

Learning from Prospectuses

SIMONA ABIS

Columbia University

ANDREA M. BUFFA

University of Colorado Boulder

APOORVA JAVADEKAR

Indian School of Business

ANTON LINES

Columbia University

December 24, 2021*

Abstract

We study qualitative information disclosures by mutual funds when investors learn from such disclosures in addition to past performance. We show theoretically that fund managers with specialized strategies optimally choose to disclose detailed strategy descriptions, while those with standardized strategies provide generic descriptions. Generic descriptions lead to benchmarking errors by investors who confuse factor returns and skill, resulting in higher fund flow uncertainty. While all managers dislike this uncertainty, those with above-average factor exposures also benefit from the errors on balance and thus grow larger. We find evidence for this trade-off in the data, using a comprehensive dataset of fund prospectuses: funds with more informative descriptions are smaller and more specialized, exhibit higher flow-performance sensitivity, and show lower correlation between size and flow volatility. Investors in these funds make fewer benchmarking errors, and the effects are more pronounced for funds with shorter return histories.

*Contacts: simona.abis@gsb.columbia.edu, buffa@colorado.edu, apoorva_javadekar@isb.edu, and anton.lines@gsb.columbia.edu. For helpful comments and discussions, we thank seminar participants at Columbia University, Carnegie Mellon University, National University of Singapore, University of Bristol, University of Washington, and Johns Hopkins University.

Mutual fund prospectuses provide you with important information so you understand how the fund works and can easily compare it with other funds. If you wish to make an informed investment decision, you should read the prospectus before buying or selling shares in a mutual fund.

– Securities and Exchange Commission, “How to Read a Mutual Fund Prospectus”¹

1 Introduction

Prospectuses are the primary source of information available to mutual fund investors in the United States, and the original source for many items in popular databases such as CRSP and Morningstar. The SEC requires funds to update and distribute these documents at regular intervals and specifies the kind of information they must contain. This includes typical quantitative information such as fees and past performance, as well as narrative descriptions of the fund’s objectives, strategies, and risk exposures. Judging from the regulator’s communications with investors, it clearly considers these narrative sections to be at least as important as the quantitative information.² However, while the role of fees and performance in investor decision-making has been studied extensively,³ much less is known about the economic role of the narrative disclosures.

To explore this role we focus on the “Principal Investment Strategies” (PIS) section of fund prospectuses. We construct a measure of informativeness (*financial content*) and show that it displays significant variation in the panel of PIS descriptions. This heterogeneity cannot be fully explained by the fund’s mandate or other fund-level characteristics, and does not present a strong correlation with differences in fund performance. That motivates the questions we ask in this paper: how do managers determine their optimal disclosure policies, and how do investors make use of this information? To address them, we develop an equilibrium model of the interaction between fund managers’ qualitative disclosure choices and investor learning. The model yields testable predictions which we validate empirically using a comprehensive sample of prospectus strategy descriptions.

Despite their lack of prominence in the academic literature, PIS descriptions are widely available to the public via the SEC’s Electronic Data Gathering, Analysis, and Retrieval (EDGAR)

¹https://www.sec.gov/oiea/investor-alerts-bulletins/ib_mfprospectus3.html

²For an additional example, see the SEC’s Beginner’s Guide to Mutual Funds: <https://www.sec.gov/reportspubs/investor-publications/investorpubsbeginmutualhtm.html>

³See, for example, Chevalier and Ellison (1997); Sirri and Tufano (1998); Baks et al. (2001); Pástor and Stambaugh (2002); Berk and Green (2004); Barber et al. (2005); Ivković and Weisbenner (2009); Barber et al. (2016); Berk and Van Binsbergen (2016).

system,⁴ industry databases such as Morningstar, and (indirectly) through the recommendations of financial advisors. In this section the SEC requires funds to “explain in general terms how the Fund’s adviser decides which securities to buy and sell”, but allows fund managers substantial discretion as to how much detail they ultimately choose to provide.⁵ Indeed, [Abis and Lines \(2020\)](#) document high dispersion in length and linguistic complexity in the panel of these descriptions.

Length and complexity should be positively related with the amount of information being disclosed, but they are poor proxies. For those reasons, we construct a targeted measure of informativeness: the proportion of the text of each PIS section which contains non-boilerplate financial content. Financial terms and short phrases are identified from a comprehensive list: Campbell Harvey’s Hypertextual Finance Glossary.⁶ We ensure informativeness by first removing financial terms found in more than 75% of all strategy sections,⁷ and then weighting each remaining term by the inverse of its frequency in the prospectuses of funds with linguistically similar mandates. Mandates are identified utilizing the Strategy Peer Groups (SPGs) of [Abis and Lines \(2020\)](#). SPGs are obtained by clustering PIS descriptions into 17 linguistically similar groups, through machine learning. Those steps ensures that common financial terms for every mandate are down-weighted, while more targeted descriptions assume a greater relevance.⁸

Our financial content measure is positively related to word count and linguistic complexity and is negatively related to standard measures of boilerplate language. As previously mentioned, it exhibits large heterogeneity: from 7% to 35% of the total word counts between the 1st and 99th percentiles of the measure’s distribution. Only 19.17% of this variation can be explained by a fund’s mandate, the quarter of analysis or other common fund-level controls; while not presenting a strong correlation with fund performance. These observations motivate the premise of our model which, in keeping with our measure’s construction, is to be intended *within* each mandate.

Consistent with the SEC’s view (expressed in the introductory quotation) that a prospectus provides information on appropriate comparisons for the fund, we hypothesize that prospectuses help investors to estimate the correct performance benchmark (i.e., a fund’s exposure to passive factors). While the typical assumption is that fund strategies are perfectly known,⁹ investors

⁴<https://www.sec.gov/edgar/searchedgar/mutualsearch.html>

⁵PIS corresponds to item 9 of N-1A mandatory disclosures: <https://www.sec.gov/files/formn-1a.pdf>

⁶Glossary link: <https://people.duke.edu/~charvey/classes/wpg/glossary.htm>

⁷That only removes the stemmed terms: invest, fund, compani, secur, market, stock, asset, cap, manag.

⁸For example, *Quantitative* funds might need to utilized more detailed terminology in order to explain their strategy. Hence, if constructing our measure relative to the terminology utilized by the universe of funds, all quantitative funds might appear to choose to disclose more information.

⁹A notable exception is the recent paper by [Franzoni and Schmalz \(2017\)](#), who show that fund flows become less sensitive to performance when aggregate risk-factor realizations are extreme, because in those states investors’ ability to learn about managerial skill is reduced.

in our framework must simultaneously infer exposure to the benchmark and managerial skill by reading the fund’s strategy description *and* observing its returns. The more information a manager provides about her strategy, the more accurately investors perceive the benchmark—and thus the manager’s skill. The manager, in turn, takes anticipated investor behavior into account when choosing her disclosure policy.

The full intuition behind the model is as follows. Since management fees are proportional to fund size, risk averse managers prefer larger funds but dislike fund flow uncertainty. We consider a competitive capital market à la Berk and Green (2004), where funds experience decreasing returns to scale and investors supply capital until expected returns in excess of passive benchmarks are equal to zero. A fund’s total return is comprised of an active component, driven by the skill of the manager, and a passive component, driven by a common risk factor. Investors read prospectuses in order to learn about the fund’s exogenous loading on this common factor (i.e., its benchmark), which would allow them to disentangle active returns (i.e., skill) from factor realizations. Investors then allocate capital based on their perception of active returns. For simplicity, we assume that a fund’s prospectus can be either of two types: *detailed* or *generic*. Detailed prospectuses are fully revealing of the fund’s factor loading, while generic prospectuses provide only a signal of the average loading across all funds in the same mandate.

There are three channels through which investor learning influences fund flows. The first two affect flow *uncertainty*, and go in opposite directions. First, investors in funds with generic prospectuses make mistakes when estimating fund skill, leading to increased flow uncertainty. This is due to the partial information conveyed by the generic signal as well as the volatility of the common factor, part of which is misperceived as active return volatility. The further a fund’s true factor loading is from the cross-sectional average, the greater the investors’ misperception, and thus the greater the increase in flow uncertainty. Second, because investors understand that they are more likely to make such mistakes with generic prospectuses, they will optimally choose to react less strongly to perceived performance, leading, all else equal, to decreased flow uncertainty. The third channel affects the average *level* of flows. If a fund with a generic prospectus has an above-average factor loading and the factor return is also expected to be positive (as is typically the case), then investor mistakes will tend to be beneficial to the manager. For example, if the true factor loading is 1.2 but investors estimate it to be 0.9, then the excess perceived loading of 0.3 times the positive factor realization will be treated as active return by investors, and their estimate of skill will be upward-biased. Thus, among funds with generic prospectuses, those with above-average factor loadings will tend to receive higher flows, while those with below-average factor loadings will tend to receive lower flows.

In equilibrium, a manager’s optimal prospectus choice is characterized by two thresholds: if the fund’s factor loading lies between the thresholds (*standardized* strategy), the manager will choose a generic prospectus, whereas if the loading lies outside of the threshold region (*specialized* strategy), she will choose a detailed prospectus. As a function of the demeaned factor loading, the direct uncertainty cost of investor mistakes is quadratic and centered at zero, while the expected level of flows is linear and increasing (for positive expected factor realizations), and the effect of flow-performance sensitivity is constant. Thus, for factor loadings far below or far above the cross-sectional average, the direct uncertainty cost dominates and detailed prospectuses are optimal. There exists a small region below the average loading where both the direct uncertainty cost and the decrease in the expected level of flows are small enough that the benefit of a lower flow-performance sensitivity wins out and generic prospectuses become optimal. However, most of the optimal-generic region lies above the average loading, where the combined benefits of increased average flows and lower flow-performance sensitivity outweigh the direct uncertainty cost.

Because most funds with generic prospectuses have above-average factor loadings, these funds will tend to be larger in equilibrium. And for the same reason, the correlation between size and total flow uncertainty among these funds will be positive overall, reflecting the central trade-off of the model: managers will tolerate higher total flow uncertainty only if it comes with higher expected flow levels. Funds with detailed prospectuses do not face the same trade-off (the higher volatility cost always dominates), so their size is uncorrelated with their flow uncertainty.

In the empirical section of the paper, we start by testing the model’s main prediction that qualitative disclosures should be more detailed for funds with more specialized strategies. We measure detail by using our financial content measure as previously described. To measure specialization, we construct a measure based on the distance of each fund’s estimated factor betas from the average betas of all funds in the same category. Although the model assumes a single-factor structure for tractability, for the empirical proxy we follow the recent asset pricing literature and use a six-factor structure (i.e., the five factors of [Fama and French \(2015\)](#) plus momentum). Specifically, our specialization measure is the sum of squared normalized betas, where the normalization subtracts the within-category mean and divides by the within-category standard deviation.

We then run a quarterly panel regression of financial content on specialization, controlling for standard fund attributes at the beginning of each quarter: (log) total net assets, age, expense and turnover ratios, and the six factor betas (included to account for any differences not captured by the SPGs). We also use several different fixed-effects specifications: quarter, category-quarter, and fund-level. Depending on the specification, a one standard deviation increase in specialization is associated with an increase in financial content of 5-16% relative to its own standard deviation.

The estimated coefficients are all statistically significant, usually at the 1% level. In this and all subsequent regressions, we double-cluster standard errors by fund and time to account for correlated unobservables in both the time series and cross section.

The above result could lend itself to alternative explanations. On the one hand, it might seem mechanical as more specialized strategies might require more detailed explanations to achieve the same clarity. On the other hand, it might be surprising as we might expect specialized funds to be more secretive about their strategies. That contrast highlights the importance of the model in clarifying the mechanisms at play, which we explore in the remaining empirical tests.

First, we test the prediction that funds with generic strategy descriptions should grow larger as they accumulate advantageous investor mistakes. To do this we run a quarterly panel regression of log TNA on financial content, controlling for the usual fund-level attributes and category-quarter fixed effects.¹⁰ We use the continuous measure of financial content as well as tercile dummies, to be closer to the model's assumption of a binary prospectus choice (the bottom tercile represents generic descriptions, the top tercile detailed ones). In both cases, funds with less detailed descriptions are on average larger; the effect is statistically significant at the 1% level. For example, log TNA is -0.203 lower for the top tercile of financial content compared to the bottom, which corresponds approximately to a 20% decline in dollar TNA. We also split the sample by median fund age, to account for variation in the length of time that generic funds have had to accumulate investor mistakes. While the effect is present in both subsamples, it is larger in magnitude for the older funds (-0.249 versus -0.176), consistent with the mechanism of the model.

Second, we examine the correlation between fund size and flow volatility (as a proxy for flow uncertainty), which is predicted to be positive for funds with generic strategy descriptions and zero for funds with detailed descriptions. Since the model is stylized and ignores many other channels that could potentially affect flow volatility, we should not expect the latter correlation to be strictly zero in practice. Rather, if our model describes a meaningful part of reality, we should find a *higher* correlation among the generic-prospectus group than the detailed-prospectus group. We test this prediction in the usual quarterly panel regression setup with category-quarter fixed effects, regressing flow volatility on log TNA, financial content tercile dummies, and the interactions between log TNA and the tercile dummies (as well as the usual controls). The coefficient on the top-tercile interaction term is -1.771 (significant at the 5% level), indicating that the correlation is 17% lower for funds with detailed descriptions.

Third, if the learning channel of our model agrees with reality, we should expect to find that flows respond more strongly to the active component of returns for funds with detailed descriptions,

¹⁰Fund and quarter fixed-effects specifications show similar results.

where investors know they can more accurately discern the skill of the manager. To measure investors’ perceptions of active returns, we employ a methodology similar to Barber et al. (2016). Separately for each fund, we run a multivariate regression of fund flows on the fund’s returns and the six factor portfolio returns. Assuming that flows only chase perceived active returns, these can be inferred from the estimated coefficients (see section 4.2.2). Then, in the full panel with category-quarter fixed effects, we regress quarter $t + 1$ fund flow on quarter t perceived active return (PAR), the financial content tercile dummies, and the interaction between PAR and the financial content dummies (along with the usual controls, quarter t flow, and squared PAR to capture convexity in the flow-performance sensitivity). We find the expected positive relationship between PAR and next-quarter flows, as well as a positive interaction coefficient (significant at the 5% level), indicating that investors respond more strongly—about twice as strongly—to active returns when funds write detailed strategy descriptions. This result is particularly telling in excluding the mechanical alternative explanation that funds with more specialized strategies must disclose more to achieve a similar level of clarity. In that case, the signal conveyed by detailed or generic prospectuses about a fund’s strategy would be of similar precision. Hence, we would not expect differences in flow-performance sensitivity across terciles of financial content.

Although fund age is not explicitly in our model, we also examine the flow-performance relationship separately in subsamples split by median age, following the logic that the information in narrative descriptions should be most important for funds with shorter returns histories. Indeed, we find that the above effect is entirely driven by the young funds. For old funds, prospectus detail has no effect on the flow-performance relationship. Moreover, for young funds, flow-performance sensitivity operates almost exclusively via the prospectus channel, with an insignificant baseline sensitivity for funds with generic prospectuses.

Lastly, we directly examine whether investors are better able to infer funds’ true benchmark factor loadings if they write detailed prospectuses. Again using the Barber et al. (2016) methodology to back out perceived factor betas, we compare these to “objective” betas estimated using full-sample, fund-level regressions of raw returns on factor returns. Consistent with the mechanism of our model, we find that these betas are 16% more similar for funds with detailed prospectuses, and the effect is again driven by funds of below-median age.

1.1 Related Literature

Our paper contributes to two broad, interrelated areas of research: the study of learning by mutual fund investors and/or managers, and textual analysis of fund disclosures.

The study of learning by *investors* dates back to at least [Chevalier and Ellison \(1997\)](#) and [Sirri and Tufano \(1998\)](#), who show that capital flows respond positively to past fund performance. Theoretical analysis of this behavior originates with [Baks et al. \(2001\)](#), who examine partial-equilibrium optimal allocation across funds when investors have Bayesian prior beliefs about fund alpha. In a richer equilibrium setting, [Berk and Green \(2004\)](#) show that investor learning about skill drives net alphas to zero when the market for asset management is perfectly competitive, fund size is endogenous, and funds experience decreasing returns to scale. This seminal framework has been extended in a number of ways that are relevant for our work. Some extensions allow learning about common components of skill within fund families ([Brown and Wu \(2016\)](#)) or different funds managed by the same portfolio manager ([Choi et al. \(2016\)](#)). Other extensions allow the precision of learning to vary with macroeconomic uncertainty ([Starks and Sun \(2016\)](#)), average idiosyncratic risk ([Harvey and Liu \(2019\)](#)), or investor uncertainty about funds' exposure to an aggregate risk factor ([Franzoni and Schmalz \(2017\)](#)). [Buffa and Javadekar \(2020\)](#) endogenize fund manager strategy choice. Our paper also extends the baseline [Berk and Green \(2004\)](#) model. Like the aforementioned prior work, we allow investors to learn about both idiosyncratic and common components of skill within fund groups. The key innovation of our framework is that factor loadings are not common knowledge, but need to be learned through qualitative fund disclosures. We are the first to analyze funds' disclosure choice and investors' learning from this choice in a rational competitive equilibrium model. This theoretical innovation generates novel predictions for the relationships between strategy specialization, disclosure informativeness, fund size, flow volatility, and flow-performance sensitivity.

Another branch of the literature studies learning by fund *managers*. In a noisy rational expectations equilibrium (NREE) model, [Kacperczyk et al. \(2016\)](#) show that managers with limited information processing capacity choose to learn about the systematic component of asset returns in recessions and the idiosyncratic component in expansions. [Gârleanu and Pedersen \(2018\)](#) combine a similar NREE setup with an imperfectly competitive market for investor capital, subject to search frictions. Unlike in the competitive market models based on [Berk and Green \(2004\)](#), informed managers in this model outperform in equilibrium, and the overall efficiency of asset prices is a decreasing function of search costs. Our work is complementary to this branch of the literature, as we assume instead that skill is homogeneous, and that managers internalize investors' optimal behavior directly rather than through a learning process.

The second broad area to which our paper contributes is the empirical examination of mutual fund qualitative disclosures, using textual analysis. The use of textual analysis in asset pricing goes back to [Tetlock \(2007\)](#), the earliest applications in the mutual fund literature regard shareholder letters: [Hillert et al. \(2016\)](#), find that a negative tone predicts lower fund flows, while writing

in "plain English" predicts higher flows; [Hwang and Kim \(2017\)](#) find that for closed-end funds low "readability" is associated with a greater discount to net asset value; [Zhang \(2020\)](#) through a machine learning approach finds that informed managers experience higher flows. More recently, a few other papers examine mutual fund prospectus text, as we do. [Abis \(2020\)](#) uses supervised machine learning on fund strategy descriptions to identify and test hypotheses about quantitative funds. [Kostovetsky and Warner \(2020\)](#) measure the textual uniqueness of strategy descriptions relative to other funds in the same Morningstar category, and find that this uniqueness positively predicts fund flows in the first three years after inception. [Abis and Lines \(2020\)](#) use strategy descriptions to construct interpretable peer groups for active mutual funds, and use them to assess whether funds behave in line with their stated objective. [Akey et al. \(2021\)](#) use strategy descriptions to categorize ETFs and index funds in order to study the activeness of ostensibly passive investments. Relative to these papers, we contribute by examining managers' strategic choice of how much to disclose, rather than taking disclosure as given. We also focus on the level of detail of the content rather than tone or writing style.

Additionally, several contemporaneous studies analyze prospectus summary sections (or separate "summary prospectus" documents). [Krakow and Schäfer \(2020\)](#) use textual uniqueness relative to other funds in the same family to measure disclosure informativeness, arguing that more similar text indicates greater use of family-specific boilerplate language. They find that more informative funds exhibit higher idiosyncratic risk and better performance, and have weakly higher fund flows. By contrast, using a more direct measure of disclosure detail, we find that funds that disclose more information are actually smaller on average. [Sheng et al. \(2021\)](#) find that summary risk descriptions largely line up with academic risk factors, and that more informative risk disclosures predict worse future performance but, unlike the other studies, have no effect on flows. [Tucker and Xia \(2020\)](#) find that summary prospectuses generally have low readability but, in contrast to [Hwang and Kim \(2017\)](#), lower readability predicts better future performance. [Tucker et al. \(2020\)](#) examine the tone of these summaries. They report that risk tone covaries positively with performance (conditional only on high readability), while strategy tone covaries negatively with performance (conditional only on low readability). The relationship between tone and measured fund risk is ambiguous. ? study the summary prospectuses of 38 S&P 500 index funds and find that those of high-fee funds display greater complexity.

Overall, the difficult-to-interpret and often contradictory nature of many existing empirical results highlights the need for stronger theoretical guidance, which is a major contribution of our paper. Like [Abis \(2020\)](#) and [Abis and Lines \(2020\)](#), we also analyze the *full* content of funds' Principal Investment Strategy descriptions. Prospectus summaries are much shorter than the full sections and contain more boilerplate language. Even the longer descriptions provided by

Morningstar, used by [Kostovetsky and Warner \(2020\)](#), are significantly shorter than the full text (70 words versus over 300 words, on average). Many of the differences in our results compared to the prior literature can be explained by this improvement in the data.

The remainder of the paper is organized as follows: section 2 describes the financial content measure; section 3 describes the model and its equilibrium predictions; section 4 lays out our empirical methodology and the tests of the model’s predictions; and section 5 concludes. Appendix A contains proofs of model propositions; and appendix B provides examples of prospectuses with high and low financial content.

2 Prospectus Informativeness

In this section we provide additional details about fund prospectuses and describe the construction of our financial content measure.

2.1 Institutional Background

The electronic disclosure of fund prospectuses is mandated by the SEC to all mutual funds through the EDGAR system. The PIS section, studied in this paper, corresponds to item 9 of the N-1A mandatory disclosure form.¹¹ In that section funds are required to both: “*describe the Fund’s principal investment strategies, including the particular type or types of securities in which the Fund principally invests or will invest*”; as well as “*explain in general terms how the Fund’s adviser decides which securities to buy and sell*”. The latter requirement was introduced in the 1998 amendment of mandatory disclosures and made mandatory as of December 1999. For those reasons our analysis of PIS descriptions starts in January 2000.

As stated in the “Administration of the Form N-1A requirements” section on how to compile the N-1A form, the requirements are “*intended to promote effective communication between the Fund and prospective investors*”. In accordance with the SEC’s “Plain English Rule” of 1998 (Rule 421(d)), that section further mandates that prospectuses be written utilizing “*concise, straightforward, and easy to understand language*”, which should be understood by “*an average or typical investor who may not be sophisticated in legal or financial matters*”. Finally, it states that the “*disclosure in the prospectus should be designed to assist an investor in comparing and contrasting the Fund with other funds*”.

¹¹Form: <https://www.sec.gov/files/formn-1a.pdf>

In accordance with those objectives, the SEC has dedicated a section of their website to instruct investors on “How to read mutual fund prospectuses”.¹² There, as reported in the introductory quote to this paper, investors are encouraged to consider the information contained in prospectuses when making investment decisions, particularly when comparing funds to their peers.

The information contained in prospectuses is widely available to investors not only through the EDGAR system but also in its dissemination through popular databases. For instance, much of the information contained in CRSP or Morningstar is obtained through these regulatory disclosures. An excerpt of the PIS section is also provided through the Morningstar database. Additionally, investors are directly sent a copy of these prospectuses through physical or electronic mail by their brokerage or advisory firms, when they possess shares in a fund. Finally, the majority of the retail capital managed by these funds goes through financial advisory firms. Hence, investors might be aware of the soft information contained in prospectuses by discussing with their financial advisors, who will have likely considered it.

Being prospectuses a legally binding document, funds must be consistent with those disclosures in all their communications with perspective investors. For those reasons, we consider that the information contained in prospectuses might reach investors through different channels and focus our analysis on its legally mandated source: the full prospectus as disclosed through the EDGAR system. Most papers which study these disclosures focus on either summary prospectuses, which provide substantially less and more boilerplate information (e.g., [Krakow and Schäfer \(2020\)](#), [Tucker and Xia \(2020\)](#), [Hwang and Kim \(2017\)](#)); or on the excerpts provided by Morningstar (e.g., [Kostovetsky and Warner \(2020\)](#)), which are significantly shorter (they contain an average of 70 words vs. the 300 words average in our data). In this paper, as in [Abis \(2020\)](#) and [Abis and Lines \(2020\)](#), we consider instead the full PIS section. Considering the full disclosure is especially important when studying its informativeness, as we do in this paper.

Utilizing the same panel of PIS sections as we do, [Abis \(2020\)](#) shows that PIS descriptions are highly auto-correlated within fund. With an average cosine similarity of 95% between any two consecutive prospectuses and an average similarity of 67% between prospectuses even at 25 prior lag. That indicates that funds do update these disclosures, but updates are infrequent. [Abis and Lines \(2020\)](#) further show that there exists substantial variation in the length, linguistic complexity and content of these sections. In order to assess differences in content they utilize a machine learning algorithm to cluster funds with linguistically similar descriptions into 17 distinct Strategy Peer Groups (SPGs): *Large Cap, Mid Cap, Small Cap, Fundamental, Quantitative, Long Term, Intrinsic Value, Defensive, Dividends, New Products & Services, PE-Ratio, Competitive*

¹²Link: https://www.sec.gov/oiea/investor-alerts-bulletins/ib_mfprospectus1.html

Advantage, Foreign (ADR), Foreign (Emerging Markets), Fixed Income, Derivatives, and Tax-Managed.¹³ They show that, whereas there aren't differences in risk-adjusted performance across SPGs, funds belonging to the same SPG display significantly greater similarity in risk factor exposures, stock characteristics, returns, and portfolio holdings. These similarities are in line with their stated objective. For instance, funds belonging to the "Dividend" SPG, hold stocks with the highest dividend yield as compared to funds in any of the other SPGs.¹⁴

2.2 Financial Content

Construction. To represent the level of detail of each prospectus we construct a continuous measure of informative *financial content* (*FinCon*). We start from *Campbell Harvey's Hypertextual Finance Glossary*, which provides a comprehensive list of finance-related words and short phrases.¹⁵ Each PIS section is represented by a vector of stemmed words: the *strategy vector*.¹⁶ From this vector, we construct another vector—the *financial vector*—as the subset of words in the *strategy vector* that can also be found in the Harvey dictionary. Since the prospectus is comprised mostly of financial text, the simple ratio between the length of the *financial vector* and that of the *strategy vector* would not necessarily identify *informative* financial content. Thus, we first exclude from the *financial vector* all words that appear in more than 75% of all PIS descriptions,¹⁷ and assign a weight to the remaining terms according to the inverse of their frequency among all funds with linguistically similar strategy descriptions, measured using the SPGs of [Abis and Lines \(2020\)](#). It is important to control for common language among similarly described strategies as certain broad strategy types might require more detailed disclosures to provide a general description of the fund's objective (e.g., Quantitative strategies). Additionally, terms which are common for some broad strategy types might be uncommon in the universe of descriptions. Both of these issues could bias our measure of informative content.

Hence, we define our financial content measures as follows:

$$FinCon_{i,t}^a = \frac{1}{L_{i,t}^S} \sum_{w=1}^{L_{i,t}^F} \frac{1}{(1 + Coverage_w)^a}, \quad (1)$$

¹³Note that while all funds hold at least 80% of their assets in common stock on average, some funds' most distinctive features are secondary asset classes that make up the remaining 20%.

¹⁴For a full description of the clustering algorithm (*k-means*), further details on the strategy labels and the strategies they represent, as well as additional validation tests, see [Abis and Lines \(2020\)](#).

¹⁵<https://people.duke.edu/~charvey/Classes/wpg/glossary.htm>

¹⁶Stop words, symbols, numbers, websites, and punctuation are excluded. Words are stemmed using the Porter stemmer.

¹⁷The only stemmed terms excluded are: invest, fund, compani, secur, market, stock, asset, cap, manag

where $Coverage_w$ indicates the fraction of all PIS sections in the same SPG as fund i that contain financial term w .¹⁸ $L_{i,t}^F$ and $L_{i,t}^S$ represent the length of the *financial vector* and *strategy vector*, respectively, for fund i in month t . The variable a controls the strength of the “penalty” applied to commonly used terms; the greater a , the greater the penalty. Note that setting $a = 0$ is equivalent to not applying any weighting. We compute the measure for $a = [0, 1, 2, 4]$, but use $a = 4$ in our main tests (results are robust to alternative specifications). Appendix B provides an example of high- and low-content prospectuses.

As a robustness, we also construct a version of our measure in which financial terms are weighted by their inverse frequency in the whole corpus of prospectuses. Results are robust to this alternative specification even if, as would be expected with a less precise measure, some results have a lower explanatory power.

Interpretation. To validate our interpretation of this measure, we examine how it relates to the length, textual complexity, and boilerplate content of the PIS sections. Length is measured using a simple word count. Textual complexity is measured using the Flesch-Kincaid grade-level complexity score (Kincaid et al. (1975)), which is calibrated to show the estimated number of years of schooling required to understand the text. Boilerplate is a continuous measure $\in [0, 1]$, constructed as follows. First we collect all four-word phrases (4-grams) from the corpus of prospectuses in our sample, and rank them according to their overall frequency in the corpus. The top 0.1% of these 4-grams (607 distinct phrases) are considered to be boilerplate language. Figure 4 displays the top 100 4-grams, ranked by frequency. Next, we construct the boilerplate measure $Boiler_{i,t}$ at the fund-month level by computing the percentage of all words each fund’s *strategy vector* that belong to the vector of boilerplate phrases.¹⁹

$$Boiler_{i,t} = \frac{L_{i,t}^B}{L_{i,t}^S}, \quad (2)$$

where $L_{i,t}^S$ is the length of the fund’s *strategy vector* and $L_{i,t}^B$ is the length of the subset of the *strategy vector* that also appears in the boilerplate vector at time t .

We consider length, complexity and boilerplate to be indirect measures of informational content. For example, a longer description with less boilerplate is likely to contain more relevant

¹⁸When an individual word is part of a multi-word phrase that also appears in the Harvey glossary, $Coverage_w$ is determined using the largest lexical group. For example, “index” appears in the glossary both as itself and as part of “index arbitrage”. If the phrase “index arbitrage” appears in a prospectus, the value of $Coverage_w$ for both “index” and “arbitrage” will be the fraction of all PIS sections containing “index arbitrage”. The coverage of both words is counted separately because the denominator $L_{i,t}^S$ also counts all words separately.

¹⁹Note: words must appear in the correct order to be considered part of a boilerplate phrase.

content, and more content is also likely to increase the complexity of the text. (However, longer descriptions would not necessarily indicate increased informational content if the additional text is too simple or too common among prospectuses.)

Thus, we assess the validity of *FinCon* as a measure of prospectus detail by running the following regressions:

$$Y_{i,t} = a + b \text{FinCon}_{i,t}^a + \gamma' X_{i,t} + \varepsilon_{i,t}. \quad (3)$$

where $Y_{i,t}$ is either length, complexity, or boilerplate ($Length_{i,t}$, $Complexity_{i,t}$, $Boiler_{i,t}$) for fund i in month t . $X_{i,t}$ is a vector of control variables: log of fund size and age, turnover and expense ratios, net fund flows and flow volatility, and additional style controls constructed as the average loading of the stocks held by fund i at time t on the Fama and French (2015) and Carhart (1997) size, value, momentum, investment and profitability factors.

The results are reported in Table I. Columns 1-4 show that all variants of *financial content* for $a > 0$ are positively related to PIS length and textual complexity, and negatively related to boilerplate, indicating that our measure is effective at capturing the presence of relevant financial content (i.e., non-standardized financial terminology) in strategy descriptions. Consistent with our intuition, the measure is more effective when financial terms are penalized more heavily according to their frequency in the PIS descriptions of funds in the same SPG (i.e., higher a).

Note that *FinCon* captures many desirable characteristics of the other indirect measures. It incorporates length via the count of financial terms, but is inversely related to the count of non-financial terms. It also negatively incorporates boilerplate by removing the most frequent financial terms within each broad strategy type (SPG), and weighting each term by the inverse of its frequency in the SPG. Finally, it captures the presence of complex financial terminology.

Heterogeneity. The top-left panel of Figure 1 reports the distribution of our financial content measure in the panel of PIS descriptions. We observe a large heterogeneity in the amount of financial content: it varies from 7% to 35% of the total words count between the 1st and the 99th percentiles of the measure’s distribution, with a standard deviation of 0.055.

In exploring the source of this heterogeneity we run the following regression:

$$\text{FinCon}_{i,t} = \alpha + \gamma X_{i,t} + \delta_{SPG,t} + \varepsilon_{i,t}. \quad (4)$$

where $X_{i,t}$ is the same list of fund controls as described in (3) for fund i in quarter t and $\delta_{SPG,t}$ are $SPG \times Quarter$ fixed-effects.

The intuition behind that regression is that commonalities in funds’ characteristics, in their broad strategies, or in time-trends might drive the reported heterogeneity. It is worth noting that fund controls include the average exposure of stocks held to the market, size, book-to-market, momentum, investment and profitability factors. It hence controls for non-textual methods of defining fund styles/mandates. We are interested in $(1 - R^2)$ of that regression, indicating the percentage of variation in financial content which is *not* explained by those simple characteristics.

The regression’s R^2 is 19.17%, indicating that 80.83% of the variation in financial content cannot be explained by these common components. The bottom-right panel of Figure 1 displays the distribution of the unexplained portion of *FinCon*: $\varepsilon_{i,t}$ in (4). We still observe a large heterogeneity in financial content, which now varies from 7.44% to 32.45% of the total words count between the 1st and the 99th percentiles of the horthogonalized measure’s distribution.²⁰ The horthogonalized distribution has a standard deviation of 0.0496. In order to decompose that effect, Figure 1 shows the same horthogonalized distribution when only including SPG fixed-effects (top-right panel); only quarter fixed-effects (bottom-left panel); and quarter \times SPG fixed-effects (bottom-right panel). All three panels exclude fund-level controls. We observe that most of the explanatory power comes from SPGs, whereas quarter fixed-effect and controls have a much smaller impact in explaining the variation in financial content.

To further assess the drivers of this variation, column 1 of Table II reports the results of the regression described in (4). We observe a negative and significant relation between fund size and financial content, always significant at the 1% level. Financial content is also related to fund style as measured by the exposure of stocks held to the size, book-to-market and momentum factors. There exists a positive correlation between financial content and fees and a negative correlation with turnover ratio, albeit only marginally significant.

To further understand the drivers of this unexplained variation, columns 2-5 add performance at time $(t+1)$ to the explanatory variables in (4) (i.e., we relate beginning-of-quarter financial content and characteristics to end-of-quarter performance). In columns 2 and 4 we measure performance as the difference between fund returns and their benchmark (*Alphas*). Benchmarks are obtained with 12-months rolling regressions of fund returns on the market, size, book-to-market and momentum factors in column 2; and those same factors plus the profitability and investment factors in column 4. All rolling regressions are run utilizing daily returns. In columns 3 and 5 performance is

²⁰The reported statistics add back the measure’s unconditional mean of 0.18 to the distributions reported in Figure 1.

measured as *Value Added*, by multiplying the 4-factor and 6-factor *Alphas* by the log of Total Net Assets. We observe that including performance does not significantly increase the explanatory power of this regression (the R^2 goes from 19.17% to 19.25%). Additionally, aside for a marginally significant effect in column 4, performance does not seem to be correlated with financial content.

3 A Model of Prospectus Choice

Given the large unexplained variation in financial content reported in Section 2.2, in this section we propose a learning model to guide our understanding of the mechanisms which might incentivize fund managers to disclose more or less information in their strategy descriptions.

3.1 Economic Setting

We consider an economy in which a continuum of investors delegate their money to a continuum of asset management funds. Both groups of agents have a measure of one.

Funds. Each fund is characterized by a fund strategy, denoted by β_i , capturing the fund’s exposure to a single risk factor. Each strategy is defined by a common component, b , and a fund-specific component, γ_i :

$$\beta_i = b + \gamma_i. \tag{5}$$

The common component b can be interpreted as the average or most common strategy within an investment category, while γ_i captures the deviation from the average. A larger γ_i , in absolute value, indicates a more specialized strategy.²¹ Fund i is run by a fund manager whose managerial skill is denoted by α_i , quantifying the additional return that fund i is able to generate over and above the return due to the risk exposure associated with its fund strategy.

We formally model the net (of fees) return generated by fund i at time t as follows:

$$r_{i,t} = \alpha_i + \beta_i F_t + \varepsilon_{i,t} - \mathcal{C}(q_{i,t-1}) - f, \tag{6}$$

²¹In this paper we treat the investment strategy of a fund manager as exogenous. [Buffa and Javadekar \(2020\)](#) endogenize this choice in a model of fund managers’ self-selection.

where F_t is the risk factor, β_i is the fund strategy, α_i is the skill of the fund manager, and $\varepsilon_{i,t}$ is the fund's idiosyncratic risk. $\mathcal{C}(q)$ captures returns to scale as a function of fund assets q , and f denotes the (exogenous) fees charged by each fund per dollar invested. We assume that $\varepsilon_{i,t}$ and F_t are independent and distributed as

$$\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \text{and} \quad F_t \sim \mathcal{N}(\mu_F, \sigma_F^2), \quad (7)$$

and maintain, as in Berk and Green (2004), that $\mathcal{C}(q) = c \cdot q$, where the fund has decreasing returns to scale (DRS) if $c > 0$ and increasing returns to scale if $c < 0$.²² We interpret the return that fund i generates by taking exposure to the risky factor as the return of the fund's *true benchmark*, and we denote it by $r_{i,t}^B \equiv \beta_i F_t$.

Investors. Fund investors do not observe α_i and β_i ; that is, they face uncertainty about the fund manager's skill, as well as the fund strategy. Uncertainty about the latter means that the investors are uncertain about how to evaluate the fund performance. Specifically, investors have priors on α_i and on the two components of β_i :

$$\alpha_i \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2), \quad b \sim \mathcal{N}(\bar{b}, \sigma_b^2), \quad \gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2), \quad (8)$$

where $\alpha_i \perp b \perp \gamma_i$ for any i , $\alpha_i \perp \alpha_j$ and $\gamma_i \perp \gamma_j$ for any $j \neq i$. Investors observe over time the net performance of their fund, $r_{i,t}$, as well as the factor return F_t , and form posterior beliefs about α_i and β_i . We refer to the corresponding posterior means, denoted by $\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,t}$, as fund i 's *perceived skill* and *perceived strategy*, respectively.

Fund prospectuses. The fundamental role of prospectuses is to provide investors with information about fund strategies. To keep the model parsimonious, we assume that fund managers can only write two types of prospectuses: *generic* or *detailed*. We denote the choice set of prospectus types by $\mathcal{P} = \{g, d\}$. Generic disclosures provide information about those characteristics of the fund strategy that are common to all funds within the same broad investment category. Detailed disclosures, instead, provides more comprehensive information about the specifics of the fund strategy that are unique to the fund. In our setting, while a generic prospectus perfectly reveals the common component of the fund's strategy (b), a detailed prospectus perfectly reveals the full

²²For simplicity, we assume that the returns to scale parameter c is common across all fund strategies.

strategy (i.e., the sum of the two components b and γ_i).²³ Formally, we assume that reading a generic prospectus is equivalent to receiving a signal

$$s_i^g = b, \tag{9}$$

while reading a detailed prospectus is equivalent to receiving a signal

$$s_i^d = b + \gamma_i. \tag{10}$$

For ease of exposition, in what follows, we refer to investors who receive signal s_i^g (i.e., read a generic prospectus) as \mathcal{I}^g , and those who receive signal s_i^d (i.e., read a detailed prospectus) as \mathcal{I}^d .

Additional assumptions. We make two additional assumptions that allow us to keep the model as parsimonious as possible and to focus on the equilibrium relationship between fund strategy and fund prospectus.

First, we assume that fund managers are uncertain about their investment skill when writing their fund prospectuses, and thus use the prior distribution of α_i in (8) when evaluating their expected utilities. This makes their optimal choice of disclosure independent of their skill α_i , and only dependent of their fund strategy β_i .

Second, while fund managers fully internalize the way fund investors learn and allocate their capital, we assume fund investors are less sophisticated. Specifically, we assume that they only learn directly from the informational content of the prospectus that they read, without internalizing the equilibrium relationship between fund strategy and prospectus choice. In other words, investors do not update their posteriors over fund strategy as a result of the equilibrium choice of prospectus.

Timeline. We consider a timeline with four dates:

- $t = 0$: funds choose which prospectus to adopt;
- $t = 1$: investors read the fund prospectuses and allocate their initial capital;
- $t = 2$: fund returns materialize, investors update their beliefs and re-allocate their capital;
- $t = 3$: fund returns materialize, funds are liquidated, the economy ends.

²³Our results remain qualitatively unaltered if the two signals were to be noisy and not perfectly revealing. What matters is that a generic prospectus does not reveal any information about γ_i .

3.2 Allocation of Capital and Prospectus Choice

We consider a competitive capital market as in Berk and Green (2004): risk neutral investors provide or withdraw capital to active funds up to the point at which their *perceived* net active returns are driven to zero in expectation. The equilibrium allocation of capital determines the optimal fund size, $q_{i,t}$, for $t = 1, 2$ and $\mathcal{I}^p \in \{\mathcal{I}^g, \mathcal{I}^d\}$, such that

$$\mathbb{E}_t^{\mathcal{I}^p} [\hat{r}_{i,t+1}^A] = 0, \quad (11)$$

where

$$\hat{r}_{i,t+1}^A = (\alpha_i + \varepsilon_{i,t+1}) + (\beta_i - \hat{\beta}_{i,t})F_{t+1} - c \cdot q_{i,t} - f. \quad (12)$$

It follows that the equilibrium size of a fund adopting a prospectus of type p is equal to

$$q_{i,t}(p) = \frac{\hat{\alpha}_{i,t}(p) - f}{c}, \quad (13)$$

where $\hat{\alpha}_{i,t}(p) \equiv \mathbb{E}_t^{\mathcal{I}^p}[\alpha_i]$ is the perceived skill of the fund manager from the investors' perspective after having observed the performance of the fund at time t . The flow of capital to fund i at time t is defined as the relative change in fund size: $flow_{i,t} \equiv (q_{i,t} - q_{i,t-1})/q_{i,t-1}$.

Taking into account how investors' learning affects flows, each fund manager chooses the prospectus that maximizes her ex-ante expected utility over managerial fees, $q_{i,t} \times f$:

$$p_i^* = \arg \max_{p \in \mathcal{P}} \mathbb{E}_0^{\mathcal{M}_i} [v_i(q_{i,1}(p)f) + \delta \cdot v_i(q_{i,2}(p)f)], \quad (14)$$

where the superscript \mathcal{M}_i indicates manager i 's information set. In what follows, we assume that $v_i(\cdot)$ takes the form of mean-variance preferences with the coefficient of risk aversion normalized to 1.

3.3 Investors' Learning

We next discuss investors' learning, and characterize the posterior distribution of managerial skill as a function of prospectus type.

Learning from prospectuses. At $t = 1$, after reading the prospectuses, investors refine their priors on fund strategies. Since generic prospectuses only reveals the common component of fund strategies b , investors \mathcal{I}^g form posterior (normal) distributions over β_i , with

$$\mathbb{E}_1^{\mathcal{I}^g}[\beta_i] = b \quad \text{and} \quad \text{Var}_1^{\mathcal{I}^g}[\beta_i] = \sigma_\gamma^2. \quad (15)$$

Investors \mathcal{I}^d , instead, perfectly learn β_i . Therefore their posterior mean and variance are equal to

$$\mathbb{E}_1^{\mathcal{I}^d}[\beta_i] = \beta_i \quad \text{and} \quad \text{Var}_1^{\mathcal{I}^d}[\beta_i] = 0. \quad (16)$$

Since at $t = 1$ fund returns have not yet materialized, investors \mathcal{I}^g and \mathcal{I}^d do not update their beliefs about managerial skill at that time. As such, the mean and variance of perceived skill are equal to those of the investors' prior, regardless of the prospectus.

Learning from returns. At $t = 2$, the investors observe the fund return $r_{i,2}$ and the factor return F_2 . Since the existing fund size $q_{i,1}$, the fund fee f , and the DRS/scale parameter c are also known, the investors can back out the fund gross return $R_{i,2} \equiv r_{i,2} + c \cdot q_{i,1} + f = \alpha_i + \beta_i F_2 + \varepsilon_{i,2}$. Using this information, they update their beliefs about managerial skill, and simultaneously their beliefs about fund strategy if the fund has adopted a generic prospectus. Given a prospectus of type $p \in \mathcal{P}$, the perceived managerial skill and its posterior variance are, respectively, equal to

$$\hat{\alpha}_{i,2}(p) \equiv \mathbb{E}_2^{\mathcal{I}^p}[\alpha_i | R_{i,2}, F_2] = (1 - \lambda_p^\alpha) \bar{\alpha} + \lambda_p^\alpha \underbrace{(R_{i,2} - \mathbb{E}_1^{\mathcal{I}^p}[\beta_i] F_2)}_{\text{perceived gross active return}}, \quad (17)$$

where

$$\lambda_p^\alpha = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \text{Var}_1^{\mathcal{I}^p}[\beta_i] F_2^2}, \quad (18)$$

and $\text{Var}_2^{\mathcal{I}^p}[\alpha_i | R_{i,2}, F_2] = \sigma_\alpha^2 (1 - \lambda_p^\alpha)$. Intuitively, the best forecast of a fund manager's skill is a weighted average of the investor's prior belief and the perceived active return, where the weight λ_p^α is given by the ratio of the covariance of managerial skill with the perceived active returns, and the variance of the latter.²⁴

²⁴Similarly, the fund strategy has a posterior mean and variance at $t = 2$ equal to

$$\begin{aligned} \hat{\beta}_{i,2}(p) &= (1 - \lambda_p^\beta) \mathbb{E}_1^{\mathcal{I}^p}[\beta_i] + \lambda_p^\beta \left(\frac{R_{i,2} - \bar{\alpha}}{F_2} \right), \\ \hat{\sigma}_{\beta_i,2}^2(p) &= \text{Var}_1^{\mathcal{I}^p}[\beta_i] (1 - \lambda_p^\beta), \quad \text{where} \quad \lambda_p^\beta = \frac{\text{Var}_1^{\mathcal{I}^p}[\beta_i] F_2^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \text{Var}_1^{\mathcal{I}^p}[\beta_i] F_2^2}. \end{aligned}$$

3.4 Equilibrium

Since the estimated managerial skill at $t = 1$ is independent of the manager's chosen prospectus (it is equal to the prior mean $\bar{\alpha}$), the optimal prospectus choice depends only on fees earned at $t = 2$. Formally, it is the solution to the following problem:

$$p_i^* = \arg \max_{p \in \mathcal{P}} \mathbb{E}_0^{\mathcal{M}_i}[\hat{\alpha}_{i,2}(p)] - \frac{f}{2c} \text{Var}_0^{\mathcal{M}_i}[\hat{\alpha}_{i,2}(p)] \quad (19)$$

$$= \arg \max_{p \in \mathcal{P}} \bar{\alpha} + \mathbb{1}_{\{p=g\}} \mathbb{E}_0^{\mathcal{M}_i}[\lambda_g^\alpha F_2] \gamma_i - \frac{f}{2c} \text{Var}_0^{\mathcal{M}_i} [\lambda_p^\alpha ((\alpha_i - \bar{\alpha}) + \varepsilon_i + \mathbb{1}_{\{p=g\}} F_2 \gamma_i)]. \quad (20)$$

When $p = d$, evaluating the expression inside the arg max in equation (20) yields

$$\bar{\alpha} - \frac{f \sigma_\alpha^4}{2c(\sigma_\alpha^2 + \sigma_\varepsilon^2)}, \quad (21)$$

whereas when $p = g$, it yields

$$\bar{\alpha} + \mathbb{E}_0^{\mathcal{M}_i}[\lambda_g^\alpha F_2] \gamma_i - \frac{f}{2c} \text{Var}_0^{\mathcal{M}_i} [\lambda_g^\alpha ((\alpha_i - \bar{\alpha}) + \varepsilon_i + F_2 \gamma_i)], \quad (22)$$

where

$$\lambda_g^\alpha = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\gamma^2 F_2^2}. \quad (23)$$

Since λ_g^α depends on the factor return at time 2, it is random from the perspective of the fund manager at time 0. A manager chooses a generic (detailed) prospectus if (22) is larger (smaller) than (21), which, as we discuss below, depends on the fund strategy γ_i .

When fund managers care about the uncertainty of their fees, they face the following trade-off. On the one hand, adopting a generic prospectus reduces the transparency of the fund strategy, which is beneficial to the manager if it induces investors to mistakenly estimate a higher skill than the manager actually possesses. This occurs, on average, when the fund's exposure to the risk factor is larger than the cross-sectional average ($\gamma_i > 0$) and the factor return is expected to be positive ($\mu_F > 0$), or when the exposure is smaller than average ($\gamma_i < 0$) and the factor return is expected to be negative ($\mu_F < 0$). In both cases, investors allocate excessive capital to the fund, increasing its management fees. On the other hand, investor mistakes result in a more uncertain allocation of capital, and consequently more uncertain fees, which reduces the risk-

Since our economy ends at $t = 3$, the quantities $(\hat{\beta}_{i,2}(p), \hat{\sigma}_{\beta_i,2}^2(p))$ are not relevant for the equilibrium analyzed here.

averse manager's utility. This increase in uncertainty is particularly severe when the fund's factor exposure γ_i is sufficiently far from the cross-sectional average. Intuitively, unless the fund's strategy is too specialized, the costs of transparency outweigh the benefits, and the fund is incentivized to choose a generic prospectus.

Although the moments in (22) are not available in closed-form, we are able to establish several useful propositions (all proofs can be found in appendix A).

Proposition 1 (Prospectus Choice). *A generic prospectus is optimally adopted if and only if the fund strategy is not too specialized, $\underline{\gamma} < \gamma_i < \bar{\gamma}$, where the two thresholds $(\underline{\gamma}, \bar{\gamma})$ always exist and are such that $\underline{\gamma} < 0 < \bar{\gamma}$, with $|\bar{\gamma}| > |\underline{\gamma}|$ when $\mu_F > 0$ and $|\bar{\gamma}| < |\underline{\gamma}|$ when $\mu_F < 0$. The fraction of fund managers adopting a generic prospectus is given by*

$$m = \Phi(\bar{\gamma}/\sigma_\gamma) - \Phi(\underline{\gamma}/\sigma_\gamma), \quad (24)$$

where $\Phi(\cdot)$ is the cumulative density function of a standard normal distribution.

Proposition 1 formally shows that the endogenous distribution of prospectus types across fund strategies is characterized by two thresholds of the fund-specific factor exposure γ_i . A generic prospectus is the optimal choice for any fund with γ_i between the two thresholds of opposite signs, and a detailed prospectus is optimal otherwise. The graph in figure 2 provides a graphical illustration of the optimal prospectus choice: the equilibrium utility of the fund manager is plotted as a function of the fund-specific component of fund strategy (γ_i), with the two thresholds $(\underline{\gamma}, \bar{\gamma})$ shown as dotted vertical lines. The region between the thresholds lies mostly above $\gamma_i = 0$, reflecting the fact that investor mistakes (induced by the generic prospectus choice) only provide a benefit if the fund's exposure to the common factor is above average (i.e., if $\gamma_i > 0$).²⁵

Proposition 2 (Fund Size). *The cross-sectional average expected fund size at time 2 is equal to $(\bar{\alpha} - f)/c$ among funds adopting a detailed prospectus, and*

$$\frac{\bar{\alpha} - f}{c} + \frac{\sigma_\gamma}{c} \left(\frac{\phi(\underline{\gamma}/\sigma_\gamma) - \phi(\bar{\gamma}/\sigma_\gamma)}{m} \right) \mathbb{E}[\lambda_g^\alpha F_2] \quad (25)$$

among funds adopting a generic prospectus, where $\phi(\cdot)$ is the probability density function of a standard normal distribution, and m is as in (24). It follows that the expected fund size associated

²⁵In this parameterization of the model, the expected factor return is positive. If it were negative, the region between the thresholds would be reflected about the y-axis.

with generic prospectuses is larger, and exhibits greater cross-sectional dispersion, than the expected fund size associated with detailed prospectuses.

Proposition 2 characterizes equilibrium fund size as a function of prospectus choice. Two key findings emerge. First, the expected fund size for generic prospectuses is higher on average than for detailed prospectuses. This is because the average mistake made by investors, $\mathbb{E}[\gamma_i | \underline{\gamma} < \gamma_i < \bar{\gamma}]$, is positive when the factor return is expected to be positive, and negative when the factor return is expected to be negative. The intuition is that managers who optimally adopt a generic prospectus tend to benefit from investors' mistakes, and this benefit comes in the form of higher expected fund flows, which lead to higher AUM all else equal. Second, expected fund size is more heterogeneous across funds who adopt generic prospectuses. This reflects the heterogeneity in mistakes made by investors, on average, when evaluating managers' performance. The more specialized the strategy, the greater the expected mistake. The top-left panel in figure 3 illustrates these findings.

Proposition 3 (Flow-Performance Sensitivity). *The flow-performance sensitivity at time 2,*

$$\frac{\partial flow_{i,2}}{\partial \hat{r}_{i,2}^A} = \frac{\lambda_p^\alpha}{\bar{\alpha} - f}, \quad (26)$$

is always larger for funds adopting detailed prospectuses, irrespective of the factor return F_2 . Moreover, the difference in flow-performance sensitivity between the two prospectus types is increasing in F_2 .

The result in proposition 3 highlights the core mechanism of our model—investor learning. The active returns of a fund with a detailed prospectus provide a more precise signal about the manager's skill, since the uncertainty about how to benchmark her performance is completely eliminated. This implies that a given active return will trigger a greater flow of capital (inflow or outflow) for a fund with a detailed prospectus than a fund with a generic prospectus. The difference in flow-performance sensitivity between the two prospectus types increases with the magnitude of the factor return because, when there is uncertainty about the benchmark (as in the case of generic prospectuses), a larger positive or negative return makes it more difficult to separate skill (whose prior variance is constant) from factor exposure. The top-right panel of figure 3 illustrates this result.

Proposition 4 (Fund Flow Uncertainty). *The cross-sectional average of fund flow uncertainty at time 2 is equal to $\sigma_\alpha^2 / ((\bar{\alpha} - f)\sqrt{\sigma_\alpha^2 + \sigma_\varepsilon^2})$ for funds adopting a detailed prospectus, and*

$$\frac{1}{m(\bar{\alpha} - f)} \int_{\underline{\gamma}/\sigma_\gamma}^{\bar{\gamma}/\sigma_\gamma} \sqrt{\mathbb{E}[(\lambda_g^\alpha)^2](\sigma_\alpha^2 + \sigma_\varepsilon^2) + \text{Var}[\lambda_g^\alpha F_2] \sigma_\gamma^2 z^2} \phi(z) dz \quad (27)$$

for funds adopting a generic prospectus, where $\phi(\cdot)$ is the probability density function of a standard normal distribution, and m is as in (24).

For funds adopting detailed prospectuses, fund flow uncertainty, which is measured by the standard deviation of fund flows, is driven only by the idiosyncratic component of returns and the uncertainty about managerial skill, $\varepsilon_i + (\alpha_i - \bar{\alpha})$. Although these two sources of uncertainty also plays a role for funds adopting generic prospectuses, here it is also affected by the mistakes that investors make when estimating the fund strategies. These mistakes constitute an additional source of uncertainty, $F_2 \gamma_i$, which tends to increase overall uncertainty of capital flows. However, this is not always the case, as the higher uncertainty is also dampened by a lower flow-performance sensitivity, λ_g^a , reflecting the slower learning rate of the investors who read generic prospectuses.

The bottom-left panel in figure 3 provides an example in which the dampening effect of lower flow-performance sensitivity dominates when γ_i is close to 0 (i.e., when investors' mistakes in inferring the fund strategy are small). In this case, the fund flow uncertainty associated with a generic prospectus becomes lower than the fund flow uncertainty associated with a detailed prospectus. It follows that the cross-sectional average of fund flow uncertainty might be higher or lower for generic prospectuses, depending on the distribution of fund strategies.

4 Empirical Analysis

4.1 Data

We focus on US active equity mutual funds, combining data from four different sources: (i) information about mutual fund characteristics and returns is obtained from the CRSP Survivorship-Bias-Free Mutual Fund dataset; (ii) holdings data are from CRSP and Thomson CDA/Spectrum Mutual Fund Holdings; (iii) stock-level information is from the CRSP security file; and (iv) textual data from mutual fund prospectuses are obtained from the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) of the SEC. After combining these datasets, our final sample con-

sists of 2,995 unique funds and 86,056 fund–quarter observations, from January 2000 to December 2017.

We restrict the sample by applying various filters. First, we limit the focus of our analysis to equity funds, excluding international funds, sector funds, index funds, and underlying variable annuities. Second, we exclude observations prior to each fund’s first offer date to avoid incubation bias (Evans (2010)), as well as funds with less than \$5 million in Total Net Assets (TNA) (Kacperczyk et al. (2008)) and funds with fewer than 12 months of observations. Finally, we remove funds that hold fewer than 10 stocks or less than 80% of their assets (excluding cash) in US common stocks, on average.

While our analysis is conducted at the fund level, all CRSP data is reported at the share class level. Hence, at each point in time, we aggregate information across all share classes belonging to the same fund, taking the sum of TNA and the average of all other variables (e.g. fees, returns, turnover, ...), weighted by lagged TNA. Following Abis (2020), we identify share classes of the same fund by constructing a comprehensive fund identifier, using the CRSP Class Group identifier, the WFICN identifier in the MFLinks linking table, and fund names. This choice is particularly relevant for matching returns and characteristics to funds’ holdings. In fact, the MFLinks linking table excludes many new funds in recent years (Zhu (2020); Shive and Yun (2013)); for that reason we use the Thomson CDA/Spectrum Mutual Fund Holdings dataset from January 2000 to August 2008 and the CRSP Mutual Fund Holdings dataset from September 2008 to December 2017. We then forward-fill holdings to the monthly frequency.²⁶

We follow Abis and Lines (2020) for the collection and processing of prospectuses data. Although the EDGAR system has been active since 1994, reliable data can be obtained only as of 2000, which is when our sample starts. We are able to match 27,353 prospectuses to funds of interest. We focus on a specific section of prospectuses, the *Principal Investment Strategies (PIS)* section, in which funds are obliged (by law) to disclose the types of assets they hold and the key criteria used in selecting those assets. Since any material change to the management or strategy of a fund must be reported to both the SEC and fund investors, for any month in which a prospectus is not available, we forward-fill using the latest available prospectus.

²⁶Monthly holdings are already available for 42% of the final sample. 90% of the data is forward-filled for at most 1 quarter and 99% is forward-filled for at most 2 quarters. Maximum forward-filling is restricted to 1 year.

4.2 Measurement

To test the predictions of our model, we require empirical proxies for four conceptual quantities: (1) the investment category/mandate of each fund; (2) the amount of detail in fund prospectuses; (3) the objective and investor-perceived benchmark for each fund; (4) the degree of specialization of a fund’s strategy relative to its peers; (5) and the uncertainty about fund flows.

We proxy for uncertainty in fund flows by utilizing the 12-months rolling volatility in net fund flows (i.e.; in the change in Total Net Assets that cannot be explained by returns).

The following subsections describe the construction of all other proxies.

4.2.1 Fund Mandates and Financial Content

Section 2 describes our measure of mandates (SPGs) and of the level of detail of strategy descriptions (FinCon). In the context of the model, *low* levels of *financial content* correspond to *generic* prospectuses, and *high* levels correspond to *detailed* prospectuses. To emphasize this correspondence, when using *FinCon* as an explanatory variable in regressions, we construct dummy variables for low, medium, and high *FinCon*, based on terciles of its distribution.

4.2.2 Objective and Perceived Benchmarks

For our purposes, a fund’s *objective* benchmark is a composite factor portfolio constructed from its true exposure to each relevant factor, while its *perceived* benchmark is the factor portfolio actually used by investors to distinguish active from passive performance, as revealed by their capital allocation decisions.

We construct an estimate of each fund’s *objective benchmark* using the vector of factor loadings from a regression of raw monthly fund returns $r_{i,t}$ on factor returns F_t over the lifetime of the fund:

$$r_{i,t} = \alpha_i + \beta_i^{ol} F_t + \varepsilon_{i,t}. \quad (28)$$

The estimated vector of coefficients $\hat{\beta}_i^{ol}$ represents the objective benchmark of fund i . In our baseline specification, we use the six-factor model of Carhart (1997) and Fama and French (2015).

In line with recent work by Berk and Van Binsbergen (2016) and Barber et al. (2016), we use fund flows to infer investors’ perception of mutual funds’ benchmarks. The idea is to select, for

each fund, the vector of loadings which best explains variations in the investor's capital flows. In a linear regression model, it is the benchmark that minimizes the mean-square error of regressing fund flows on the active returns computed with that benchmark.

Fund flows are constructed using the standard definition:

$$flow_{i,t+1} \equiv \frac{TNA_{i,t+1} - TNA_{i,t}(1 + r_{i,t+1})}{TNA_{i,t}}. \quad (29)$$

Then, assuming a factor structure with constant loadings, we estimate the linear model:

$$flow_{i,t+1} = a_i + b_i r_{i,t} - b'_{F,i} F_t + \varepsilon_{i,t+1}, \quad (30)$$

and obtain the *perceived benchmark* of fund i as the vector

$$\hat{\beta}_i^p = \frac{1}{\hat{b}_i} \hat{b}_{F,i}. \quad (31)$$

It follows that fund i 's perceived active returns (PAR) are obtained as follows:

$$\hat{r}_{i,t}^A = r_{i,t} - \hat{\beta}_i^{p'} F_t. \quad (32)$$

Armed with estimates of the objective and perceived benchmarks, we can also compute a measure of distance, or *misalignment*, between them:

$$BenchMisalign_i = \sqrt{\sum_{k=1}^K \left(\frac{(\hat{\beta}_i^o(k) - \hat{\beta}_i^p(k)) - \text{mean}(\hat{\beta}_i^o(k) - \hat{\beta}_i^p(k))}{\text{std}(\hat{\beta}_i^o(k) - \hat{\beta}_i^p(k))} \right)^2} \quad (33)$$

where $\hat{\beta}_i^o(k)$ and $\hat{\beta}_i^p(k)$ denote the k -th element of the objective and the perceived benchmark, respectively, and K is the number of factors characterizing the benchmarks. The operators $\text{mean}(\cdot)$ and $\text{std}(\cdot)$ represent the cross-sectional mean and standard deviation, respectively, and are used to normalize the differences between objective and perceived factor loadings so that they are all on the same scale. Larger distances imply greater misalignment between the true and perceived benchmarks.

4.2.3 Fund Strategy Specialization

Our measure of strategy specialization is the sum of squared normalized objective factor loadings, estimated from equation 28. The loadings are normalized across all funds assigned to the same SPG as fund i at time t .²⁷:

$$\hat{\beta}_{i,t}^{norm} = \frac{\hat{\beta}_{i,t}^o - \text{mean}(\hat{\beta}_{i,t}^o)}{\text{std}(\hat{\beta}_{i,t}^o)}, \quad (34)$$

where $\text{mean}(\hat{\beta}_{i,t}^o)$ and $\text{std}(\hat{\beta}_{i,t}^o)$ represent, respectively, the cross-sectional average and standard deviation of the vector of objective factor loadings across all funds belonging to the same SPG as fund i at time t . Betas are winsorized at the 1% level before normalization.

Our measure of specialization is then:

$$\text{Specialization}_{i,t} = \sum_{k=1}^K \left(\hat{\beta}_{i,t}^{norm}(k) \right)^2, \quad (35)$$

where K is the number of factors. This measure summarizes deviations from the average exposure to each factor among funds belonging to the same SPG.

4.3 Results

4.3.1 Prospectus Choice

Proposition 1 shows that managerial incentive to disclose more information via prospectuses is positively related to the degree of specialization of the fund's strategy. In order to test this prediction, we run variants of the following regression:

$$\text{FinCon}_{i,t} = a + b\text{Specialization}_{i,t} + \gamma' X_{i,t-1} + f(t, \eta_i, \zeta_{SPG}) + \varepsilon_{i,t}, \quad (36)$$

where $\text{FinCon}_{i,t}$ measures financial content in the prospectus of fund i at time t (see subsection ??) and $\text{Specialization}_{i,t}$ is as defined in subsection 4.2.3. $X_{i,t-1}$ is a vector of demeaned fund level control variables: the log of fund size and age, the turnover and expense ratios, net fund flows and flow volatility, and additional style control variables constructed as the average loading of the stocks held by fund i at time t on the Fama and French (2015) and Carhart (1997) size,

²⁷In unreported robustness tests, we also construct the specialization measure relative to 27 clusters of funds sorted by size, book-to-market and momentum, finding similar results.

value, momentum, investment and profitability factors. In all subsequent regressions, we use the same set of fund-level controls, only omitting variables when they appear on the left-hand side. We always use contemporaneous variables for endogenous relationships predicted by the model, while exogenous controls are always lagged one quarter. We further add quarter, SPG-quarter, or fund fixed effects, and cluster standard errors at the quarter and fund level. The $f(\cdot)$ operator defines how the fixed effects are combined in different specifications. Fund mandates (SPGs) are constructed as described in section 4.2.1.

Table III displays the results of these regressions. All specifications show that higher specialization is associated with a significantly higher level of financial content in strategy descriptions, confirming the main prediction of the model.²⁸ The results are similar across the different fixed effect specifications, which correspond to “within mandate” (column 1), “across funds” (columns 2), and “within fund” (columns 3) identification. The magnitudes of the effects are sensible: a one-standard-deviation increase in specialization is associated with an increase in financial content ranging from 5% (columns 1 and 2) to 16% (column 3) relative to its own standard deviation. The results are statistically significant at the 1% level for the “within mandate” and “across funds” specifications, and at the 5% level for the “within fund” specification.

4.3.2 Fund Size

The second prediction of the model is that, within each investment mandate, funds with generic prospectuses should be larger. We test this prediction by running the following regression:

$$\ln(TNA_{i,t}) = a + bFinCon_{i,t} + \gamma'X_{i,t-1} + \iota_{SPG,t} + \varepsilon_{i,t} \quad (37)$$

where $\ln(TNA_{i,t})$ is the natural logarithm of the size of fund i at time t . We log-transform this variable due to its high skewness. $X_{i,t-1}$ is a vector of the standard fund-level control variables, and $\iota_{SPG,t}$ are SPG-quarter fixed effects (the other fixed effect combinations are omitted here for brevity, but show similar results).

Table IV reports the results, with financial content included in the regressions either as the continuous variable $FinCon$ (column 1), or as dummy variables corresponding to the middle (Medium $FinCon$) and top tercile (High $FinCon$) of the variable’s distribution (columns 2-4). Columns 1 and 2 report coefficient estimates for the full sample, while columns 3 and 4 report the results from the subsamples of young (below-median age) and old (above-median age) funds,

²⁸In untabulated results, we confirm that this relationship is robust to measuring financial terminology using dummies for middle and top terciles instead of the continuous variable.

respectively. Subsampling by age is included in the table to test whether the relationship between financial content and size differs for funds depending on how long they have had to accumulate additional flows from advantageous investor mistakes in skill attribution.

Across all specifications, the negative and significant \hat{b} coefficients indicate that fund size decreases when financial content is greater (i.e., for more detailed prospectuses), as predicted by the model. Recall that the dependent variable is the natural logarithm of total net assets; therefore the coefficients can be interpreted as approximate percentage changes in dollar TNA. Accordingly, our estimates indicate that high *FinCon* funds are about 20% smaller overall relative to low *FinCon* funds (column 2), and the result is significant at the 1% level. For Young funds (column 3), the result is present but weaker in both magnitude (detailed funds are about 17% smaller) and statistical significance (5% level). For old funds, those with generic strategy descriptions have had more time to accumulate the benefits of investor mistakes than those with detailed descriptions, which can account for the larger difference of about 25% between them.

4.3.3 Flow Volatility

As shown in the bottom-right panel of figure 3, the model predicts that an increase in a fund's exposure to the common risk factor should be associated with an increase in both size and flow volatility within the subset of funds with *generic* prospectuses. This result reflects the central trade-off of the model: investors' performance attribution mistakes are beneficial to managers with more positive exposure to the factor, but greater errors also result in higher volatility. On the other hand, for funds with detailed prospectuses (in equilibrium, those with sufficiently extreme exposures to the factor), the volatility drawback always dominates and both fund size and flow volatility remain constant.

Theoretically, the most straightforward test of this prediction would be to regress size and flow volatility on funds' exposure to the common factor, separately for funds with detailed and generic prospectuses. However, while the model assumes a single-factor structure for tractability, in reality many different factors may be relevant to investors. The outcome of such a regression would be highly sensitive to the specification of the factors and difficult to implement in practice. In previous tests, we have relied on an indirect measure of specialization—the sum of squared factor exposures normalized within investment mandates. However, this measure is unsigned, and so is unsuitable for testing predictions based on the *direction* of factor exposures.

Our solution is to measure the cross-sectional correlation between fund size and flow volatility for funds with detailed or generic prospectuses: the model predicts that, in equilibrium, this

correlation should be positive for funds with generic prospectuses and zero for funds with detailed prospectuses. In reality, there may be other forces affecting size and flow volatility that lie outside of the model, so we would not necessarily expect the correlation for detailed prospectuses to be precisely zero in practice. However, if the mechanism underlying the model is relevant in reality, the correlation should be higher for funds with more generic prospectuses.

We test this prediction by running the following regression:

$$\begin{aligned}
 FlowVol_{i,t} = & a + b_0 \ln(TNA_{i,t}) + b_1(\ln(TNA_{i,t}) \times Med-FinCon_{i,t}) + \\
 & b_2(\ln(TNA_{i,t}) \times High-FinCon_{i,t}) + Med-FinCon_{i,t} + High-FinCon_{i,t} + \\
 & \gamma' X_{i,t-1} + \iota_{SPG,t} + \varepsilon_{i,t}.
 \end{aligned} \tag{38}$$

$FlowVol_{i,t}$ is the 12-months-rolling net flow volatility and $\ln(TNA_{i,t})$ is the natural logarithm of total net assets for fund i at time t . Funds in the top and middle terciles of our financial content measure are indicated by the dummy variables $High-FinCon$ and $Med-FinCon$, respectively. The vector $X_{i,t-1}$ contains the usual fund-level control variables, and $\iota_{SPG,t}$ indicates SPG-quarter fixed effects.

Table V reports the estimated coefficients. Column 1 shows a version of the regression without interaction terms for the subsample of funds with generic strategy descriptions (bottom tercile of $FinCon$) and column 2 shows the same for funds with detailed prospectuses (top tercile of $FinCon$). Column 3 reports coefficients from the full-sample regression described in equation 38, including the interaction terms. For both types of funds, the relationship between fund size and flow volatility is significantly positive.²⁹ However, column 3 shows that this relationship is weaker for funds with detailed prospectuses, as indicated by the negative interaction coefficient of -1.771 for the $High-FinCon$ dummy. Consistent with the main mechanism of the model, the association between fund size and flow volatility is 17% smaller for funds with detailed prospectuses.

4.3.4 Flow-Performance Sensitivity

A key implication of the mechanism of our model is that investors with access to more detailed strategy descriptions should be better able to infer the appropriate benchmark for the fund, learn about fund skill, and allocate their capital accordingly. In practical terms, funds' flow-performance

²⁹It is worth noting that this relationship is not simply a mechanical artifact of increasing size, as flow volatility is computed from monthly percentage flows (i.e., scaled by total assets).

sensitivity (FPS) should be increasing in the amount of detail (financial content) in their prospectuses. To test this prediction, we run the following regression:

$$\begin{aligned}
 Flow_{i,t+1} = & a + b_0 \hat{r}_{i,t}^A + b_1(\hat{r}_{i,t}^A \times Med-FinCon_{i,t}) + \\
 & b_2(\hat{r}_{i,t}^A \times High-FinCon_{i,t}) + \delta Flow_{i,t} + \xi(\hat{r}_{i,t}^A)^2 + \\
 & \gamma' X_{i,t} + f(\iota_t, \eta_i, \zeta_{SPG}) + \varepsilon_{i,t},
 \end{aligned} \tag{39}$$

where *Med-* and *High-FinCon*_{*i,t*} are dummy variables associated with the middle and top terciles of the financial content of fund *i*'s prospectus in quarter *t*. $\hat{r}_{i,t}^A$ is fund *i*'s perceived active return (PAR) at time *t*, obtained as the difference between the fund's total return and its perceived benchmark (see section 4.2.2). $X_{i,t}$ are demeaned fund-specific control variables, and we also include past flow ($Flow_{i,t}$) and squared PAR ($(\hat{r}_{i,t}^A)^2$), the latter to capture potential convexity in the flow-performance relationship. ι_t are quarter fixed-effects, η_i are fund fixed effects, ζ_{SPG} are SPG fixed effects, and the $f(\cdot)$ operator defines how the fixed effects are combined.

The coefficient of interest is \hat{b}_2 , which indicates by how much the FPS of fund *i* in quarter *t* changes as the financial content measure increases from the bottom to the top tercile of its cross-sectional distribution. The coefficient \hat{b}_0 represents the FPS for the funds in the bottom tercile of FinCon, while $\hat{b}_0 + \hat{b}_2$ represents the FPS for funds in the top tercile. Hence, \hat{b}_2 captures the incremental effect.

Table VI presents the results of estimating this regression, which are consistent with the learning mechanism in our model. The results are consistent across the different fixed effects specifications. Column 1 (“within mandate”) shows a positive sensitivity to lagged perceived active returns (coefficient of 0.015) for funds with generic prospectuses, while for funds with detailed prospectuses, the FPS doubles: the coefficient on the interaction term between the top *FinCon* tercile and lagged PAR is 0.0155. This confirms our model prediction that flow-performance sensitivity is stronger for funds whose prospectuses have higher financial content. We note that compared to the existing literature, we do not rely on objective benchmarks to compute active returns in these regressions. Instead we use perceived active returns, which reflect the way investors infer the relative performance of mutual funds in practice.

In columns 2-3 we test the interaction between FPS and financial content in alternative panel specifications: across and within funds, respectively. That is, in column 2 we absorb quarter fixed effects (“across funds”), and in column 3 we absorb fund fixed effects (“within fund”). Both columns have larger FPS for funds with more detailed prospectuses. In particular, column 3 suggests that when the *same* fund increases the financial content or the informativeness of its prospectus, it

experiences higher flow-performance sensitivity. This result mitigates the concern that higher FPS might be driven by some other unobservable, time-invariant characteristics of the funds that write more detailed prospectuses.

4.3.5 Benchmark Misalignment

Another key aspect of the learning mechanism in our model is that investors should be able to identify the objective/true benchmark of a fund more accurately if a fund’s prospectus contains detailed information about its strategy. This implies that perceived and objective benchmarks should be more aligned (i.e., factor loadings should be more similar) for funds with detailed prospectuses. To test this implication, we use the measure of *benchmark misalignment* defined in section 4.2.2, equation 33.

We then estimate the relationship between benchmark misalignment and financial content using variants of the following regression:

$$BenchMisalign_{i,t} = a + bFinCon_{i,t} + \gamma' X_{i,t-1} + \iota_{SPG,t} + \varepsilon_{i,t}. \quad (40)$$

$X_{i,t-1}$ are the usual fund-specific control variables and $\iota_{SPG,t}$ are SPG-quarter fixed-effects.

Table VII presents the results. Column 1 reports the estimated relationship, which is in line with the mechanism of our model. We find that funds with more detailed strategy descriptions (i.e., high-*FinCon* funds) have lower misalignment, with a coefficient of -0.174 , corresponding to a decrease of about 16% relative to the funds with generic descriptions.

4.3.6 Age Effects

Although not an explicit prediction of the model, if our proposed learning mechanism is indeed behind the documented empirical findings, one might expect the results on flow-performance sensitivity and benchmark misalignment to be more pronounced for younger funds. Without a long return history, investors may know less about the strategy of young funds, and in these cases the information communicated via prospectuses may be particularly helpful. On the other hand, prospectus information may be less relevant for older funds with longer return histories.

As reported in columns 4-5 of table VI, we find that the financial content of fund prospectus increases FPS only for younger funds (below-median age). In fact, flows have almost no relationship with perceived active returns among young funds that adopt generic prospectuses. This suggests

that investors find it harder to benchmark the performance of these funds. For the group of older funds (i.e., fund age above median), instead, we find that there is no additional FPS associated with detailed prospectuses.

Strong age effects can also be found when analyzing benchmark misalignments. In columns 2-3 of table VII, we find that detailed prospectuses are associated with lower benchmark misalignment only for the subsample of young funds. For older funds with enough performance history, prospectus information does not appear to make a material difference to investors trying to infer fund strategies. Overall, our findings confirm that fund prospectuses are an important source of information for investors, and that this information is particularly helpful when only a short history of performance is available, consistent with the key mechanism of our model.

5 Conclusion

In this paper we study the interplay between investor learning and fund managers' incentives to disclose information through prospectuses. We show both theoretically and empirically that these incentives vary depending on how specialized the manager's strategy is relative to her peers. In deciding how much information to disclose, the manager faces a trade-off. Greater information disclosure provides a clearer signal about the fund's strategy to investors, and therefore about the appropriate benchmark. In response to this signal, investors perceive the fund's true benchmark more accurately, learn faster about the manager's skill, and therefore exhibit a higher flow-performance sensitivity. Less detailed disclosure, on the other hand, causes investors to make more mistakes in determining the correct benchmark for the fund, which leads to higher net flows volatility, but also greater fund size. In equilibrium, managers with more specialized strategies are incentivized to disclose more information, while managers with more standardized strategies are incentivized to obscure information by writing more generic prospectuses. We confirm our model's predictions and find evidence for its core learning mechanism via textual analysis of Principal Investment Strategy descriptions from fund prospectuses. Lastly, we show that investors' learning from prospectuses is greater for young funds, reinforcing the proposed theoretical mechanism.

APPENDIX

A Proofs

Lemma A.1 (Truncated Normal Distribution). *Let X be normally distributed with mean μ_X and variance σ_X^2 , $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$. Define $Y \equiv (X | \underline{x} < X < \bar{x})$. It follows that Y has a truncated normal distribution with density equal to*

$$h_Y(y) = \begin{cases} \sigma_X^{-1} \frac{\phi((y-\mu_X)/\sigma_X)}{\Phi(\bar{z})-\Phi(\underline{z})} & \text{if } \underline{x} < y < \bar{x} \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A.1})$$

and first and second moments equal to

$$\mathbb{E}[Y] = \mu_X + \sigma_X \frac{\phi(\underline{z}) - \phi(\bar{z})}{\Phi(\bar{z}) - \Phi(\underline{z})}, \quad (\text{A.2})$$

$$\text{Var}[Y] = \sigma_X^2 \left(1 + \frac{\underline{z}\phi(\underline{z}) - \bar{z}\phi(\bar{z})}{\Phi(\bar{z}) - \Phi(\underline{z})} - \left(\frac{\phi(\underline{z}) - \phi(\bar{z})}{\Phi(\bar{z}) - \Phi(\underline{z})} \right)^2 \right), \quad (\text{A.3})$$

where $\underline{z} \equiv (\underline{x} - \mu_X)/\sigma_X$, $\bar{z} \equiv (\bar{x} - \mu_X)/\sigma_X$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and the cumulative distribution function of a standard normal distribution, respectively.

Proof. See [Johnson et al. \(1994\)](#). □

Proof of Proposition 1. A fund manager chooses a generic prospectus over a detailed prospectus when (22) is larger than (21). Let $\Psi(\gamma_i)$ denote the difference between (22) and (21), multiplied by $2c/f$:

$$\Psi(\gamma_i) = \frac{2c}{f} \mathbb{E}[\lambda_g^\alpha F_2] \gamma_i - \text{Var}[\lambda_g^\alpha ((\alpha_i - \bar{\alpha}) + \varepsilon_i + F_2 \gamma_i)] + \frac{\sigma_\alpha^4}{(\sigma_\alpha^2 + \sigma_\varepsilon^2)}, \quad (\text{A.4})$$

where, for brevity, we omitted the subscript 0 and superscript \mathcal{M}_i from the moments' operators. Defining the variable $z_i \equiv (\alpha_i - \bar{\alpha}) + \varepsilon_i$, which is normally distributed with mean 0 and variance $(\sigma_\alpha^2 + \sigma_\varepsilon^2)$, the variance term in (A.4) can be written as

$$\begin{aligned} \text{Var}[\lambda_g^\alpha (z_i + F_2 \gamma_i)] &= \text{Var}[\lambda_g^\alpha F_2] \gamma_i^2 + \text{Var}[\lambda_g^\alpha z_i] + 2\text{Cov}[\lambda_g^\alpha z_i, \lambda_g^\alpha F_2] \gamma_i \\ &= \text{Var}[\lambda_g^\alpha F_2] \gamma_i^2 + \mathbb{E}[(\lambda_g^\alpha)^2] (\sigma_\alpha^2 + \sigma_\varepsilon^2), \end{aligned} \quad (\text{A.5})$$

where the second equality follows from the independence of z_i and λ_g^α and the fact that $\mathbb{E}[z_i] = 0$:

$$\begin{aligned}\text{Var}[\lambda_g^\alpha z_i] &= \mathbb{E}[(\lambda_g^\alpha)^2 (z_i)^2] - \mathbb{E}[\lambda_g^\alpha z_i]^2 \\ &= \mathbb{E}[(\lambda_g^\alpha)^2] \mathbb{E}[(z_i)^2] - (\mathbb{E}[\lambda_g^\alpha] \mathbb{E}[z_i])^2 \\ &= \mathbb{E}[(\lambda_g^\alpha)^2] (\sigma_\alpha^2 + \sigma_\varepsilon^2),\end{aligned}\tag{A.6}$$

and

$$\begin{aligned}\text{Cov}[\lambda_g^\alpha z_i, \lambda_g^\alpha F_2] &= \mathbb{E}[z_i (\lambda_g^\alpha)^2 F_2] - \mathbb{E}[\lambda_g^\alpha z_i] \mathbb{E}[\lambda_g^\alpha F_2] \\ &= \mathbb{E}[z_i] \text{Cov}[\lambda_g^\alpha, \lambda_g^\alpha F_2] \\ &= 0,\end{aligned}\tag{A.7}$$

given that $\text{Var}[XY] = \mathbb{E}[X^2] \mathbb{E}[Y^2] - \mathbb{E}[X]^2 \mathbb{E}[Y]^2$ when X and Y are independent. Therefore, $\Psi(\gamma_i)$ is a quadratic function of γ_i ,

$$\Psi(\gamma_i) = a_0 + 2a_1 \gamma_i + a_2 \gamma_i^2,\tag{A.8}$$

with coefficients

$$a_0 = ((\lambda_d^\alpha)^2 - \mathbb{E}[(\lambda_g^\alpha)^2]) (\sigma_\alpha^2 + \sigma_\varepsilon^2),\tag{A.9}$$

$$a_1 = (c/f) \mathbb{E}[\lambda_g^\alpha F_2],\tag{A.10}$$

$$a_2 = -\text{Var}[\lambda_g^\alpha F_2].\tag{A.11}$$

The equation $\Psi(\gamma_i) = 0$ admits two distinct (real) solutions $(\underline{\gamma}, \bar{\gamma})$ since its discriminant $(a_1^2 - a_2 a_0)$ is strictly positive. This is because $\lambda_d^\alpha > \lambda_g^\alpha(F_2)$ for any $F_2 \neq 0$, and $\lambda_d^\alpha = \lambda_g^\alpha(F_2)$ for $F_2 = 0$, thus making $(\lambda_d^\alpha)^2 > \mathbb{E}[(\lambda_g^\alpha)^2]$. Since $\partial^2 \Psi(\gamma_i) / \partial \gamma_i^2 = -2 \text{Var}[\lambda_g^\alpha F_2] < 0$, $\Psi(\gamma_i)$ is concave, implying that $\Psi(\gamma_i) > 0$ (i.e., a generic prospectus is optimal) for $\underline{\gamma} < \gamma_i < \bar{\gamma}$, and negative (i.e., a detailed prospectus is optimal) otherwise. Since the coefficient $a_0 > 0$, it follows that $\bar{\gamma} > 0$ and $\underline{\gamma} < 0$. When $\mathbb{E}[\lambda_g^\alpha F_2] > 0$, the coefficient a_1 is positive, it implies that $|\bar{\gamma}| > |\underline{\gamma}|$. When, instead, $\mathbb{E}[\lambda_g^\alpha F_2] < 0$, $a_1 < 0$ and $|\bar{\gamma}| < |\underline{\gamma}|$.

Given a measure one of fund managers, the fraction of managers adopting a generic prospectus, m , is equal to $\mathbb{P}(\underline{\gamma} < \gamma_i < \bar{\gamma}) = \Phi(\bar{\gamma}/\sigma_\gamma) - \Phi(\underline{\gamma}/\sigma_\gamma)$. \square

Proof of Proposition 2. Given the equilibrium fund size in (13), the expected fund size associated with detailed and generic prospectuses are equal to

$$\mathbb{E}[q_{i,2}(d)] = \frac{\mathbb{E}[\hat{\alpha}_{i,2}(d)] - f}{c} = \frac{\bar{\alpha} - f}{c}, \quad (\text{A.12})$$

$$\mathbb{E}[q_{i,2}(g)] = \frac{\mathbb{E}[\hat{\alpha}_{i,2}(g)] - f}{c} = \frac{\bar{\alpha} - f}{c} + \frac{1}{c}\mathbb{E}[\lambda_g^\alpha F_2]\gamma_i, \quad (\text{A.13})$$

respectively. Since $\mathbb{E}[q_{i,2}(d)]$ does not depend on the fund strategy γ_i , the cross-sectional average is equal to $(\bar{\alpha} - f)/c$. Since, instead, $\mathbb{E}[q_{i,2}(g)]$ depends on the fund strategy γ_i , the cross-sectional average is equal to

$$\frac{\bar{\alpha} - f}{c} + \frac{1}{c}\mathbb{E}[\lambda_g^\alpha F_2]\mathbb{E}[\gamma_i|\underline{\gamma} < \gamma_i < \bar{\gamma}], \quad (\text{A.14})$$

where the last term is the mean of a truncated normal distribution. By the results in Lemma A.1,

$$\mathbb{E}[\gamma_i|\underline{\gamma} < \gamma_i < \bar{\gamma}] = \sigma_\gamma \frac{\phi(\underline{\gamma}/\sigma_\gamma) - \phi(\bar{\gamma}/\sigma_\gamma)}{\Phi(\bar{\gamma}/\sigma_\gamma) - \Phi(\underline{\gamma}/\sigma_\gamma)} \quad (\text{A.15})$$

and (25) obtains.

Since the function $\Lambda(F_2) = \lambda_g^\alpha F_2 = \sigma_\alpha^2 F_2 / (\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\gamma^2 F_2^2)$ is symmetric around 0 (i.e., $\Lambda(-F_2) = -\Lambda(F_2)$), it follows that $\mathbb{E}[\lambda_g^\alpha F_2]$ is positive if $\mu_F > 0$, and is negative if $\mu_F < 0$. This is because the probability density function of the factor return F_2 , $h_{F_2}(\cdot)$, is such that $h_{F_2}(x) > h_{F_2}(-x)$ for $x > 0$ if $\mu_F > 0$, and $h_{F_2}(x) < h_{F_2}(-x)$ for $x > 0$ if $\mu_F < 0$. Since when $\mathbb{E}[\lambda_g^\alpha F_2] > 0$, $|\bar{\gamma}| > |\underline{\gamma}|$, and when $\mathbb{E}[\lambda_g^\alpha F_2] < 0$, $|\bar{\gamma}| < |\underline{\gamma}|$, it follows that $\phi(\underline{\gamma}/\sigma_\gamma) - \phi(\bar{\gamma}/\sigma_\gamma)$ has the same sign as $\mathbb{E}[\lambda_g^\alpha F_2]$, making the cross-sectional average of the expected fund size associated with a generic prospectus always larger than that associated with a detailed prospectus.

While the cross-sectional dispersion (variance) of $\mathbb{E}[q_{i,2}(d)]$ is equal to zero, that of $\mathbb{E}[q_{i,2}(g)]$ is positive and equal to

$$\frac{\sigma_\gamma^2}{c^2}\mathbb{E}[\lambda_g^\alpha F_2]^2 \left(1 + \frac{(\underline{\gamma}/\sigma_\gamma)\phi(\underline{\gamma}/\sigma_\gamma) - (\bar{\gamma}/\sigma_\gamma)\phi(\bar{\gamma}/\sigma_\gamma)}{\Phi(\bar{\gamma}/\sigma_\gamma) - \Phi(\underline{\gamma}/\sigma_\gamma)} \left(\frac{\phi(\underline{\gamma}/\sigma_\gamma) - \phi(\bar{\gamma}/\sigma_\gamma)}{\Phi(\bar{\gamma}/\sigma_\gamma) - \Phi(\underline{\gamma}/\sigma_\gamma)} \right)^2 \right), \quad (\text{A.16})$$

where the last term corresponds to $\text{Var}[\gamma_i|\underline{\gamma} < \gamma_i < \bar{\gamma}]$ and is obtained by Lemma A.1. \square

Proof of Proposition 4. Fun size at time 1 and time 2 are equal to $(\bar{\alpha} - f)/c$ and $(\hat{\alpha}_{i,2}(p) - f)/c$, respectively. Therefore, fund flows at time 2 associated with a fund prospectus p , $q_{i,2}(p)/q_{i,1} - 1$, are equal to

$$flow_{i,2}(p) = \frac{\hat{\alpha}_{i,2}(p) - \bar{\alpha}}{\bar{\alpha} - f}. \quad (\text{A.17})$$

Fund flow volatility is therefore equal to

$$\begin{aligned} \sqrt{\text{Var}[flow_{i,2}(p)]} &= \frac{1}{\bar{\alpha} - f} \sqrt{\text{Var}[\hat{\alpha}_{i,2}(p)]} \\ &= \frac{1}{\bar{\alpha} - f} \sqrt{\text{Var}[\lambda_p^\alpha(z_i + \mathbf{1}_{\{p=g\}}F_2\gamma_i)]}, \end{aligned} \quad (\text{A.18})$$

where $z_i \equiv (\alpha_i - \bar{\alpha}) + \varepsilon_i$ and the second equality follows from (20). When the prospectus is detailed ($p = d$), the fund flow volatility and its cross-sectional average (since $\sqrt{\text{Var}[flow_{i,2}(d)]}$ does not depend on γ_i) are both equal to

$$\sqrt{\text{Var}[flow_{i,2}(d)]} = \frac{1}{\bar{\alpha} - f} \sqrt{\text{Var}[\lambda_d^\alpha z_i]} = \frac{\lambda_d^\alpha}{\bar{\alpha} - f} \sqrt{(\sigma_\alpha^2 + \sigma_\varepsilon^2)}. \quad (\text{A.19})$$

When the prospectus is generic ($p = g$), the fund flow volatility is equal to

$$\begin{aligned} \sqrt{\text{Var}[flow_{i,2}(g)]} &= \frac{1}{\bar{\alpha} - f} \sqrt{\text{Var}[\lambda_g^\alpha(z_i + F_2\gamma_i)]} \\ &= \frac{1}{\bar{\alpha} - f} \sqrt{\mathbb{E}[(\lambda_g^\alpha)^2](\sigma_\alpha^2 + \sigma_\varepsilon^2) + \text{Var}[\lambda_g^\alpha F_2]\gamma_i^2}, \end{aligned} \quad (\text{A.20})$$

where the second equality follows from (A.5). Since $\sqrt{\text{Var}[flow_{i,2}(g)]}$ depends on γ_i , the cross-sectional average is obtained by integrating $\sqrt{\text{Var}[flow_{i,2}(g)]}$ over γ_i in the interval $[\underline{\gamma}, \bar{\gamma}]$, thus yielding (27). \square

Proof of Proposition 3. Fund flows at time 2 associated with a fund prospectus p are as in (A.17), where, based on (17), $\hat{\alpha}_2(p) = (1 - \lambda_p^\alpha)\bar{\alpha} + \lambda_p^\alpha \hat{r}_{i,2}^A$. It follows that the flow-performance sensitivity $fps_{i,2}(p) \equiv \partial flow_{i,2}(p) / \partial \hat{r}_{i,2}^A$ is equal to

$$\frac{\lambda_p^\alpha}{\bar{\alpha} - f}. \quad (\text{A.21})$$

Since $\lambda_d^\alpha \geq \lambda_g^\alpha$ for any realization of the factor return F_2 , the flow-performance sensitivity is always larger for funds adopting detailed prospectuses. Moreover, since λ_g^α decreases with F_2 , σ_γ , and

σ_b (while λ_d^α is independent of them), the difference in flow-performance sensitivity across funds adopting detailed and generic prospectuses increases in those quantities. \square

B Financial Content and Boilerplate Measures Examples

In order to better understand the measures of financial content utilized, consider the two examples below. They have been chosen to be of same length (100 words), which is below the median PIS size. One of them is ranked in the top 10% of the financial content measure distribution while the other belongs to the bottom 10%. In both examples, the underlined words indicate the *boilerplate vector* while words highlighted in red indicate the *financial content vector*. Note that the two might be overlapping, i.e. some financial terms might be part of the boilerplate vector.

Low financial content example: *Crsp_cl_grp: 2000960; Date: 2004-12*
Fund: American AAdvantage Funds: American AAdvantage Large Cap Value Fund

Full text before pre-processing:

Ordinarili, at least 80% of the Fund's net asset plu amount borrow invest purpos invest equiti secur larg market cap unit state compani compani gener market cap similar market cap compani

Pre-processed text with financial words in red and boilerplate 4-grams underlined:

ordinarili least **fund net asset plu** amount **borrow invest** purpos **invest equiti secur** larg **market cap unit** state compani compani gener **market cap** similar market cap compani

russel time invest russel measur perform largest unit state compani base total market
cap januari market cap compani russel rang xxx million xxx billion fund invest may
includ common stock prefer stock secur convert unit state common stock unit state dollar
denomin american depositari receipt unit state dollar denomin foreign stock trade unit
state exchang collect refer stock manag current alloc fund asset among four invest sub
advis barrow hanley mewhinney strauss inc

Financial Content: 8.35%; Boilerplate: 55%

In this example there exists a large overlap between the boilerplate and financial vectors. Indeed, after removing the most frequent financial terms in the corpus and down-weighting other common financial terms, we obtain a financial content measure of 8.35%, compared to 55% of all utilized terms being classified as boilerplate.

High financial content example: *Crsp_cl_grp: 2002444 Date: 2004-12*

Fund: Delaware Group Adviser Funds: Delaware US Growth Fund

Full text before pre-processing:

We invest primarily in common stocks. Our goal is to identify companies whose earnings are expected to grow faster than the U.S. economy in general. We look for stocks with low dividend yields, strong balance sheets and high expected earnings growth rates as compared to other companies in the same industry. Whether companies provide dividend income and how much income they provide will not be a primary factor in the Fund's selection decisions. We may also invest up to 20% of the Fund's assets in debt securities and bonds. We may also invest in convertible bonds, preferred stock and convertible preferred stocks, provided that these instruments, when aggregated with the Fund's investments in debt securities and bonds, do not exceed 35% of the Fund's assets. Under normal circumstances, the Fund will invest at least 80% of its net assets in U.S. investments. This policy is not a fundamental investment policy and can be changed without shareholder approval. However, shareholders will be given notice at least 60 days prior to any such change.

Pre-processed text with financial words in red and boilerplate 4-grams underlined:

invest primarili common stock goal identifi compani whose earn expect grow faster unit
state economi gener look stock low dividend yield strong balanc sheet high expect earn

growth rate compar compani industri whether compani provid dividend incom much incom provid primari factor fund select decis may also invest fund asset debt secur bond may also invest convert bond prefer stock convert prefer stock provid instrument aggreg fund invest debt secur bond exceed fund asset normal circumst fund invest least net asset unit state invest polici fundament invest polici chang without sharehold approv howev sharehold given notic least day prior chang

Financial Content: 27.18%; Boilerplate: 19%

In this example there is minimal overlap between the financial vector and the boilerplate vector. The boilerplate vector only represents 19% of the overall word count of the strategy description. On the other hand, there is an extensive presence of financial terminology. After applying the same weights, we obtain a financial content measure of 27.18%.

References

- Abis, S. (2020). Man vs. machine: Quantitative and discretionary equity management. Working paper.
- Abis, S. and A. Lines (2020). Text-based mutual fund peer groups. Working paper.
- Akey, P., A. Z. Robertson, and M. Simutin (2021). Closet active management of passive funds. Working paper.
- Baks, K. P., A. Metrick, and J. Wachter (2001). Should investors avoid all actively managed mutual funds? A study in bayesian performance evaluation. *Journal of Finance* 56(1), 45–85.
- Barber, B. M., X. Huang, and T. Odean (2016). Which factors matter to investors? Evidence from mutual fund flows. *Review of Financial Studies* 29(10), 2600–2642.
- Barber, B. M., T. Odean, and L. Zheng (2005). Out of sight, out of mind: The effects of expenses on mutual fund flows. *The Journal of Business* 78(6), 2095–2120.
- Berk, J. and R. Green (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112(6), 1269–1295.
- Berk, J. B. and J. H. Van Binsbergen (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics* 119(1), 1–23.
- Brown, D. P. and Y. Wu (2016). Mutual fund flows and cross-fund learning within families. *Journal of Finance* 71(1), 383–424.
- Buffa, A. M. and A. Javadekar (2020). The allocation of talent across mutual fund strategies. Working paper.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance* 52(1), 57–82.
- Chevalier, J. and G. Ellison (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy* 105(6), 1167–1200.
- Choi, D., B. Kahraman, and A. Mukherjee (2016). Learning about mutual fund managers. *Journal of Finance* 71(6), 2809–2860.
- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52(3), 1035 – 1058.

- Evans, R. B. (2010). Mutual fund incubation. *The Journal of Finance* 65(4), 1581–1611.
- Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3–56.
- Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics* 116(1), 1–22.
- Franzoni, F. and M. C. Schmalz (2017). Fund flows and market states. *Review of Financial Studies* 30(8), 2621–2673.
- Gârleanu, N. and L. H. Pedersen (2018). Efficiently inefficient markets for assets and asset management. *Journal of Finance* 73(4), 1663–1712.
- Harvey, C. R. and Y. Liu (2019). Cross-sectional alpha dispersion and performance evaluation. *Journal of Financial Economics* 134(2), 273–296.
- Hillert, A., A. Niessen-Ruenzi, and S. Ruenzi (2016). Mutual fund shareholder letters: flows, performance, and managerial behavior. Working paper.
- Huang, J. C., K. D. Wei, and H. Yan (2012). Investor learning and mutual fund flows. Working paper.
- Hwang, B.-H. and H. H. Kim (2017). It pays to write well. *Journal of Financial Economics* 124(2), 373–394.
- Ivković, Z. and S. Weisbenner (2009). Individual investor mutual fund flows. *Journal of Financial Economics* 92(2), 223–237.
- Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions*. John Wiley and Sons.
- Jones, C. S. and J. Shanken (2005). Mutual fund performance with learning across funds. *Journal of Financial Economics* 78(3), 507–552.
- Kacperczyk, M., C. Sialm, and L. Zheng (2008). Unobserved actions of mutual funds. *Review of Financial Studies* 21(6), 2379–2416.
- Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds’ attention allocation. *Econometrica* 84(2), 571–626.

- Kincaid, P. J., R. P. Fishburne, R. L. J. Rogers, and B. S. Chissom (1975). Derivation of new readability formulas (automated readability index, fox count and flesh reading ease formula) for navy enlisted personnel. *Institute for Simulation and Training* 56.
- Kostovetsky, L. and J. B. Warner (2020). Measuring innovation and product differentiation: Evidence from mutual funds. *Journal of Finance* 75(2), 779–823.
- Krakov, N. J. and T. Schäfer (2020). Mutual funds and risk disclosure: Information content of fund prospectuses. Working Paper 20-54.
- Pástor, L. and R. F. Stambaugh (2002). Investing in equity mutual funds. *Journal of Financial Economics* 63(3), 351–380.
- Sheng, J., N. Xu, and L. Zheng (2021). Do mutual funds walk the talk? A textual analysis of risk disclosure by mutual funds. Working paper.
- Shive, S. and H. Yun (2013). Are mutual funds sitting ducks? *Journal of Financial Economics* 107(1), 220–237.
- Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *Journal of Finance* 53(5), 1589–1622.
- Starks, L. T. and S. Y. Sun (2016). Economic policy uncertainty, learning and incentives: Theory and evidence on mutual funds. Working paper.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62(3), 1139–1168.
- Tucker, A. M. and Y. Xia (2020). The promise & perils of plain english mutual fund disclosures. *Working Paper*.
- Tucker, A. M., Y. Xia, and S. Smelcer (2020). It ain't just what funds disclose (it's the way that they do it). Working paper.
- Zhang, A. L. (2020). Uncovering mutual fund private information with machine learning. Working paper.
- Zhu, Q. (2020). The missing new funds. *Management Science*.

Figure 1: Financial Content

The top left Panel of this Figure shows the estimated density function for our de-meaned Financial Content (FinCon) measure in the panel of PIS descriptions, constructed as in equation (1). De-meaning is done to facilitate comparison between panels. The distribution's mean was 0.18. The top right Panel shows the same measure, horthogonalized with respect to SPG fixed-effects. The bottom left Panel instead shows the horthogonalized measure with respect to SPG \times Quarter fixed-effects. Finally in the bottom-right Panel we also include fund-specific control variables to horthogonalize the financial content measure.

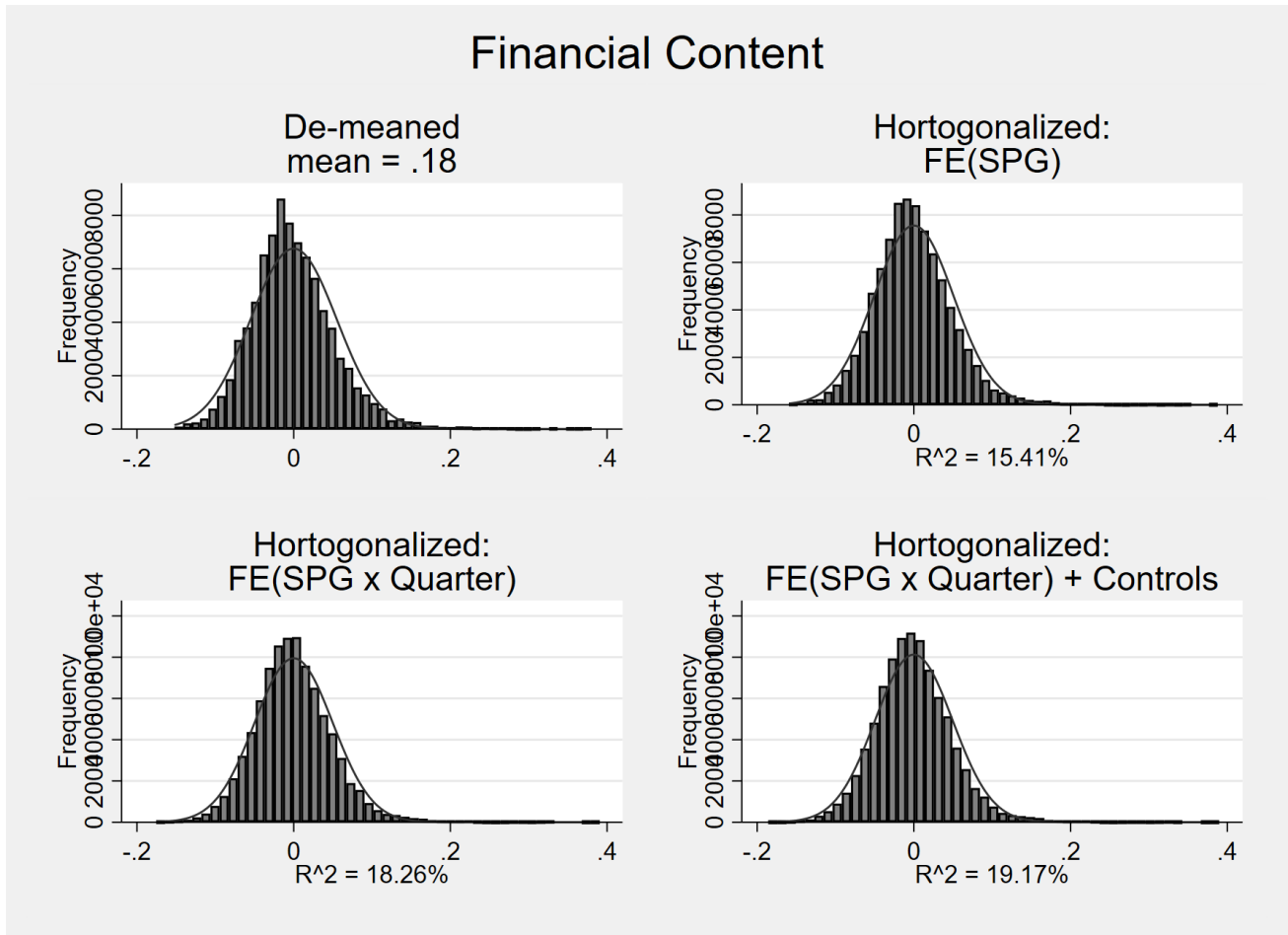


Figure 2: Equilibrium

In this figure we plot a fund manager's expected utility $\mathbb{E}[v_i(q_{i,2} \cdot f)]$ as a function of the fund strategy γ_i . Parameter values are: $\mu_F = 0.05$, $\sigma_F = 0.1$, $\sigma_\varepsilon = 0.15$, $\bar{\alpha} = 0.02$, $\bar{b} = 1$, $\sigma_\alpha = \sigma_b = \sigma_\eta = \sigma_\gamma = 0.25$, $f = 0.01$, $c = 0.001$.

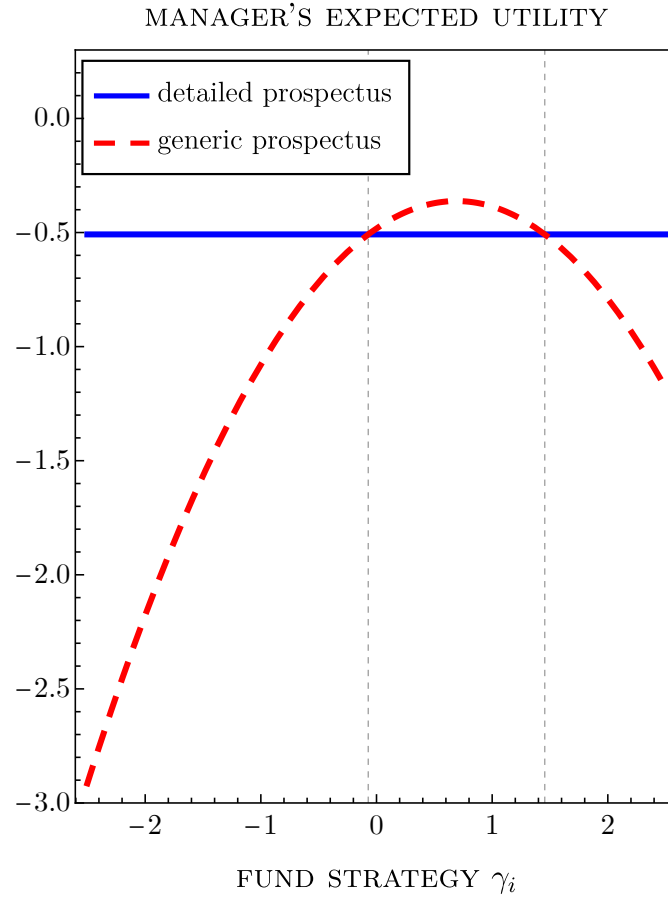


Figure 3: Comparative Statics

In this figure we plot the expected fund size $\mathbb{E}[q_{i,2}]$ (top-left panel), and the fund flow volatility $\text{Var}[q_{i,2}/q_{i,1} - 1]$ (bottom-left panel) at time 2, as a function of the fund strategy γ_i . The top-right panel plots the flow-performance sensitivity $\lambda_p^\alpha/(\bar{\alpha} - f)$, as a function of the factor return F_2 , while the bottom-right panel plots the fund flow volatility as a function of the expected fund size. Parameter values are as in Figure 2.

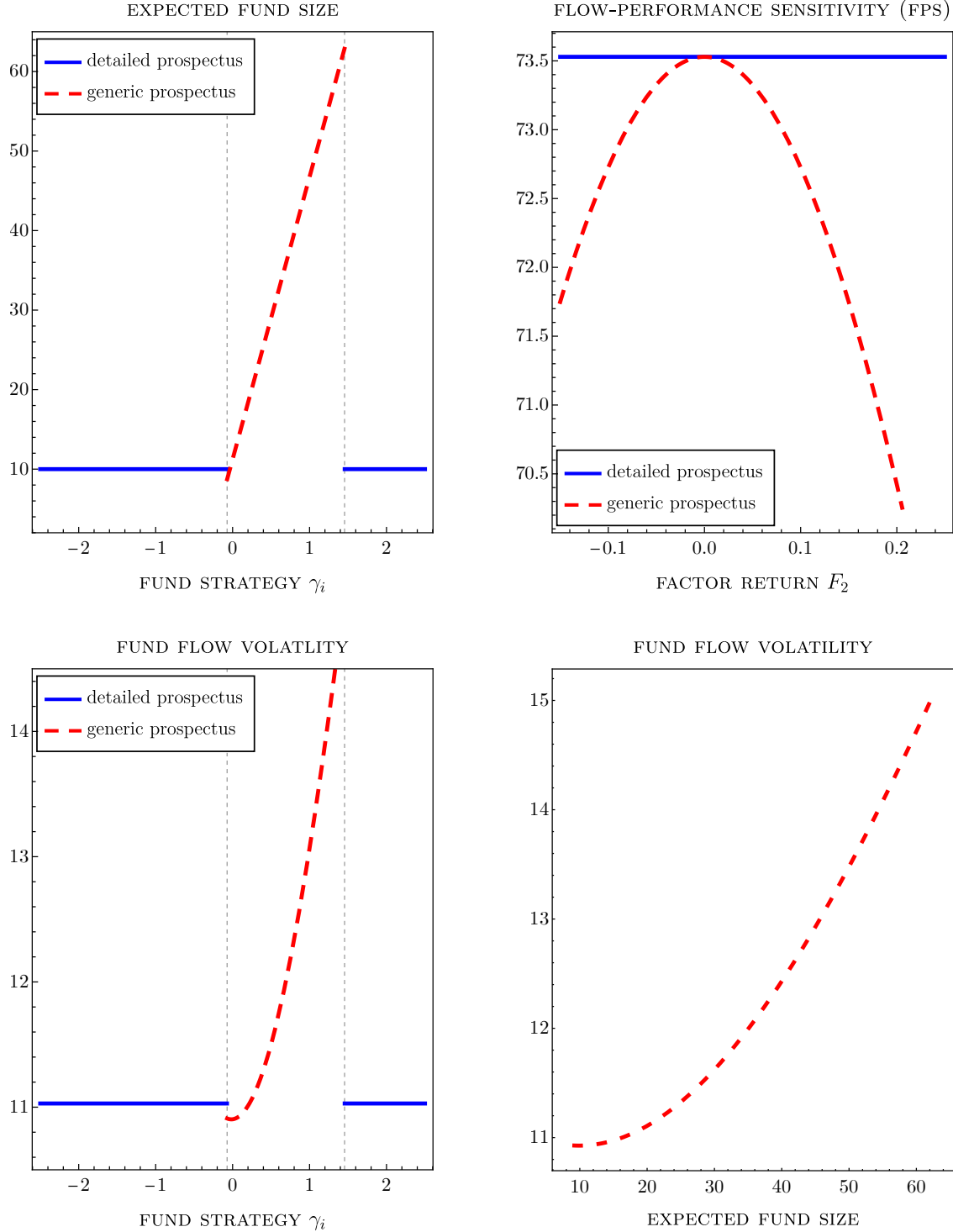


Figure 4: Boilerplate Phrases

This figure contains a list of the top 100 boilerplate phrases (stemmed), ordered by frequency.

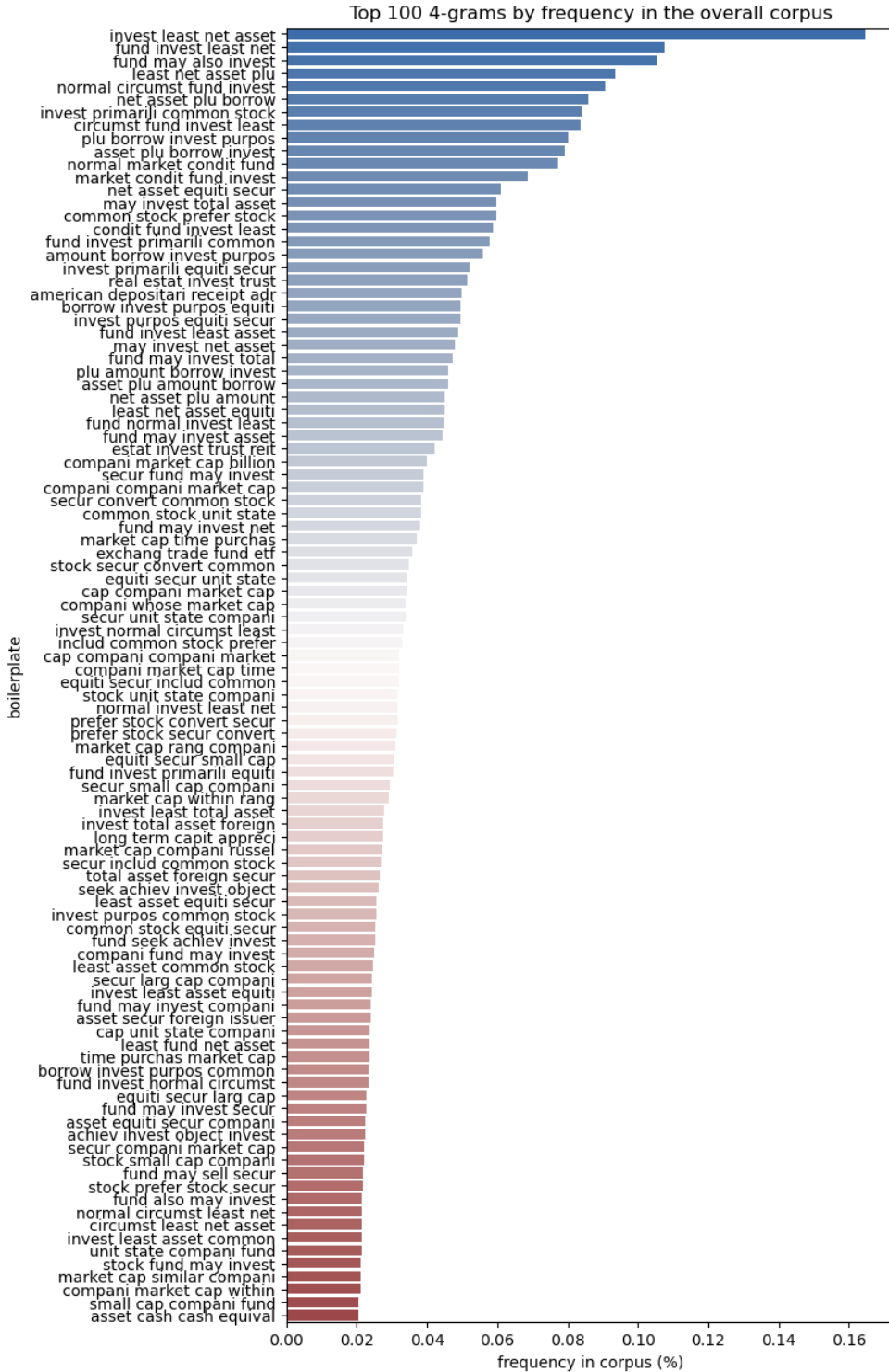


Table I: Financial Content and Text Length/Complexity/Boilerplate

This table shows the association between the financial content (FinCon) of funds' strategy descriptions, and the length, complexity, and boilerplate percentage of those descriptions. The dependent variable, FinCon is constructed as in equation (1), with varying penalties (a) for common usage of financial terms. Column 1 displays results for the unweighted ratio of financial terms to total terms ($a = 0$), while columns 2-4 weight each term by the inverse of their frequency among other PIS descriptions. The independent variables of interest are: (i) the percentage of boilerplate in each strategy description, as defined in equation (2); (ii) the complexity of the strategy descriptions, measured using the Flesch-Kincaid grade-level complexity score; and the length (word count) of the descriptions. Separate regressions are run for each of these three independent variables. All regressions contain fund-level control variables: fund age, size, expense ratio, turnover ratio, net flow, flow volatility, as well as additional style controls, measured as the weighted average betas of stocks held by the fund, where the betas are estimated using the combined six factors of Fama and French (2015) and Carhart (1997). All regressions also include month and mandate (SPG) fixed effects. The regression only includes fund-month combinations for which prospectuses are available (i.e., no forward filling). Standard errors are clustered at the fund and month level. t -statistics are in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Financial Content			
	$a = 0$ (1)	$a = 1$ (2)	$a = 2$ (3)	$a = 4$ (4)
Intercept	1.054*** (45.42)	1.086*** (41.98)	1.120*** (38.69)	1.186*** (33.72)
Boilerplate	-0.0220 (-0.75)	-0.150*** (-4.68)	-0.272*** (-7.75)	-0.492*** (-11.88)
Intercept	0.967*** (39.67)	0.962*** (36.04)	0.961*** (32.51)	0.964*** (27.13)
Complexity	0.00561*** (5.26)	0.00609*** (5.12)	0.00653*** (4.91)	0.00726*** (4.47)
Intercept	1.046*** (46.99)	1.039*** (41.99)	1.034*** (37.34)	1.030*** (30.46)
Length	0.00000724 (0.51)	0.0000330** (2.11)	0.0000586*** (3.34)	0.000107*** (4.99)
SPG+Month FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Obs	27,353	27,353	27,353	27,353

Table II: Financial Content and Fund Characteristics

This table shows the association between the financial content (FinCon) of funds' strategy descriptions, and fund-level characteristics. The dependent variable, FinCon is constructed as in equation (1). The independent variables are the following fund-level characteristics: fund age, size, expense ratio, turnover ratio, net flow, flow volatility, as well as additional style controls, measured as the weighted average betas of stocks held by the fund, where the betas are estimated using the combined six factors of Fama and French (2015) and Carhart (1997). Models (2)-(6) also include measures of fund performance at quarter end. Models (2) and (4) include fund *Alphas* constructed as the difference between fund returns and their benchmark. Benchmarks are obtained with 12-months rolling regressions of fund returns on the market, size, book-to-market and momentum factors in Model (2); and those same factors plus the profitability and investment factors in Model (4). All rolling regressions are run utilizing daily returns. In models (3) and (5) performance is measured as value added, by multiplying the 4-factor and 6-factor alphas by the log of TNA. All regressions also include month and mandate (SPG) fixed effects. Standard errors are clustered at the fund and month level. *t*-statistics are in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Financial Content				
	All (1)	CAR Alpha (2)	CAR VAdd (3)	FF6 Alpha (4)	FF6 VAdd (5)
Ln(TNA)	-0.00177*** (-3.31)	-0.00177*** (-3.25)	-0.00175*** (-3.24)	-0.00177*** (-3.25)	-0.00169*** (-3.13)
Ln(Age)	-0.000211 (-0.19)	-0.000158 (-0.14)	-0.000159 (-0.14)	-0.000159 (-0.14)	-0.000164 (-0.14)
ExpRatio	0.468* (1.94)	0.488** (2.00)	0.470* (1.91)	0.526** (2.16)	0.497** (2.02)
TurnRatio	-0.00195* (-1.75)	-0.00197* (-1.75)	-0.00196* (-1.74)	-0.00197* (-1.75)	-0.00196* (-1.74)
Style(MKT)	-0.00211 (-1.57)	-0.00206 (-1.52)	-0.00205 (-1.51)	-0.00203 (-1.49)	-0.00203 (-1.50)
Style(SMB)	-0.00646*** (-6.04)	-0.00645*** (-6.02)	-0.00645*** (-6.01)	-0.00643*** (-6.00)	-0.00643*** (-6.01)
Style(HML)	0.00278*** (3.29)	0.00275*** (3.27)	0.00274*** (3.26)	0.00273*** (3.26)	0.00273*** (3.26)
Style(MOM)	-0.00270** (-2.05)	-0.00269** (-2.03)	-0.00269** (-2.03)	-0.00270** (-2.04)	-0.00269** (-2.04)
Style(RMW)	0.000759 (1.27)	0.000777 (1.30)	0.000775 (1.29)	0.000749 (1.25)	0.000758 (1.26)
Style(CMA)	0.000866 (1.59)	0.000852 (1.56)	0.000850 (1.56)	0.000859 (1.58)	0.000855 (1.57)
CARAlpha		-0.00952 (-0.88)			
CARVadd			-0.000619 (-0.33)		
FF6Alpha				-0.0221* (-1.94)	
FF6Vadd					-0.00252 (-1.25)
Constant	0.193*** (28.74)	0.193*** (28.52)	0.193*** (28.50)	0.193*** (28.53)	0.193*** (28.41)
Adj R2	0.1917	0.1925	0.1925	0.1925	0.1925
Obs	86,056	83,493	83,493	83,493	83,493

Table III: Specialization and Prospectus Choice

This table documents the relationship between fund strategy specialization and financial content (FinCon) in prospectus strategy descriptions. Columns 1-3 estimate the relationship for all funds in panel regressions with different identifying variation. Column 1 estimates the panel regression within strategy mandates (SPG) at fixed points in time (quarters). Column 2 estimates the regression across funds (i.e., with quarter fixed effects), while column 3 estimates the regression within funds (i.e., fund fixed effects). Strategy specialization is measured by the dispersion in objective factor loadings with respect to other funds in the same Strategy Peer Group (SPG), as defined in Section 4.2.3. All regressions contain fund-level control variables: fund age, size, expense ratio, turnover ratio, net flow, flow volatility, as well as additional style controls, measured as the weighted average betas of stocks held by the fund, where the betas are estimated using the combined six factors of Fama and French (2015) and Carhart (1997). Standard errors are clustered at the fund and time level. t -statistics are in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Financial Content		
	Within Mandate (1)	Across Funds (2)	Within Fund (3)
Specialization	0.00254*** (2.82)	0.00283*** (2.84)	0.00849** (2.60)
Ln(TNA)	-0.00161*** (-3.02)	-0.00189*** (-3.35)	-0.000283 (-0.34)
Ln(Age)	0.000235 (0.21)	0.00101 (0.84)	-0.0175*** (-9.48)
Expense Ratio	0.412* (1.74)	0.384 (1.45)	0.111 (0.31)
Turnover Ratio	-0.00224** (-2.05)	-0.00370*** (-3.19)	0.000756 (0.78)
Style (MKT)	-0.00225* (-1.69)	-0.00302** (-2.00)	-0.000432 (-0.50)
Style (SMB)	-0.00695*** (-6.60)	-0.0103*** (-9.06)	-0.000947* (-1.74)
Style (HML)	0.00281*** (3.42)	0.00370*** (3.89)	-0.000957* (-1.88)
Style (MOM)	-0.00246* (-1.97)	-0.00309** (-2.16)	-0.00144* (-1.73)
Style (RMW)	0.00102* (1.75)	0.00179*** (2.76)	0.000707* (1.87)
Style (CMA)	0.000935* (1.75)	0.00142** (2.31)	-0.000978*** (-2.78)
Constant	0.176*** (81.00)	0.176*** (73.93)	0.163*** (22.82)
SPG×Quarter FE	Y		
Quarter FE		Y	
Fund FE			Y
Adj R2	0.19	0.06	0.55
Obs	83,670	83,671	83,670

Table IV: Fund Size

This table documents the relationship between financial content and fund size. Fund size is measured as the natural logarithm of assets under management (Ln(TNA)). In column 1 the independent variable of interest is financial content (FinCon), as defined in equation (1). In column 2 the independent variable of interest is a dummy variable that takes value of 1 when FinCon is in the top tercile of its distribution. Fund level control variables are: fund age, expense ratio, turnover ratio, net flow, flow volatility, and style, measured as the weighted average of the beta of stocks held with respect to the six factors of Fama and French (2015) and Carhart (1997). All regressions include month fixed effects and SPG fixed effects. Standard errors are clustered at the fund and time level. *t*-statistics are in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Fund Size			
	(1)	(2)	Age Group	
			Below Median	Above Median
	(1)	(2)	(3)	(4)
Financial Content	-1.659*** (-3.27)			
Medium FinCon		-0.00940 (-0.19)	-0.0447 (-0.68)	0.0124 (0.18)
High FinCon		-0.203*** (-3.23)	-0.176** (-2.20)	-0.249*** (-2.84)
Ln(Age)	0.823*** (20.56)	0.823*** (20.60)	0.728*** (12.81)	0.780*** (8.43)
Expense Ratio	-146.7*** (-14.43)	-146.2*** (-14.44)	-116.0*** (-9.93)	-175.8*** (-13.21)
Turnover Ratio	-0.285*** (-7.01)	-0.285*** (-6.99)	-0.168*** (-3.55)	-0.437*** (-7.35)
Style (MKT)	0.0619 (1.50)	0.0621 (1.50)	0.115** (2.23)	0.0217 (0.37)
Style (SMB)	0.0304 (0.89)	0.0301 (0.88)	0.0114 (0.27)	0.0440 (0.87)
Style (HML)	-0.00441 (-0.16)	-0.00634 (-0.23)	0.0303 (0.86)	-0.0420 (-1.09)
Style (MOM)	-0.0302 (-0.72)	-0.0297 (-0.71)	0.0173 (0.30)	-0.0804 (-1.44)
Style (RMW)	-0.0467** (-2.51)	-0.0460** (-2.47)	-0.0307 (-1.31)	-0.0702** (-2.65)
Style (CMA)	0.00234 (0.13)	0.00139 (0.08)	0.0163 (0.72)	-0.0120 (-0.47)
Constant	5.639*** (186.93)	5.639*** (187.00)	5.537*** (107.35)	5.688*** (91.98)
SPG×Quarter FE	Y	Y	Y	Y
Adj R2	0.30	0.30	0.18	0.31
Obs	83,670	83,670	40,366	43,294

Table V: Fund Flow Volatility

This table presents the estimated coefficients from a panel regression of fund flow volatility on fund size ($\text{Ln}(\text{TNA})$) and other fund-level control variables: expense ratio, turnover, age, and style betas estimated from the [Fama and French \(2015\)](#) model plus momentum. The regression is run separately for generic (column 1) and detailed (column 2) prospectuses, where prospectus detail is measured using the financial content (FinCon) measure defined in equation (1). Low FinCon (bottom tercile) indicates a generic strategy description, while high FinCon (top tercile) indicates a detailed strategy description. The regression in column 3 is run for all funds. Standard errors are clustered at the fund and time level. Fixed effects are reported at the bottom of the table. t -statistics are in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Fund Flow Volatility		
	Low FinCon (1)	High FinCon (2)	All Funds (3)
$\text{Ln}(\text{TNA})$	10.37*** (13.05)	8.806*** (16.40)	10.40*** (12.99)
$\text{Ln}(\text{TNA}) \times \text{Medium FinCon}$			-0.298 (-0.35)
$\text{Ln}(\text{TNA}) \times \text{High FinCon}$			-1.771** (-2.01)
Medium FinCon			1.756 (0.41)
High FinCon			9.693** (2.19)
$\text{Ln}(\text{Age})$	-2.698** (-2.50)	-1.009 (-1.37)	-1.525** (-2.51)
Expense Ratio	-178.7 (-1.07)	-15.54 (-0.13)	-95.67 (-1.03)
Turnover Ratio	-0.612 (-1.10)	0.239 (0.43)	0.177 (0.51)
Style (MKT)	-1.666* (-1.79)	-0.588 (-0.85)	-0.997* (-1.89)
Style (SMB)	-1.523** (-2.32)	-1.402*** (-2.78)	-1.529*** (-3.66)
Style (HML)	-0.956* (-1.72)	-0.160 (-0.25)	-0.247 (-0.58)
Style (MOM)	-0.898 (-1.20)	-1.182 (-1.49)	-1.007** (-2.11)
Style (RMW)	0.606* (1.73)	0.810** (2.02)	0.357 (1.46)
Style (CMA)	-0.535 (-1.43)	-0.291 (-0.71)	-0.328 (-1.28)
Constant	13.49*** (28.52)	13.68*** (32.70)	13.68*** (42.01)
SPG \times Quarter FE	Y	Y	Y
Adj R2	0.20	0.17	0.16
Obs	28,035	27,459	83,561

Table VI: Flow-Performance Sensitivity

This table shows the estimated sensitivity of fund flows to perceived active returns (PAR) in a panel regression framework, and how this sensitivity varies for different levels of prospectus financial content (FinCon). Perceived active returns are computed by subtracting investors' perceived benchmark returns from funds' total returns (see section 4.2.2). The dummy variables Medium FinCon and High FinCon identify funds in the middle and top terciles of FinCon, respectively. Young and old funds are identified depending on whether their age is, respectively, below or above the median age in a given quarter. Additional style controls are included as in all other tables, but omitted here for brevity. Standard errors are clustered at the fund and time level. Fixed effects are reported at the bottom of the table. t -statistics are in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Quarterly Flows (%)				
	Within Mandate	Across Funds	Within Fund	Young Funds	Old Funds
	(1)	(2)	(3)	(4)	(5)
PAR	0.0150*** (2.88)	0.0156*** (2.97)	0.0158*** (2.71)	0.00820 (1.05)	0.0223*** (3.63)
PAR \times Medium FinCon	0.00747 (1.15)	0.00688 (1.07)	0.00873 (1.34)	0.0101 (1.03)	0.00285 (0.41)
PAR \times High FinCon	0.0155** (2.19)	0.0154** (2.14)	0.0179** (2.24)	0.0410*** (3.37)	-0.00288 (-0.38)
Medium FinCon	-0.00110 (-1.05)	-0.00110 (-1.06)	-0.0000395 (-0.03)	-0.00231 (-1.50)	0.0000263 (0.02)
High FinCon	-0.00145 (-1.17)	-0.00143 (-1.22)	0.0000642 (0.04)	-0.00215 (-1.10)	-0.000578 (-0.43)
Lagged Flows	0.429*** (34.19)	0.431*** (34.09)	0.368*** (31.13)	0.430*** (27.45)	0.415*** (26.98)
PAR ²	0.00888*** (2.78)	0.00851*** (2.70)	0.0285*** (3.91)	0.00805 (1.40)	0.00956*** (3.19)
Constant	-0.00442*** (-17.15)	-0.00441*** (-17.05)	-0.00454*** (-3.61)	-0.00919*** (-8.49)	-0.00920*** (-12.53)
Controls	Y	Y	Y	Y	Y
SPG \times Quarter FE	Y			Y	Y
Quarter FE		Y			
Fund FE			Y		
Adj R2	0.24	0.22	0.27	0.26	0.22
Obs	82,389	82,390	82,389	39,915	42,455

Table VII: Benchmark Misalignment

This table estimates the relationship between benchmark misalignment and financial content (FinCon) of fund prospectuses. Benchmark misalignment is measured as the distance between a fund's true (objective) benchmark and the benchmark perceived by investors; the former is estimated using a linear factor regression, and the latter is inferred from fund flows (see section 4.3.5). We construct dummy variables for each tercile of FinCon, and include the middle and top tercile dummies in the regression. In columns 2 and 3, the panel analysis is run separately for young and old funds, identified by whether their age is, respectively, below or above the median age in each quarter. Standard errors are clustered by fund and time. Fixed effects are reported at the bottom of the table. *t*-statistics are in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Benchmark Misalignment		
	All Funds (1)	Young Funds (2)	Old Funds (3)
Medium FinCon	-0.167** (-2.36)	-0.328*** (-3.23)	-0.0282 (-0.31)
High FinCon	-0.174** (-2.10)	-0.367*** (-3.27)	0.0203 (0.18)
Ln(TNA)	-0.0760*** (-2.85)	-0.112*** (-3.34)	-0.0334 (-0.85)
Ln(Age)	0.0421 (0.59)	-0.0240 (-0.31)	0.146 (0.77)
Expense Ratio	2.331 (0.17)	3.240 (0.21)	5.115 (0.23)
Turnover Ratio	0.0239 (0.45)	-0.0683 (-1.07)	0.143 (1.66)
Style (MKT)	-0.0186 (-0.37)	-0.0278 (-0.41)	-0.0319 (-0.43)
Style (SMB)	0.00615 (0.13)	-0.0226 (-0.40)	0.0367 (0.49)
Style (HML)	-0.0996** (-2.01)	-0.116** (-2.07)	-0.0818 (-1.08)
Style (MOM)	-0.0137 (-0.27)	-0.0585 (-0.87)	0.00487 (0.07)
Style (RMW)	-0.0131 (-0.49)	-0.00486 (-0.16)	-0.0112 (-0.32)
Style (CMA)	-0.0226 (-0.82)	-0.0342 (-1.11)	-0.0108 (-0.27)
Constant	1.217*** (27.20)	1.190*** (17.31)	1.128*** (10.68)
SPG×Quarter FE	Y	Y	Y
Adj R2	0.03	0.05	0.03
Obs	83,670	40,366	43,294