund flows. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the number of words devoted to that topic. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). Columns 1-4 present results with no variable transformations. Columns 5-8 present results where both the dependent and the main independent variables have been standardized (z-scores). *Fund controls* (not shown) are for each fund letter's total word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only.

Dependent Variables:	Net Flow$_{i,t\to t+1}$ (%)				z(Net Flow$_{i,t\to t+1}$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Word Count$_{i,t}$	0.0006***	0.0005**	0.0005**	0.0005***				
	(0.0002)	(0.0002)	(0.0002)	(0.0002)				
z(Word Count$_{i,t}$)					0.1670***	0.1266**	0.1216**	0.1408***
					(0.0479)	(0.0524)	(0.0515)	(0.0524)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls$_t$			✓				✓	
Year-month FEs	✓			✓	✓			✓
Fund FEs		✓	✓	✓		✓	✓	✓
N	1,253	1,253	1,253	1,253	1,253	1,253	1,253	1,253
R^2	0.19	0.27	0.27	0.40	0.19	0.27	0.27	0.40

*Clustered (Fund & Year-month) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

D.2 Fractional Specifications

The second alternative measure of the amount of detail communicated is the fraction of the fund's letter devoted to a discussion of risk. That is, rather than controlling for the total word count of the letter implicitly, the total word count is explicitly normalized as the denominator of the letter fraction independent variable.

Using this measure, Table IA.6 provides evidence on the effect of the amount of detail communicated on various fund flows and under various fixed effect specifications. The estimated directions of the effect remain unchanged when compared to the baseline regressions in the main paper: an increase in the amount of detail communicated about risk attracts inflows to the fund over the subsequent month.

Table IA.6: Effect of the fractions of fund letters devoted to risk discussions on S&P 500 index fund flows. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk using the fraction of words devoted to that topic in the whole letter. Controls (not shown) are for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only. Each flow variable is the next month's flow, in percentage points of TNA.

Dependent Variables:	Net Flow (1)	Inflow (2)	Outflow (3)
Fraction of Letter$_{i,t}$ (%)	0.0282**	0.0268*	2.2×10^{-5}
	(0.0122)	(0.0145)	(0.0038)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	1,253	1,253	1,253
R^2	0.41	0.78	0.85

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

D.3 Log-Log Specifications, With a Structural Interpretation

Table IA.7 presents the results of panel regressions of log flows on the log word count communicated about risk. Thus the main paper's baseline specifications are modified by taking the log of the dependent variable.

All the coefficient estimates are positive and significant. Estimates in columns 1-4 are based on net flows, which therefore exclude samples with negative net flows. For this reason, estimates in columns 5-8 are preferred: these are based on gross inflows, and so fewer samples are discarded during the estimation.

A Structural Interpretation The estimated coefficients in Table IA.7 admit a structural interpretation. Consider the following simple framework for the effect of the number of words about risk communicated, $w_t \geq 0$, on an investor's anxiety multiplier:

$$a_{t+1} = \max\left\{\frac{a_t}{1+w_t^k}\,,\ 1\right\}, \quad k > 0. \tag{32}$$

This relation fulfills a number of desirable properties. First, it trivially fulfills the GSV condition that $a_{t+1} \geq 1$. Second, when there is no communication, $w_t = 0$, the investor's anxiety multiplier remains unchanged: $a_{t+1} = a_t$. Third, anxiety is decreasing in the extent of communication: $\frac{\partial a_{t+1}}{\partial w_t} < 0$ (away from the boundary of 1, for $w_t > 0$). To interpret the coefficient estimates, substitute (32) into (6) to obtain the following simple relation, assuming beliefs are not updated:[24]

$$\text{Net Flow} = w_t^k. \tag{33}$$

Since

$$\log(\text{Net Flow}) = k \log(w_t), \tag{34}$$

the estimated coefficients on the log number of words in Table IA.7 columns 4-8 are also estimates for the constant k in Eqn. (32). Therefore, $\hat{k} \approx 0.1$, and the relation

$$a_{t+1} = \max\left\{\frac{a_t}{1+\sqrt[10]{w_t}}\,,\ 1\right\} \tag{35}$$

is a concise framework for capturing the dampening effect of more detailed communication on the investor's effective risk aversion.

[24] Sections 6.2 and 6.3 provide extensive evidence that investors do not update their beliefs based on this communication.

Other Possible Structural Interpretations Other structural interpretations of the estimates in Table IA.7 are possible. Consider, for example, augmenting the simple framework (32) with a time-varying process b_t, as follows:

$$a_{t+1} = \max\left\{\frac{a_t}{1 + b_t w_t^k}\ ,\ 1\right\}, \quad k > 0,\ b_t > 0. \tag{36}$$

The above relation continues to exhibit the same desirable properties as before; however, it now follows that

$$\text{Net Flow} = b_t\, w_t^k \tag{37}$$

$$\Rightarrow \qquad \log(\text{Net Flow}) = k\, \log(w_t) + \log(b_t). \tag{38}$$

In this case, the additional parameters b_t cannot be identified based on the regressions in Table IA.7 alone. For example, one cannot rule out the possibility that $b_t := a_t$ might enter into the framework (36). Such parameters b_t may capture other influences on the effectiveness of communication: $b_t > 1$ would strengthen the effect, whereas $b_t < 1$ would moderate it. Importantly, however, if the term $\log(b_t)$ is absorbed by panel regression time effects, or by the additive combination of time effects and fund fixed effects, then k is still successfully identified as $\hat{k} \approx 0.1$.

Table IA.7: Log-log estimates of effect of communication about risk on S&P 500 index fund flows. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Gross dollar flows are extracted from mandatory SEC filing documents (N-SAR, N-PORT), and then normalized as a percentage of lagged total net assets. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only. To aid a structural interpretation, flow variables are expressed in fractions, not percentage points.

Dependent Variables:	Log Net Flow$_{i,t\to t+1}$				Log Gross Inflow$_{i,t\to t+1}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Word Count$_{i,t}$	0.1555**	0.3384***	0.3511***	0.3839***	0.1903**	0.1142**	0.1154**	0.1241**
	(0.0781)	(0.1010)	(0.1029)	(0.1453)	(0.0880)	(0.0490)	(0.0486)	(0.0590)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls$_t$			✓				✓	
Year-month FEs	✓			✓	✓			✓
Fund FEs		✓	✓	✓		✓	✓	✓
N	391	391	391	391	1,142	1,142	1,142	1,142
R^2	0.40	0.41	0.42	0.69	0.21	0.67	0.67	0.72

Clustered (Fund & Year-month) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

E Further Identification Details

E.1 Instrumental Variables Strategy

This appendix supplements Section 4.2.2 in the main text with further discussions of the underlying assumptions, as well as additional empirical evidence.

Evidence of Information Diffusion Within Fund Families The high and significantly non-zero estimates in columns (1) and (3) of Table IA.8 show that the amount of detail provided in S&P 500 index fund letters is positively correlated with the average amount of detail in similar fund letters that are written by domestic active equity funds, who are likely to conduct fundamental research. This is consistent with information spillovers within fund families. I provide a second piece of evidence by linking these spillovers to fund family incentives, which may encourage cooperation between fund managers. For this additional test, I include a family-level index of the extent to which families engage in cooperative behavior, in columns (2) and (4) of Table IA.8. This index was computed by Evans, Prado, and Zambrana (2020) based on a variety of incentives faced by managers, such as the structure of fund management and the design of compensation contracts. If the observed information spillovers are due to cooperation within fund families, one would expect a positive interaction term. That is indeed the case: the positive and significant interaction terms in columns (2) and (4) suggest that information spillovers are stronger when funds within the family have a greater incentive to cooperate.

Table IA.8 also makes clear that the instrument predicts the (log of the) number of words within the fund letters of S&P 500 funds in the same fund families. The results are robust to different fixed effects specifications, and to including a host of controls that will be employed in future analyses. The coefficients in columns (1) and (3) are positive and significantly different to zero: the instrument is relevant. Furthermore, F-stats around the 300 level suggest the instrument is also strong enough to test for coefficient significance even at the 99% level.[25]

Source of Variation Variation in w_A arises due to variation based on the signals uncovered by fund managers as part of their day-to-day operations. Different fund families would have a different mix of fund strategies; even within the same class of strategies, it is clear from the high volume of trade in financial markets that considerable heterogeneity in signal structures (and their realizations) must exist. S&P 500 indexing is a common fund strategy for fund families to offer, and therefore my sample of S&P 500 funds will be subject to heterogeneous

[25]According to Lee, McCrary, Moreira, and Porter (2022, Table 3), a first-stage F-stat of 104.67 is necessary to test for significance at the 95% level without adjusting standard errors. Similarly, an F-stat of 252.34 is necessary to test at the 99% level.

information spillovers from the other funds within those fund families, driving variation in the instrument.

Supplementary Results This appendix concludes with a number of additional empirical results: Tables IA.9 and IA.10 apply the instrumental variables strategy to analyze the response of gross inflows and gross outflows (respectively) to the amount of detail provided. The results are similar to the un-instrumented results presented in Section 4.1.

Table IA.8: Instrument relevance, and evidence of information spillovers. The amount of detail about risk is the log word count devoted to that topic. The regressions in this table document a high correlation between the amount of detail about risk contained in S&P 500 index fund letters (dependent variable) written by fund i at time t, and the corresponding amount of detail across active domestic equity sibling funds (only) within the same fund family, aggregated as a simple mean. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. Columns (1) & (3) show the unconditional strength of this relationship, across different fixed effects specifications, for the same sample used in future analyses. Columns (2) & (4) additionally include an interaction with an index for the strength of cooperation between sibling funds in the focal fund's overall family, as measured by Evans, Prado, and Zambrana (2020); their dataset ends in December 2015, leading to a smaller sample size in those columns.

Dependent Variable:	Risk Detail$_{i,t}$			
	(1)	(2)	(3)	(4)
Sibling Mean Risk Detail$_{i,t}$	0.8495***	0.6102***	0.8124***	0.6411***
	(0.0335)	(0.0857)	(0.0407)	(0.0939)
Sibling Mean Risk Detail$_{i,t}$		0.3308**		0.3727**
× Family Cooperation Index$_{i,t}$		(0.1324)		(0.1439)
Family Cooperation Index$_{i,t}$		-1.801**		-2.289***
		(0.7939)		(0.8547)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓		
Fund FEs			✓	✓
N	1,155	640	1,155	640
R^2	0.94	0.94	0.95	0.95
F-stat	253.4	95.1	294.8	118.2

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.9: Instrumented effect of the amount of detail communicated about risk on the inflows of S&P 500 index funds. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic.The instrument for the main covariate used in the 2SLS specifications is constructed from the log word counts (about the same topic of risk) of the letters written by active domestic equity sibling funds within the focal index fund's overall family. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month returns, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only. Columns 1-3 show un-instrumented panel specifications. Columns 4-6 show 2SLS specifications, with F-stats corresponding to the first stage; the first-stage regressions are shown in Table IA.8.

Dependent Variable:	Inflow$_{i,t\rightarrow t+1}$ (%)							
Regression:	OLS				2SLS IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Detail$_{i,t}$	0.2272	0.4901***	0.4932***	0.4709**	0.4932**	0.5845***	0.5879***	0.6686***
	(0.2277)	(0.1731)	(0.1742)	(0.2021)	(0.2456)	(0.1972)	(0.1984)	(0.2490)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls$_t$			✓				✓	
Year-month FEs	✓			✓	✓			✓
Fund FEs		✓	✓	✓		✓	✓	✓
N	1,155	1,155	1,155	1,155	1,155	1,155	1,155	1,155
R^2	0.26	0.74	0.74	0.79	0.26	0.74	0.74	0.79
F-stat (IV 1st-stage)					253.4	294.8	236.2	361.3

Clustered (Fund & Year-month) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.10: Instrumented effect of the amount of detail communicated about risk on the gross outflows of S&P 500 index funds. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. The instrument for the main covariate used in the 2SLS specifications is constructed from the log word counts (about the same topic of risk) of the letters written by active domestic equity sibling funds within the focal index fund's overall family. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month returns, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only. Columns 1-3 show un-instrumented panel specifications. Columns 4-6 show 2SLS specifications, with F-stats corresponding to the first stage; the first-stage regressions are shown in Table IA.8.

Dependent Variable:	Outflow$_{i,t\rightarrow t+1}$ (%)							
Regression:	OLS				2SLS IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Detail$_{i,t}$	-0.0772	0.0194	0.0203	0.0503	0.0297	-0.0448	-0.0444	0.0296
	(0.1676)	(0.0820)	(0.0823)	(0.0971)	(0.1928)	(0.0982)	(0.0984)	(0.1221)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls$_t$			✓				✓	
Year-month FEs	✓			✓	✓			✓
Fund FEs		✓	✓	✓		✓	✓	✓
N	1,155	1,155	1,155	1,155	1,155	1,155	1,155	1,155
R^2	0.25	0.83	0.83	0.85	0.25	0.83	0.83	0.85
F-stat (IV 1st-stage)					253.4	294.8	236.2	361.3

Clustered (Fund & Year-month) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

E.2 Corner Bunching Strategy

This section is an informal exposition of CCN's methodology for identifying treatment effects by exploiting corner bunching, and uses similar notation to theirs. Refer to the papers by Caetano, Caetano, and Nielsen (forthcoming) and Caetano, Caetano, Nielsen, and Sanfelice (2023) for more formal expositions.

The goal is to estimate the causal effect β of a treatment variable X on an outcome Y, in settings where the treatment variable is constrained to be non-negative, $X \geq 0$.

Assumption 1 (Linearity) The treatment and outcome are related by a linear model:

$$Y \;=\; \beta X \;+\; Z'\gamma \;+\; \underbrace{\delta\eta + \varepsilon}_{\text{Unobservable}}\;, \qquad \mathbb{E}[\varepsilon|X, Z, \eta] = 0, \tag{39}$$

where Z is a vector of observable controls, and η is an unobservable confounder. Note, therefore, that X is exogenous when $\delta = 0$ but endogenous otherwise.

Assumption 2 (Corner bunching) The (unobserved) unconstrained choice of X^* is related to the (observed) constrained treatment variable X by the controls Z and confounder η:

$$X = \max\{0, X^*\} = \max\{0, Z'\pi + \eta\}. \tag{40}$$

Also, the non-negativity constraint is binding for part of the population:

$$0 < \mathbb{P}(X^* < 0) < 1; \tag{41}$$

i.e. part of the population exhibits no variation in X at the corner value $X = 0$, but *does* have variation in the unobserved confounder η that drives the endogeneity. As CCN put it, "Our approach relies on the fact that we have a glimpse of how η affects Y separately from the effect of X only at the bunching point [...] At that location, we can guarantee that the treatment will not vary while the unobservables will."

From these first two assumptions, it follows that:

$$\begin{aligned}
\mathbb{E}[Y|X, Z] &= (\beta + \delta)X + Z'(\gamma - \pi\delta) + \delta\mathbb{E}[X^*|X^* \leq 0, Z]\,\mathbb{1}\{X = 0\} && (42)\\
&= \beta X + Z'(\gamma - \pi\delta) + \delta\Big(\underbrace{X + \mathbb{E}[X^*|X^* \leq 0, Z]\,\mathbb{1}\{X = 0\}}_{\text{Correction term}}\Big). && (43)
\end{aligned}$$

That is, the causal effect β may be estimated if an endogeneity correction term is also estimated and included in a regression. The coefficient on this correction term is δ; therefore, testing for $\delta = 0$ is equivalent to testing whether X is exogenous (see Eqn. (39)).

In order to generate the correction term, note that since X and $\mathbb{1}\{X = 0\}$ are observed, only $\mathbb{E}[X^*|X^* \leq 0, Z]$ must be identified. For this purpose, a shape restriction is required. I now review a classic parametric shape restriction, a semiparametric version of it, and a weaker nonparametric shape restriction.

Assumption 3a (Tobit shape restriction) If we assume $\eta|Z \sim \mathcal{N}(Z'\mu, \sigma^2)$, then

$$\begin{aligned} \mathbb{E}[X^*|X^* \leq 0, Z] &= \mathbb{E}[Z'\pi + \eta|X^* \leq 0, Z] && (44) \\ &= -Z'(\pi+\mu) - \sigma\lambda\left(\frac{-Z'(\pi+\mu)}{\sigma}\right), && (45) \end{aligned}$$

where $\lambda(x) = \frac{\text{Gaussian PDF } \phi(x)}{\text{Gaussian CDF } \Phi(x)}$; i.e. the function $\lambda(\cdot)$ is the Inverse Mills Ratio. Therefore, the parameters μ, σ^2 can be estimated by a Tobit regression of the treatment X onto the controls Z with censoring below 0.

Assumption 3b (Semiparametric Tobit shape restriction) If we assume $\eta|Z \sim \mathcal{N}(Z'\mu_z, \sigma_z^2)$; i.e. allowing μ, σ to be functions of $Z = z$, then

$$\mathbb{E}[X^*|X^* \leq 0, Z = z] = -Z'\pi + \mu_z - \sigma_z\lambda\left(\frac{-(Z'\pi+\mu_z)}{\sigma_z}\right). \tag{46}$$

The parameters μ_z, σ_z^2 may be estimated by a Tobit regression of X onto a constant with censoring below 0, for the subset of observations where $Z = z$.

Assumption 3c (Nonparametric Symmetric Tails shape restriction) We can make the nonparametric assumption that the distribution $\eta|Z$ has symmetric tails. Many distributions satisfy this assumption (including the Gaussian, Student's t and Uniform distributions). Note, however, that this assumption does *not* assume the entire distribution is symmetric, and so is quite general. Following from this weak assumption,

$$\begin{aligned} \mathbb{E}[X^*|X^* \leq 0, Z = z] \; &= F^{-1}_{X|Z=z}\big(1 - F_{X|Z=z}(0)\big) && (47) \\ &\quad - \mathbb{E}\Big[X \,\Big|\, X \geq F^{-1}_{X|Z=z}\big(1 - F_{X|Z=z}(0)\big), Z = z\Big]. \end{aligned}$$

Therefore, the empirical CDF can be used to identify the correction term, assuming that $\mathbb{P}(X = 0|Z = z) < 0.5$.

Assumptions 3a, 3b & 3c above offer alternative distributional assumptions for identifying the endogeneity correction term that is required by Equation (43), for the ultimate purpose of identifying the treatment effect β. All three shape restrictions are used in the empirical analysis of Section 4.2.3 in the main paper.

Given Assumptions 1, 2 and 3 (a, b, or c), CCN show that the terms β, $(\gamma - \pi\delta)$ and δ in Equation (43) are identified, and furthermore that standard errors are consistently estimated by the bootstrap.

CCN also discuss the challenges of estimation when the vector of controls Z is high-dimensional. Given the large number of fixed effects and controls required in the current paper's setting, I closely follow their suggested approach of discretizing Z into K=50 clusters. I also verify that the estimates are robust to cluster counts in the neighborhood of $K = 50$.

F Ruling Out Further Alternative Explanations

Section 6 rules out a number of potential alternative explanations for the paper's main results. This appendix continues to evaluate and rule out additional potential explanations.

F.1 Are Investors Being Educated Rather Than Reassured?

I now assess whether the communication related to risk actually educates investors rather than simply reassuring them. While both channels involve information provision, they have distinct implications.

First, I note it is likely that investors in S&P 500 index funds already possess a relatively high level of financial literacy. Bailey, Kumar, and Ng (2011) find that unsophisticated investors exhibiting "strong behavioral biases [...] tend to avoid" index mutual funds altogether. Nonetheless, I directly consider this potential explanation.

To conduct this test, I employ state-level financial literacy scores that have been computed by Lusardi, Bumcrot, and Lin (2013) based upon geographical surveys. I aggregate these state-level scores to fund letter readership measures. Then I assess how the strength of the main flow-detail regressions vary with the financial literacy of fund letter readers. Panel regression results are displayed in Table IA.11; these incorporate an interaction term for below-median levels of financial literacy among fund letter readerships.

If an educational mechanism is at play, the coefficients on the interaction term in columns (1) and (2) should be positive and significantly different to zero. Instead, neither is significant, and the signs across specifications are not consistent with one another. Therefore, these results do not support an interpretation in which investors are educated by information provision (instead of reassurance).

Table IA.11: Interaction between communication detail & readers' financial literacy, and the relation with S&P 500 index fund flows. The amount of detail contained in the letter written by fund i at time t about risk is measured using the log of the number of words devoted to that topic. The indicator captures a fund letter readership with a below-median level of financial literacy; a baseline effect is estimated in addition to the interaction shown in the table. Financial literacy is measured at a state level by Lusardi, Bumcrot, and Lin (2013), and then aggregated to a fund letter readership level. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, total fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only.

Dependent Variables:	Net Flow	Inflow	Outflow
	(1)	(2)	(3)
Risk Detail$_{i,t}$	0.5103***	0.6247**	-0.0093
	(0.1528)	(0.2627)	(0.0906)
Risk Detail$_{i,t}$	0.0444	-0.1173	0.0069
$\times$ $\mathbb{1}${Low Financial Literacy}$_{i,t}$	(0.1378)	(0.1701)	(0.0958)
Fund Controls$_{i,t}$	✓	✓	✓
Market Controls$_t$	✓	✓	✓
Fund FEs	✓	✓	✓
N	901	901	901
R^2	0.29	0.68	0.80

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

F.2 Does More Communication Shroud Risks?

I now consider as an explanation for my results that a relatively larger fraction of text shrouds this attribute, and makes it less salient. Such a mechanism would be most closely related to that of Célérier and Vallée (2017) and Vokata (2023), but at odds with the classic interpretation of more communication about risk signaling a higher level of risk (for example, Hassan, Hollander, van Lent, and Tahoun (2019)). If writing more about risk diminishes the salience of this attribute in the eyes of a letter reader, this may lead her to perceive or assume a lower level of risk, and therefore unintentionally take more risk than in the counterfactual of full awareness of this important attribute.

I rule out this explanation by directly measuring the complexity of the text discussing risk, and controlling for this complexity. Table IA.12 repeats this paper's main flow-detail panel regressions, but additionally includes four controls for the textual complexity of the text. Whatever measure is used (either separately or jointly), I continue to find the main effect is robust and of the same order of economic magnitude as in the baseline.

Table IA.12: Communication detail & textual complexity, and the relation with S&P 500 index fund flows. The amount of detail contained in the letter written by fund i at time t about risk is measured using the log of the number of words devoted to that topic. In addition to this main independent variable, I control first separately and then jointly for four alternative measures of textual complexity. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, total fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Risk Detail$_{i,t}$	0.5044***	0.5021***	0.4811**	0.5049***	0.4632**	0.4175**
	(0.1808)	(0.1847)	(0.1845)	(0.1810)	(0.1911)	(0.1827)
Flesch Reading Ease$_{i,t}$		-0.0013				0.0285
		(0.0091)				(0.0339)
Flesch-Kincaid Grade Level$_{i,t}$			0.0248			0.0231
			(0.0320)			(0.1865)
FOG Index$_{i,t}$				0.0002		0.0001
				(0.0006)		(0.0017)
SMOG Index$_{i,t}$					0.0539	0.1492
					(0.0507)	(0.1811)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓	✓
N	1,154	1,154	1,154	1,154	1,154	1,154
R^2	0.44	0.44	0.44	0.44	0.44	0.44

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

F.3 Does More Communication About Risks Attract New Investors?

A main result of the paper is that more detailed communication attracts more inflows. If the additional assets are invested by existing clients of the fund, then my interpretation of investors' effective risk aversion being reduced is justified. Indeed, the fact that fund letters (as well as the annual and semi-annual reports that contain them) are distributed to existing investors supports this interpretation. However, if the detailed communication additionally attracts new clients to invest with the fund, then the main channel is potentially confounded empirically. Therefore, this appendix explicitly tests whether new investors are attracted by more detailed communication about risk, and finds no evidence that this occurs.

I introduce a novel measure of the number of new investors in a fund during the month t in which a fund letter is filed, and during the subsequent month $t+1$ (which is the same period over which I compute fund flows in this paper's main analyses). A new investor is defined as an investor who has acquired a prospectus for the fund during this period, and has also *not* acquired a prospectus for the same fund during some past period beginning in month $t-k$ and ending in month $t-1$; i.e. over a k month lookback window. Similarly to Grice and Guecioueur (2023), a prospectus acquisition is defined as a download from the SEC's EDGAR website, and an investor is identified by her IP address. Grice and Guecioueur (2023) discuss that EDGAR is a widely-used source of prospectus information for investors (and is also indexed by search engines) and that this sample of investors are representative of the broader US population.

The test is to compare whether more communication about risks attracts any new investors above a baseline average. I implement this test by predicting the number of new investors using Poisson regressions with fixed effects (Cohn, Liu, and Wardlaw, 2022). Fund and year-month fixed effects control for fund-level and time-specific averages, and I additionally measure the mean monthly new investor count over the lookback window and include this variable as an explicit control. The specifications are completed with fund-level controls that are identical to those in the main paper's baseline (OLS) panel regressions.

Table IA.13 presents the results of such Poisson panel regressions, varying the lookback window k across specifications. As might be expected, the historical rate of arrival of new investors is a positive and (mostly) significant predictor of the number of new investors concurrent with the dissemination of a fund letter. However, the amount of detail communicated about risk is not: coefficients in the first row are not significantly different to zero. Moreover, the negative sign is the opposite to that which we would expect, if more communication about risk attracts additional investors to the fund.

These results strongly suggest that the effect of risk detail communicated acts on the existing base of investor clients, rather than attracting ones.

Table IA.13: Detail communicated about risk vs. new investors in S&P 500 index funds. This table shows the estimates of Poisson panel regressions. The number of new investors in fund i over months t and $t+1$, measured using prospectus acquisition decisions on the SEC EDGAR website, is the common dependent variable of interest. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. In addition, the historical monthly mean number of new investors is included as a control. Existing and new investors are determined based on prospectus acquisition decisions made over the trailing k month window prior to the filing of each fund letter at time t. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	New Investors in Fund i During $\{t, t+1\}$				
	(1)	(2)	(3)	(4)	(5)
$\text{Detail}_{i,t}$	-0.0181	-0.0285	-0.0338	-0.0582	-0.0538
	(0.0496)	(0.0620)	(0.0659)	(0.0690)	(0.0714)
Mean Monthly New $\text{Investors}_{i,t}$	0.0168^{***}	0.0117^{***}	0.0147^{***}	0.0079	0.0104^{*}
	(0.0028)	(0.0033)	(0.0039)	(0.0060)	(0.0055)
Lookback k (months)	12	24	36	48	60
Fund $\text{Controls}_{i,t}$	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓
N	862	836	802	721	635

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

F.4 Do Managers Transmit Signals That Pander to Investor Beliefs?

One strand of the literature on behavioral persuasion posits that communication that is targeted to agree with a recipient's existing beliefs can persuade her to take an action favored by the sender of a message (Mullainathan and Shleifer, 2005). To test whether such a mechanism is currently at play, I examine whether the distance between prior perceived risk and the level signaled by fund letters is associated with flows.

I modify the regressions of Table 10 to incorporate the distance between investors' prior perceived risk and the signal conveyed by the fund letter, rather than a signed update:

$$\text{Distance from Prior}_{i,t} = |\text{Letter Signal}_{i,t} - \text{Readership Prior}_{i,t}|, \tag{48}$$

where the first term is defined by Eqn. (8) and the second term is defined by Eqn. (23).

Table IA.14 presents the results of such panel regressions. Under a mechanism in which distance between the recipient's beliefs and the message attracts inflows to the fund – the desired action of the fund manager who sends the message – one would expect be a negative and significant coefficient on the distance covariate. While negative in sign, the large standard errors and lack of significance of these coefficients are not consistent with the proposed pandering mechanism.[26] Furthermore, the main coefficient estimate on the amount of detail communicated is largely unchanged across specifications, and remains strongly significant throughout.

[26] GSV discuss that fund managers may pander to investor beliefs when designing securities to attract investor clients in the first place. Such pandering in security design is distinct from pandering in persuasive communication.

Table IA.14: Interaction between communication detail & distance of communicated risk from readers' priors, and the relation with S&P 500 index fund flows. This table shows the relationship between net flows and both communication detail and the distance of the signaled level of risk from fund letter readers' priors. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. The letter's signal about risk is measured based on the net sentiment of the text discussing risk, and is defined by Eqn. (8). The prior perceived level of risk is measured based on local Google searches, and is defined by Eqn. (23) The distance term is the absolute difference between the two, as defined by Eqn. (48). Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)		
	(1)	(2)	(3)
Risk Detail$_{i,t}$	0.5207***	0.5077***	0.4875**
	(0.1854)	(0.1762)	(0.2040)
Distance from Prior$_{i,t}$		-0.0747	-0.1619
		(0.1365)	(0.3355)
Risk Detail$_{i,t}$ × Distance from Prior$_{i,t}$			0.0171
			(0.0590)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	901	898	898
R^2	0.45	0.45	0.45

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

F.5 Are Managers Conveying Opportunities to Time the Market?

I now consider the possibility that managers are conveying a buying opportunity in their letters, to which investors respond. In other words, it is possible that managers may be informing investors that expected returns are high, and investors are behaving accordingly.

To exclude this explanation, I discard every letter that contains the characters "opportun" (as in the words opportune and opportunity) or the phrase "chance to" from the sample. A tenth of letters are thus eliminated. I then repeat the paper's main regressions of the association between risk detail and the level of risk conveyed and fund flows. Table IA.15 displays the results of this analysis. The estimated coefficient on the amount of detail conveyed about risk remains positive, significant, and of a similar magnitude to the paper's main results (for example, Table 2 columns 1-4).

Table IA.15: Robustness of communication effect to excluding discussions of opportunities. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. The level of risk conveyed is measured based on the net sentiment of the text discussing risk, and is defined by Eqn. (8). The prior perceived level of risk is measured based on local Google searches, and is defined by Eqn. (23) The distance term is the absolute difference between the two, as defined by Eqn. (48). Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only, and only the letters that do not discuss an opportunity.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.3981***	0.4863***	0.4912***	0.5408***
	(0.1204)	(0.1634)	(0.1632)	(0.1794)
Risk Level$_{i,t}$	0.0078	0.0142	0.0175	-0.0265
	(0.1216)	(0.0957)	(0.0960)	(0.1110)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Market Controls$_t$			✓	
Year-month FEs	✓			✓
Fund FEs		✓	✓	✓
N	1,034	1,034	1,034	1,034
R^2	0.23	0.31	0.31	0.44

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

F.6 Does More Communication Reduce Perceived Ambiguity?

Perhaps increased communication about future risks acts to decrease investors' perceived ambiguity about the distribution of future returns? If this does not alter the perceived level of risk (or even increases it to some small extent) it is possible that the net effect of this communication will be to reduce the uncertainty that investors perceive concerning the unknown payoff of the asset. From the perspective of investors, this could then make the asset more desirable to hold and thus drive investor inflows into the index fund.

Measuring investors' perceived ambiguity about the payoff of the S&P 500 market portfolio is not a simple task, as it requires a measure that is independent of perceived risk and ambiguity aversion (i.e. the strength of any potential preference against ambiguity). Brenner and Izhakian (2018) propose one such measure of the perceived level of ambiguity, which they base on the expected volatility of probabilities across the relevant outcomes:

$$\mho^2 = \int \mathbb{E}[\varphi(r)]\,\mathbb{V}\mathrm{ar}[\varphi(r)]\,dr, \tag{49}$$

where $\varphi(\cdot)$ is a probability density function and r is a rate of return. As Brenner and Izhakian (2018, pp. 504) explain, "the intuition of $\mho^2$ is that, as the degree of risk can be measured by the volatility of returns, so too can the degree of ambiguity be measured by the volatility of the probabilities of returns."[27]

Brenner and Izhakian (2018, pp. 510-512) show how to estimate a monthly timeseries of the perceived ambiguity of the S&P 500 portfolio using high-frequency realized returns. In this appendix, I employ their measure to test the role played by ambiguity perceptions in investors' response to communication by fund managers.

Table IA.16 repeats the paper's main flow-versus-risk detail panel regressions, and now includes interactions with the level of perceived ambiguity, in aggregate. The effect of increased communication about risk on flows remains positive and significant, and is now attenuated by an increase in the level of perceived ambiguity: when ambiguity is perceived to be higher in the month prior to the arrival of the flow, the magnitude of the inflow is reduced.

This result is not consistent with increased communication acting to reduce the perceived level of ambiguity. If that were the case, one would expect communication to have a greater effect when investors need it the most – when ambiguity is the highest (i.e. positive interaction terms). At the very least, one could also argue that a similar effect for different levels of ambiguity (i.e. insignificant interaction terms) could also be consistent with ambiguity being reduced in all cases. However, the fact that both interaction terms are negative and significant,

[27] As a theoretical foundation, Brenner and Izhakian (2018) show that $\mho^2$ enters into the uncertainty premium of the risky asset if the investor makes her decisions under Expected Utility with Uncertain Probabilities.

indicating that communication is less effective when ambiguity is high, suggests an ambiguity-alleviating mechanism is unlikely to be at play.

Table IA.16: Aggregate ambiguity levels, and the differential effects of the level of detail communicated about risk on S&P 500 index fund flows. The amount of detail contained in the letter written by fund i at time t about risk is measured using the log of the number of words devoted to that topic. The indicator variable $\mathbb{1}\{\text{High Ambiguity}\}_t$ is 1 when the aggregate ambiguity perception of the market is above the median level, and zero otherwise; a baseline effect is estimated in addition to the interaction shown in the table. The variable Ambiguity_t is rank-standardized, and thus runs from 0 to 1; a baseline effect is estimated in addition to the interaction shown in the table. The level of perceived ambiguity is Brenner and Izhakian's (2018) $\mho^2$ measure, as defined in Eqn. (49). Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net $\text{Flow}_{i,t\rightarrow t+1}$ (%)		
	(1)	(2)	(3)
$\text{Risk Detail}_{i,t}$	0.5228***	0.7613***	0.6221***
	(0.1846)	(0.2430)	(0.2142)
$\text{Risk Detail}_{i,t} \times \text{Ambiguity}_t$		-0.5013**	
		(0.2379)	
$\text{Risk Detail}_{i,t} \times \mathbb{1}\{\text{High Ambiguity}\}_t$			-0.1938*
			(0.1138)
$\text{Fund Controls}_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	901	901	901
R^2	0.45	0.45	0.45

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

G Additional Analyses

G.1 Cross-Sectional Drivers of the Amount of Communication

Table IA.17 displays the cross-sectional relationships between the amount of detail provided about risk in fund letters (the primary independent variable used throughout the paper) and important fund-level and family-level attributes. When the total length of each letter is controlled for (column 2), the only other significant relation is the positive association with the size of the entire family to which a fund belongs. Since communicating effectively about risk incurs some cost, this empirical fact is consistent with a greater ease of bearing this cost across a larger fund family; for example, larger fund families are more likely to have greater legal, editorial and other administrative resources available internally.

Table IA.17: Cross-sectional determinants of the amount of detail provided about risk. This table shows the cross-sectional associations between the amount of communication about risk conveyed by a fund letter, and various fund, family and letter characteristics. The sample consists of S&P 500 index funds only.

Dependent Variable:	Risk word count (log)	
	(1)	(2)
Fund age (years)	0.0140*	6.8×10^{-5}
	(0.0080)	(0.0025)
Fund total fee (%)	0.0526	-0.1094
	(0.2280)	(0.1186)
Fund size (log)	0.0266	-0.0038
	(0.0577)	(0.0279)
Family age (years)	-0.0430	0.0007
	(0.0325)	(0.0133)
Family size (log)	0.1246*	0.0745***
	(0.0717)	(0.0268)
Total word count (log)		0.8966***
		(0.0392)
Year-month FEs	✓	✓
N	1,155	1,155
R^2	0.26	0.67

Clustered (Fund & Year-month) std. errs. in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

G.2 When Is Reassurance the Most Effective?

This appendix provides further evidence about which situations anxiety-alleviating communication is the most effective in. The cross-sectional results in Section 5.1 showed that, at any given point in time, anxious investors respond more strongly to the same amount of communication than less-anxious investors. The results in this appendix supplement the main paper's findings.

When the VIX is High I first turn to analyzing within-fund timeseries relationships. While the VIX index is nominally a measure of implied volatility, it is commonly referred to by financial media as a "fear index" or "fear gauge."[28] Kaplanski and Levy (2010) argue that the VIX captures investor anxiety in addition to a belief about short-term volatility. To the extent that the VIX is high when investors overall are anxious, one would therefore expect the effect of reassuring communication to be most pronounced during months when the VIX is high. This is precisely what is shown by the within-fund panel regression estimates in Table IA.18 columns (2) and (4): flows respond more strongly to the same amount of detail communicated about risk during periods when the VIX is above its median.

When Investors Are Less Sophisticated Next, I return to the cross-section and analyze the relationship between reassurance and investor sophistication. To measure the relative sophistication of S&P 500 index funds' investor bases, I calculate the 12b-1 distribution and marketing fees charged to investors by the fund. While not a perfect proxy for the funds' distribution methods, studies by Bergstresser, Chalmers, and Tufano (2008) and Del Guercio and Reuter (2014) find that these 12b-1 fees are higher for broker-sold funds than for direct-sold ones. And, as Del Guercio and Reuter (2014, pp. 1675) put it, "experienced and knowledgeable investors are likely to self select into direct-sold funds." The regression estimates in Table IA.19 column (2) show that investors who pay higher 12b-1 fees respond more strongly to the same amount of detail communicated about risk. To the extent that the level of 12b-1 fees charged is a measure of the sophistication of a funds' client base, this analysis shows that less sophisticated investors respond more strongly to reassuring communication about risk, although even the most sophisticated investors (who invest in funds that charge no 12b-1 fee at all) continue to respond positively and significantly.

[28] For example: `https://www.ft.com/content/5c840fbe-9949-43ab-9058-633023031582`

Table IA.18: Communication-driven flows and the level of the VIX. The amount of detail contained in the letter written by fund i at time t about risk is measured using the log of the number of words devoted to that topic. The indicator variable $\mathbb{1}\{\text{High VIX}\}_t$ is 1 during months when the level of the VIX index is above its long-run median; a baseline effect is estimated in addition to the interaction shown in the table. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. *Market controls* (not shown) are for the changes in the VIX and in the subjective aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)) from t to $t+1$; note these are not compatible with time effects. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.4747***	0.3492**	0.4754***	0.3492**
	(0.1529)	(0.1384)	(0.1528)	(0.1380)
Risk Detail$_{i,t}$ × $\mathbb{1}\{\text{High VIX}_t\}$		0.2335**		0.2301**
		(0.1150)		(0.1162)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Market Controls$_t$			✓	✓
Fund FEs	✓	✓	✓	✓
N	1,155	1,155	1,155	1,155
R^2	0.31	0.31	0.31	0.31

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.19: Communication-driven flows and 12b-1 marketing & distribution fees levied. The amount of detail contained in the letter written by fund i at time t about risk is measured using the log of the number of words devoted to that topic. The 12b-1 fee is a Total Net Asset-weighted average across the share classes of the fund, in basis points; a baseline effect is estimated in addition to the interaction shown in the table. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, total fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t \to t+1}$ (%)	
	(1)	(2)
Risk Detail$_{i,t}$	0.3893***	0.3259***
	(0.1238)	(0.1203)
Risk Detail$_{i,t}$ × 12b-1 Fee$_{i,t}$		0.0223**
		(0.0093)
Fund Controls$_{i,t}$	✓	✓
Year-month FEs	✓	✓
N	1,155	1,155
R^2	0.21	0.22

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

G.3 Do Investors Benefit By Bearing More Risk?

Money doctors should benefit the investors in their fund (Gennaioli, Shleifer, and Vishny, 2015), and therefore the influence of their communication upon investors – which is to encourage risk-taking – should be beneficial to these investors. I now confirm that this is indeed the case for a representative investor.

I begin by quantifying objective measures of the performance of the incremental capital that investors deploy to the stock market in response to more detailed communication about risk in fund letters: for each month t, I aggregate up all communication driven-inflows across all funds as a measure of the additional stock of capital deployed to the stock market. I then multiply these by the return of the S&P 500 in excess of the fees levied by the fund and the T-bill yield, for a particular holding period $t \rightarrow t + k$. Such a timeseries can be interpreted as the incremental excess return earned by the representative investor, net-of-fees, over and above the risk-free rate.[29]

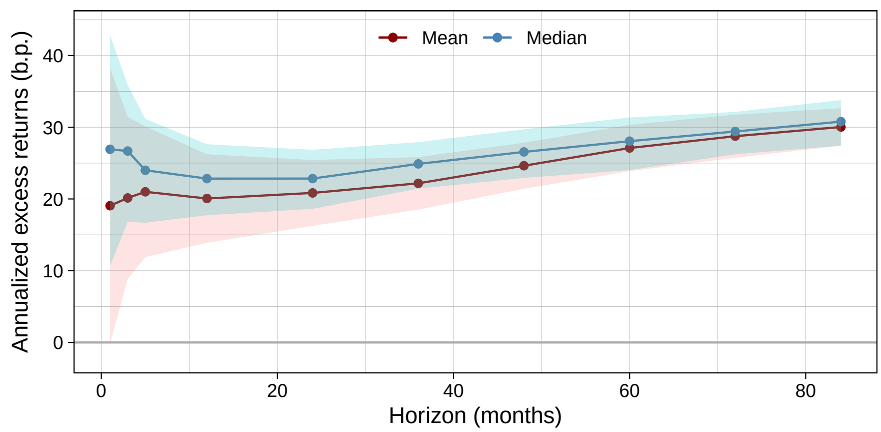
Figure IA.5(a) displays the unconditional mean and median of these incremental excess returns, for various holding periods k, each annualized for comparability. The average returns are positive, and the displayed 95% confidence bands confirm the quantities are significantly different to zero, irrespective of holding period.

Figure IA.5(b) displays the annualized Sharpe Ratios that summarize the same incremental excess returns. Once again, they are positive, and the 95% confidence bands confirm the quantities are significantly different to zero, irrespective of holding period.

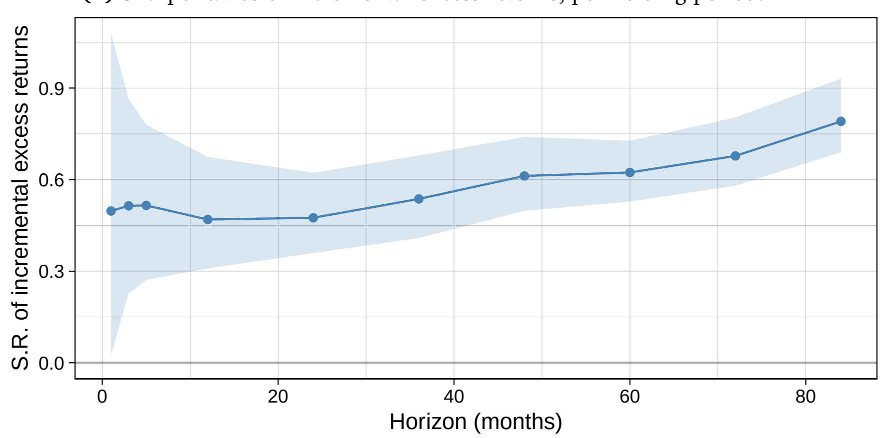
[29] This analysis assumes that holding cash is the alternative to bearing stock market risk. Such an assumption is consistent with the absence of within-equity rebalancing in Tables IA.29 and IA.30.

Figure IA.5: Performance of aggregate communication-driven incremental excess returns. Averages (sub-figure a) and Sharpe Ratios (sub-figure b) of the timeseries of the incremental excess returns earned by a representative investor who takes additional stock market risk in response to all fund-level communication about risk, for various holding periods. The excess returns are over and above the risk-free rate (T-bill yield) and net of the fees levied by the S&P 500 index mutual funds. All shaded areas represent 95% confidence intervals for the summary quantities, calculated using t-tests (for the means), Wilcoxon tests (for the medians) or quantile bootstraps with 1,000 replicates each (for the Sharpe Ratios).

(a) Average annualized incremental excess returns, per holding period.



(b) Sharpe Ratios of incremental excess returns, per holding period.



G.4 Further Results on Communication & Readership Anxiety

Table IA.20 recasts the split-sample regression of Table 7 to use interactions. Whether an indicator for above-median anxiety is used (first row) or a rank-percentile (third row), the interaction terms are positive and statistically significant, confirming that the effect of communication is concentrated among anxious investors. Furthermore, the baseline effect of anxiety (shown in the second and fourth rows) is to predict a decrease in risk-taking, consistent with the prior literature (for example, Guiso, Sapienza, and Zingales (2018)). Communication therefore overcomes anxious investors' inclination to take less risk.

Table IA.21 additionally includes a number of demographic controls for the fund's letter readership base, and find the interaction terms remain positive and statistically significant. The estimated coefficients are similar to those in Table IA.20.

Table IA.20: Readership anxiety levels interacted with the amount of detail about risk versus S&P 500 index fund flows. Risk Detail measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Risk Level measures the level of risk conveyed by the letter, calculated according to Eqn. (8). The indicator for $\mathbb{1}\{\text{High Anxiety}\}_{i,t}$, in columns 1-2, is one when fund i's readership base has an above-median anxiety level in the cross-section (i.e. for the given year-month t) and zero otherwise. The variable $\text{Anxiety}_{i,t}$ is the rank-percentile of the cross-sectional anxiety level of fund i's readership base for year-month t. The aggregate anxiety level of a fund's readership base is calculated using Google Trends data, as described in Section 5.1. *Local economy controls* (not shown) are for each fund letter readership's contemporaneous asset-weighted exposure to local economic activity & employment (measured by the Philadelphia Fed's State Coincident Index) and to local inflation rates (measured at a state level by Hazell, Herreno, Nakamura, and Steinsson (2022)). *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net $\text{Flow}_{i,t\to t+1}$ (%)			
	(1)	(2)	(3)	(4)
$\text{Risk Detail}_{i,t} \times \mathbb{1}\{\text{High Anxiety}\}_{i,t}$	0.3838***	0.3817***		
	(0.1205)	(0.1227)		
$\mathbb{1}\{\text{High Anxiety}\}_{i,t}$	-1.630***	-1.632***		
	(0.5863)	(0.5931)		
$\text{Risk Detail}_{i,t} \times \text{Anxiety}_{i,t}$			0.0059***	0.0058***
			(0.0021)	(0.0022)
$\text{Anxiety}_{i,t}$			-0.0233**	-0.0231**
			(0.0105)	(0.0106)
$\text{Risk Detail}_{i,t}$	0.2336*	0.2234*	0.1335	0.1273
	(0.1362)	(0.1316)	(0.1545)	(0.1529)
$\text{Risk Level}_{i,t}$		0.1130		0.1055
		(0.0794)		(0.0789)
Local Economy $\text{Controls}_{i,t}$	✓	✓	✓	✓
Fund $\text{Controls}_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	1,239	1,237	1,239	1,237
R^2	0.40	0.40	0.40	0.40

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.21: Readership anxiety levels interacted with the amount of detail about risk versus S&P 500 index fund flows, including demographic controls. The variable definitions, included controls, and sample used are identical to those of Table IA.20, with the additional inclusion of demographic controls. *Demographic controls* (not shown) are for the (log) geographic distance to the headquarters of the fund company, (log) regional median age of the population, regional fraction of the population with a Bachelor's degree, (log) regional mean household income, regional fraction of the voting population that voted for a Democrat president, regional fraction of the population living in an urban area; each fund-level demographic variable is a mean investment-weighted average of the fund letter readership of state-level geographic variables.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$ × $\mathbb{1}$\{High Anxiety\}$_{i,t}$	0.3792***	0.3772***		
	(0.1189)	(0.1208)		
$\mathbb{1}$\{High Anxiety\}$_{i,t}$	-1.639***	-1.645***		
	(0.5709)	(0.5762)		
Risk Detail$_{i,t}$ × Anxiety$_{i,t}$			0.0057***	0.0056**
			(0.0021)	(0.0022)
Anxiety$_{i,t}$			-0.0230**	-0.0228**
			(0.0103)	(0.0105)
Risk Detail$_{i,t}$	0.2399*	0.2303*	0.1447	0.1395
	(0.1367)	(0.1321)	(0.1581)	(0.1565)
Risk Level $_{i,t}$		0.1160		0.1081
		(0.0814)		(0.0810)
Demographic Controls$_{i,t}$	✓	✓	✓	✓
Local Economy Controls$_{i,t}$	✓	✓	✓	✓
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	1,236	1,234	1,236	1,234
R^2	0.40	0.40	0.40	0.40

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H Further Robustness Checks

This appendix contains extended and alternative specifications to verify the robustness of the paper's main analyses.

H.1 Shannon's Entropy as an Alternative Measure of Risk Detail

In the main paper, the amount of detail provided about risk in a fund letter is measured using the (log) number of words written about this topic. In this appendix, I test the following alternative measure of risk detail,

$$H_{i,t} = -\sum_k \mathbb{P}(\text{word}_k) \log_2 \mathbb{P}(\text{word}_k), \tag{50}$$

which is the Shannon's entropy of the discrete empirical distribution of the frequency of all k words that discuss risk, computed for each letter disseminated by fund i during month t.

Table IA.22 compares the association between flows and each measure of risk detail, with identical controls for each specification. The coefficient estimates in column (2) show that the Shannon's entropy of words describing risk in each letter positively and significantly predicts next-month flows. This analysis confirms that the information content of risk descriptions drives risk-taking.

Table IA.22: Measuring risk detail using Shannon's entropy. The covariates each measure the amount of detail contained in the letter written by fund i at time t about risk, either using the log of the number of words devoted to that topic (column 1) or the Shannon's entropy of the distribution of word frequencies discussing risk (column 2), as in Eqn. (50). Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)	
	(1)	(2)
Log word count$_{i,t}$	0.5044***	
	(0.1808)	
Shannon's entropy $H_{i,t}$		0.5203***
		(0.1890)
Fund Controls$_{i,t}$	✓	✓
Year-month FEs	✓	✓
Fund FEs	✓	✓
N	1,154	1,154
R^2	0.44	0.44

Clustered (Fund & Year-month) std. errs.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H.2 Persistence of Communication-Driven Flows

Table IA.23 presents the results of panel regressions of net flows over the next month, the next quarter, and the next five-month period (i.e. before the dissemination of the next fund semi-annual report) versus the level of detail communicated about risk, with the usual controls and fixed effects. The positive and significant association in each specification highlights the persistence of communication-induced fund flows over multiple horizons.

Table IA.23: Detail communicated about risk vs. net flows over various horizons into S&P 500 index funds. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only. Each flow variable is measured in percentage points of TNA.

Dep. Vars.:	Net Flow$_{i,t\rightarrow t+1}$ (1)	Net Flow$_{i,t\rightarrow t+3}$ (2)	Net Flow$_{i,t\rightarrow t+5}$ (3)
Detail$_{i,t}$	0.4955***	0.4107*	0.8200**
	(0.1834)	(0.2460)	(0.4016)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	1,130	1,130	1,130
R^2	0.43	0.50	0.53

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H.3 Past Flows

This appendix verifies that the paper's baseline results are not driven by any potential serial correlation in fund flows over time.

Table IA.24 augments the baseline specification with an additional control for the fund flow in the month *prior* to the dissemination of a fund letter. The positive and significant association between risk detail and *next* month's fund flow is robust to the inclusion of this additional control.

Table IA.25 presents the results of a placebo test. The dependent variables are the fund flows (net flow, gross inflow and gross outflow in columns 1-3, respectively) in the month immediately prior to the dissemination of a fund letter, rather than the month immediately following it. As expected, all estimated coefficients on the amount of communication about risk are not significantly different to zero.

Table IA.24: Including past fund flows as a control. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Specification (2) includes the past month's net flow as a control variable. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). *Common fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only. Each flow variable is measured in percentage points of TNA.

Dependent Variable:	Net Flow$_{i,t+1}$ (%)	
	(1)	(2)
Risk Detail$_{i,t}$	0.5044***	0.4953***
	(0.1808)	(0.1838)
Net Flow$_{i,t-1}$ (%)		0.0364
		(0.0751)
Common Fund Controls$_{i,t}$	✓	✓
Year-month FEs	✓	✓
Fund FEs	✓	✓
N	1,155	1,135
R^2	0.44	0.45

Clustered (Fund & Year-month) std. errs.
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.25: Placebo test on past flows. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. All dependent variables are fund flows (measured in percentage points of TNA) for the month *prior* to the dissemination of fund letters to existing investors. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). Gross dollar flows are extracted from mandatory SEC filing documents (N-SAR, N-PORT), and then normalized as a percentage of lagged total net assets. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variables:	Net Flow$_{i,t-1}$	Gross Inflow$_{i,t-1}$	Gross Outflow$_{i,t-1}$
	(1)	(2)	(3)
Risk Detail$_{i,t}$	0.0006	0.1999	0.0502
	(0.0959)	(0.1679)	(0.1371)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	1,135	1,135	1,135
R^2	0.44	0.75	0.82

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H.4 Controls for Risk and Past Returns

This appendix verifies the robustness of the paper's baseline results to various controls for fund risk exposures and past returns.

The cross-section of S&P 500 index mutual funds deliver near-identical risk exposures, by construction, and therefore the funds' delivered risk and returns are implicitly controlled for throughout the main paper. Nevertheless, I now explicitly add a control for risk exposures, which I measure using the CAPM β^{MKT} exposure of each fund, in columns 2-4 of Table IA.26. The economic magnitudes of the effect of communicated risk detail on flows remain very similar to the baseline estimate in column 1.

In Table IA.27, I vary the horizon of past returns (and their squares) included as controls, from the baseline of the past month's return (column 1) going up to the past year's return (column 4). Once again, the economic magnitudes of the effect of communicated risk detail on flows remain very similar to the baseline estimate in column 1.

Table IA.26: Controlling for fund risk exposure. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). *Common fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. In addition, specifications 2-4 include each fund's CAPM β^{MKT} as an explicit control for its risk exposure, estimated from monthly fund returns over a rolling window (10, 5, and 2 years, respectively). The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.5044***	0.5186***	0.5134**	0.5210***
	(0.1808)	(0.1952)	(0.1992)	(0.1961)
Estimation window for $\beta^{MKT}_{i,t}$ control	No control	10 years	5 years	2 years
Common Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	1,155	1,044	1,044	1,043
R^2	0.44	0.43	0.42	0.42

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.27: Controlling for various horizons of past fund return terms. The main covariate measures the amount of detail contained in the letter written by fund i at time t about risk, using the log of the number of words devoted to that topic. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (11). *Common fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's fees & (log) size, and for the overall fund family's (log) age and (log) size. In addition, each specification includes its own set of *past return controls*, consisting of the fund's return over the prior k months and the square of this term. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.5044***	0.5110***	0.5189***	0.5248***
	(0.1808)	(0.1778)	(0.1810)	(0.1816)
Horizon k for past return controls	1 month	3 months	6 months	12 months
Common Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	1,155	1,155	1,155	1,155
R^2	0.44	0.43	0.43	0.43

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H.5 Subsample Analysis When the Communicated Risk Level Is High

Table IA.28 shows the results of repeating the paper's baseline panel regression (column 1) on the subset of high-risk letters (columns 3 and 4). This analysis confirms that one of the major empirical results of the paper (greater risk detail attracting higher flows) is robust to high-risk states. This finding complements the earlier empirical analyses shown in Table 4 of the main paper.

Table IA.28: Adding the level of risk communicated to the amount of detail, and their joint relation with S&P 500 index fund flows. This table incorporates the level of risk conveyed by the discussion about risk into the baseline flow-detail panel regression. The amount of detail is measured as the log of the number of words devoted to discussion of risk. The level of risk is measured as the negative of the within-fund-standardized net sentiment of the text devoted to discussing risk, using the Loughran and McDonald (2011) sentiment dictionary; note that this measure covaries positively with the level of the VIX. Column (1) repeats the baseline flow-detail panel regression. Column (2) shows the relationship between net flows, communication detail and the level of risk inferred from the communicated information of the communication. Columns (3) and (4) repeat the baseline specification of column (1) on the sub-sample of fund letters that convey a high level of risk; a high-level of risk is measured as below-mean sentiment in column (3) and below-median sentiment in column (4). Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.5044***	0.4976***	0.4489*	0.7101**
	(0.1808)	(0.1768)	(0.2702)	(0.2948)
Risk Level$_{i,t}$		0.1043		
		(0.1123)		
Risk Level Sample	All	All	Above-Mean	Above-Median
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	1,155	1,137	570	629
R^2	0.44	0.43	0.63	0.56

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H.6 Further Tests to Rule Out Within-Family Rebalancing

Tables IA.29 and IA.30 split the baseline sample of active domestic equity funds into funds taking above-market and below-market risks, respectively, when testing whether within-family rebalancing can explain the paper's main effects. This classification is based funds' rolling CAPM β^{MKT} exposures.

Table IA.29: Testing for communication-driven rebalancing flows out of high-risk active funds. This table shows the relationship between outflows from active domestic equity funds and the amount of detail about risk contained in the letters of their sibling S&P index tracker funds (columns 1-2) or in their own letters (columns 3-4). The amount of detail is measured as the log of the number of words devoted to forward-looking risk-related statements. Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of active domestic equity funds with CAPM $\beta^{MKT} > 1$ and that belong to fund families that also contain an S&P 500 index tracker fund.

Dependent Variable:	Sibling Outflow$_{i,t\to t+1}$ (%)			
	(1)	(2)	(3)	(4)
S&P 500 Index Fund Detail$_{i,t}$	-0.0240	-0.1138		
	(0.2758)	(0.1245)		
Sibling Active Fund Detail$_{i,t}$			-0.0247	-0.1770
			(0.2645)	(0.1303)
Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓		✓	
Fund FEs		✓		✓
N	4,502	4,502	4,502	4,502
R^2	0.40	0.79	0.41	0.79

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.30: Testing for communication-driven rebalancing flows out of low-risk active funds. This table shows the relationship between outflows from active domestic equity funds and the amount of detail about risk contained in the letters of their sibling S&P index tracker funds (columns 1-2) or in their own letters (columns 3-4). The amount of detail is measured as the log of the number of words devoted to forward-looking risk-related statements. Controls (not shown) are for each fund letter's total (log) word count, for the fund's prior month return, square of this return, fees & (log) size, and for the overall fund family's (log) age and (log) size. The sample consists of active domestic equity funds with CAPM $\beta^{MKT} < 1$ and that belong to fund families that also contain an S&P 500 index tracker fund.

Dependent Variable:	Sibling Outflow$_{i,t\rightarrow t+1}$ (%)			
	(1)	(2)	(3)	(4)
S&P 500 Index Fund Detail$_{i,t}$	-0.1756	-0.1053		
	(0.1858)	(0.1247)		
Sibling Active Fund Detail$_{i,t}$			-0.1019	-0.0854
			(0.2061)	(0.1135)
Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓		✓	
Fund FEs		✓		✓
N	4,621	4,621	4,621	4,621
R^2	0.38	0.81	0.37	0.81

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H.7 Fund-Level Timeseries Variation in Communication

Table IA.31 analyzes timeseries variation at an individual fund level, rather than at an index level, as in the main text. The VIX continues to be positively and significantly associated with fund-level communication measures.

Table IA.31: Timeseries variation in communication about risk contained in S&P 500 index fund letters. This table displays the results of fund-level panel timeseries regressions of the amount of detail communicated about risk (using the log of the number of words devoted to that topic), and the level of risk conveyed (using the sentiment-based measure of the level). The sample consists of S&P 500 index funds only.

Dependent Variables:	Risk Detail$_{i,t}$	Risk Level$_{i,t}$
	(1)	(2)
Log VIX$_t$	0.1905*	0.2297**
	(0.1005)	(0.1156)
Fund FEs	✓	✓
N	1,154	1,154
R^2	0.64	0.07

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H.8 Measuring Information Spillovers Over an Extended Sample

Table IA.32 extends the sample of fund family cooperation indices and verifies the baseline results on information spillovers continue to hold.

Table IA.32: Information spillovers within fund families, using forward-filled cooperation index values. The regressions in this table document a high correlation between the amount of detail about risk contained in S&P 500 index fund letters (dependent variable) written by fund i at time t, and the corresponding amount of detail across active domestic equity sibling funds (only) within the same fund family, aggregated as a simple mean. *Fund controls* (not shown) are for each fund letter's total (log) word count, for the fund's prior month returns, fees & (log) size, and for the overall fund family's (log) age and (log) size. Columns (1) & (3) show the unconditional strength of this relationship, across different fixed effects specifications. Columns (2) & (4) additionally include an interaction with an index for the strength of cooperation between sibling funds in the focal fund's overall family, as measured by Evans, Prado, and Zambrana (2020) – since their dataset ends in December 2015, the last observation for each family has been forward-filled to the end of my sample, where possible.

Dependent Variable:	Detail$_{i,t}$			
	(1)	(2)	(3)	(4)
Sibling Mean Detail$_{i,t}$	0.8495***	0.5969***	0.8124***	0.6595***
	(0.0335)	(0.0740)	(0.0407)	(0.0926)
Sibling Mean Detail$_{i,t}$		0.3951***		0.3320**
× Family Cooperation Index$_{i,t}$		(0.1101)		(0.1456)
Family Cooperation Index$_{i,t}$		-2.099***		-1.540*
		(0.6596)		(0.7864)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓		
Fund FEs			✓	✓
N	1,155	1,083	1,155	1,083
R^2	0.94	0.93	0.95	0.93
F-stat	253.4	103.2	294.8	113.1

Clustered (Fund & Year-month) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

I Regular Expressions to Replicate Textual Measures

This appendix provides the regular expressions used to detect fund letters and to classify individual statements. I have two goals in doing so. First, I aim to facilitate the replication and extension of this work. Second, I hope to illustrate the simplicity and transparency of this approach, which requires no estimation.

Internet Appendix B.1 validates these manually-defined measures by comparing them to other approaches.

I.1 Extracting Fund Letters From Reports

I provide two regular expressions that can be applied to the extracted text of N-CSR(S) shareholder reports. The separate regular expressions may detect overlapping text (for example if one is subsumed by the other in the report), so the detected text output of each should be merged appropriately.

The first regular expression detects letter portions that are formatted as fund letters:

```
’(?:(?:Dear|Fellow|To ?(?:a|A)ll|To ?(?:a|A)ll ?(?:o|O)ur|To ?(?:o|O)ur)(?:
 ?(?:m|M)ost)?(?: ?(?:(?:a|A)ppreciated|(?:c|C)herished|(?:f|F)ellow|(?:g|G
)reat(?:est)?|(?:h|H)onored|(?:e|E)steemed|(?:p|P)rized|(?:r|R)espected|(?:
v|V)alued))? ?(?:(?:c|C)lient|(?:c|C)ontributor|(?:c|C)ustomer|(?:i|I)
nvestor|(?:o|O)wner|(?:p|P)artner|(?:s|S)hareholder|(?:s|S)tockholder|(?:s|
S)ubscriber|(?:s|S)ponsor|(?:u|U)nit.?(?:h|H)older)s?(?:,|\\:|\\n)?|(?:DEAR
|FELLOW|TO ?ALL|TO ?ALL ?OUR|TO ?OUR)(?: ?MOST)?(?: ?(?:APPRECIATED|
CHERISHED|FELLOW|GREAT(?:EST)?|HONORED|ESTEEMED|PRIZED|RESPECTED|VALUED))?
 ?(?:CLIENT|CONTRIBUTOR|CUSTOMER|INVESTOR|OWNER|PARTNER|SHAREHOLDER|
STOCKHOLDER|SUBSCRIBER|SPONSOR|UNIT.?HOLDER)S?(?:,|\\:|\\n)?|(?:Dear|Fellow
|To)(?:.{0,100})(?:(?:c|C)lient|(?:c|C)ontributor|(?:c|C)ustomer|(?:i|I)
nvestor|(?:o|O)wner|(?:p|P)artner|(?:s|S)hareholder|(?:s|S)tockholder|(?:s|
S)ubscriber|(?:s|S)ponsor|(?:u|U)nit.?(?:h|H)older)s?(?:,|\\:|\\n)|(?:DEAR|
FELLOW|TO)(?:.{0,100})(?:CLIENT|CONTRIBUTOR|CUSTOMER|INVESTOR|OWNER|PARTNER
|SHAREHOLDER|STOCKHOLDER|SUBSCRIBER|SPONSOR|UNIT.?HOLDER)S?(?:,|\\:|\\n)|
Welcome,|LETTER|Letter from|PHOTO|Photo|MESSAGE|Message from|CHATTER|
Chatter)(.*?)(?:(?:(?:Yours ?(?:(?:c|C)ordially|(?:f|F)aithfully|(?:r|R)
espectfully|(?:s|S)incerely|(?:t|T)ruly))|(?:(?:In|With) ?(?:(?:a|A)ll ?)
?(?:(?:o|O)ur ?)?(?:(?:a|A)ppreciation|(?:g|G)ratitude|(?:p|P)artnership
|(?:r|R)espect))|(?:(?:(?:b|B)est|(?:k|K)ind|(?:w|W)arm)(?:est)? ?(?:(?:r|R
)egards|(?:w|W)ishes))|Many ?(?:t|T)hanks|Stay ?(?:(?:s|S)afe|(?:h|H)ealthy
)|Until ?(?:n|N)ext ?(?:t|T)ime|(?:(?:(?:Very ?)?(?:(?:a|A)ppreciatively
|(?:c|C)ordially|(?:f|F)aithfully|(?:r|R)espectfully|(?:t|T)ruly|(?:s|S)
incerely))(?: ?(?:(?:(?:y|Y)ours)|(?:(?:s|S)ubmitted)))?)(?:,|\\n))|(?:(?:
```

```
Best|Yours|Truly|Sincerely|Respectfully|Cordially)(?:,|\\n))|CFA|PhD|Chief|
CHIEF|Officer|OFFICER|Founder|FOUNDER|PRESIDENT|CHAIRMAN|Head ?of|HEAD ?OF|
Chairman ?of|CHAIRMAN OF|Investor ?Relations|INVESTOR ?RELATIONS|Managing ?
Director|MANAGING ?DIRECTOR|Managing ?Partner|MANAGING ?PARTNER|Registered
?(?:r|R)epresentative|REGISTERED ?REPRESENTATIVE|UNDERSTANDING ?YOUR|
ANALYZING ?YOUR|VALUE ?OF|TOTAL ?RETURNS|FEE|SUMMARY|SCHEDULE|STATEMENT|
INVESTING|BENCHMARK|COMPARISON|DISCLOSURE|(?:p|P)rospectus|PROSPECTUS|
Important ?(?:i|I)nformation|IMPORTANT ?INFORMATION|Past ?performance|PAST
?PERFORMANCE|does not constitute)'
```

The second regular expression detects letter portions that are formatted as a section in the shareholder report:

```
'(?:(?:manage(?:ment|r).?s?.? ?)?discussion ?(?:(?:and ?|\\& ?|\\+ ?)
analysis ?)?of ?(?:(?:fund|portfolio) ?)?performance|manage(?:ment|r).?s?.?
 ?discussion ?(?:and ?|\\& ?|\\+ ?)analysis|(?:(?:fund|portfolio) ?)?
performance ?(?:review|discussion|analysis)|m ?a ?n ?a ?g ?e ?(?:m ?e ?n ?t
|r ?).{0,10}(?:a ?n ?a ?l ?y ?s ?i ?s|d ?i ?s ?c ?u ?s ?s ?i ?o ?n|r ?e ?v
?i ?e ?w)|(?:(?:fund|portfolio) ?)?(?:management|manager|managers|manager.s
|managers.) ?(?:analysis|discussion|review)|(?:questions? ?(?:and|\\&|\\+)
?answers?|q ?(?:and|\\&|\\+) ?a|interviews?)(?: ?with)?|markets? ?(?:
commentar|perspective|outlook|view)|(?:portfolio|fund)? ?:((?:management|
manager|managers|manager.s|managers.) ?)?(?:commentar|perspective|outlook|
view))(.*?)(?:(?:(?:Yours ?(?:(?:c|C)ordially|(?:f|F)aithfully|(?:r|R)
espectfully|(?:s|S)incerely|(?:t|T)ruly))|(?:(?:In|With) ?(?:(?:a|A)ll ?)
?(?:(?:o|O)ur ?)?(?:(?:a|A)ppreciation|(?:g|G)ratitude|(?:p|P)artnership
|(?:r|R)espect))|(?:(?:(?:b|B)est|(?:k|K)ind|(?:w|W)arm)(?:est)? ?(?:(?:r|R
)egards|(?:w|W)ishes))|Many ?(?:t|T)hanks|Stay ?(?:(?:s|S)afe|(?:h|H)ealthy
)|Until ?(?:n|N)ext ?(?:t|T)ime|(?:(?:(?:Very ?)?(?:(?:a|A)ppreciatively
|(?:c|C)ordially|(?:f|F)aithfully|(?:r|R)espectfully|(?:t|T)ruly|(?:s|S)
incerely))(?: ?(?:(?:(?:y|Y)ours)|(?:(?:s|S)ubmitted)))?)(?:,|\\n))|(?:(?:
Best|Yours|Truly|Sincerely|Respectfully|Cordially)(?:,|\\n))|u ?n ?d ?e ?r
?s ?t ?a ?n ?d ?i ?n ?g ? ?y ?o ?u ?r|c ?o ?m ?p ?a ?r ?i ?n ?g ? ?y ?o ?u
?r|p ?o ?r ?t ?f ?o ?l ?i ?o ? ?s ?u ?m ?m ?a ?r|p ?e ?r ?f ?o ?r ?m ?a ?n
?c ?e ? ?s ?u ?m ?m ?a ?r|s ?c ?h ?e ?d ?u ?l ?e ? ?o ?f|f ?e ?e ? ?s ?c ?h
 ?e ?d ?u ?l ?e|s ?t ?a ?t ?e ?m ?e ?n ?t ? ?o ?f|f ?e ?e ? ?s ?t ?a ?t ?e
?m ?e ?n ?t|v ?a ?l ?u ?e ? ?o ?f ? ?p ?o ?r ?t ?f ?o ?l ?i ?o|t ?o ?t ?a ?
l ? ?r ?e ?t ?u ?r ?n ?s|e ?x ?p ?e ?n ?s ?e ?s ? ?p ?a ?i ?d|c ?o ?m ?p ?a
 ?r ?i ?s ?o ?n ? ?o ?f|u ?n ?d ?e ?r ?s ?t ?a ?n ?d ?i ?n ?g ? ?y ?o ?u ?r
|a ?n ?a ?l ?y ?z ?i ?n ?g ? ?y ?o ?u ?r|f ?u ?n ?d ? ?e ?x ?p ?e ?n ?s ?e|
g ?l ?o ?s ?s ?a ?r ?y|n ?o ?t ?e ?s|d ?i ?s ?c ?l ?o ?s ?u ?r ?e|d ?o ?e ?
s ? ?n ?o ?t ? ?c ?o ?n ?s ?t ?i ?t ?u ?t ?e|p ?l ?e ?a ?s ?e ? ?n ?o ?t ?e
```

```
|i ?m ?p ?o ?r ?t ?a ?n ?t ? ?i ?n ?f ?o ?r ?m ?a ?t ?i ?o ?n|f ?o ?r ?e ?g
 ?o ?i ?n ?g ? ?i ?n ?f ?o ?r ?m ?a ?t ?i ?o ?n|integral ?part|return ?
includes|registered ?under|past ?performance|investment ?company ?act
|----|====)'
```

I.2 Classifying Sentences Within Fund Letters

The following regular expressions should be applied to individual sentences. Sentences are defined as grammatically-valid phrases, as detected using the open-source StanfordNLP toolkit (Manning et al., 2014).

Detecting Boilerplate Sentences

```
'40 act|\\.com|accompanying (note|statement)|account(ed|ing)|accrual|(
additional|detailed|further) information|(advised to|please) (contact|
consult|email|visit)|admin|adhere to|agent|board|officer|treasurer|trustee|
amortiz|(after|before|pre|post)-?tax|appendix|arrear|audit|accordance with|
accretion|additional charge|advisory|affiliat|aggregate cost|amount is|
annualize|arrangement|as follows|attached|benchmark|(beyond|outside) our
control|board considered|board of directors|books|calculated (based|by|on|
using)|capital contribution|cash equivalent|charged to|class shares are (
offered|sold) at|(cease|commence)d?( investment)? (operation|trading)|
collateral|committee|compliance|consolidated|contained herein|(contact|
consult|email|visit) y?our|contingent|contractual|control over|constitute|
correspondence|course of business|custodian|deferred|deem|denominated|
depreciation|derivative|determined in accordance|differ|disclosure|
distribut|effective rate|enter(ing)? into|examples? (above|below|is|are)|
exceeds the value|exempt from|exercise price|fair valu|filed|(feeder|master
) (fund|portfolio)|for example|foregoing|gaap|(gross|net|total) change in|
if any|in determining the|index is (a |an |comprised|based|designed|managed
|offers|unmanaged|provides)|include.? but are not limited to|illiquid asset
|in writing|incur|independent (consultant|director)|indemni(t|f)|
information (is provided on an|is not|on|presented|reflects|shown)|integral
 part of|input|invest up to|investment company act|involve risk|is
calculated|is shown in|lease|left blank|liabilit|lipper|may buy and sell|
may differ from|may (include|involve|result)|(may|should) not be (
considered|construed|relied|depended|taken)|may (also )?(invest in|hold)|
maximum (exposure|commitment)|means? the?|meaningful|methodology|
morningstar|must be (accompanied|preceded)|no significant change|not (
eliminate|indicative|intended|meant|reflected|warranted)|not? (assurance|
guarantee)s? (of|that|the)|not be relied|notes to|obligat|offered at|
```

```
organized as|payable|potential for gain|proprietary to|principal (loss|risk
)|purchases? (and|or) sale|principally traded|recognized on|records|redeem|
redemption|refers to|registered under|reimburse|repurchase|representative
of|required to|represents the|respectively|responsibility of|restricted
security|restrictions on|rights reserved|risk of loss|rounds to|schedules?
of|sales (charge|expense)|see (accompanying|note|the|page)|seeks to|settle|
share class|stated as|statements? of|structured as|subject to|(such|these|
valuation) models|tax.exempt|tax (advantage|benefit|expense|jurisdiction|
position|purpose|return)|thank you|the adviser|
the funds? (do|did|had no|have no|invest in|may|offer|value)|total
investment|transfer|unanticipated change|under (normal|typical) (circum|
condition|course of|market)|underlying security|up.to.date|valued (using|
based|by)|views expressed|waiver|we welcome your|when applicable|will seek
to|www\\.'
```

Detecting Risk-Related Sentences

```
'alarm|alert|capitulat|catastroph|correction|choppy|collaps|crash|crisis|
danger|disaster|disorder|expose|forebod|hang over|impend|insecurity|
instability|jeopard|hazard|menace|overhang|precarious|rare event|risk|
uncertain|unpredictable|unstable|storm|stress|squall|systemic|threat|
tempest|trouble|tumult|turbulen|uneas|warn|wary|wild|worr|volatil|\\b(loom|
rough|tail)\\b|afraid|anxiety|anxious|apprehensi|caution|cautious|
consternation|daunt|dismay|distress|doom|fear|fright|gloom|nervous|rattle|
shake|shook|terror|trepid|defended|defenceless|immune to|in the firing line
|guarded|mercy of|open to attack|prey to|powerless|prone to|protected|safe|
sensitive to|shielded|vulnerable|wide open to|(black|white) swan|peso
problem'
```

Detecting Forward-Looking Sentences

```
'anticipat|belie(f|ve)|conceiv|conjecture|consider|diagnos|expect|foresee|
estimat|guess|hope|interpret|look|plan|project|predict|think|\\b(develop|
unfold|regard|see|view)\\b|we (continue|remain|favou?r)|horizon|timeline|(
short|medium|near|long).term|(over the|next|coming|forthcoming|future|
subsequent) (period|quarter|year|month|meeting|date|release)s?|ahead|
attitude|forward|future|going|moving|opinion|optimis|outlook|perspectiv|
pessimis|prognosis|realis|stance|chance|danger|indicat|hazard|likelihood|
risk|possib|probab|prospect|signal|spell|threat|warn|(can|cannot|can't|will
|won't|not|should|shouldn't|would|wouldn't|to) (affect|arise|change|cascade
|cause|climb|come (about|off|to)|continue|crop up|(de|in)crease|(de|en)
```

```
velop|enlarge|emerge|ensue|expand|fall|feed|follow|guide|grow|happen|hold|
lead|lengthen|materiali|modify|occur|pan out|release|remain|pass|persist|
push|raise|rise|spill|stretch|surface|take place|transpire|turn out|widen|
update)|(ready|(position|prepar)(ed|ing)) (for|to)|bullish|bearish|(over|
under).?valu|cheap|expensive|richly'
```

Detecting Performance- and Return-Related Sentences

```
'performance|to.date|\\bytd\\b|((our|the (etf|fund|portfolio|trust|share
class)).*?(achieve|accrue|affected|assisted|cumulat|earned|experienced|gain
|fail|fell|helped|hit|hurt|impacted|loss|lost|perform|realize|return|showed
|success|succeed|strong|support|weak|win|won))'
```

References in Internet Appendix

Bailey, W., Kumar, A., & Ng, D. (2011). Behavioral biases of mutual fund investors. *Journal of Financial Economics*, *102*(1), 1–27.

Bergstresser, D., Chalmers, J. M., & Tufano, P. (2008). Assessing the costs and benefits of brokers in the mutual fund industry. *Review of Financial Studies*, *22*(10), 4129–4156.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*.

Brenner, M., & Izhakian, Y. (2018). Asset pricing and ambiguity: Empirical evidence. *Journal of Financial Economics*, *130*(3), 503–531.

Caetano, C., Caetano, G., Nielsen, E., & Sanfelice, V. (2023). The effect of maternal labor supply on children: Evidence from bunching.

Caetano, C., Caetano, G., & Nielsen, E. R. (forthcoming). Correcting for endogeneity in models with bunching. *Journal of Business & Economic Statistics*. https://doi.org/10.1080/07350015.2023.2252471

Célérier, C., & Vallée, B. (2017). Catering to investors through security design: Headline rate and complexity. *Quarterly Journal of Economics*, *132*(3), 1469–1508.

Cohn, J. B., Liu, Z., & Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, *146*(2), 529–551.

Del Guercio, D., & Reuter, J. (2014). Mutual fund performance and the incentive to generate alpha. *Journal of Finance*, *69*(4), 1673–1704.

Evans, R. B., Prado, M. P., & Zambrana, R. (2020). Competition and cooperation in mutual fund families. *Journal of Financial Economics*, *136*(1), 168–188.

Gennaioli, N., Shleifer, A., & Vishny, R. (2015). Money doctors. *Journal of Finance*, *70*(1), 91–114.

Giannetti, M., & Wang, T. Y. (2016). Corporate scandals and household stock market participation. *Journal of Finance*, *71*(6), 2591–2636.

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

Grice, R., & Guecioueur, A. (2023). Mutual fund market structure and company fee competition: Theory and evidence. *Available at SSRN*, *4449026*.

Guiso, L., Sapienza, P., & Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, *128*(3), 403–421.

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics*, *134*(4), 2135–2202.

Hazell, J., Herreno, J., Nakamura, E., & Steinsson, J. (2022). The slope of the Phillips Curve: Evidence from US states. *Quarterly Journal of Economics*, *137*(3), 1299–1344.

Hillert, A., Niessen-Ruenzi, A., & Ruenzi, S. (2021). Mutual fund shareholder letters: Flows, performance, and managerial behavior. *Available at SSRN*, *2524610*.

Kaplanski, G., & Levy, H. (2010). Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics*, *95*(2), 174–201.

Kim, D., Wang, Q., & Wang, X. (2022). Geographic clustering of institutional investors. *Journal of Financial Economics*, *144*(2), 547–570.

Lee, D. S., McCrary, J., Moreira, M. J., & Porter, J. (2022). Valid t-ratio inference for iv. *American Economic Review*, *112*(10), 3260–3290.

Li, F. (2010). The information content of forward-looking statements in corporate filings — A naïve Bayesian machine learning approach. *Journal of Accounting Research*, *48*(5), 1049–1102.

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

Lusardi, A., Bumcrot, C. B., & Lin, J. (2013). The geography of financial literacy. *Numeracy*, *6*.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Association for Computational Linguistics (ACL) System Demonstrations*, 55–60.

Mullainathan, S., & Shleifer, A. (2005). Persuasion in finance. *National Bureau of Economic Research Working Paper*, *11838*.

Vokata, P. (2023). Salient attributes and household demand for security designs. *Fisher College of Business Working Paper*, (2023-03-007).