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

Ahmed Guecioueur*

September 10, 2025

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

Abstract

I show that communication by asset managers can foster trust and encourage their investor clients to bear greater risk. Specifically, voluntary transparency about risk in discretionary investor letters leads to higher fund flows. To establish causality, I focus on index funds and exploit the presence of corner bunching using a control function approach. Various channels including learning, shrouding, and marketing cannot explain this result. Instead, the evidence supports a trust-building mechanism in which voluntary transparency about risk reduces investors' effective risk aversion, consistent with the "money doctors" theory proposed by Gennaioli, Shleifer, and Vishny (2015).

Keywords: Fund Flows, Trust, Risk Aversion, Persuasion, Text Data

JEL Codes: D14, D83, D91, G11, G20, G41

*University of Washington, Foster School of Business, email: ahmedgc@uw.edu

First version: June 2023. I would like to thank the participants at various seminars and conferences, including NBER Behavioral Finance, WFA and FIRS, for their helpful feedback. I am grateful to my doctoral advisor, Joël Peress, and my dissertation committee for their invaluable guidance, as well as to many others who have offered helpful comments, including (and not limited to) Pat Akey, Bronson Argyle, Carolina Caetano, Gavin Cassar, Aditya Chaudhry, Caitlin D. Dannhauser, Tilman Fries, Nicola Gennaioli, Sergei Glebkin, Markus Ibert, Mark Kamstra, Alessandro Previtero, Vesa Pursiainen, Paola Sapienza, Yang Song, Boris Vallée, Petra Vokata and Scott Yonker. Finally, I am grateful to Henu Park for valuable research assistance, and to Yehuda Izhakian for sharing data. An earlier version of this paper was circulated under the title "Money Doctors and Their Prognoses."

1 Introduction

Investing is inherently risky, and many individuals seem reluctant to bear the risk that is necessary for building financial wealth.[1] To the extent that behavioral biases are responsible for this hesitation, existing remedies to encourage risk-taking center on altering investors' choice sets (e.g., Thaler and Sunstein, 2008; Calvet, Célérier, Sodini, and Vallée, 2023). Yet a vast asset management industry often intermediates investors' holdings. Can financial intermediaries influence investors' attitudes toward risk-taking?

I show that communication by financial intermediaries can foster trust and thereby encourage risk-taking. As Gennaioli, Shleifer, and Vishny (2015) argue, asset managers play the important role of granting their investors "peace of mind" and the trust and confidence to overcome their anxieties and take risks, including through communication.[2] In this context, trust-building reduces investors' effective risk aversion. I demonstrate this mechanism by examining the effect of letters that are written voluntarily by asset managers to their existing clients – a widespread industry practice (Cassar, Gerakos, Green, Hand, and Neal, 2018; Hillert, Niessen-Ruenzi, and Ruenzi, 2025). Letters that foster trust and reduce investors' effective risk aversion should produce inflows as investors become more willing to take risks, offering a natural laboratory to test for the trust-building effect of communication.

My key hypothesis is that voluntary transparency about risk builds trust with investors. The logic is straightforward: when asset managers openly discuss the key attribute that investors tend to dislike – risk itself – this transparency should foster trust.[3] This yields a testable prediction: that letters with more detailed risk discussions should generate larger fund inflows in response, as investors' effective risk aversion is dampened. Furthermore, this hypothesis sharply distinguishes between the roles of trust and learning: if communication simply conveys information about risk exposures, then more detailed discussions should *reduce* or reverse fund flows (as investors will perceive or notice a higher level of risk). But if transparency has

[1] Stock market investors display a low intensive margin of risk-taking: 40% of all stock market participants hold less than 20% of their financial wealth in stocks (Calvet, Célérier, Sodini, and Vallée, 2023). In contrast, realistic lifecycle models prescribe a much higher share of financial wealth be held in the risky asset (e.g., Cocco, Gomes, and Maenhout, 2005). The current paper focuses on investment in diversified large-cap equity portfolios such as the S&P 500 as the risky asset, which is generally considered beneficial for households.

[2] Trust is known to encourage stock market participation (Guiso, Sapienza, and Zingales, 2008; Giannetti and Wang, 2016) and may even be a prerequisite for economic development (Arrow, 1972; Fukuyama, 1995). Prior work shows investors' trust can be broken (Kostovetsky, 2016; Giannetti and Wang, 2016; Gurun, Stoffman, and Yonker, 2018), leading to decreases in risk-taking. I demonstrate how asset managers can build trust.

[3] My study focuses on voluntary, free-form communication – not mandatory risk disclosures contained in prospectuses (or other filings). Providing information about future *performance* is discouraged under regulations relating to the Investment Company Act (Rule 34b-1, 17 C.F.R. § 270.34b-1) and the Securities Act (Rule 156, 17 C.F.R. § 230.156). Communication about *risk* is neither discouraged nor mandated in the letters that I examine.

the effect of building trust, then greater risk detail should instead *increase* flows into the fund.

I show that letters with a more detailed voluntary discussion of risks do indeed encourage higher fund flows, consistent with a trust-building role for communication. I isolate the effect of this communication by focusing on asset managers holding the same portfolio; namely, S&P 500 index mutual funds that deliver an identical exposure to the stock market, yet whose letters to investors vary in content. I find that letters containing more detailed forward-looking information about risks – in the sense of dedicating a larger proportion of the letter to discussing this topic – are followed by a higher fund flow immediately after they are disseminated to investors, even though the actual risk exposure of the underlying portfolio is the same.[4] Furthermore, the reaction of investors is not explained by whether the letter conveys a high or low level of risk.

The effect of voluntary transparency on flows is economically significant: a one standard deviation (SD) growth in the amount of detail voluntarily communicated about risk predicts a 0.09–0.17 SD growth in assets under management during the subsequent month. This is equivalent to an additional 0.28–0.50 percentage points of assets, on average, and up to one-third of the magnitude of the well-known association between fees and flows.

Pinning down the effect of communication can often be challenging, but my focus on quasi-identical S&P 500 index funds addresses many potential concerns. To begin with, the passive and standardized nature of these index funds rules out the possibility that investors react to communication because they are learning about private information, such as funds' investment strategies or their managers' skill. Furthermore, any differences in the funds' portfolios, risk and returns are minimal, implicitly controlling for these potential drivers of flows. I also explicitly control for observable drivers of fund flows highlighted by prior research, most notably past performance, fees and other fund and family characteristics. Importantly, I incorporate time effects, which control for investors' prior beliefs about risk and return, 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 persistent differences between funds, the families they belong to, and their investor bases.

To further mitigate concerns that other variables might be driving my results – such as unobservable marketing activities or other content within the fund letters – I employ an additional

[4]By linking *changes* in risk-taking (i.e., flows) to changes in trust, I follow a common approach in the literature (Kostovetsky, 2016; Giannetti and Wang, 2016; Gurun, Stoffman, and Yonker, 2018) that sidesteps the difficult empirical problem of measuring *levels* of trust and risk-taking for specific investor-asset manager pairs.

identification strategy that exploits a natural feature of my treatment variable: the distribution of the fraction of a letter discussing risk exhibits corner bunching at the lower end of the support. This pattern enables me to adjust for selection bias using a control function methodology developed by Caetano, Caetano, and Nielsen (2024) specifically for settings with corner bunching. I am thus able to recover a causal effect from a naturally occurring feature of the data. When correcting for potential endogeneity, the relationship between risk detail and flows remains significant and becomes larger in magnitude; estimated coefficients on the correction term are also consistent with a negative selection bias in the baseline estimate. This analysis strengthens the causal interpretation that voluntary transparency increases fund flows.

The positive relationship between risk detail and fund flows might seem puzzling if one's prior is that investors should be learning about portfolio risk from fund letters. However, the direction of the effect aligns well with a trust-based mechanism: transparency that fosters trust should encourage willingness to bear risk. Another fact that points to a role for trust is that investor bases who are more dependent on financial intermediaries (as revealed by payment of higher 12b-1 marketing and distribution fees) show a greater sensitivity to this effect and respond more strongly to the same amount of transparency. Furthermore, investors are rewarded for their trust, earning a positive and significant Sharpe ratio, net of fund fees, for holding periods greater than a month.

I next provide evidence that trust in my setting acts through the economic mechanism of shifting investors' risk aversion. I arrive at this conclusion by first deriving a parsimonious test for this channel based on portfolio choice theory, which predicts that trust-building effects should be stronger when investors' prior risk aversion is higher. After formalizing this prediction, I use an empirical measure of the representative investor's relative risk aversion coefficient by Bekaert, Engstrom, and Xu (2022) to implement the analysis. I find that in market states where investors are more risk averse, the relationship between transparency and risk-taking strengthens. This precisely matches what we should expect if trust-building reduces investors' risk aversion.

I complement the evidence on risk aversion with an additional cross-sectional analysis based on investors' baseline levels of anxiety. I construct a novel measure of investor anxiety attitudes using geographic search behavior and letter readership data,[5] then analyze whether the effect of trust-building communication depends on investors' prior anxiety. I find that the effect of

[5]My investor anxiety measure is constructed by geographically locating fund letter readerships using the locations of filing downloaders from the SEC's EDGAR website, then inferring their aggregate anxiety attitudes from the intensity of local Google searches for anxiety-related terms (such as "stressed," "anxiety," "stop worrying," and a variety of other, related terms detected by Google's search engine). The resulting fund-level measure represents an asset-weighted average of these local anxiety attitudes.

transparent communication is concentrated among letter readerships with high baseline levels of anxiety. This directional effect suggests that trust most likely soothes the anxiety component of risk aversion (Guiso, Sapienza, and Zingales, 2018).

A final set of tests systematically evaluates and rejects a dozen potential explanations for the main transparency–risk taking effect that I uncover. Most notably, various learning channels, including learning about risk, the return distribution, or tracking error, find no support in the data. My tests also reject other, non-learning channels, including shrouding risks through complexity and marketing to attract new investors. This battery of tests strengthens confidence that the observed effect of communication by funds to their investors operates through trust-building, rather than alternative mechanisms.

Economists have long sought to understand the origins of trust attitudes (e.g., Fisman and Khanna, 1999; Alesina and La Ferrara, 2002; Algan and Cahuc, 2010). I contribute by showing that transparency in communication fosters trust and influences financial decision-making in an important economic setting. My paper also sheds light on the role of communication in economic interactions. Persuasive communication is a widespread phenomenon (McCloskey and Klamer, 1995), and I show it can act to shift a recipient's effective risk aversion.[6] This finding is a useful complement to our current understanding that persuasion can exploit or create inaccurate beliefs (e.g., Mullainathan and Shleifer, 2005; Hu and Ma, forthcoming). I show that persuasion can in fact benefit recipients, by helping to overcome psychological frictions and thus encouraging them to bear diversified financial risk.

My results suggest that trust-building communication acts to soothe the anxiety component of risk aversion. This finding is consistent with the "money doctors" theory of Gennaioli, Shleifer, and Vishny (2015), in which asset managers can alleviate investors' anxiety about taking risk. Anxiety and negative affect have long been understood to play a role in financial decision-making (e.g., Kamstra, Kramer, and Levi, 2003; Wang and Young, 2020). Specifically, anxiety has been shown to alter the decision-maker's risk aversion (Kuhnen and Knutson, 2011; Giorgetta et al., 2012; Bassi, Colacito, and Fulghieri, 2013; Cohn, Engelmann, Fehr, and Maréchal, 2015; Guiso, Sapienza, and Zingales, 2018). Financial anxiety produces important real spillovers (Haushofer and Fehr, 2014; Engelberg and Parsons, 2016; Lin and Pursiainen, 2023; Kaur, Mullainathan, Oh, and Schilbach, 2025), and it is thus important to understand how to alleviate such anxieties. I uncover a means by which asset managers soothe investors.

This paper shows that trust-building communication is a determinant of investor flows into

[6] In the taxonomy of DellaVigna and Gentzkow (2010), I uncover a form of preference-based persuasion (e.g., Galbraith, 1958; Becker and Murphy, 1993), complementing more common belief-based forms of persuasion.

mutual funds, and thus part of the value added by asset managers (Berk and Van Binsbergen, 2015). Understanding the drivers of investor flows to mutual funds has long been an important research question (e.g., Sirri and Tufano, 1998; Chevalier and Ellison, 1999), and my results provide further evidence on the behavioral drivers of fund flows (e.g., Cooper, Gulen, and Rau, 2005; Bailey, Kumar, and Ng, 2011; Ben-David, Li, Rossi, and Song, 2022).

Finally, this paper advances the literature on information transmission in the mutual fund industry, which has so far focussed on understanding what disclosures reveal about the strategies or skills of active managers. Sheng, Xu, and Zheng (forthcoming) and Abis, Buffa, Javadekar, and Lines (2022) examine the accuracy of mandatory disclosures in prospectuses. Hillert, Niessen-Ruenzi, and Ruenzi (2025) and Cao, Yang, and Zhang (2025) study whether shareholder letters are informative about manager skill or future performance, and whether investors can thus learn from this communication. The current paper focuses on letters addressed to investors in homogeneous and passive S&P 500 index mutual funds to uncover a behavioral channel that is distinct from rational learning mechanisms: my approach isolates the trust-building effect of voluntary communication.

I describe the institutional setting further in Section 2. Section 3 identifies the trust-building effect of communication on investors' risk-taking. Section 4 provides complementary evidence that this communication alleviates investors' anxiety. Section 5 evaluates and rules out a number of potential alternative explanations, including various learning mechanisms. Section 6 concludes and discusses the wider implications of my findings. The Appendix contains tables and figures. An Internet Appendix contains additional analyses and data checks.

2 Institutional Setting and Data

2.1 Fund Letters

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 may also contain free-form text communications to shareholders. Such communications do not follow a standardized format, are completely optional, and may contain unaudited statements. I label these discussions "fund letters."[7,8]

Fund letters are often formatted as just that, beginning with phrases such as "Dear investors" and ending with a signoff; refer to Internet Appendix B for an example. However, such free-form voluntary text 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.

I define a customized set of patterns that identify the start and end locations of letters based on common phrases and section headings. To do this, I manually sample from N-CSR and N-CSRS filings and read through them to identify fund letters, then I define patterns that match the start and end of detected letters. 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. This procedure yields a set of letter-detecting patterns that is simple yet effective, consistently identifying letters in over nine out of ten annual reports and eight out of ten semi-annual reports every year throughout my sample. This high rate of detection indicates that discretionary letters are a very common form of communication by funds and their managers.

As part of regular and high-profile shareholder reports, the letters I examine provide a vehicle for funds to broadcast messages to their investor bases. Because fund letters are entirely voluntary and can take flexible formats, they represent discretionary communication by fund managers, making them an ideal setting to test for a trust-building mechanism in communication. Furthermore, by focusing on next-month flow responses immediately following letter dissemination, I can cleanly measure investors' reactions to this voluntary communication while avoiding the challenge of identifying longer-lasting trust effects that would be confounded over long horizons.

Extracting Individual Fund Letters From Reports. Annual and semi-annual reports are published at the level of a legal entity, as identified by a Central Index Key (CIK), and this is not necessarily at the level of an individual fund.[9] To deal with cases where one CIK nests several funds, I match letters in the report to individual funds by detecting fund names or fund tickers in the contents of each letter. Letters that do not contain a fund name or ticker are matched to every fund associated with that report (which may just be a single fund).

[7]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.

[8]The literature also uses the more general phrase "shareholder letter" (Hillert, Niessen-Ruenzi, and Ruenzi, 2025) 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.

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

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.[10]

2.2 Extracting and Quantifying Risk Discussions in Fund Letters

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

2.2.1 Extracting Sentences That Discuss Risk

Extracting Valid Non-Boilerplate Sentences. To prepare the text corpus, I first extract grammatically valid sentences from fund letters using Stanford CoreNLP tools (Manning et al., 2014); this is the only step of my procedure that relies upon machine learning tools. I next define a set of patterns that match to non-informative boilerplate text. I construct this set of patterns by randomly sampling from valid sentences and defining patterns that match sentences that are legal disclaimers, then remove these boilerplate sentences from the database.

The remaining steps in my text extraction methodology likewise screen sentences using patterns of words or phrases that are manually defined by sampling and reading through the database of fund letters. The key advantage of such a bag-of-words approach is ensuring the textual analysis procedure is adapted to the corpus of fund letters, rather than a different setting where phrases can have entirely different meanings: Loughran and McDonald (2011) argue persuasively that adapting textual analysis procedures to a financial context is crucial.

Screening for Sentences that Discuss Risk. I use two sets of word and phrase patterns to screen for sentences that discuss the topic of risk. An important challenge in isolating information provision about risks is excluding discussions that simply recap past events: fund letters frequently discuss past portfolio holdings, returns, and economic and world events from recent months. However, investors' portfolio choices are forward-looking, and this paper focuses on communication about the unknown, yet-to-be-realized risks that investors dislike. I therefore screen separately for sentences that contain risk-related words and phrases, as well as forward-looking words and phrases, and extract sentences that jointly meet both criteria. Figure I graphically illustrates this conjunction.

[10]The SEC announced reforms to the format of N-CSR and N-CSRS filings in 2022. My sample period running from 2006-2021 benefits from a consistent regulatory regime, while also capturing major financial and economic shocks (the global financial crisis and the Covid shock).

[Insert Figure I here]

Sentence-level matching is preferable to counting forward-looking and risk-related words separately across the letter: such an approach would struggle to capture whether these concepts appear together, since individual words rarely convey both meanings. Sentences provide the natural unit for identifying where forward-looking and risk-related content intersect.

Validating the Matching Patterns. In Internet Appendix D, I compare my customized set of words and phrases to those taken from other contexts, or that are encoded in off-the-shelf software. I find that my customized approach detects a comprehensive set of forward-looking and risk-related statements, while 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.

2.2.2 Quantifying the Amount of Risk Detail Present in a Letter's Text

The key measure used in this paper is the fraction of a fund letter devoted to discussing risk:

$$\text{Detail Fraction}_{i,t} = \frac{\text{Number of words about risk}_{i,t}}{\text{Total number of words in letter}_{i,t}}. \tag{1}$$

Similar share-based measures are typically used to measure objective risk exposures; for example, Hassan, Hollander, van Lent, and Tahoun (2019) measure firms' exposure to political risk by the share of their earnings calls that discuss that topic. By contrast, the current paper focuses on homogeneous S&P 500 tracker funds that have near-identical risk exposures by construction. This allows me to interpret Eqn. (1) as variation in how funds subjectively and voluntarily discuss the *same* risk exposure when communicating with their investor clients.

2.2.3 Quantifying the Level of Risk Conveyed by Text

As well as my main measure of transparency, it is also important to quantify the level of risk: i.e., whether the letter implies risk is high or low. This signal is potentially independent to the amount of detail provided. I measure the level of risk conveyed by a fund letter using two approaches. Since each approach is a reasonable *ex ante* measure, I remain agnostic about which to prefer and so employ both in my subsequent analyses.

Sentiment-Based Approach. My first approach measures the level of the risk conveyed using the net sentiment of the text that discusses risk. 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:

$$\text{Net sentiment of risk text} = \frac{\text{Positive word count} - \text{Negative word count}}{\text{Total word count}} \tag{2}$$

Finally, the level of risk implied by the sentiment of this text is simply the negative of the net sentiment, standardized within each fund to account for any fund-specific mean sentiment due to writing styles. My empirical results are not dependent on this standardization:

$$\text{Risk Level} = -z(\text{Net sentiment of risk text}). \tag{3}$$

Contextual Approach. My 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:

$$\text{Risk Level} = \text{High risk word count} - \text{Low risk word count} \tag{4}$$

2.2.4 Variation in Communication About Risk

Cross-Sectional Dispersion. Figure II summarizes the cross-sectional dispersion in the amount of detail communicated about risk (by the fraction of the letter) and the net sentiment of the same risk-related text, respectively. Cross-sectional variation in both variables is considerable, even though S&P 500 funds by construction deliver a common risk exposure to their investor clients. As one might expect from discussions of risk, the tone is on average net-negative, but both positive-sentiment and negative-sentiment letters coexist.

[Insert Figure II here]

Time-Series Variation & Determinants. There is also considerable time series variation. 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 low, at -0.05 (-0.05). As I show next, market conditions are an important determinant of this time-series variation.

[Insert Table I here]

Table I analyzes time-series indices of two risk detail measures and two risk level measures. Each measure shown in the table is an equally weighted cross-sectional year-month t mean of the within-fund-standardized (i.e. z-scores) equivalents; every measure is thus a simple time-varying index. The primary driver of all four time series is the contemporaneous level of the VIX. Within-fund variation in their communication to investors about risk is thus informative, in the sense that fund letters provide more detail about risk at precisely the moments when the VIX is higher. The communication is also informative about the level of risk: managers write more pessimistically about risk when the VIX is high, as one might expect if the VIX proxies for investors' perceived risk,[11] or for investors' demand for information about the fund's risk exposure.

2.3 Mutual Fund Classifications and Characteristics

My source of mutual fund and fund family characteristics is the CRSP US Survivor-Bias-Free Mutual Fund Database (MFDB).

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 CRSP MFDB. 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 (i.e., before screening for index funds) to avoid the undue influence of outliers.

My main results focus on communication by S&P 500 index mutual funds. A supplementary result relies on measuring communication by active funds; to identify these, I first classify index

[11] Note that the VIX embeds both expected volatility and a risk premium. Therefore, when controlling for perceived risk in subsequent analyses, I avoid relying upon the VIX.

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.

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:

$$\text{Net Flow}_{i,t\to t+1}(\$) \;=\; \text{TNA}_{i,t+1} \;-\; (1+R_{i,t\to t+1})\times \text{TNA}_{i,t}. \tag{5}$$

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. A similar normalization to the net flow (below) is applied to gross inflows and gross outflows:

$$\text{Net Flow}_{i,t\to t+1}(\%) \;=\; \frac{\text{Net Flow}_{i,t\to t+1}(\$)}{\text{TNA}_{i,t}} \times 100\%. \tag{6}$$

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. These steps occur before screening for S&P 500 index funds.

S&P 500 Index Funds. To select S&P 500 index mutual funds for my main sample, I compare fund-level returns to the returns on the S&P 500 index and discard funds with a correlation of less than 98%. I then 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 197 S&P 500 index mutual funds (over the entire sample) for which SEC EDGAR identifiers are also available. In my main tests, I focus on fund-months where a fund letter was disseminated to investors; summary statistics for my main sample are shown in Internet Appendix A.

3 Transparent Communication and Investor Risk-Taking

This section shows that more detailed voluntary communication about risks encourages investor risk-taking. Section 3.1 quantifes the relationship between risk transparency and S&P 500 index fund flows. I show that the detail provided about risk plays a distinct role, beyond the level of risk that letters might convey. Section 3.2 confirms the presence of a causal effect by exploiting corner bunching in the treatment variable. Section 3.3 and Section 3.4 interpret the effect and present evidence consistent with a trust-driven willingess to bear risk by investors.

3.1 Relationship Between Communication and Fund Flows

I begin by analyzing the relationship between net flows into S&P 500 index mutual funds and the amount of information about risk that funds voluntarily transmit to their existing investors. 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. The baseline panel regression is the following:

$$\text{Net Flow}_{i,t\rightarrow t+1}(\%) = \beta \times \text{Risk Detail}_{i,t} + \Gamma \times \text{Fund Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}. \qquad (7)$$

That is, I measure the strength of the relationship β between the degree of voluntary transparency about risks during a month t and the normalized percentage flow into the fund from the end of month t to the end of month $t+1$; due to the monthly frequency of the flows dataset, this is the most immediate period following the dissemination of a fund letter. The baseline specification (Eqn. (7)) includes fund fixed effects π_i to capture any time-invariant unobservable fund-, family- and investor base-specific characteristics that might drive flows, as well as year-month time effects ϕ_t that capture any time-varying characteristics of the common portfolio, such as its actual risk exposure, and any common shocks that might drive flows. I also 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. Γ is a vector of coefficients on the set of controls.

[Insert Table II here]

Baseline Relationship. Table II column 1 shows the result of estimating the panel regression (Eqn. (7)). The estimated coefficient value $\widehat{\beta} \approx 0.02$ is positive and significant at the 95% level, with double-clustered standard errors. This estimate establishes a key result: when a larger share of a fund's letter is devoted to discussing risk, net flows into the fund are higher in the month immediately following the dissemination of this letter. Since the choice of text content examined in this paper is voluntary, this result shows that voluntary transparency about risks by a fund is positively related to investors' willingness to bear risk in that fund.

Including the Level of Risk Conveyed by Fund Letters. The baseline result above is that more transparent and informative discussion of risk predicts higher subsequent flows. However, investors might also be reacting to whether the letter characterizes risk as being high or low going forward. This could be conveyed through the tone with which a letter discusses risk, or even by explicit descriptions like risk being "high" or "low." Indeed, both features of the text are correlated with market conditions (see Table I). I test for this possibility by augmenting the baseline specification with a control for the level of risk conveyed (wherever letters discuss risk at all):

$$\begin{aligned}\text{Net Flow}_{i,t\to t+1}(\%) = {} & \beta_1 \times \text{Risk Detail}_{i,t} + \beta_2 \times \text{Risk Level}_{i,t} \\ & + \Gamma \times \text{Fund Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}.\end{aligned} \tag{8}$$

Table II columns 2, 4 and 5 show estimates of Eqn. (8), using three different measures of the risk level conveyed by fund letters. These measures are defined based on either the Loughran and McDonald (2011) sentiment of the text discussing risk, or the presence of descriptive words, and are described further in Section 2.2.3. The results are similar irrespective of the measure used: the baseline coefficient estimate $\widehat{\beta}_1 \approx 0.3$ remains positive, significant, and has increased in magnitude with the inclusion of the risk level control. As for the coefficient estimated on the control itself, $\widehat{\beta}_2$, it is not significantly different to zero. These results suggest that investors respond to the amount of detail provided about risk (i.e., the transparency of the fund letters) rather than any signals transmitted about the actual level of risk.

I next test whether the level of risk matters by attenuating the importance of transparency, rather than entering as a direct control. To do so, I augment the panel specification with interaction effects between the risk detail and an indicator for conveying a high risk level (once again, through the sentiment of the text, or via the presence of certain descriptors):

$$\begin{aligned}\text{Net Flow}_{i,t\to t+1}(\%) = {} & \beta_1 \times \text{Risk Detail}_{i,t} + \beta_2 \times \mathbb{1}\{\text{High Risk Level}\}_{i,t} \\ & + \beta_3 \times \mathbb{1}\{\text{High Risk Level}\}_{i,t} \times \text{Risk Detail}_{i,t} \\ & + \Gamma \times \text{Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}.\end{aligned} \tag{9}$$

Table II columns 3 and 6 show estimates of Eqn. (9). Neither the interaction $\widehat{\beta}_3$ nor the high-risk indicator alone $\widehat{\beta}_2$ is statistically significant, and the interaction effects are in any case small. This indicates that the positive impact of risk detail on flows does not differ systematically depending on whether the risk discussion conveys a relatively high or low level of risk.

Overall, Table II shows that flows into S&P 500 index mutual funds are consistently higher when funds communicate more detail about risk in their voluntary letters. This effect is robust to controlling for various textual measures of the risk level conveyed in the same fund letters, and is not moderated by whether the discussion frames risk as especially high.

[Insert Figure III here]

Figure III visualizes the relationship between risk detail and fund flows in a binscatter plot. The displayed net flow is residualized against the level of risk (defined using sentiment) and the fund controls and fixed effects employed in prior specifications, in order to include them as controls in the binscatter. Net flows are monotonically increasing in risk detail quintiles.

Economic Significance. The relationship between flows and transparency about risks is an economically meaningful one. This is most clear when regressing standardized net flows against the standardized risk detail, the results of which are displayed in Table III. The estimates show that a one standard deviation (SD) increase in the detail provided about risk predicts a next-month increase in the fund's assets in the range of 0.09–0.17 SDs in magnitude; i.e., 0.28–0.50 percentage points of assets (on average), depending on the controls employed. By way of comparison, the well-known, strong and negative association between flows and total fees has a magnitude in the range of 0.44–0.52 SDs. The magnitude of the flow-detail relationship is thus approximately one-sixth to one-third of the magnitude of the association between quantities and prices, demonstrating the economic importance of trust-building communication.

[Insert Table III here]

Robustness of the Transparency–Risk Taking Relationship. I conduct a battery of robustness tests in Internet Appendix H, and find no change to the direction of the association between risk detail and fund flows. Encouragingly, the magnitude mostly remains similar, and sometimes increases. The robustness tests are as follows. I explicitly *(i)* control for (the near-identical) fund risk exposures, *(ii)* control for different trailing past return windows, *(iii)* control for past fund flows, and *(iv)* control for Morningstar ratings (Ben-David, Li, Rossi, and Song, 2022). I next *(iv)* employ a different measure of risk detail: the Shannon's entropy of the distributions of words employed in each risk discussion. I also *(v)* repeat the analysis when focussing on the sub-sample of letters that explicitly communicate a high level of risk. I next *(vi)* verify the presence of a pure within-fund effect, consistent with an ongoing relationship between a fund and its investor base over time. I finally *(vii)* confirm the relationship between fund flows and the features of communication about risk is similar when flows are measured using gross inflows instead of net flows.

3.2 Identification of a Causal Effect Using Corner Bunching

I now confirm that the flow–risk detail relationship is a causal effect.

The central pillar of this paper's identification strategy is its focus on quasi-identical S&P 500 index funds. Because these funds passively replicate the same benchmark, important confounding channels – such as signaling managers' stock-picking abilities – are absent by construction. Furthermore, the empirical design controls for a variety of fund-level and time-varying confounders, both observed and unobserved, through the inclusion of fixed effects.

That said, the possibility of unobserved shocks that vary at the fund–month level remains a potential concern. For instance, contemporaneous changes in distribution networks, marketing intensity, or idiosyncratic letter content could affect both transparency and flows. To address these residual unobservables, I exploit an empirical feature of the treatment variable's distribution.

[Insert Figure IV here]

Figure IV visualizes the distribution of my key treatment variable, defined in Eqn. (1) as the fraction of each detected fund letter devoted to discussing risk. A striking feature of this empirical distribution is the mass of observations bunched at the corner value of 0% (highlighted by the red bar in the histogram). This bunching at the lower support produces a discontinuity in the empirical cumulative distribution function that can be exploited for identification.[12]

Recent methodological contributions by Caetano, Caetano, and Nielsen (2024) have been developed to exploit precisely this pattern of corner bunching in order to identify the treatment effect. Their key insight is that, while the observed treatment intensity is censored at the corner, the unobservables that threaten identification are not. By modeling the shape of the treatment distribution around the corner, one can construct a control function to capture the latent selection component. Intuitively, the excess mass at the corner value of 0% reveals the types of funds and months in which the latent propensity to be transparent is low, and therefore allows for the econometrician to control for this selection directly. Further details of the econometric methodology are described in Internet Appendix E.

[Insert Table IV here]

I augment this paper's baseline regression (7), which relates the risk detail provided in a fund letter to subsequent fund flows. The main coefficient of interest is β, and I once more include the usual fund controls and additive year-month & fund fixed effects. The modified specification includes a control function $\lambda(\cdot)$ that corrects for potential endogeneity, and is specified following using Caetano, Caetano, and Nielsen's (2024) methodology. This control

[12] My setting has parallels to a fuzzy regression discontinuity design: in the standard case, treatment propensity jumps at an interior cutoff of a separate running variable. By contrast, the discontinuity here is at the corner of the support, and the treatment itself doubles as the running variable.

function exploits the presence of corner bunching in the treatment variable Risk $\text{Detail}_{i,t}$:

$$\begin{aligned}\text{Flow}_{i,t\to t+1}(\%) = {} & \beta \times \text{Risk Detail}_{i,t} + \delta \times \overbrace{\lambda\big(\text{Risk Detail}_{i,t};\, \text{Fund Controls}_{i,t},\, \pi_i,\, \phi_t\big)}^{\text{control function}} \\ & + \Gamma \times \text{Fund Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}. \end{aligned} \tag{10}$$

Table IV reports the results. Column (1) estimates Eqn. (10) without the control function. The resulting estimate, $\widehat{\beta} = 0.014$, is similar in magnitude to my baseline panel regressions: positive, statistically significant, but still potentially biased by unobserved fund–month shocks.

The specifications reported in columns 2–4 incorporate the control function $\lambda(\cdot)$ constructed according to Caetano, Caetano, and Nielsen's (2024) methodology. This control function exploits the presence of corner bunching in the treatment variable, Risk $\text{Detail}_{i,t}$, to correct for potential endogeneity in β. Specification 2 constructs $\lambda(\cdot)$ under a Tobit shape restriction; specification 3 relaxes this to a semiparametric Tobit; and specification 4 exploits a still weaker, nonparametric restriction that the distribution's tails are symmetric. Each column also reports estimates of δ, the coefficient on the correction term, which provides a direct measure of the bias arising from selection on unobservables.

The estimates $\widehat{\delta}$ are consistently negative and statistically significant. This pattern indicates that the baseline specification is subject to a downward bias due to selection on unobservables. Once the bias is corrected, the causal treatment effect estimates $\widehat{\beta}$ in columns 2–4 are substantially larger than the baseline association in column 1, ranging from 0.035 to 0.050. The implied causal effect is thus between 2.5–3.5 times the magnitude of the uncorrected estimate, depending on the shape restriction used to implement the control function.

3.3 Trust-Building Transparency

Having shown that transparency about risks drives fund flows, I now explain how this evidence aligns with a trust-related channel.

Trust-Building Rather Than Learning. My focus on voluntary transparency about risks provides a sharp test of potential channels. Since risk is inherently undesirable, investors who perceive or notice a higher level of risk should be less willing to bear that risk; under a learning view, providing more detail about risks should therefore decrease fund flows. Empirically, the effect takes the opposite direction, however.

While this result may appear puzzling from a learning perspective, it aligns naturally with a trust-building channel. The very undesirability of risk makes voluntary disclosure about this attribute a credible signal that can build investor trust. Gennaioli, Shleifer, and Vishny (2015) suggest that trust increases investors' willingness to bear risk, which is consistent with voluntary risk transparency attracting fund flows.

Section 4 elaborates this mechanism in greater detail with additional empirical evidence. Section 5.1 systematically rules out various learning-based channels.

Reliance on Financial Intermediation. I now conduct a cross-sectional test for the role of trust. Since trust is a key ingredient that sustains relationships between clients and their chosen financial intermediaries, it should matter most for clients who are especially reliant on intermediaries for financial decision-making. That is, investors who are most dependent on intermediaries should exhibit the greatest sensitivity to trust-enhancing transparency.

[Insert Table V here]

I implement this test by using mutual funds' 12b-1 fees to proxy for investor reliance on financial intermediaries: Del Guercio and Reuter (2014, pp. 1682) find that broker-sold index funds charge substantially higher 12b-1 fees than direct-sold ones. Higher 12b-1 fees are thus a proxy for investor bases that are less self-directed and more reliant on intermediaries in their financial decision-making. Table V augments the baseline flow–risk detail regression with an interaction term for the fund's 12b-1 fee. The results align with my cross-sectional prediction: the relationship between risk disclosure and flows is strongest among funds with higher 12b-1 fees, consistent with a trust-building role for transparency in fund communication.

3.4 Further Understanding Investor Risk-Taking

Testing for a Flight to Safety. Greater voluntary transparency about risks in fund letters disseminated by S&P 500 index funds causes higher flows into these funds. I now assess whether this fact could (paradoxically) reveal that investors are actually taking *less* risk: this could occur if investors perceive higher risk levels from the disclosures, or if detailed risk information otherwise reduces their willingness to bear risk. I find this is not the case.

An ideal test would exploit individual investor-level data rather than aggregate fund-level flows, but this limitation is mitigated by the structure of investors' switching costs: mutual fund investors face high switching costs across fund families but lower costs for within-family switches (Massa, 2003). Therefore, testing whether intra-family flows into index funds are rebalanced from riskier funds within the same family captures the most frictionless setting, where such risk-reducing behavior would most likely be apparent if it exists.

[Insert Table VI here]

My test is to measure the predictive association between the risk detail in S&P 500 index funds' letters disseminated during a month t and the gross outflows from riskier funds within the same family during the subsequent month $t+1$. I thus test whether outflows from riskier funds are occurring simultaneously with flows into the less-risky S&P 500 index funds that comprise my main sample. If communication-driven flows into S&P 500 funds are due to outflows from riskier funds, then a higher amount of risk detail should predict higher outflows from the riskier funds, just as it is known to predict higher flows into the less-risky S&P 500 index funds.

The negative and significant coefficients in columns 1–2 of Table VI show that this is not the case. In fact, the direction is the opposite to what one might expect under the rebalancing hypothesis: a higher risk detail in S&P 500 fund letters predicts *lower* outflows from riskier sibling funds. A similar result obtains when using the risk detail contained in the sibling funds' letters to predict the sibling funds' outflows, in columns 3–4.

Based on this test, I interpret higher transparency-flows into S&P 500 index funds as evidence of greater-risk taking by their existing investors, to whom fund letters are disseminated.

Investors Benefit By Bearing More Risk. Having established that investor flows are indeed consistent with higher risk-taking, I now quantify the gains that accrue to investors. At first blush, one might think it obvious that it should benefit investors to place their savings in a broadly-diversified market portfolio, particularly since their outside option appears to be holding cash or another less-risky alternative. However, this may not be the case if the timing of fund letters (and variation in how they discuss risk) leads investors to mis-time the stock market cycle. Furthermore, the mutual funds in my sample levy fees on investors' holdings, which must be taken into account.

[Insert Figure V here]

I aggregate up all flows in my sample of S&P 500 index funds that are attributable to fund letters' transparency about risk in the prior month, and then quantify the Sharpe ratio of holding these flows in the stock market over a variety of horizons, accounting for fund fees. Figure V presents the resulting annualized Sharpe ratios. The Sharpe ratio is always positive and significant except in the case of a short-term holding period of only 1 month, in which case it remains positive but statistically indistinguishable from zero. Given that mutual fund holdings tend to be held over multiple years (Sirri and Tufano, 1998), I conclude that investors do indeed benefit from responding to trust-building transparency in fund letters.

Having established the role of transparency in building trust, I conduct further tests of the economic mechanism in Section 4 and Section 5.

4 Inspecting the Mechanism

This section inspects the economic mechanism through which trust-building communication encourages risk-taking. I provide evidence consistent with a hypothesis that investors' effective risk aversion is decreased by this communication. Furthermore, the anxiety component of risk aversion (Guiso, Sapienza, and Zingales, 2018) is important for explaining my main effect. I begin the section by laying out why a risk aversion channel may be responsible, and then conduct two empirical tests of this mechanism.

4.1 A Parsimonious Test for a Risk Aversion Mechanism

A standard view holds that communication affects decisions by shifting beliefs. This view includes games in which senders strategically choose signals while recipients decide whether to update their beliefs accordingly. However, communication may influence economic decisions through channels beyond expectation formation. Experimental evidence indicates that information can provide psychological benefits such as confidence and security even when it does not change decision-makers' underlying beliefs (Eliaz and Schotter, 2010, pp. 316).

Gennaioli, Shleifer, and Vishny (2015) theorize that trusted money managers provide their investors with "peace of mind" and reduce "investor anxiety about taking risk." Building on both this argument and the insight that communication can build confidence, I propose that trust-building communication reduces investors' effective risk aversion. By providing reassurance or confidence, trust-building voluntary transparency about risks thus affects investors' willingness to bear risk. I propose a parsimonious test of this hypothesis.

The Investor's Risk-Taking Decision. Consider the case of an investor who can access a riskless asset that returns R^f_{t+1} and a risky asset with excess return R^e_{t+1} above the riskless rate. I interpret the risky asset as the market portfolio, since this paper's empirical analyses focus on 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, and abstract away from the drivers of her initial delegation decisions to assume that the investor is paired with a particular asset manager through whom she invests in the risky asset: the setting is thus one of strengthening an existing relationship rather than marketing to new investors.

Assume the investor has a mean-variance objective and γ_t captures her degree of effective risk aversion.[13] She therefore seeks to maximize the objective function

$$U(x_t) = R^f_{t+1} + x_t \mathbb{E}_t[R^e_{t+1}] - \frac{\gamma_t}{2} x_t^2 \, \mathbb{V}\text{ar}_t(R^e_{t+1}), \tag{11}$$

and as a result will optimally allocate the fraction

$$x_t = \frac{\mathbb{E}_t[R^e_{t+1}]}{\gamma_t \, \mathbb{V}\text{ar}_t(R^e_{t+1})} \tag{12}$$

of her portfolio to the risky asset, based upon her risk aversion and beliefs about the payoff. Under my hypothesis, trust-building communication shifts γ_t, the effective risk aversion.[14]

My earlier empirical results cannot be explained by revisions to investors' beliefs.[15] Therefore, in what follows I assume the beliefs $\mathbb{E}_t$ and $\mathbb{V}\text{ar}_t$ remain constant in order to simplify the derivations. Given that fund letters are addressed to existing investors, and since I focus on investors in S&P 500 index mutual funds, we can safely assume the investor already holds some fraction of her financial wealth in the risky asset ($x_t \neq 0$). Under these two assumptions, the identity linking flows and changes to risk aversion takes the following clean form:

$$\text{Flow}_{t\to t+1} \triangleq \frac{x_{t+1} - x_t}{x_t} = \frac{\gamma_t}{\gamma_{t+1}} - 1. \tag{13}$$

Eqn. (13) highlights that changes to risk aversion generate fund flows as the investor rebalances from the riskless asset to the risky asset; i.e., takes greater risk.

I now derive the parsimonious test for whether a risk aversion-reducing mechanism underlies the trust-building effect of providing more detailed risk discussions in fund letters. First, I rewrite Eqn. (13) to allow for risk aversion γ_{t+1} to be a function of prior risk aversion γ_t and the amount of transparency τ provided about risk:

$$\text{Flow}_{t\to t+1} = \frac{\gamma_t}{\gamma_{t+1}(\tau, \gamma_t)} - 1. \tag{14}$$

[13] The assumptions of mean-variance preferences and a static horizon are not crucial. Under widely used CARA-Normal assumptions, the behavior is identical. 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 (12) is an important component of dynamic portfolio choice.

[14] Gennaioli, Shleifer, and Vishny (2015, pp. 96) similarly model trust through shifts in the investor's effective risk aversion parameter, which they label "anxiety."

[15] I carefully test for the presence of various learning channels in Section 5.1.

Next, I compute the partial derivative of the flow with respect to the transparency τ:

$$\frac{\partial \text{Flow}_{t\to t+1}}{\partial \tau} = \underbrace{\frac{-1}{\gamma_{t+1}(\tau,\gamma_t)^2}}_{<0} \underbrace{\frac{\partial \gamma_{t+1}(\tau,\gamma_t)}{\partial \tau}}_{\text{Unobserved}} \gamma_t. \tag{15}$$

The goal is to test whether risk aversion γ_{t+1} is decreasing in the amount of transparency τ that the asset manager provides about risk; i.e., whether the term $\frac{\partial \gamma_{t+1}(\tau,\gamma_t)}{\partial \tau}$ has a negative sign. If so, it follows immediately from Eqn. (15) that the relationship between the flow-detail sensitivity $\frac{\partial \text{Flow}_{t\to t+1}}{\partial \tau}$ and the prior level of risk aversion γ_t should be positive, under such a scenario. This leads to a testable prediction for whether a risk aversion-decreasing effect is present:

Test for Risk Aversion Mechanism. *The same amount of transparency τ provided about risks will produce a larger fund flow when the prior risk aversion γ_t is higher.*

An important advantage of the above test is that it does not rely on obtaining measurements of both the initial risk aversion γ_t and subsequent risk aversion γ_{t+1} levels: while the former can be proxied using baseline measurements of risk aversion, it would not be feasible to measure the latter for the subset of investors receiving fund letters. Conducting the test requires only measuring the $\text{Flow}_{t\to t+1}$, the amount of risk detail τ, and the prior risk aversion γ_t.

4.2 Result of the Risk Aversion Test

I empirically capture the baseline risk aversion γ_t with a monthly time series of the relative risk aversion of a representative investor, constructed by Bekaert, Engstrom, and Xu (2022). This test therefore exploits time-series variation in the strength of the effect of transparency on risk-taking across market states. If trust-building communication acts through a risk aversion-based mechanism, we would expect to see a stronger effect on flows during so-called bad times, when the aggregate risk aversion measure γ_t is higher.

To conduct this test, I augment the baseline panel regression with an interaction with $\mathbb{1}\{\text{High Risk Level}\}_t$, an indicator that takes the value 1 when the Bekaert, Engstrom, and Xu (2022) measure of relative risk aversion is above its long-run median:

$$\begin{aligned}\text{Net Flow}_{i,t\to t+1}(\%) = {} & \beta_1 \times \text{Risk Detail}_{i,t} + \beta_2 \times \mathbb{1}\{\text{High Risk Level}\}_t \\ & + \beta_3 \times \mathbb{1}\{\text{High Risk Level}\}_t \times \text{Risk Detail}_{i,t} \\ & + \Gamma \times \text{Controls}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}.\end{aligned} \tag{16}$$

If trust-building communication reduces investors' effective risk aversion, my parsimonious test based on portfolio choice intuition predicts a positive coefficient β_3 on the interaction term.

[Insert Table VII here]

Table VII reports estimates of Eqn. (16) in columns 2–3, and the baseline non-augmented panel regression in column 1 for comparison. The estimate $\widehat{\beta}_2 < 0$ reveals that a high level of risk aversion coincides with lower fund flows: investors are less willing to take risk when they are more risk averse, in accordance with basic economic intuition.

My test for the mechanism hinges on the estimated sign of β_3, the coefficient on the interaction term. Columns 2–3 of Table VII show that $\widehat{\beta}_3 > 0$, which is precisely the sign predicted by a risk aversion channel. While investors are initially less willing to take risk during periods of high risk aversion, trust-building communication has a stronger effect at precisely these moments. In other words, investors' anxiety about taking risk is higher in such states but is also soothed more effectively.

The loss of significance on $\widehat{\beta}_1$ (which nevertheless remains positive in sign) suggests that the soothing effect of communication is concentrated during these times when investors are more risk averse. Trust-building communication therefore acts as a stabilizing force that counteracts the effects of time-varying risk aversion.

The results of this mechanism test are robust to the inclusion of additional controls (in column 3) for month-to-month changes in expected returns and the VIX. Internet Appendix H.6 repeats the test using a different proxy for investors' prior risk aversion (constructed from observable financial conditions) and finds the qualitative results are unchanged.

4.3 Supplementary Analysis Using Local Anxiety Attitudes

To complement the main mechanism test, I present cross-sectional evidence that suggests a role for the anxiety component of investors' risk aversion. An anxiety-based mechanism would aliign with both Gennaioli, Shleifer, and Vishny (2015)'s money doctors theory, which explicitly models investor anxiety in the context of trusted money managers, and Guiso, Sapienza, and Zingales (2008)'s laboratory evidence showing that anxiety constitutes an important component of risk aversion in financial decision-making more generally. I provide evidence that the effect of risk transparency on risk-taking is concentrated among funds whose investors have a high baseline level of anxiety.

This analysis relies on a fund-level measure of investors' baseline anxiety attitudes. I begin by constructing a measure based on local search activity on Google for anxiety-related terms, and combine this geographic index with the geolocations of fund letter readers in order to arrive at a fund letter readership measure of anxiety attitudes. My main assumption is that overall anxiety attitudes spill over into investors' anxiety about bearing risk; this is supported both by experimental evidence (e.g., Kuhnen and Knutson, 2011; Giorgetta et al., 2012; Guiso, Sapienza, and Zingales, 2018) and by empirical analysis of mutual fund flows (Wang and Young, 2020).

[Insert Table VIII here]

SEC EDGAR Data on Letter Readership Locations. The SEC makes mutual funds' filings available for download from its EDGAR website, which is the only comprehensive and free source of such documents. These filings include the N-CSR and N-CSRS reports containing the fund letters that this paper focuses on. The SEC makes usage logs for EDGAR filing downloads available to researchers, and I use this dataset to geolocate where the readers of fund letters contained in this filings were located at the time of each download.

Merging with the SEC EDGAR usage logs dataset shortens my sample – which now begins in 2006 and ends in June 2017 – but the sample still includes the global financial crisis. I follow Lee, Ma, and Wang (2015) in identifying and removing robots from the usage logs, and de-anonymize IP addresses using the lookup table provided by Chen, Cohen, Gurun, Lou, and Malloy (2020). Similarly to Grice and Guecioueur (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.

In using EDGAR data, I assume that the geographic distributions of filing downloads from EDGAR are a reasonable proxy for the geographic distributions of the readers of fund letters. EDGAR links to filings often appear in search engine results, and may be linked to directly from funds' websites, making it a natural source of filing downloads to utilize for this purpose. Importantly, I do *not* assume that all fund letter readers must be using EDGAR. One might be concerned that readership locations inferred from this dataset are somehow unrepresentative of investors who might read fund letters through other means, so I describe my dataset of readership geolocations in Internet Appendix F. Reassuringly, the fund letter reader locations inferred from EDGAR downloads are geographically distributed very similarly to the overall population of the United States, and are not concentrated in regions where we would expect financial professionals to cluster (i.e., New York and Massachusetts).

Google Trends Data on Anxiety Attitudes. I measure geographic variation in anxiety attitudes using revealed search activity on Google. Google's 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. "Psychological Stress" (labelled by Google), which includes search terms such as "stressed", "what is stress", "anxiety" & "relieve stress."
2. "Worry" (labelled by Google), which includes search terms such as "worry about", "anxiety", "stop worrying" & "bible worry."

The example search terms above reveal that the indices capture broad anxiety attitudes and a demand for reassurance. 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. Specifically, I take the sum of both anxiety indices.[16] Then, for this geographic state s-month t anxiety measure, I aggregate across fund letter readers j to construct a panel of the prior anxiety attitudes of the readers of the letter issued by fund i at month t:

$$\text{Readership Anxiety}_{i,t} = \frac{\sum_j \text{Investor Assets}_{j,t} \times \text{Geographic Anxiety}_{j,t}}{\sum_j \text{Investor Assets}_{j,t}}. \tag{17}$$

In constructing the weighted average (17), I proxy for the aggregate financial assets held by investors 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); these vary at an annual frequency. Figure VI illustrates my aggregation procedure.

[Insert Figure VI here]

Split-Sample Analysis. My cross-sectional analysis continues its focus on the sample of S&P 500 index mutual funds i that disseminate a letter during year-month t. Each cross-section t is split into two halves: fund letters whose readership has an above-median level of anxiety, and fund letters whose readership has a below-median level of anxiety, according to Eqn. (17). I report regression results separately for each sub-sample. Specifically, I estimate the baseline panel regression (7) with an augmented set of controls for each fund letter readership's (mean asset-weighted) sociodemographic attributes – based upon Census Bureau geographic surveys – as well as each readership's 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. This rich set of controls limits the statistical power of my analysis, but the direction of any heterogeneity can be informative.

[16] Internet Appendix G charts geographic variation in the SVIs of the above anxiety topics, which highlights broad similarity in the patterns revealed by each topic separately.

[Insert Table VIII here]

Table VIII shows the results of my cross-sectional test. Fund letters that are read by low-anxiety investors have only a weak impact on flows (columns 1–2). Instead, the risk transparency–fund flow relationship is concentrated among high-anxiety investors, in columns 3–4. These directional results point to the anxiety component of investors' risk aversion being soothed by trust-building communication.

5 Evaluating Potential Alternative Mechanisms

I now consider a number of potential alternative explanations for my main empirical findings. Going through the proposed mechanisms in turn, I conduct additional empirical tests that rule them out as potential explanations. Most notably, the evidence is not consistent with a variety of learning-based mechanisms.

The current section describes each of the potential alternative channels that I rule out, together with a description of each empirical strategy. The analyses themselves comprise over a dozen panel regressions, so these are presented separately in Internet Appendix I.

5.1 Learning

Given that learning provides the most obvious explanation for how information affects investor decisions, I begin by systematically testing a variety of learning-related mechanism.

5.1.1 Bayesian Updating About Risk: Receiving a Directional Signal

To fix ideas, 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.[17]

The investor receives a 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:

$$\sigma^2 | s \sim \mathcal{N}\left(\alpha s + (1-\alpha) m_p \ , \ \frac{1}{\rho_s + \rho_p}\right), \tag{18}$$

[17]The conclusions from this example apply more generally. In fact, the current argument using Normal-Normal updating carries through exactly if the *log* of the variance is assumed to be Normally distributed, which is a standard assumption for guaranteeing a positive variance. I present the Normal-Normal case for ease of exposition.

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:

$$\Delta(\sigma^2) = [\alpha s + (1-\alpha)m_p] - m_p = (s - m_p)\frac{\rho_s}{\rho_s + \rho_p}. \tag{19}$$

As Equation (19) 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 4, the difference $(s - m_p)$ should be proportional to her *reduction* of her allocation to the risky asset. Changes in risk perceptions can thus drive fund flows.

Given the above, perhaps greater transparency about risk attracts inflows because it conveys a lower level of risk σ^2 going forward? In fact, the evidence runs counter to this possibility. Table I indicates that risk detail is positively related to the VIX, which is an objective measure of the S&P 500's implied volatility. It is therefore possible to draw inferences about the level of risk – a greater degree of risk transparency correlates with higher implied volatility – but this effect would reverse the flow-risk detail relationship described in Table II. Because a high risk detail indicates objective risk is higher, it should predict a *decrease* in fund flows, yet the effect is precisely the opposite. It is therefore implausible that investors' flows are due to them learning about risk from the degree of transparency about risk in fund letters.

Another feature of how fund letters discuss risk is their sentiment or tone, or explicit descriptors of whether risk is high or low. Once again, these features of the text convey information about the objective level of risk (see Table I). In this case, however, investors appear to ignore this signal, as the coefficients on risk level in Table II are all insignificant.

In sum, investors are provided with the opportunity to learn about risks from the information conveyed in fund letters, and yet they either act counter to this information (in the case of risk detail conveyed) or ignore it (in the case of risk level conveyed). Investors do not act upon this information in a manner consistent with learning about risk from fund letters.

In the above discussion, I assumed that fund investors' priors (m_p) are adequately controlled for by an additive combination of fund fixed effects and time effects. This allows priors to vary across funds and time heterogeneously. To allay any concerns that even this flexible specification is too restrictive, I introduce a measure of prior risk perceptions that can vary within funds over time as well as across funds. My conclusions are robust to this richer specification. The details of this analysis are in Internet Appendix I.1.

5.1.2 Bayesian-Like Updating About Risk: Gaining Precision

I next devise an additional, complementary test for whether investors might be learning about the risk of the fund. The intuition behind this new test (which I derive below) is that if investors are updating their beliefs upon receipt of a signal, then investors should update more strongly whenever their prior precision is weaker. 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.

Consider once more the Bayesian investor who learns about the variance parameter σ^2 of the return distribution, as described in Section 5.1.1. Define the strength of her update from the prior mean m_p to the posterior as the magnitude of the signed update (Eqn. 19):

$$\left|\Delta(\sigma^2)\right| \;=\; \left|s-m_p\right| \frac{\rho_s}{\rho_s+\rho_p}. \tag{20}$$

It follows immediately that the strength of the update is decreasing in the strength of the prior:

$$\frac{\partial}{\partial \rho_p}\left|\Delta(\sigma^2)\right| \;=\; -\left|s-m_p\right| \frac{\rho_s}{(\rho_s+\rho_p)^2} \;<\; 0. \tag{21}$$

I introduce a measure of investors' prior precision ρ_p, and test for a negative interaction between the strength of the update $|\Delta(\sigma^2)|$, captured by the magnitude of the inflow, and the prior precision of investors' risk perception ρ_p. I find this interaction is not significantly different to zero, and therefore it is unlikely investors are updating their beliefs about perceived risk. Details of this test are presented in Internet Appendix I.2.

DellaVigna and Gentzkow (2010, pp. 654) argue this test applies to the entire class of belief-based persuasion 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)." Based on their argument, it is unlikely that any form of belief-based persuasion is taking place as regards fund letters' transparency about risk.

5.1.3 Learning About the Return: Gaining Precision as a Bayesian

Instead of updating their beliefs about risk directly, could the results be explained by investors learning about general properties of the return distribution? After all, discussing risk also provides a more precise signal about the overall return.

Specifically, consider an investor who learns about the return r (or payoff, or fundamental value) of an asset in general – instead of learning solely about the variance or some risk-related parameter. The investor starts with a Gaussian prior on the return, $r_p \sim \mathcal{N}(m_p, {}^1\!/_{\rho_p})$, in which ρ_p denotes the prior precision. The investor receives a 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:

$$r|s \sim \mathcal{N}\left(\alpha s + (1-\alpha)m_p \ , \ \frac{1}{\rho_s + \rho_p}\right), \tag{22}$$

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 $\mathbb{V}\mathrm{ar}[r|s] = {}^1\!/_{(\rho_s+\rho_p)}$ has decreased in comparison to the prior variance ${}^1\!/_{\rho_p}$. Uncertainty is thus always reduced, irrespective of the actual value (realization) of the signal s that the investor receives. As usual, we assume the investor chooses between the riskless asset and this risky asset with subjectively perceived moments $\mathbb{E}[r|s]$ and $\mathbb{V}\mathrm{ar}[r|s]$.

In this scenario, the receipt of *any* signal about the return will mechanically reduce the investor's perceived uncertainty $\mathbb{V}\mathrm{ar}[r|s]$ and hence potentially lead her to increase her allocation to the risky asset. Could this mechanism explain my results?

I test for this possibility by repeating my entire text extraction procedure to measure the amount of forward-looking detail that fund letters provide about returns and performance (rather than risk), and then assessing whether a higher amount of detail about returns predicts higher inflows. The logic behind this test is simple: if investors are learning about returns from *risk* discussions, then they should be learning about returns from *return* discussions also. One might even expect the effect to be more pronounced in this scenario. However, I do not find any evidence that flows respond to the amount of detail provided about returns and performance. This placebo test therefore conflicts with the possibility that investors are learning about returns through more detailed discussion of the risk feature of the return distribution. Details of this test are presented in Internet Appendix I.3.

5.1.4 Learning About the Return: Predicting the Return

One might argue that investors may be using the amount of risk detail in fund letters to infer whether expected returns are high (or low). Furthermore, such a learning channel might not necessarily involve Bayesian-like updating. To test for this possibility, I first aggregate letters' risk detail similarly to the indexes constructed for Table I. I then assess whether this index of communicated risk detail predicts future realized returns of the underlying S&P 500 index over 1-month, 3-month and 5-month horizons (i.e., until the arrival of the next fund letter). I do not detect any signficance. Furthermore, an index of the level of risk conveyed by the text (another feature that investors might use as a predictor) is also insignificant, with an opposing sign to risk detail. These results suggest there is nothing to learn about returns from the risk-related information conveyed by fund letters. The test results are presented in Internet Appendix I.4.

5.1.5 Other Forms of Learning

Learning About Manager Skill. An important advantage of focussing on S&P 500 index funds is the absence of active management: since funds passively track a common index, there is no role for stock-picking skill or for managers to hold private information, and thus a limited role for communication to act as a signal of skill or ability. One might argue that minimizing a passive fund's tracking error is a form of manager skill, however, and thus letters might be signalling this ability. I therefore check whether transparency about risk might attract flows because it signals a low tracking error in the future. I test for this by using my measure of risk detail as a predictor for the fund's tracking error over a variety of horizons, and find it to be a poor predictor: the direction of the relationship is inconsistent, typically in contradiction to what we might expect under this hypothesis, and never significant at the 95% level. The test results are presented in Internet Appendix I.5.

Being Educated By Information Provision. Information provision might educate investors to make better financial decisions. For instance, low-financial literacy investors could be educated on the desirability of bearing diversified market risk. Could transparency about risks be educational? Risk discussions in fund letters typically have a negative tone, and this is hard to square with the possibility that the text is extolling the benefits of risk-taking. In addition, I develop an explicit empirical test for an educational effect: under this mechanism, the effect should be stronger among investors with a lower baseline level of financial literacy. However, I do not find any significant heterogeneity on inflows. By contrast, the interaction coefficient is significant and positive for predicting *outflows*, which is inconsistent with an educational mechanism. Details of this test are presented in Internet Appendix I.6.

Learning About Opportunities. Might fund letters contain more information about risk because they are describing buying opportunities to investors? This alternative explanation would rationalize the arrival of fund flows following risk discussions, if those risk discussions talk about risk as an opportunity. To assess this possibility, I re-estimate the baseline flow-risk detail relationship on a sample of fund letters that completely omit any discussion of opportunities (to be conservative). I implement this test by discarding every fund letter that contains the characters "opportun" (as in the words opportune and opportunity) or the phrase "chance to" from the sample, and re-estimating variants of Eqn. (8). The results remain qualitatively and quantitatively similar, indicating this channel is not responsible for the main effect. The test results are presented in Internet Appendix I.7.

Ambiguity Alleviation. Perhaps communication acts to decrease investors' perceived ambiguity about the distribution of future returns? Receiving a signal about risk might increase the perceived level of risk, which is one form of uncertainty, but it could also decrease the amount of Knightian uncertainty perceived by investors.[18] If investors are averse to such ambiguity, a decrease in perceived ambiguity could plausibly make investors more willing to hold the asset, driving inflows. Testing this alternative explanation requires a measure of perceived ambiguity of the S&P 500 index return, which Brenner and Izhakian (2018) compute. Using their measure, I find the effect of providing more information about risk does not significantly interact with the prior level of perceived ambiguity, as one would expect for an ambiguity-alleviating channel. Details of this test are presented in Internet Appendix I.8.

5.2 Shrouding

Risk is typically considered an undesirable attribute from investors' perspective. For instance, portfolio choice decisions typically model risk-averse investors, as in Eqn. (11), leading to a negative relationship between perceived risk and the optimal level of risk-taking. It is for this reason that transparency about risk can usefully distinguish between a trust-building channel and learning channels, since a higher level of perceived risk produces opposite predictions to a higher level of trust.

The salience, or visibility, of a fund's risk exposure may also be an important channel, and in this paper's setting it acts in a similar direction to a learning channel: since risk is a negative attribute, more detailed risk discussions that make this attribute more salient or noticeable should reduce investors' willingess to invest in the fund that makes risk a more visible attribute to the consumer. One might therefore argue a salience-based mechanism could explain my results through the shrouding or obfuscation of this undesirable attribute (Gabaix and Laibson, 2006; Bordalo, Gennaioli, and Shleifer, 2013). The argument is that talking more about risk could make it harder for the investor to discern the level of risk, and thereby increases her willingess to bear this risk.

However, there are two reasons shrouding cannot explain my results. *First*, the interaction between risk detail and the level of risk conveyed in the text is insignificant in Table II columns 3 and 6. Talking more about risk neither eases nor impedes investors' ability to extract the signal about risk (conveyed by the sentiment of the discussion or by explicit descriptors). Moreover, the coefficient on the risk level variable remains consistently insignificant, and has the wrong sign for an investor who dislikes risk.

[18]This channel is related to the gain in precision a Bayesian experiences upon receipt of a signal (see Section 5.1.3) but is potentially more general.

Second, I follow the literature on shrouding (e.g., Célérier and Vallée, 2017; deHaan, Song, Xie, and Zhu, 2021) and explicitly measure the complexity of the text discussing risk in fund letters. This allows for the possibility that letters obfuscate risk by employing complex language that obscures the information content of their risk discussions. Controlling for complexity in the flow-risk detail predictive regression thus allows me to disentangle the role of transparency from that of obfuscation. I find that my baseline results survive when I control for four standard measures of textual complexity employed in the accounting literature. The very feature of the text that should shroud risks (its complexity) is therefore not responsible for my main effect. The test results are presented in Internet Appendix I.9.

5.3 Marketing

This paper shows that more transparent voluntary discussions about risk (targeted at existing investors) encourage greater risk-taking, which manifests as inflows immediately following the dissemination of these fund letters. One might think that attracting flows from investors using communication is a form of marketing to new investors, akin to advertising; however I now show that the evidence does not line up with such an explanation.

Attracting New Investors with Transparency. The marketing-based alternative hypothesis is that transparency about risks attracts new investors, rather than encouraging existing investors to bear more risk. This could be the case if potential new investors read the fund letters targeted at existing investors. While this result would not necessarily conflict with a trust-based interpretation, it could throw doubt on the interpretation of greater risk-taking by *existing* investors. I therefore test explicitly for the presence of this effect by measuring the rate of arrival of new investors into funds via prospectus readership. I find no significant relationship between this measure and the detail funds provide about risk in their letters addressed to their existing investors. Details of this test are presented in Internet Appendix I.10.

Correlation with Overall Marketing Intensity. Another potential objection is that transparency about risks in fund letters might merely covary with marketing intensity, and so the inflows that I document are due to the fund's general marketing activities, which may be unobserved. However, there are four reasons that this possiblity is unlikely. *First*, if marketing activities are time-invariant or slow-moving, the fund fixed effects I use in my baseline panel regressions should absorb this effect. *Second*, if unobserved marketing activities vary at the fund-month-level, then the corner bunching-based identification strategy I presented in Section 3.2 likewise corrects for potential endogeneity that is due to this omitted variable. *Third*, my baseline results are robust to controlling for funds' 12b-1 fees as proxies for marketing intensity (see Roussanov, Ruan, and Wei (2021)). *Fourth*, I conduct an additional robustness test using another, independent measure of time-varying marketing intensity: the share of labor employed in marketing functions (see Chen, Jiang, and Xiaolan (2022)). Including this as an explicit control does not affect my main results, suggesting once again that marketing activities are a separate channel. Both robustness tests are presented in Internet Appendix I.11.

5.4 Pandering

A self-interested sender seeking to persuade a receiver may do so by tailoring their message to fit the receiver's prior beliefs (Mullainathan and Shleifer, 2005). Under such a pandering mechanism, funds might persuade investors to take more risk by framing risk communications to align with investors' pre-existing risk perceptions. To concretely test whether communication about risk panders to investors' priors, I introduce a measure of the distance of the message from the receiver's priors. Under a pandering-based explanation, one would expect messages that align most closely with investors' priors to generate the strongest responses. However, I do not find any significant heterogeneity; moreover, the interaction takes the opposite sign. Details of this test are presented in Internet Appendix I.12.

6 Discussion and Conclusion

In this paper, I show that asset managers who voluntarily provide more transparency about risk – the very quantity that investors dislike – encourage inflows from investors, as this voluntary transparency fosters trust. By shedding light on this effect, I provide evidence that asset managers act as "money doctors" (Gennaioli, Shleifer, and Vishny, 2015) to investors by actively taking steps to reduce their effective risk aversion (in this case, through trust-building communication). Investors benefit from this communication, in the case of gaining exposure to the broadly-diversified S&P 500-indexed equity portfolio managed by the funds in my sample.

I examine a straightforward aspect of communication: its transparency about a negative attribute that the recipients dislike (i.e., risk exposure). In other settings such as interactive communications, building trust might be also be achievable through nonverbal cues (e.g., Gorodnichenko, Pham, and Talavera, 2023) in addition to transparency. Likewise, investigating whether communication can build trust in the absence of a human touch remains an open question (see Laudenbach and Siegel, 2025).

I find that letters with more detailed risk discussions alleviate investors' anxiety. It should not be surprising that anxious readers would select into reading these discussions: acquiring information has long been known to psychologists as a coping strategy for anxiety (Lazarus and Folkman, 1984). More generally, Loewenstein, Weber, Hsee, and Welch (2001) argue that cognitive processes and emotional processes are interlinked in the face of risk. Information-avoidant investors who select out of reading fund letters should be unaffected by their contents, so fund letters are effectively a treatment targeted at those who would benefit from them. Still, it would be useful to understand how asset managers engage with avoidant clients.

As an unconditional effect, the medical literature has long understood the beneficial effects of communication: Hayward (1975) finds that surgical patients report lower anxiety (and even pain) when provided with information, and a meta-analysis by Hall, Roter, and Katz (1988) finds the amount of information provided strongly predicts patient satisfaction. The analogy between "money doctors" and medical doctors is not merely rhetorical.

My findings suggest that the asset managers who provide information and the investors who consume it are engaged in a mutually-beneficial ongoing interaction, beyond the initial investment delegation decision. Understanding how economic relationships are actively sustained over an extended period is an interesting and broad question.

Relatedly, economic actors appear to be aware of the virtue and possibility of fostering trust: In one survey, 93% of firm executives agreed that the "ability to build and maintain trust improves the bottom line"; in another survey, 70% of firm directors agreed that "enhancing shareholder communications (e.g., disclosures or reporting) can have a positive impact on stakeholder trust."[19] Given these prevailing attitudes, it would be useful to rigorously examine whether communication can build trust outside the asset management industry as well.

[19] Sources: https://www.pwc.com/us/en/library/trust-in-business-survey.html and https://corpgov.law.harvard.edu/2023/04/11/using-transparency-to-build-trust-a-corporate-directors-guide/

Appendix

Tables

Table I: Time-Series Variation in Fund Letters' Risk Details and Risk Levels

This table reports time-series regressions relating cross-fund indices of fund letters' risk descriptions to market-wide conditions. 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 (4) (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 portfolio, the independent variables measure aggregate expected returns (proxied by the AAII survey employed by Greenwood and Shleifer (2014)), and the perceived ambiguity of the S&P 500 portfolio (measured by Brenner and Izhakian (2018)). The sample of fund letters used to construct the dependent variables are written by S&P 500 index funds only.

Dependent Variables:	Risk Detail Index$_t$		Risk Level Index$_t$	
Measure:	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*

Table II: Relationship Between Letters' Risk Details, Risk Levels, and the Subsequent-Month S&P 500 Index Fund Flows

The panel regressions in this table describe the relationship between how risk is discussed in the voluntary letter disseminated by a fund i during year-month t and the subsequent flow into the fund, and show that flows are higher when communication about risk is more transparent. The two (potentially independent) textual features of the risk discussions used as independent variables are the amount of detail provided about risk, and the level of risk conveyed by each discussion. The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter written by fund i at time t discussing risk; see Eqn. (1). Column (1) shows the key relationship between Risk Detail$_{i,t}$ and the subsequent Net Flow$_{i,t\rightarrow t+1}$. Columns (2), (4) and (5) augment this regression with the secondary covariate Risk Level$_{i,t}$, measured in one of three ways: as the negative of the within-fund-standardized net sentiment of the text that discusses risk, using the Loughran and McDonald (2011) sentiment dictionary (column 2), or the net score of contextual risk words defined by Equation (4) (column 4), or a modified score that counts only high-risk words (column 5). Columns (3) and (6) include an interaction between the Risk Detail$_{i,t}$ and an indicator $\mathbb{1}$\{High Risk Level\}$_{i,t}$ for whether the letter conveys a high level of risk, defined as a below-median sentiment (column 3), or 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, total fee & (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$_{i,t\,i,t}$	0.0156**	0.0275***	0.0262***	0.0293***	0.0289***	0.0303**
	(0.0068)	(0.0100)	(0.0095)	(0.0105)	(0.0106)	(0.0118)
Risk Level$_{i,t}$		0.0711		0.0102	0.0163	
		(0.0786)		(0.0140)	(0.0171)	
$\mathbb{1}$\{High Risk Level\}$_{i,t}$			0.0144			0.5685
			(0.2141)			(0.4029)
Risk Detail$_{i,t}$			0.0040			-0.0186
× $\mathbb{1}$\{High Risk Level\}$_{i,t}$			(0.0082)			(0.0167)
Risk Level measure		Sentiment	Sentiment	High-Low Words	High Words	High Words
$\mathbb{1}$\{High Risk Level\} threshold			Median			0
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓	✓
N	2,238	1,569	1,569	1,579	1,579	1,579
R^2	0.37	0.40	0.40	0.41	0.41	0.41

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

Table III: Assessing the Economic Significance of the Flow–Risk Detail Relationship

The first covariate is the total annual fee charged by the fund i at time t. The second covariate Risk Detail$_{i,t}$ measures the fraction of the letter written by fund i at time t discussing risk. The third covariate is the negative of the net sentiment of this risk discussion, calculated by Eqn. (2). All independent variables are in units of standard deviation after having been transformed by z-scoring over the entire sample (hence the $\sigma(\cdot)$ labels). In columns 4–6, the dependent variable is also standardized. *Fund controls* (not shown) are for the fund's prior month return, square of this return & (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$_{i,t\to t+1}$ (%)			σ(Net Flow$_{i,t\to t+1}$)		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
σ(Total Fee$_{i,t}$)	-1.582**	-1.588**	-1.327**	-0.5201**	-0.5218**	-0.4360**
	(0.6472)	(0.6361)	(0.6126)	(0.2127)	(0.2091)	(0.2014)
σ(Risk Detail$_{i,t}$)		0.2829**	0.5073***		0.0930**	0.1667***
		(0.1238)	(0.1856)		(0.0407)	(0.0610)
σ(-Risk Net Sentiment$_{i,t}$)			0.0647			0.0213
			(0.0693)			(0.0228)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓	✓
N	2,238	2,238	1,588	2,238	2,238	1,588
R^2	0.37	0.37	0.41	0.37	0.37	0.41

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

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

Table IV: Corner Bunching Control Function Estimates of the Effect of Risk Detail on Fund Flows

This table reports estimates of Eqn. (10), which augments the baseline flow–risk detail regression with a control function $\lambda(\cdot)$ constructed from corner bunching in the treatment variable in order to correct for endogeneity. Columns 2–4 report results from various specifications of the control function, following Caetano, Caetano, and Nielsen (2024): a Tobit restriction, a semiparametric Tobit, and a nonparametric symmetry restriction. The coefficient β captures the effect of Risk Detail$_{i,t}$ (the fraction of the letter (%) written by fund i at time t discussing risk) on subsequent fund flows. The coefficient δ is on the correction term $\lambda(\cdot)$; it absorbs the component of variation in risk detail that is correlated with unobservables and therefore provides a direct estimate of the bias in the uncorrected regression. The procedures for estimating the endogeneity correction terms used in columns 2–4 are described in Internet Appendix E. *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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	0.014***	0.049***	0.050***	0.035***
	(0.005)	(0.017)	(0.014)	(0.013)
δ		-0.027*	-0.029***	-0.017*
		(0.014)	(0.01)	(0.01)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	2,238	2,238	2,238	2,238

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

Table V: Interaction of Communicated Risk Detail With 12b-1 Fees

This table shows that the positive effect of providing more transparency about risks on flows is higher for those funds that charge higher 12b-1 fees; i.e., the funds with investors that have a higher reliance on the services of financial intermediaries. The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. Actual 12b-1 Fee$_{i,t}$ is the fee recorded in the CRSP Mutual Fund Database. Effective 12b-1 Fee$_{i,t}$ is an alternative measure: the sum of the actual 12-b1 fee and the annualized front load, following Roussanov, Ruan, and Wei (2021). Fees are aggregated up from the share class level to the level of the mutual fund i, as described in Section 2.3. *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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)	(3)
Risk Detail$_{i,t}$	0.0156**	0.0077	0.0098*
	(0.0068)	(0.0051)	(0.0056)
Actual 12b-1 Fee$_{i,t}$ (bps)		0.0153	
		(0.0204)	
Risk Detail$_{i,t}$ × Actual 12b-1 Fee$_{i,t}$		0.0021**	
		(0.0009)	
Effective 12b-1 Fee$_{i,t}$ (bps)			-0.0252
			(0.0285)
Risk Detail$_{i,t}$ × Effective 12b-1 Fee$_{i,t}$			0.0007**
			(0.0003)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	2,238	2,238	2,238
R^2	0.37	0.38	0.37

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

Table VI: Testing for Any Communication-Driven Rebalancing Flows Out of Sibling Active Funds

Columns (1) and (2) show 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 500 index tracker funds. As a robustness check, the analysis is repeated for the active funds' letters' own risk detail measure in columns (3) and (4). Sibling active funds are defined as non-index domestic equity funds within the same family as the S&P 500 index funds that this study focuses on; columns (2) and (4) impose a further filter by requiring these active funds to also have a high market beta. The amount of risk detail is consistently measured as the letter fraction (%) discussing risk. *Fund controls* (not shown) are for the active fund's prior month return, square of this return, total fee & (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$_{i,t}$	-0.0086**	-0.0120***		
	(0.0040)	(0.0046)		
Sibling Active Fund Risk Detail$_{i,t}$			-0.0080**	-0.0135***
			(0.0039)	(0.0048)
Sibling Active Funds	All	$\beta^{MKT} > 1$	All	$\beta^{MKT} > 1$
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FES	✓	✓	✓	✓
Fund FES	✓	✓	✓	✓
N	14,788	7,662	14,788	7,662
R^2	0.77	0.78	0.77	0.78

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

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

Table VII: Interaction of Communicated Risk Detail With the Prior Risk Aversion Level

This table shows that the positive effect of providing more transparency about risks on flows is higher during times when investors' prior risk aversion levels are on average higher. Based on the test derived in Section 4.1, this result is consistent with transparency alleviating investors' effective risk aversion. The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. Risk aversion is measured by Bekaert, Engstrom, and Xu (2022) as a monthly, market-wide estimate of the coefficient of relative risk aversion of the representative investor; the indicator $\mathbb{1}\{\text{High Risk Aversion}\}_t$ takes the value 1 when their risk aversion measure is currently above its median. *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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 $t+1$.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)		
	(1)	(2)	(3)
Risk Detail$_{i,t}$	0.0135**	0.0045	0.0044
	(0.0065)	(0.0058)	(0.0058)
$\mathbb{1}\{\text{High Risk Aversion}\}_t$		-0.3344**	-0.3222**
		(0.1457)	(0.1472)
Risk Detail$_{i,t}$		0.0146**	0.0147**
$\times$ $\mathbb{1}\{\text{High Risk Aversion}\}_t$		(0.0060)	(0.0060)
Fund Controls$_{i,t}$	✓	✓	✓
Market Controls$_t$			✓
Fund FEs	✓	✓	✓
N	2,238	2,238	2,238
R^2	0.31	0.31	0.31

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

Table VIII: Variation in the Effect of Communicated Risk Detail Across Letter Readerships' Anxiety Attitudes

This table shows that the effect of providing more transparency about risks on flows is concentrated among more anxious readers, as revealed by local Google search activity. The main covariate Risk $\text{Detail}_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. The level of risk conveyed is calculated according to Eqn. (3). The sample consists of fund letters disseminated during each year-month t, and is further split into fund letter readerships with below-median anxiety levels (columns 1–2), and those with anxiety levels at the median or greater (columns 3–4) within each cross-section t. 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)). *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 asset-weighted average of the fund letter readership of state-level geographic variables. *Fund controls* (not shown) are for the 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 $\text{Flow}_{i,t\rightarrow t+1}$ (%)			
Sub-sample:	Low Anxiety		High Anxiety	
	(1)	(2)	(3)	(4)
Risk $\text{Detail}_{i,t}$	0.0093	0.0142	0.0199**	0.0374**
	(0.0111)	(0.0119)	(0.0092)	(0.0171)
Risk $\text{Level}_{i,t}$		0.0563		0.2332*
		(0.1035)		(0.1188)
Local Economy $\text{Controls}_{i,t}$	✓	✓	✓	✓
Demographic $\text{Controls}_{i,t}$	✓	✓	✓	✓
Fund $\text{Controls}_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	809	563	923	671
R^2	0.53	0.62	0.50	0.52

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

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

Figures

Figure I: Screening for Sentences that Discuss Risk

Statements about risk are defined as valid English sentences that are matched by both forward-looking and risk-related patterns. The percentage figures denote the fraction of all sentences in the overall fund letter corpus that are matched by each type of pattern.

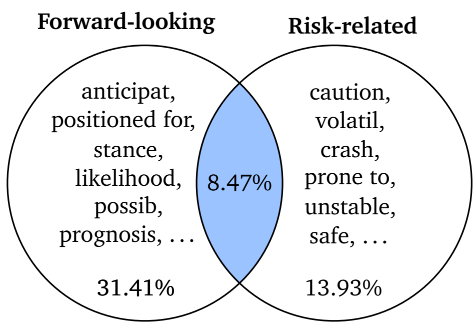
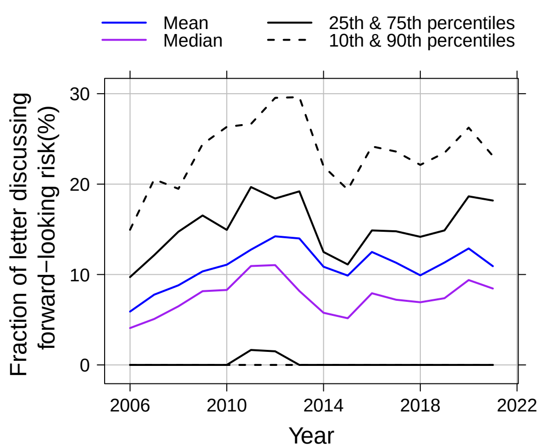


Figure II: Cross-Sectional Dispersion in How Fund Letters Discuss Risk

These charts summarize dispersion in the fraction of a fund letter that discusses risk (left-hand side) and the net sentiment of the words in that discussion (right-hand side). The net sentiment is computed using the Loughran and McDonald (2011) dictionary, according to Eqn. (2).



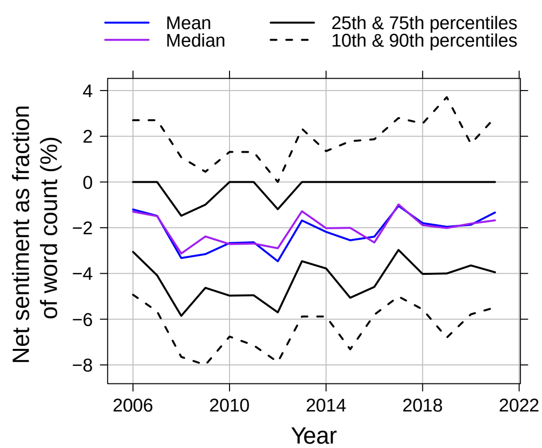


Figure III: Binned Scatterplot of Risk Detail vs Residualized Net Flow

The x-axis shows Risk $\text{Detail}_{i,t}$, which measures the fraction of the letter (%) written by fund i at time t discussing risk. The y-axis shows the corresponding Net $\text{Flow}_{i,t\to t+1}$ (%) after it has been residualized against controls. The set of controls comprises fund fixed effects, year-month fixed effects, the Risk $\text{Level}_{i,t}$ conveyed by the letter (computed using net sentiment scores), and other fund controls: the 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.

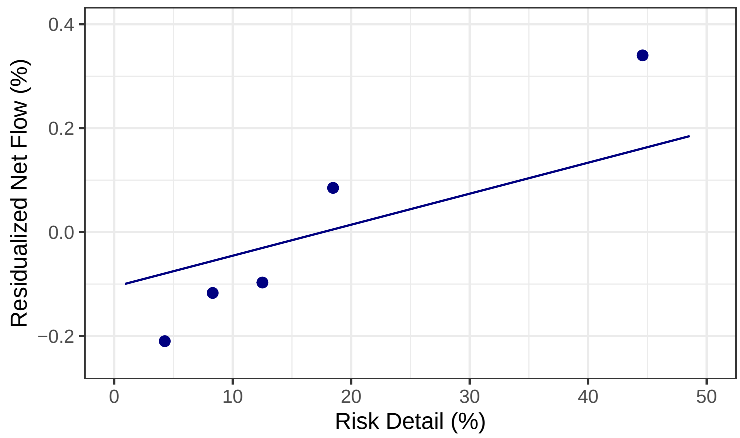
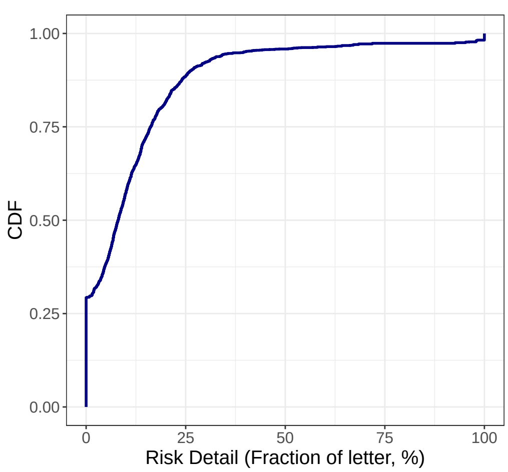


Figure IV: Corner Bunching in the Treatment Variable Used for Identification

These charts visualize the distribution of Risk $\text{Detail}_{i,t}$, which measures the fraction of the letter (%) written by fund i at time t discussing 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 also visible in the empirical CDF at the 0% value. The presence of this feature of the empirical distribution is used in the corner bunching identification strategy described in Section 3.2.



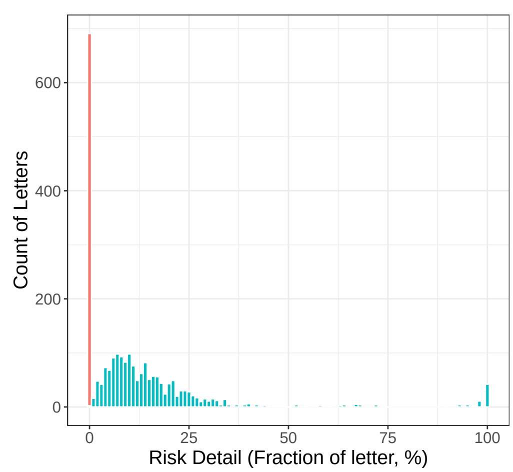


Figure V: Performance of Aggregate Communication-Driven Incremental Excess Returns

Sharpe ratio of the total additional holdings of S&P 500 index mutual funds that have been induced by risk transparency in fund letters (y-axis), assuming a variety of holding periods before those initial flows are withdrawn (x-axis). The excess returns used in the Sharpe ratio calculations are net of the fees levied by the S&P 500 index mutual funds who are disseminating each fund letter. The risk-free rate is computed using T-bills. Sharpe ratios are annualized. The shaded area represents the 95% confidence interval for each Sharpe ratio, calculated using quantile bootstraps with 1,000 replicates each.

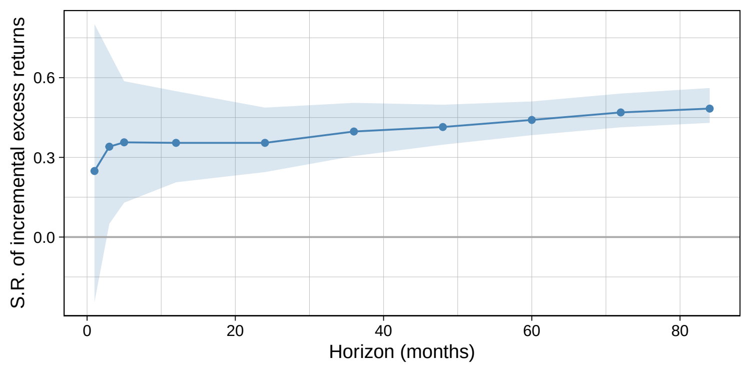
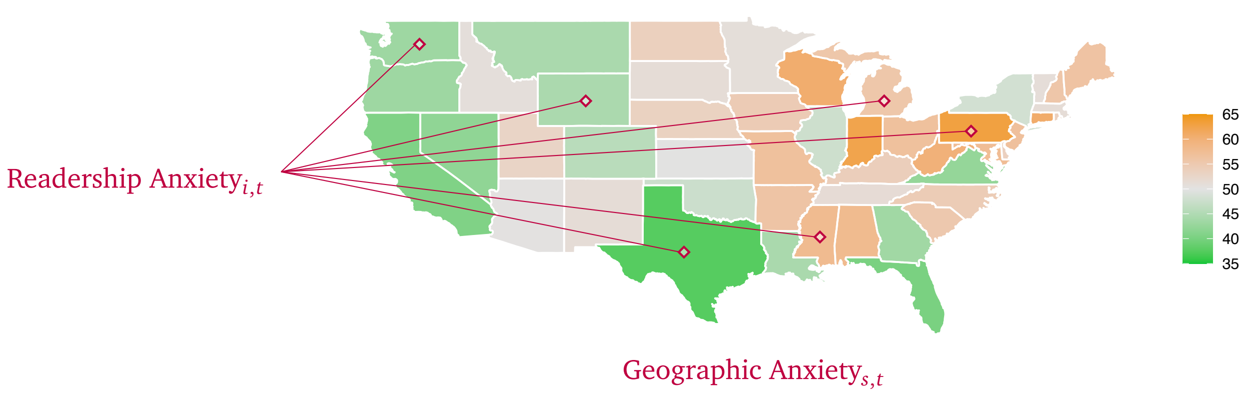


Figure VI: Illustration of Geographic Anxiety and its Aggregation to Readership Anxiety

This chart summarizes geographic variation in local anxiety attitudes revealed through Google search activity. Specifically, it displays means of cross-sectional rank percentiles for the combination of two Google Search Volume Indices for the topics of "psychological stress" and "worry", which Google themselves have defined. This index of anxiety attitudes, Geographic Anxiety$_{s,t}$, varies over time as well as geographically. By construction, the chart colors display persistent geographic variation: orange states denote a cross-sectionally higher level of anxiety; green states denote a lower level of anxiety; the legend is for mean cross-sectional rank-percentiles. The figure also illustrates the procedure of aggregating from Geographic Anxiety$_{s,t}$ to a fund-level measure of Readership Anxiety$_{i,t}$ using Eqn. (17).



References

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

Alesina, A., & La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, *85*(2), 207–234.

Algan, Y., & Cahuc, P. (2010). Inherited trust and growth. *American Economic Review*, *100*(5), 2060–92.

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

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

Bassi, A., Colacito, R., & Fulghieri, P. (2013). 'O sole mio: An experimental analysis of weather and risk attitudes in financial decisions. *Review of Financial Studies*, *26*(7), 1824–1852.

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.

Bekaert, G., Engstrom, E. C., & Xu, N. R. (2022). The time variation in risk appetite and uncertainty. *Management Science*, *68*(6), 3975–4004.

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., & Van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, *118*(1), 1–20.

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. (2024). Correcting for endogeneity in models with bunching. *Journal of Business & Economic Statistics*, *42*.

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. (2025). Managerial risk assessment and fund performance: Evidence from textual disclosure. *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.

Chen, J., Jiang, W., & Xiaolan, M. Z. (2022). The economics of mutual fund marketing. *Available at SSRN*, *4277000*.

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.

Cocco, J. F., Gomes, F. J., & Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. *Review of Financial Studies*, *18*(2), 491–533.

Cohn, A., Engelmann, J., Fehr, E., & Maréchal, M. A. (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review*, *105*(2), 860–885.

Cooper, M. J., Gulen, H., & Rau, P. R. (2005). Changing names with style: Mutual fund name changes and their effects on fund flows. *Journal of Finance*, *60*(6), 2825–2858.

deHaan, E., Song, Y., Xie, C., & Zhu, C. (2021). Obfuscation in mutual funds. *Journal of Accounting and Economics*, 101429.

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

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

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.

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

Fisman, R., & Khanna, T. (1999). Is trust a historical residue? Information flows and trust levels. *Journal of Economic Behavior & Organization*, *38*(1), 79–92.

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. Free Press.

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, *121*(2), 505–540.

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

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.

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.

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

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.

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.

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.

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.

Hillert, A., Niessen-Ruenzi, A., & Ruenzi, S. (2025). Mutual fund shareholder letters: Flows, performance, and managerial behavior. *Management Science*, *71*(5), 4453–4473.

Hu, A., & Ma, S. (forthcoming). Persuading investors: A video-based study. *Journal of Finance*.

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

Kamstra, M. J., Kramer, L. A., & Levi, M. D. (2003). Winter blues: A SAD stock market cycle. *American Economic Review*, *93*(1), 324–343.

Kaur, S., Mullainathan, S., Oh, S., & Schilbach, F. (2025). Do financial concerns make workers less productive? *Quarterly Journal of Economics*, *140*(1), 635–689.

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

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. (2025). Personal communication in an automated world: Evidence from loan repayments. *Journal of Finance*, *80*(1), 515–559.

Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer.

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.

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

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.

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.

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

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

Sheng, J., Xu, N., & Zheng, L. (forthcoming). Do mutual funds walk the talk? A textual analysis of risk disclosure by mutual funds. *Review of Financial Studies*.

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

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

A Summary Statistics

Table IA.1 summarizes monthly S&P 500 market-level variables over my study's sample period of 2006-2011.

My main sample is a panel of fund letters disseminated by S&P 500 index mutual funds i during year-months t, merged with characteristics of the funds disseminating these letters and of the families to which the funds belong. Table IA.2 summarizes this panel.

Figure IA.1 summarizes the fraction of N-CSR and N-CSRS filings that contain a fund letter.

Table IA.1: Summary of S&P 500 Characteristics

This table summarizes market time series during the period 2006-2021, all at a monthly frequency.

Market Variable	N	Mean	Std. Dev.	25th Percentile	Median	75th Percentile
AAII Survey Expected Return (%) (Greenwood and Shleifer, 2014)	192	3.277	12.145	−5.569	3.991	11.536
VIX	192	19.727	8.713	13.742	17.070	23.425
VVIX	190	91.375	15.983	81.460	89.920	100.542
Ambiguity (Brenner and Izhakian, 2018)	184	0.068	0.051	0.031	0.057	0.088
Risk Aversion (Bekaert, Engstrom, and Xu, 2022)	192	3.098	0.868	2.657	2.828	3.180

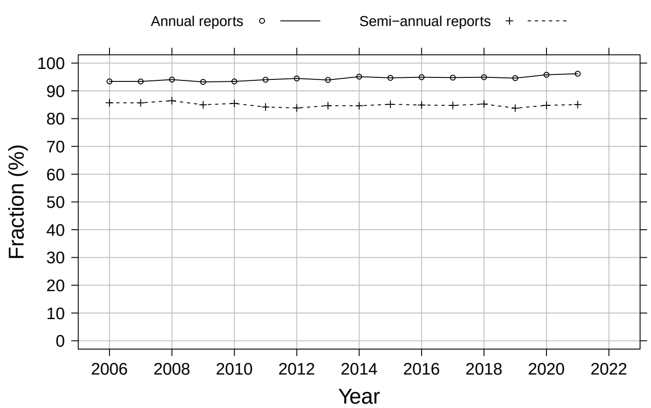
Table IA.2: Summary of Fund Letter Characteristics

This table shows summary statistics for the main sample used in the paper, which consists of letters disseminated by S&P 500 index mutual funds i during year-months t.

Fund Variable	N	Mean	Std. Dev.	25th Percentile	Median	75th Percentile
Net Flow during month $t+1$ (%)	2,353	−0.104	3.042	−1.101	−0.454	0.358
Fraction of Letter about Risk (%)	2,353	12.535	18.170	0.000	8.040	16.231
Loughran and McDonald (2011) Net Sentiment (%) of Letter Portion about Risk	1,664	−2.135	3.739	−4.387	−2.083	0.000
Return during month $t-1$ (%)	2,353	1.049	3.556	−1.449	1.641	3.676
Total Fee (%)	2,239	0.428	0.505	0.000	0.200	0.840
Log of Fund Size	2,351	20.286	2.017	18.961	20.435	21.629
Log of Family Age (in months)	2,310	5.213	0.361	4.990	5.275	5.454
Log of Family Size	2,310	11.035	2.047	9.985	11.321	12.305

Figure IA.1: Fraction of Filed Reports Containing Fund 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. A consistently high and stable fraction of filings were detected to have at least one fund letter.



B Example of a Fund Letter

Figure IA.2 shows an example fund letter addressed to the investors in one S&P 500 index fund.

Figure IA.2: 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

C Examples of Sentences Discussing Risk in Fund Letters

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 focuses on, and also from the wider overall corpus of fund letters.

Example Sentences 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 Sentences About Risk From Overall 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."*

D Validation of Textual Measures

This appendix validates my 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 a widely-used psychometrics software package that additionally suffers from a lack of transparency.

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 companies' communications.

A statement is classified as forward-looking if it contains at least one forward-looking word, as defined by Li (2010). Note that this dictionary was also employed by Hillert, Niessen-Ruenzi, and Ruenzi (2025).

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 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.3 compares the results of this paper's baseline methodology for detecting forward-looking sentences 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.4 compares the results of this paper's baseline methodology for detecting risk-related sentences 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.3: Forward-Looking Sentence Detection

Number & fraction of forward-looking sentences 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 patterns customized for the fund letter corpus	1,529,823	31.41%	—
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

Table IA.4: Risk-Related Sentence Detection

Number & fraction of risk-related sentences 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 patterns customized 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 companies' communications 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

E Corner Bunching Identification Methodology

I implement an identfication strategy based on the methodology of Caetano, Caetano, and Nielsen (2024, henceforth **CCN**). 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 their paper for a more formal exposition, as well as a companion paper (Caetano, Caetano, Nielsen, and Sanfelice, forthcoming).

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:

$$Y \;=\; \beta X \;+\; Z'\gamma \;+\; \underbrace{\delta\eta + \varepsilon}_{\text{Unobservable}}\;, \qquad \mathbb{E}[\varepsilon|X,Z,\eta] = 0, \tag{23}$$

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:

$$X = \max\{0, X^*\} = \max\{0,\, Z'\pi + \eta\}. \tag{24}$$

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

$$0 < \mathbb{P}(X^* < 0) < 1; \tag{25}$$

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.

From these first two assumptions, it follows that:

$$\mathbb{E}[Y|X,Z] \;=\; (\beta+\delta)X + Z'(\gamma - \pi\delta) + \delta\mathbb{E}[X^*|X^* \leq 0,\, Z]\,\mathbb{1}\{X = 0\} \tag{26}$$

$$\;=\; \beta X + Z'(\gamma - \pi\delta) + \delta\Big(\underbrace{X + \mathbb{E}[X^*|X^* \leq 0,\, Z]\,\mathbb{1}\{X = 0\}}_{\text{Correction term}}\Big). \tag{27}$$

That is, the causal effect β may be estimated if an correction term is also estimated and included in a regression. The coefficient on this correction term is δ. The sign and significance of δ are informative about the presence and direction of endogeneity.

In order to generate the correction term, note that since X and $\mathbb{1}\{X = 0\}$ are observed, only $\mathbb{E}[X^*|X^* \le 0, Z]$ 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

$$\mathbb{E}[X^*|X^* \le 0, Z] = \mathbb{E}[Z'\pi + \eta|X^* \le 0, Z] \tag{28}$$

$$= -Z'(\pi+\mu) - \sigma\lambda\left(\frac{-Z'(\pi+\mu)}{\sigma}\right), \tag{29}$$

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 $\eta|Z \sim \mathcal{N}\big(\mu(Z), \sigma^2(Z)\big)$; i.e. allowing μ, σ to be functions of Z, then

$$\mathbb{E}[X^*|X^* \le 0, Z] \;=\; \alpha(Z) - \sigma(Z)\lambda\left(\frac{-\alpha(Z)}{\sigma^2(Z)}\right), \tag{30}$$

$$\text{where } \alpha(Z) \;=\; Z'\pi + \mu(Z). \tag{31}$$

The parameters 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). Since this assumption does *not* assume the entire distribution is symmetric, it is quite general in scope. Following from this (weak) assumption,

$$\begin{aligned}\mathbb{E}[X^*|X^* \le 0, Z = z] \;=\;& F^{-1}_{X|Z=z}\big(1 - F_{X|Z=z}(0)\big) \\ & - \mathbb{E}\Big[X \Big| X \ge F^{-1}_{X|Z=z}\big(1 - F_{X|Z=z}(0)\big), Z = z\Big].\end{aligned} \tag{32}$$

Therefore, the empirical CDF can be used to identify the correction term, assuming that $\mathbb{P}(X = 0|Z = z) < 0.5$. The bunching probability $F_{X|Z=z}(0)$ is estimated using sample frequencies. The quantile $F^{-1}_{X|Z=z}\big(1 - F_{X|Z=z}(0)\big)$ is estimated using the empirical quantile of X among the observations with $Z = z$. And finally, the conditional trimmed mean $\mathbb{E}\Big[X\Big|X \ge F^{-1}_{X|Z=z}\big(1 -$

$F_{X|Z=z}(0)\big), Z = z\big]$ is estimated using a sample analog.

Assumptions 3a, 3b & 3c above offer alternative distributional assumptions for identifying the endogeneity correction term that is required by Equation (27), for the ultimate purpose of identifying the treatment effect β. All three shape restrictions are used in the empirical analysis of Section 3.2 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 (27) are identified, and furthermore that standard errors are consistently estimated by the bootstrap.

CCN also discuss the nuances of estimation when the vector of controls Z is high-dimensional. I closely follow their suggested approach of discretizing Z into clusters. Since my set of controls includes a variety of fixed effects, I use a large cluster size of $K = 50$; this is identical to the cluster size used in the companion paper by Caetano, Caetano, Nielsen, and Sanfelice (forthcoming).

F Validation of EDGAR-Based 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.3(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).

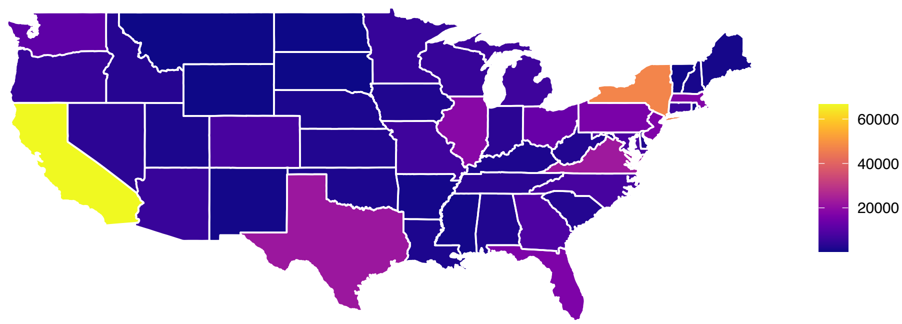
In fact, the correlation between state-level reader counts and the total state populations is 0.90 over the full sample used in this study, which runs from 2006 to the end of the SEC EDGAR log files in June 2017. At a more fine-grained unit of analysis, the correlation between county-level reader counts and the total county populations is also high, at 0.72. Figure IA.3(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.

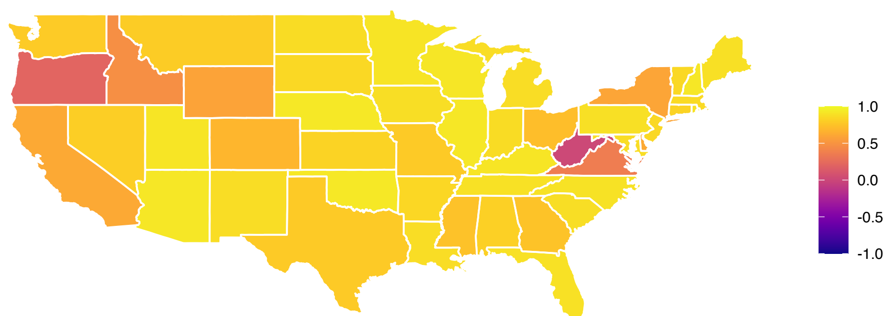
Figure IA.3: Representativity of Readership Geographic Distributions

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.



G The Geography of Anxiety

This appendix charts the geographic variation in Google search-based measures of anxiety (Figure IA.4). These geographic measures are used in the main paper to construct fund letter readership-level measures of anxiety, which is an important component of risk aversion.

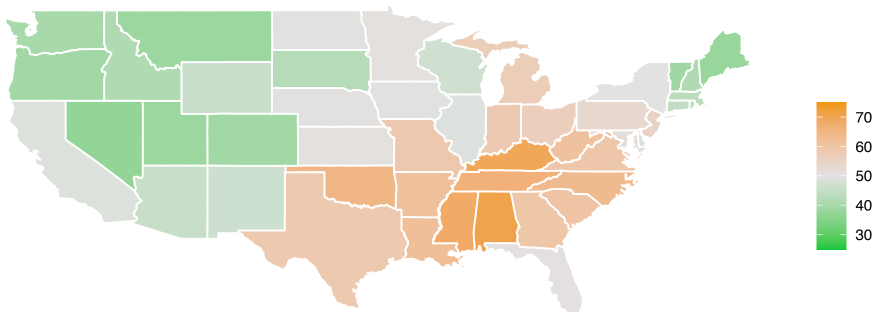
Recall that the Google search volume index (SVI) panel consists of state s and year-month t search volume indices for a particular topic k: $\text{SVI}^k_{s,t}$. To produce visualizations of search activity for a particular topic k, I first compute the rank-percentiles for each state s within a given 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.4 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) shows the average of both topic SVIs. Figure IA.4 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.

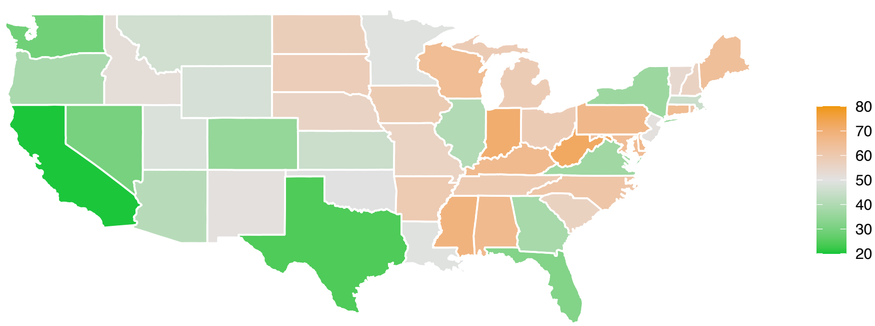
Figure IA.4: Geography of Anxiety Attitudes Revealed Through Search Interest

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. The legends are for mean cross-sectional rank-percentiles.

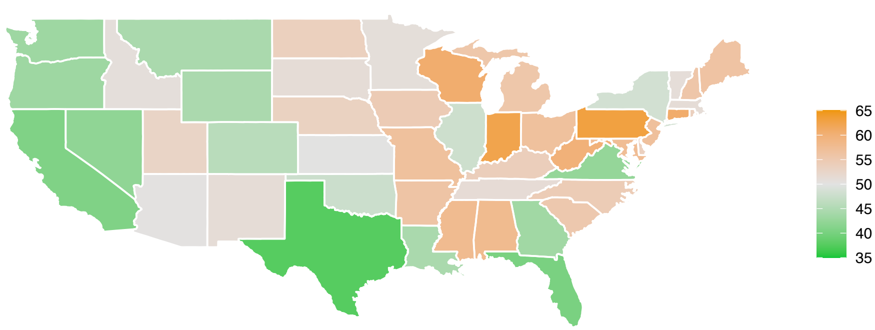
(a) "Psychological Stress" topic.



(b) "Worry" topic.



(c) Mean of both topics.



H Robustness Checks

H.1 Including Additional Fund-Level Controls

Table IA.5: Controlling for Fund Risk Exposure

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (6). *Common fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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$_{i,t}$	0.0156**	0.0168**	0.0167**	0.0170**
	(0.0068)	(0.0072)	(0.0073)	(0.0073)
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	2,238	2,023	2,023	2,014
R^2	0.37	0.34	0.34	0.34

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

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

Table IA.6: Controlling for Different Horizons of Past Fund Return

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (6). *Common fund controls* (not shown) are 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$_{i,t\to t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.0156**	0.0158**	0.0156**	0.0157**
	(0.0068)	(0.0068)	(0.0068)	(0.0068)
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	2,238	2,238	2,238	2,238
R^2	0.37	0.37	0.37	0.37

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

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

Table IA.7: Controlling for Past Fund Flows

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. In addition, specifications 2-4 include past fund flows as additional control variables; for the past month, past quarter, and past 6-month flows, respectively. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (6). *Fund controls* (not shown) are for the fund's prior month t return, square of this return, total fee & (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$_{i,t\to t+1}$			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$ (%)	0.0156**	0.0147**	0.0125**	0.0139**
	(0.0068)	(0.0068)	(0.0062)	(0.0065)
Net Flow$_{i,t-1\to t}$		0.1692**		
		(0.0665)		
Net Flow$_{i,t-3\to t}$			0.1286***	
			(0.0233)	
Net Flow$_{i,t-6\to t}$				0.0473***
				(0.0112)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	2,238	2,205	2,215	2,216
R^2	0.37	0.38	0.42	0.39

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

Table IA.8: Controlling for Morningstar Ratings

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. In addition, I include controls for each fund's Morningstar rating (also known as the Star Rating) where available, using a similar dataset of ratings to that of Ben-David, Li, Rossi, and Song (2022). The Star Rating control is implemented as a covariate in column (2), and as a set of fixed effects in column (3). *Fund controls* (not shown) are for the fund's prior month returns, 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$_{i,t}$	0.0156**	0.0224**	0.0221**
	(0.0068)	(0.0100)	(0.0110)
Morningstar controls:			
Star Rating$_{i,t}$		0.5104**	
		(0.2199)	
Star Rating FEs			✓
Other controls:			
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	2,238	1,220	1,220
R^2	0.37	0.41	0.44

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

H.2 Shannon's Entropy as an Alternative Measure of Risk Detail

Table IA.9: Using Shannon's Entropy to Measure Risk Detail

The analysis in this table confirms that the informational content of risk descriptions drives fund flows. Both the main covariates measure the amount of detail contained in the letter written by fund i at time t about risk, either using the fraction of the letter (column 1) or the Shannon's entropy of the distribution of word frequencies discussing risk (column 2), as follows:

$$H_{i,t} = -\sum_{k} \mathbb{P}(\text{word}_k) \log_2 \mathbb{P}(\text{word}_k). \tag{33}$$

This alternative measure of the risk detail 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. Net flows are calculated using the CRSP Mutual Fund Database, as in Eqn. (6). *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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)
Letter Fraction$_{i,t}$ (%)	0.0156**	
	(0.0068)	
Shannon's Entropy $H_{i,t}$		0.3120***
		(0.1098)
Fund Controls$_{i,t}$	✓	✓
Year-month FEs	✓	✓
Fund FEs	✓	✓
N	2,238	1,579
R^2	0.37	0.41

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

H.3 Sub-Sample Analysis When the Communicated Risk Level Is High

Table IA.10: Sub-Sample Analysis When Fund Letters Convey a High Level of Risk

This table modifies the sample used in the baseline flow-detail panel regression to verify results are robust to focussing only on high-risk states (as measured by letter content). The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing 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 fund letter. Columns (3) and (4) repeat the baseline specification of column (1) on the sub-sample of fund letters that communicate a high level of risk, measured as follows: a high-level of risk is denoted by below-mean sentiment (i.e., above-mean negated sentiment) in column (3) and by below-median sentiment (i.e., above-median negated sentiment) in column (4). Controls (not shown) are for each fund's prior month return, square of this return, total fee & (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$_{i,t}$	0.0156**	0.0275***	0.0307*	0.0562***
	(0.0068)	(0.0100)	(0.0167)	(0.0193)
Risk Level$_{i,t}$		0.0711		
		(0.0786)		
Risk Level Sample	All	All	Above-Mean	Above-Median
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	2,238	1,569	800	880
R^2	0.37	0.40	0.60	0.55

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

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

H.4 Focussing on Within-Fund Variation

Table IA.11: Within-Fund Flow–Risk Detail Relationship

This table repeats the baseline flow-risk detail panel regression with time effects removed, in order to confirm the main effect exists within-fund, i.e., for the important case of an ongoing relationship maintained over time between a fund and its investor base. The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. Variables labeled $\sigma(\cdot)$) have been standardized to be interpretable as z-scores. *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 $t+1$. The sample consists of S&P 500 index funds only.

Dependent Variables:	Net Flow$_{i,t\to t+1}$ (%)		σ(Net Flow$_{i,t\to t+1}$)	
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.0135**	0.0135**		
	(0.0065)	(0.0065)		
σ(Risk Detail$_{i,t}$)			0.0806**	0.0805**
			(0.0389)	(0.0390)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Market Controls$_t$		✓		✓
Fund FEs	✓	✓	✓	✓
N	2,238	2,238	2,238	2,238
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*

H.5 Gross Inflow vs. Risk Detail and Risk Level

Table IA.12: Relationship Between Letters' Risk Details, Risk Levels, and the Subsequent-Month S&P 500 Index Fund Gross Inflows

This table shows the relationship between how risk is discussed in the voluntary letter disseminated by a fund i during year-month t and the subsequent gross inflow into the fund, confirming the baseline results that use net flows. The main covariate Risk $\text{Detail}_{i,t}$ measures the fraction of the letter written by fund i at time t discussing risk; see Eqn. (1). Column (1) shows the key relationship between Risk $\text{Detail}_{i,t}$ and the subsequent Net $\text{Flow}_{i,t\rightarrow t+1}$. Columns (2), (4) and (5) augment this regression with the secondary covariate Risk $\text{Level}_{i,t}$, measured in one of three ways: as the negative of the within-fund-standardized net sentiment of the text that discusses risk, using the Loughran and McDonald (2011) sentiment dictionary (column 2), or the net score of contextual risk words defined by Equation (4) (column 4), or a modified score that counts only high-risk words (column 5). Columns (3) and (6) include an interaction between the Risk $\text{Detail}_{i,t}$ and an indicator $\mathbb{1}\{\text{High Risk Level}\}_{i,t}$ for whether the letter conveys a high level of risk, defined as a below-median sentiment (column 3), or 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, total fee & (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:	Gross $\text{Inflow}_{i,t\rightarrow t+1}$ (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Risk $\text{Detail}_{i,t\,i,t}$	0.0184*	0.0344**	0.0302**	0.0356**	0.0355**	0.0369**
	(0.0100)	(0.0138)	(0.0128)	(0.0148)	(0.0149)	(0.0163)
Risk $\text{Level}_{i,t}$		0.0110		0.0090	0.0110	
		(0.1448)		(0.0181)	(0.0224)	
$\mathbb{1}\{\text{High Risk Level}\}_{i,t}$			0.1046			0.5471
			(0.2711)			(0.5129)
Risk $\text{Detail}_{i,t}$			0.0079			-0.0194
$\times\ \mathbb{1}\{\text{High Risk Level}\}_{i,t}$			(0.0078)			(0.0186)
Risk Level measure		Sentiment	Sentiment	High-Low Words	High Words	High Words
$\mathbb{1}\{\text{High Risk Level}\}$ threshold			Median			0
Fund $\text{Controls}_{i,t}$	✓	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓	✓
N	2,238	1,569	1,569	1,579	1,579	1,579
R^2	0.74	0.77	0.77	0.77	0.77	0.77

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

H.6 An Alternative Time Series Measure of Prior Risk Aversion

Table IA.13: Interaction of Communicated Risk Detail With an Alternative Measure of the Prior Risk Aversion Level

The main covariate Risk $\text{Detail}_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. Risk aversion is proxied by the Risk sub-index of the Chicago Fed National Financial Conditions Index (Brave and Kelley, 2017). This proxy for risk aversion varies at a monthly frequency. The indicator $\mathbb{1}\{\text{High Risk Aversion}\}_t$ takes the value 1 when the risk aversion proxy is currently above its median. *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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 $t+1$.

Dependent Variable:	Net $\text{Flow}_{i,t\to t+1}$ (%)		
	(1)	(2)	(3)
Risk $\text{Detail}_{i,t}$	0.0135**	0.0073	0.0074
	(0.0065)	(0.0057)	(0.0057)
$\mathbb{1}\{\text{High Risk Aversion}\}_t$		-0.4691**	-0.4840**
		(0.2118)	(0.2088)
Risk $\text{Detail}_{i,t}$		0.0228*	0.0225*
$\times$ $\mathbb{1}\{\text{High Risk Aversion}\}_t$		(0.0117)	(0.0117)
Fund $\text{Controls}_{i,t}$	✓	✓	✓
Market Controls_t			✓
Fund FEs	✓	✓	✓
N	2,238	2,238	2,238
R^2	0.31	0.31	0.31

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

I Ruling Out Further Alternative Explanations

I.1 Bayesian Updating About Risk: Interactive Prior Risk Perceptions

This appendix presents details of the analysis described in Section 5.1.1, for the case of investor prior risk perceptions that are allowed to vary at the interacted fund i × and year-month t level.

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.

SEC EDGAR-Based Letter Readership Measure. 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 4.3 and Internet Appendix F.

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 focuses on. I then use the geo-locations of these readers j to construct an aggregate index for the prior held about risk by the readership of each fund i at each year-month t:

$$\text{Readership Prior}_{i,t} = z\left(\frac{1}{\sum_j \text{Investor Assets}_{j,t}} \sum_j \text{Investor Assets}_{j,t} \times \text{Geographic Prior}_{j,t}\right). \tag{34}$$

The above index is standardized within-fund to ensure comparability. Asset 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.

Google SVI-Based Risk Perception Measure. 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 time series 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:

$$\text{Geographic Prior}_{s,t} = \text{SVI}_{s,t}^{\text{stock market crash topic}} \tag{35}$$

Validation of the Google SVI-Based Measure. The United States-wide aggregate version of the index $\text{SVI}_t^{\text{stock market crash topic}}$ covaries strongly with the VIX. Table IA.14 compares the aggregate United States-wide version of the index, SVI_t, to the VIX index of aggregate implied stock market volatility. Columns 1-3 compare levels and include controls for a time t trend and quadratic t^2 trend: the SVI_t and VIX are positively and significantly associated in each. Columns 4-5 of the same table 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. I conclude from these validation checks that Google search interest in the topic of a stock market crash is a valid 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.

Defining the Cross-Section of Geographic Priors. Having confirmed that the Google SVI measure captures investor perceptions about stock market risk, I now make use of state s-month t cross-sectional variation in this index. Figure IA.5 charts geographic dispersion in the population's stock market risk perceptions (but does not show variation over time). Note that this measure has little overlap with the geographic anxiety attitudes shown in Internet Appendix G – the lack of overlap with anxiety attitudes (which influence risk aversion) confirms that this measure (which captures the perceived level of risk) is distinct.

Defining the Cross-Section of Belief Updates. I measure the strength of belief updating about investors' perceived level of risk that is induced by the receipt of a fund letter as follows. In close analogy to the term $s - m_p$ in Eqn. (19), the update is defined as the difference between the letter's signal about risk and investors' priors,

$$\text{Risk Level Update}_{i,t} = \text{Letter Signal}_{i,t} - \text{Readership Prior}_{i,t}, \tag{36}$$

with the first term is defined by either Eqn. (3) or (4) (depending on the measure employed), and the second term is defined by Eqn. (34).

Results. Table IA.15 shows a set of panel regression estimates that build upon the previous regressions of Table II, 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 IA.15 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).

Figure IA.5: Geography of Stock Market Risk Perceptions Revealed Through Search Interest

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. The legend is for mean cross-sectional rank-percentiles.

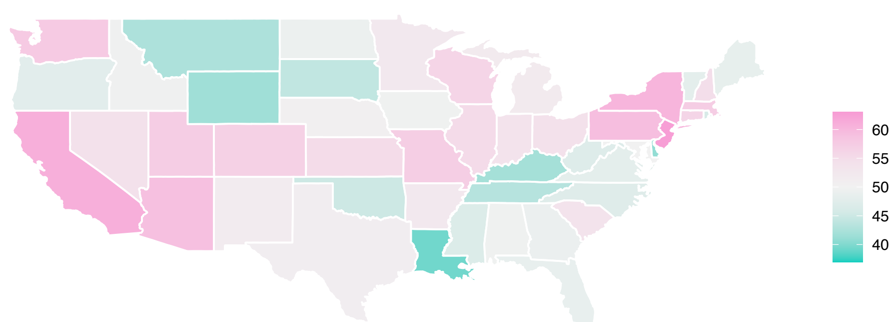


Table IA.14: Comparing Aggregate Google Search Interest to the VIX

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*

Table IA.15: Controlling for Updates in Perceived Risk Levels With Fund × Time-Varying Priors

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. 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. (3); in columns (3) and (5) it is measured as the net score of contextual risk words, as defined by Eqn. (4). The prior perceived level of risk varies at the interacted fund × year-month level and is measured based on local Google searches, as defined by Eqn. (34) The update term is the difference between the two, as defined by Eqn. (36). *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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$_{i,t}$	0.0179**	0.0281**	0.0168**	0.0283**	0.0169**
	(0.0081)	(0.0110)	(0.0082)	(0.0111)	(0.0083)
Risk Level Update$_{i,t}$		0.1394	0.1576*		
		(0.1000)	(0.0888)		
Letter Signal$_{i,t}$				0.1294	0.1346
				(0.1009)	(0.0952)
Readership Prior$_{i,t}$				-0.3424	-0.3622
				(0.3167)	(0.3473)
Risk Level measure		Sentiment	High-Low Words	Sentiment	High-Low Words
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓
N	1,755	1,219	1,745	1,219	1,745
R^2	0.34	0.37	0.34	0.37	0.34

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

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

I.2 Bayesian-Like Updating About Risk: Gaining Precision

Based on the alternative hypothesis laid out in Section 5.1.2, this appendix tests whether more detailed communication about risk interacts with the strength of investors' priors about the level of risk.

I measure the prior precision of the representative investor 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.[20]

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 IA.16 presents the results of just such an analysis. To be precise, I repeat this paper's main flow–risk 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. These results suggest the underlying persuasive mechanism is unlikely to be belief-based.

[20]Source: "How can an investor use the VVIX?", `https://www.cboe.com/us/indices/dashboard/vvix/`.

Table IA.16: Flow–Risk Detail Relationship Interacted With the *Ex Ante* Strength of the Prior About Risk

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. 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 the fund's prior month return, square of this return, total fee & (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$_{i,t\to t+1}$ (%)		Inflow$_{i,t\to t+1}$ (%)	
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.0156**	0.0129*	0.0184*	0.0146*
	(0.0068)	(0.0074)	(0.0100)	(0.0087)
Risk Detail$_{i,t}$		0.0056		0.0080
$\times\ \mathbb{1}\{\text{Low Prior Strength}\}_t$		(0.0063)		(0.0076)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
N	2,238	2,238	2,238	2,238
R^2	0.37	0.37	0.74	0.74

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

I.3 Learning About the Return: Gaining Precision as a Bayesian

Based on the alternative hypothesis laid out in Section 5.1.3, this appendix tests whether information provision about risk leads to a higher posterior precision of the return. I test for 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 extract such sentences by modifying the set of patterns used in the main text. The new set of screens is illustrated by Figure IA.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 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 IA.17 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 IA.17 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.

Figure IA.6: Screening for Sentences that Discuss Performance or Returns

Statements about performance or returns are defined as valid English sentences that are matched by both forward-looking and performance or return-related patterns. The percentage figures denote the fraction of all sentences in the fund letter corpus that are matched by each type.

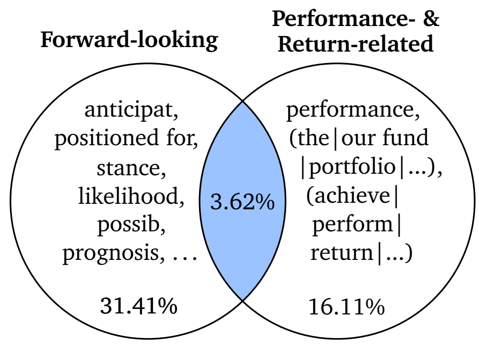


Table IA.17: Placebo Test for the Amount of Detail Communicated about Future Performance and Returns on S&P 500 Index Fund Flows

The main covariate measures the fraction of the letter written by fund i at time t about performance & returns in a forward-looking manner. *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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$_{i,t}$	0.0003	0.0079	0.0020
	(0.0080)	(0.0164)	(0.0037)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	2,238	2,238	2,238
R^2	0.37	0.74	0.83

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

I.4 Learning About the Return: Predicting the Return

Table IA.18: Placebo Test for Predicting the S&P 500 Return Using Time-Series Indices of Risk Detail and Risk Level

The dependent variables (in percentage points) are the return of the S&P 500 index from the end of month t to the end of month $t+h$, for various horizons of h. The independent variables are the equally-weighted across funds and standardized within fund number of words devoted to discussing risk, and our standard measure of the level of risk implied by the sentiment of that communication. Each independent variable is calculated as the cross-sectional mean over individual fund i variables, $\frac{1}{N_t}\sum_i z_{i,t}$, and these individual fund variables $z_{i,t}$ are pre-standardized to focus purely on within-fund variation, $z_{i,t} = (x_{i,t} - \overline{x_i})/\sigma(x_i)$. The sample of funds used to construct the independent variables consists of S&P 500 index funds only.

Dependent Variables:	S&P 500 Return$_{t\to t+1}$		S&P 500 Return$_{t\to t+3}$		S&P 500 Return$_{t\to t+5}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Risk Detail$_t$	0.3937		1.105		1.166	
	(0.6247)		(1.416)		(1.878)	
Risk Level$_t$		-0.4570		-0.1984		-1.335
		(0.7359)		(1.269)		(2.638)
N	187	187	187	187	187	187
R^2	0.001	0.003	0.003	0.0001	0.002	0.004

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

I.5 Learning About Manager Skill

Table IA.19: Placebo Test for Predicting the Tracking Error

Consistent with Chevalier and Ellison, 1997; Grinold and Kahn, 1999; Cremers and Petajisto, 2009, the tracking error is defined as the standard deviation of the difference between monthly fund returns and the S&P 500 index returns, from the end of the current month t over some future horizon of h months:

$$\text{Tracking Error}_{i,t\to t+h} = \text{StdDev}\left\{r^{Fund}_{i,t+k} - r^{S\&P\,500}_{t+k} \;\middle|\; 1 \leq k \leq h\right\}. \tag{37}$$

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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:	Tracking Error$_{i,t\to t+h}$			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.00003	-0.00006*	0.00005	0.0001
	(0.00003)	(0.00004)	(0.00006)	(0.00006)
Horizon h (months)	6	12	24	36
Fund Controls$_{i,t}$	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓
N	1,153	1,147	959	895
R^2	0.83	0.75	0.60	0.60

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

I.6 Being Educated by Information Provision

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) using geographically-targeted surveys. By matching fund letter readers' locations to geographic states, I am able to aggregate these state-level scores to fund letter readership measures. Next 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.20; 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; however, they are not significant. These results do not support an educational channel.

In fact, the positive and significant interaction coefficient in column (3) suggests the opposite effect may be present: low-financial literacy investors appear more likely to withdraw assets when receiving more communication about risk.

Table IA.20: Flow–Risk Detail Relationship Interacted With Letter Readers' Financial Literacy

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. 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 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 Variables:	Net Flow	Inflow	Outflow
	(1)	(2)	(3)
Risk Detail$_{i,t}$	0.0148*	0.0190	-0.0067
	(0.0078)	(0.0126)	(0.0043)
Risk Detail$_{i,t}$	0.0062	0.0089	0.0063**
$\times$ $\mathbb{1}\{\text{Low Financial Literacy}\}_{i,t}$	(0.0062)	(0.0081)	(0.0031)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs$_t$	✓	✓	✓
Fund FEs	✓	✓	✓
N	1,755	1,755	1,755
R^2	0.34	0.68	0.80

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

I.7 Are Letters Conveying Opportunities to Time the Market?

Table IA.21: Robustness to Excluding Any Discussion of Opportunities

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. The level of risk conveyed, Risk Level$_{i,t}$, is measured based on the net sentiment of the text discussing risk, and is defined by Eqn. (3). *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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 $t+1$. The sample consists of S&P 500 index funds only, and only those letters that do not discuss an opportunity.

Dependent Variable:	Net Flow$_{i,t\rightarrow t+1}$ (%)			
	(1)	(2)	(3)	(4)
Risk Detail$_{i,t}$	0.0106**	0.0242**	0.0245**	0.0333**
	(0.0052)	(0.0106)	(0.0105)	(0.0130)
Risk Level$_{i,t}$	0.0153	0.0095	0.0104	-0.0019
	(0.0848)	(0.0689)	(0.0690)	(0.0772)
Fund Controls$_{i,t}$	✓	✓	✓	✓
Market Controls$_t$			✓	
Year-month FEs	✓			✓
Fund FEs		✓	✓	✓
N	1,408	1,408	1,408	1,408
R^2	0.17	0.32	0.32	0.42

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

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

I.8 Does More Communication Reduce Perceived Ambiguity?

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:

$$\mho^2 = \int \mathbb{E}[\varphi(r)]\,\mathbb{V}\mathrm{ar}[\varphi(r)]\,dr, \tag{38}$$

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."[21]

Brenner and Izhakian (2018, pp. 510-512) show how to estimate a monthly time series 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.22 repeats the paper's baseline flow-versus-risk detail panel regressions, now including 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 somewhat 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. The coefficient on this interaction is not significant.

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., a positive and significant interaction).

[21] 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.

Table IA.22: Flow–Risk Detail Relationship Interacted With the Perceived Ambiguity of the S&P 500 Return

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. The indicator variable $\mathbb{1}\{\text{High Ambiguity}\}_t$ is 1 when the aggregate ambiguity perception of the market is above the median, 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. (38). *Fund controls* (not shown) are for the fund's prior month return, square of this return, total fee & (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 $t+1$. The sample consists of S&P 500 index funds only.

Dependent Variable:	Net Flow$_{i,t\to t+1}$ (%)		
	(1)	(2)	(3)
Risk Detail$_{i,t}$	0.0134**	0.0214**	0.0171**
	(0.0065)	(0.0104)	(0.0084)
Risk Detail$_{i,t}$ × Ambiguity$_t$		-0.0160	
		(0.0107)	
Risk Detail$_{i,t}$ × $\mathbb{1}\{\text{High Ambiguity}\}_t$			-0.0070
			(0.0059)
Fund Controls$_{i,t}$	✓	✓	✓
Market Controls$_t$	✓	✓	✓
Fund FEs	✓	✓	✓
N	2,238	2,238	2,238
R^2	0.31	0.31	0.31

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

I.9 Shrouding: Textual Complexity

Table IA.23: Controlling for the Textual Complexity of Risk Discussions

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. In addition, I control first separately and then jointly for four alternative measures of textual complexity, which are computed on the risk-related portions of the letters: the Flesch Reading Ease (Flesch, 1948), the Flesch-Kincaid Grade Level (Kincaid, Fishburne Jr, Rogers, and Chissom, 1975), the FOG Index (Gunning, 1952), and the SMOG Index (McLaughlin, 1969). *Fund controls* (not shown) are 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$_{i,t}$	0.0299***	0.0295***	0.0278***	0.0299***	0.0277***	0.0255**
	(0.0104)	(0.0105)	(0.0100)	(0.0105)	(0.0105)	(0.0099)
Flesch Reading Ease$_{i,t}$		-0.0055				0.0356
		(0.0072)				(0.0263)
Flesch-Kincaid Grade Level$_{i,t}$			0.0563*			0.1338
			(0.0286)			(0.1438)
FOG Index$_{i,t}$				7.16×10^{-5}		0.0005
				(0.0005)		(0.0013)
SMOG Index$_{i,t}$					0.0671*	0.0517
					(0.0380)	(0.1450)
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓	✓
N	1,579	1,579	1,579	1,579	1,579	1,579
R^2	0.41	0.41	0.41	0.41	0.41	0.41

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

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

I.10 Marketing: Attracting New Investors

This appendix tests whether new investors are attracted by voluntary transparency about risk addressed to existing investors (in fund letters).

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.24 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.24: Risk Detail Does Not Predict New Investor Counts

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 Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. 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 the fund's prior month return, square of this return, total fee & (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)
Risk Detail$_{i,t}$	-0.0032	-0.0031	-0.0036	-0.0049	-0.0053
	(0.0029)	(0.0034)	(0.0035)	(0.0036)	(0.0035)
Historical Mean of New Investors$_{i,t}$	0.0161***	0.0112***	0.0115***	0.0053	0.0049
	(0.0032)	(0.0031)	(0.0039)	(0.0039)	(0.0037)
Lookback k (months)	12	24	36	48	60
Fund Controls$_{i,t}$	✓	✓	✓	✓	✓
Year-month FEs	✓	✓	✓	✓	✓
Fund FEs	✓	✓	✓	✓	✓
N	1,640	1,560	1,488	1,335	1,172

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

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

I.11 Marketing: Potential Omitted Variable

Table IA.25: Controlling for Marketing Intensity Using a Fee-Based Measure

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. Actual 12b-1 Fee$_{i,t}$ is the fee recorded in the CRSP Mutual Fund Database. Effective 12b-1 Fee$_{i,t}$ is an alternative measure: the sum of the actual 12-b1 fee and the annualized front load, following Roussanov, Ruan, and Wei (2021). *Fund controls* (not shown) are for the fund's prior month returns, 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)	(3)
Risk Detail$_{i,t}$	0.0156**	0.0159**	0.0157**
	(0.0068)	(0.0069)	(0.0068)
Actual 12b-1 Fee$_{i,t}$ (bps)		0.0307	
		(0.0198)	
Effective 12b-1 Fee$_{i,t}$ (bps)			-0.0123
			(0.0275)
Fund Controls$_{i,t}$	✓	✓	✓
Fund FEs	✓	✓	✓
Year-month FEs	✓	✓	✓
N	2,238	2,238	2,238
R^2	0.37	0.37	0.37

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

Table IA.26: Controlling for Marketing Intensity Using a Labor-Based Measure

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. The regressions in columns (2) and (3) include additional control variables that are constructed by closely following Chen, Jiang, and Xiaolan (2022) and similarly using Form ADV filings: MKT is the share of the fund company's employees engaged in marketing- and sales-related functions, and EMP is the total number employees dislosed on Form ADV. *Fund controls* (not shown) are for the fund's prior month returns, 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$_{i,t}$	0.0156**	0.0150**	0.0150**
	(0.0068)	(0.0068)	(0.0068)
MKT$_{i,t}$		0.9924**	1.027*
		(0.4548)	(0.5848)
log(1+EMP$_{i,t}$)			-0.0071
			(0.0569)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	2,238	2,238	2,238
R^2	0.37	0.37	0.37

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

I.12 Pandering

Table IA.27: Flow–Risk Detail Relationship Interacted With the Distance of the Signal from Investors' Priors

The main covariate Risk Detail$_{i,t}$ measures the fraction of the letter (%) written by fund i at time t discussing risk. The specifications in columns 2–3 furthermore incorporate the distance between investors' prior perceived risk and the signal conveyed by the fund letter:

$$\text{Distance from Prior}_{i,t} = |\text{Letter Signal}_{i,t} - \text{Readership Prior}_{i,t}|, \tag{39}$$

where the letter's signal about risk is measured based on the net sentiment of the text discussing risk (Eqn. (3)), and the prior perceived level of risk is measured based on local Google searches (Eqn. (34)). Controls (not shown) are for the fund's prior month return, square of this return, total fee & (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)	(3)
Risk Detail$_{i,t}$	0.0179**	0.0311***	0.0248**
	(0.0081)	(0.0118)	(0.0115)
Distance from Prior$_{i,t}$		0.1510	0.0498
		(0.0934)	(0.1267)
Risk Detail$_{i,t}$ × Distance from Prior$_{i,t}$			0.0062
			(0.0045)
Fund Controls$_{i,t}$	✓	✓	✓
Year-month FEs	✓	✓	✓
Fund FEs	✓	✓	✓
N	1,755	1,219	1,219
R^2	0.34	0.37	0.37

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

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.

Bekaert, G., Engstrom, E. C., & Xu, N. R. (2022). The time variation in risk appetite and uncertainty. *Management Science*, *68*(6), 3975–4004.

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.

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

Brave, S. A., & Kelley, D. (2017). Introducing the Chicago Fed's new adjusted national financial conditions index. *Chicago Fed Letter*, *386*(2017).

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. (forthcoming). The effect of maternal labor supply on children: Evidence from bunching. *Journal of Labor Economics*.

Caetano, C., Caetano, G., & Nielsen, E. R. (2024). Correcting for endogeneity in models with bunching. *Journal of Business & Economic Statistics*, *42*.

Chen, J., Jiang, W., & Xiaolan, M. Z. (2022). The economics of mutual fund marketing. *Available at SSRN*, *4277000*.

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

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

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.

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.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221.

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*.

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

Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.

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.

Hillert, A., Niessen-Ruenzi, A., & Ruenzi, S. (2025). Mutual fund shareholder letters: Flows, performance, and managerial behavior. *Management Science*, *71*(5), 4453–4473.

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

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel* (tech. rep.).

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.

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*.

McLaughlin, G. H. (1969). Smog grading – a new readability formula. *Journal of Reading*, *12*(8), 639–646.

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