Learning from Prospectuses

SIMONA ABIS
Columbia University

ANDREA M. BUFFA
University of Colorado Boulder

APOORVA JAVADEKAR
Indian School of Business

ANTON LINES
Columbia University

October 24, 2022*

Abstract

We analyze fund managers' incentives to disclose qualitative information about their strategies, and investors' ability to learn from these disclosures. We propose a mechanism whereby investors make fewer errors in distinguishing active returns from passive factor exposures when they have access to more detailed strategy descriptions. In a formal model, we show that investor attribution errors are, on balance, more costly for managers with more specialized strategies, leading them to write more detailed descriptions. In the data, we find evidence for this prediction and support for the model's core learning mechanism, as well as new insights into the flow-performance relationship.

*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 Philip Bond, Lauren Cohen, Ron Kaniel, Alberto Rossi, Yang Song, Chester Spatt, Savitar Sundaresan, Paul Tetlock, and conference/seminar participants at AFA, SFS Cavalcade, Future of Financial Information, 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 should contain. This includes typical quantitative items 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 widely studied,[3] much less is known about the economic role of the narrative disclosures.

To explore this role, we construct a new textual measure to capture the amount of informative financial content in fund prospectuses. We focus on the "Principal Investment Strategies" (PIS) section, in which fund managers are required to "explain in general terms how the fund's adviser decides which securities to buy and sell".[4] Managers are given wide latitude in choosing how much to disclose, leading to significant variation in financial content across prospectuses. This heterogeneity motivates our main research questions: how do managers determine their optimal disclosure policies, and how do investors make use of this information? Such questions are difficult to answer simply by examining the data, as managers' incentives are not directly observable. Therefore, we develop an equilibrium model of the mutual fund industry in which managers optimally choose the level of detail in their prospectuses, taking into account investor learning, while investors use the textual information together with fund returns to infer managerial ability.

[1] See, `https://www.sec.gov/oiea/investor-alerts-bulletins/ib_mfprospectus3.html`

[2] For an additional example, see the SEC's Beginner's Guide to Mutual Funds

[3] E.g., 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).

[4] Form N-1A, item 9

Specifically, we hypothesize that detailed descriptions provide a more accurate signal of funds' exposure to passive factors, allowing for better inference of active returns. Our model delivers clear predictions for the observed financial content in PIS descriptions and its relationship to investor behavior, which we confirm using a comprehensive sample of US equity mutual fund prospectuses from 2000 to 2017.

Due to the specialized nature of PIS text, standard linguistic metrics—e.g., word counts, sentence complexity, textual uniqueness, and boilerplate language—may not fully capture informativeness. As such, we rely on Campbell Harvey's Hypertextual Finance Glossary to construct a more targeted measure of relevant *financial content*: i.e., the fraction of words/phrases in a particular PIS description that can be found in the glossary, with each word weighted by the inverse of its frequency in the overall corpus. This intuitive measure captures the domain-specific terminology of the finance industry while giving more weight to less common terms. It is correlated with the aforementioned metrics—positively related to word count, complexity, and uniqueness, and negatively related to boilerplate—but the correlations are relatively small in magnitude (maximum 0.26), indicating that our measure captures a novel dimension of informativeness. In the panel, financial content ranges from 9.8% to 39.4% of the total word counts, with a mean of 21% and a standard deviation of 5.7%. This variation is correlated with fund characteristics: funds with prospectuses in the top tercile of financial content are on average 11.8% smaller and about a year younger than those in the bottom tercile, though risk-adjusted performance does not differ.

We then turn to the model to make sense of these stylized facts. 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 *active* returns are equal to zero. Each fund's total return is comprised of an active component, driven by the ability of the manager, and a passive component, driven by exposure to a single common risk factor. Prospectuses—specifically, PIS descriptions—provide information on the fund's asset selection criteria and thus allow investors to make inferences about the factor loading. For tractability, we assume that this loading is exogenously determined, though we depart from the typical assumption that it is perfectly known.[5] Also for simplicity, we assume that prospectuses can be one of two types: *detailed* or *generic*. Detailed prospectuses are fully revealing of the fund's factor loading, while generic prospectuses only provide a signal of the average loading across all funds. Investors are Bayesian learners who rationally estimate active returns and allocate capital accordingly. Given these expected allocations, fund managers

[5] Another notable departure from the standard assumption can be found in 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 ability is reduced.

choose their prospectus type with the objective of maximizing the (mean-variance) utility of their compensation, computed as a fixed percentage of assets under management.

Prospectus choice affects capital allocation in three distinct ways. First, because investors are uncertain about the factor exposures of funds with generic prospectuses, they make errors when inferring the ability of those managers. In other words, there is a misalignment between the fund's true benchmark and the benchmark perceived by investors. This leads to higher variance in perceived active returns, greater uncertainty in expected allocation of capital, and ultimately higher compensation variance for managers (which lowers their utility due to risk aversion). Second, because generic prospectuses permit inference only about the mean factor loading among funds within the same investment mandate, managers with above-average loadings know that investor mistakes are likely to go in their favor, provided the expected factor return is positive. For this group, there is a trade-off between a larger expected fund size (and thus higher expected compensation) and greater uncertainty in size. Managers with below-average factor loadings or detailed prospectuses do not face this trade-off. Third, funds with detailed prospectuses experience greater flow-performance sensitivity (FPS), since investors understand that accurately-perceived active returns are a more informative signal of ability. All else equal, higher FPS also leads to increased uncertainty in expected fund size.

In equilibrium, the manager's optimal choice depends on her fund's (exogenous) factor loading relative to that of other funds in the same investment mandate. The key intuition is that the uncertainty cost of choosing a generic prospectus increases quadratically with the distance between the fund's loading and the cross-sectional mean, while the costs or benefits of average investor mistakes are linear in the loadings, and the effect of flow-performance sensitivity is constant. Thus, any sufficiently extreme factor loading (which we label a *specialized* strategy) will induce the manager to choose a detailed prospectus, while loadings closer to the cross-sectional mean (which we label *standardized* strategies) will induce the manager to choose a generic prospectus. Overall, a generic prospectus will be optimal when either the lower-FPS effect dominates (for below-average factor loadings), beneficial investor mistakes dominate (for above-average loadings), or a combination of these cases occurs (for above-but-close-to-average loadings). We prove that the two thresholds separating the optimal prospectus types exist and are unique.

Our model is able to account for the observed heterogeneity in financial content across fund prospectuses. In particular, using a more rigorous panel regression setup, we confirm that funds with generic prospectuses are larger on average, even controlling for their initial size at inception. The model also makes several additional falsifiable predictions that can be tested in the data.

We start by confirming the main equilibrium result: qualitative disclosures should be more detailed for funds with more specialized strategies (i.e., more extreme factor loadings). We distinguish between generic and detailed descriptions by constructing terciles of financial content, where the top tercile corresponds to detailed descriptions and the bottom to generic descriptions. To measure specialization, we compute the sum of squared normalized betas with respect to the Fama and French (2015) five-factor model plus momentum (henceforth FF6), where the normalization is performed within investment mandates. Two aspects of this measure are worth further comment. First, while the model assumes a single-factor structure for tractability, our empirical proxy follows the recent asset pricing literature and employs a six-factor structure. In this context, the single factor in the model can be thought of as a composite factor portfolio. Second, normalization within investment mandates is crucial, both in the model and in reality. Different risk factors can be relevant for funds with different mandates, thus our model's results are explicitly conditioned on the mandate. We use three alternative methods of constructing mandates, based on (i) the Fama and French (1993) three-factor model, (ii) the Daniel et al. (1997) characteristics (which embed the Morningstar equity style categories), and (iii) the text-based strategy peer groups of Abis and Lines (2020).

Using these measures, we find that a one-standard-deviation increase in specialization increases the probability of a fund having a detailed prospectus by a statistically significant 7% relative to the baseline probability.[6] However, this result, while encouraging, is not by itself sufficient to validate the model. For instance, a plausible alternative story is that funds with more complex or unusual strategies may simply require greater detail to effectively explain what they do (although the result could also be seen as surprising, since we might expect specialized funds to be more secretive about their strategies). It is therefore essential to probe the core mechanism of the model more deeply, by testing whether investors actually do learn from prospectuses in the predicted way. The model tells us that, for funds with detailed PIS descriptions, misalignment between their true benchmarks and investors' perception of these benchmarks should be smaller, and perceived active returns (PARs) should be less volatile as investors make fewer mistakes. As a result, they should be more sensitive to PARs when allocating capital.

Using a methodology similar to Barber et al. (2016), we infer perceived factor loadings as those that best explain observed investor capital flows, under the assumption that flows only chase active returns. "True" factor loadings are instead estimated from the econometrician's point of view, using full-sample regressions of fund returns on the FF6 factors. Benchmark misalignment is then

[6] While the magnitude of this effect may seem modest, we note that the empirical proxies for both specialization and financial content are likely to contain measurement error, and thus what reported here should be interpreted as a lower bound.

computed as the normalized Euclidean distance between the true and perceived factor loading vectors. Consistent with the model's predictions, we find that this distance is around 16% smaller for funds with detailed prospectuses. We also derive PARs from the perceived factor loadings, and find that PAR volatility is about 15% lower for funds with detailed descriptions. Finally, we find that the flow-performance sensitivity is about three times as strong when funds write detailed strategy descriptions. These findings are difficult to reconcile with alternative stories (such as the example mentioned at the beginning of this paragraph).

Our final set of results are not derived explicitly from the model; however, they are motivated by taking its conclusions seriously. If learning from prospectuses is an important channel for capital allocation, we would expect the relationships described in the previous paragraphs to be stronger for younger funds. These funds have shorter return histories and thus fewer opportunities for investors to learn about factor loadings and performance through traditional quantitative channels, making prospectuses comparatively more important. In support of this hypothesis, we find that the relationships are consistently stronger for funds of below-median age, and often disappear entirely for funds of above-median age.

When assessing these findings, it is worth asking how learning might manifest in practice. Given the many papers arguing against mutual fund investor sophistication (e.g. Frazzini and Lamont (2008); Barber et al. (2016); Evans and Sun (2021); Ben-David et al. (2022)), the reader may be skeptical of their ability to disentangle factor loadings from active returns. However, while investors may or may not understand academic asset pricing models, it is sensible to believe that they can distinguish, say, a dividend fund from a small-cap fund, and an extreme strategy from a standard one (e.g., see Abis and Lines (2020)). For the mechanism in our model to hold, it need only be the case that investors compare funds with similar peers (and that those peers have similar factor exposures), and they will arrive at estimates of active returns in line with those derived from factor models. This thinking is in line with the SEC's view that the function of prospectuses is to identify appropriate fund comparisons, as expressed in the introductory quotation. It is also broadly in line with earlier work showing that investors evaluate funds relative to their stated benchmarks (Sensoy (2009)), as well as survey responses (Choi and Robertson (2020)) indicating that many retail investors consider market-adjusted performance. Furthermore, recent evidence using machine learning finds that capital flows are in fact predictive of future risk-adjusted fund performance (Kaniel et al. (2022)), in contrast to the "dumb money" effect documented by Frazzini and Lamont (2008). Our paper provides incremental evidence in favor of some degree of investor sophistication, contributing to this debate.

Another concern with our analysis might be whether mutual fund investors are likely to read prospectuses in the first place. Despite receiving comparatively little attention from academics, prospectuses are widely available to the public via direct investor mailings, the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system,[7] industry databases (e.g. Morningstar), and indirectly through the recommendations of financial advisors. The majority of mutual fund assets are allocated through intermediated channels; e.g., financial advisors or institutions (Reid and Rea (2003); Holden et al. (2020)), whose due diligence includes reading funds' offering materials. Prospectus disclosures are also legally binding; hence, communication with investors through other (regulated) channels should be consistent with this information. We can therefore think of prospectuses as a proxy for investor communication more generally.

In addition to the debate over investor sophistication, we contribute to the broader theoretical and empirical literatures on capital allocation in the mutual fund industry. We extend the Berk and Green (2004) model to a novel learning environment and confirm its predictions empirically, suggesting the continued usefulness of the framework.[8] We also build on the empirical work of Chevalier and Ellison (1997) and Sirri and Tufano (1998)—and more recently Berk and Van Binsbergen (2016) and Barber et al. (2016)—by showing how qualitative signals from fund prospectuses mediate the flow-performance relationship. Lastly, we contribute to the emerging literature that uses textual analysis to study the mutual fund industry (e.g., Hillert et al. (2016); Hwang and Kim (2017); Abis (2020); Kostovetsky and Warner (2020); Abis and Lines (2020); Krakow and Schäfer (2020); Akey et al. (2021); and Sheng et al. (2021)). In particular, we show that our financial content measure is distinct from the uniqueness measure of Kostovetsky and Warner (2020). We are also the first paper in this literature to employ a theoretical model, allowing us to gain new insights into the disclosure incentives of the fund managers who generate these important documents.

2 Prospectus Informativeness

In this section, we describe our data, provide further institutional background on fund prospectuses, and discuss the construction and properties of our financial content measure.

[7] See, `https://www.sec.gov/edgar/searchedgar/mutualsearch.html`

[8] Other notable extensions to the baseline model include Brown and Wu (2016); Choi et al. (2016); Starks and Sun (2016); Harvey and Liu (2019); Franzoni and Schmalz (2017); and Buffa and Javadekar (2022).

2.1 Data

We focus on US active equity mutual funds, combining data from four different sources: (i) information about mutual fund characteristics and returns from the CRSP Survivorship-Bias-Free Mutual Fund dataset; (ii) holdings data from CRSP and Thomson Reuters (formerly CDA/Spectrum); (iii) stock-level information from the CRSP security file; and (iv) mutual fund prospectus text from the SEC's Electronic Data Gathering, Analysis, and Retrieval system (EDGAR). 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 on average hold fewer than 10 stocks or less than 80% of their assets (excluding cash) in US stocks.

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. returns, turnover, etc) weighted by lagged TNA. Following Abis (2020), we construct 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. Since MFLinks excludes many new funds in recent years (Zhu (2020); Shive and Yun (2013)), we use Thomson Reuters holdings data from January 2000 to August 2008 and CRSP holdings data from September 2008 to December 2017. We then forward-fill to the monthly frequency.[9]

Lastly, we follow Abis and Lines (2020) for the collection and processing of prospectus 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. Since any material change to the management or strategy of a fund must legally be reported to the SEC, we are able to forward-fill (using the latest available prospectus) fund-quarter observations with no prospectuses in the EDGAR database. After applying filters and matching to prospectuses, our final sample consists of $2,995$ unique funds and $86,056$ fund–quarter observations, from January 2000 December 2017.

[9]Monthly 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.

2.2 Institutional Background

The disclosure of prospectuses is mandated by the SEC for all mutual funds via the EDGAR system. The "Principal Investment Strategies" (PIS) section, studied in this paper, corresponds to item 9 of the N-1A mandatory disclosure form. In this section, funds are required to *"describe the Fund's principal investment strategies, including the particular type or types of securities in which the Fund principally invests or will invest"*, and *"explain in general terms how the Fund's adviser decides which securities to buy and sell"*. The latter requirement was introduced in 1998 and made compulsory as of December 1999.

As stated in the "Administration of the Form N-1A requirements", 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)), prospectuses must 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"*. In accordance with those objectives, the SEC introduced a section of their website to instruct investors on "How to read mutual fund prospectuses". Investors are encouraged to utilize information in prospectuses when making investment decisions, particularly when comparing funds to their peers.

Although data on mutual fund distribution channels is not readily available, perspectives papers by the Investment Company Institute (ICI) reveal that most capital flows into mutual funds occur through intermediated channels, with direct sales to end investors representing only 12% of mutual fund assets in 2002 (Reid and Rea (2003)). A more recent ICI survey of households (Holden et al. (2020)) documents that while 31% of investors say they have purchased some shares of mutual funds directly from mutual fund companies or through discount (not intermediated) brokers, only 1% of them rely exclusively on those direct channels. Additionally, a large portion of direct sales correspond to Vanguard, whose funds are mostly passive and do not constitute a large portion of funds in our active equity sample.

Given these considerations, one should think of the information contained in prospectuses as mostly consumed by institutional investors, financial advisors, or other intermediaries, who should have a sufficient level of financial literacy to draw simple inferences from its content. For instance, if a fund places great emphasis on delivering a steady stream of dividends, one might infer that its portfolio is likely to contain an above-average proportion of larger, older firms at a mature stage in their life cycle. In academic terms, that insight would equate to expecting a greater exposure to the value factor than to the growth factor.

Abis and Lines (2020) show that PIS descriptions correspond to funds' actual portfolio composition. For instance, funds placing particular emphasis on dividends in their strategy description hold stocks that are larger and older, have lower investment and less cash on the balance sheet (plausibly due to higher payout ratios) and have higher dividend yields. Finally, Abis (2020) documents that PIS descriptions are highly auto-correlated within a fund, with an average cosine similarity close to 100% between any two consecutive prospectuses and an average similarity of 61.77% even at 25 lags. This indicates that funds do update their disclosures, but only infrequently.

2.3 Financial Content

Construction. To represent the level of detail in fund prospectuses, 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. Each PIS section is represented by a vector of stemmed words, which we call the *strategy vector*.[10] From this vector, we construct another vector—the $financial\ vector$—as the subset of words and short phrases in the *strategy vector* that can also be found in the Harvey glossary. Since prospectuses are 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. For this reason, we assign a weight to each financial term according to the inverse of its frequency among all PIS descriptions. Our financial content measure is defined as

$$FinCon^a_{i,t} = \frac{1}{L^S_{i,t}} \sum_{w=1}^{L^F_{i,t}} \frac{1}{(1 + Coverage_w)^a}, \tag{1}$$

where $Coverage_w$ indicates the fraction of all PIS sections that contain financial term w, while $L^F_{i,t}$ and $L^S_{i,t}$ represent the length of the $financial\ vector$ and *strategy vector* for fund i in month t, respectively.[11] The variable a controls the strength of the "penalty" applied to commonly used terms. The greater a, the greater the penalty is, where $a = 0$ corresponds to the case with no

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

[11] When an individual term 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 $L^S_{i,t}$ also counts all words separately.

penalty. To discount common financial terms that constitute the linguistic context of describing a fund's strategy (e.g., *fund* or *invest*) but do not provide details about the strategy, we choose $a = 4$ for our baseline specification. Appendix A provides an example of high and low financial content prospectuses, based on $FinCon$. Figure 1 displays the distribution of $FinCon$ in the panel of PIS descriptions. We observe large heterogeneity in the measure: it ranges from 9.82% to 39.36% of the total words count, with a mean of 21% and a standard deviation of 5.7%.

Interpretation. To validate our interpretation of $FinCon$, we examine how it relates to other commonly used textual measures constructed from the same PIS sections. We consider the following measures: length, textual complexity, uniqueness, and boilerplate content. 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. Uniqueness is constructed following Kostovetsky and Warner (2020) as the average normalized overlap in word vectors between a particular fund and all other funds in the same Morningstar/Daniel et al. (1997) peer group. Boilerplate, similarly to Lang and Stice-Lawrence (2015), is constructed as the percentage of terms in a PIS section that can be categorized as boilerplate language. Boilerplate language comprises the most frequent 0.1% of four-word-phrases present across all prospectuses in our sample.

Figure 2 displays the correlations among the considered textual measures, with those for $FinCon$ presented in the first column. We observe a positive correlation between $FinCon$ and length (21%), complexity (9%) and uniqueness (26%),[12] and a negative correlation between $FinCon$ and boilerplate (-21%). All of the correlations go in the direction we expect, given the interpretation of our measure as financial informativeness. For example, a longer, more unique description with less boilerplate is likely to contain more relevant content, and more content is also likely to increase the complexity of the text. The low correlations are also justifiable as the other considered measures don't always capture informativeness. For instance, a longer descriptions would not necessarily indicate increased informational content if the additional text is not financially relevant, too simple, or too common among prospectuses.

$FinCon$ captures many desirable characteristics of the other measures. It incorporates length via the count of financial terms, but is inversely related to the count of non-financial terms. It also negatively (positively) incorporates boilerplate (uniqueness) by weighting each term by

[12] We construct *Uniqueness* utilizing both the full PIS text–KW (all text)–as well as only the first 70 words–KW (first 70w), to better approximate the dataset utilized by Kostovetsky and Warner (2020). Both measures display a similar correlation with $FinCon$. In what follows, for consistency, we utilize the measure constructed based on the full text.

the inverse of its frequency in all PIS descriptions. Finally, it captures the presence of complex financial terminology. the modest correlations with other measures, however, indicate that *FinCon* captures a novel dimension of informativeness. That is likely because *FinCon* explicitly excludes non-financial portions of the text and downweights common financial terms. Hence, it measures content in the portion of the PIS description that is likely to be most informative. Other measures either focus on the full text (length and complexity) or quantify the portion of text that is common across descriptions (boilerplate and uniqueness), without explicitly accounting for the amount of financial detail present in the remainder. Given this interpretation, it is not surprising that the largest absolute correlation is between uniqueness and boilerplate (-56%).

Financial Content and Fund Characteristics. In this section we provide summary statistics about how *FinCon* covaries with fund-level characteristics and managerial skill. Table I reports the mean of each characteristic of interest for top and bottom terciles of FinCon, after controlling for the other caracteristics. We also report the differences between the means, and the statistical significance of these differences.[13] We focus on four fund characteristics: size (i.e., log total net assets), age, expense ratio and turnover ratio; and on two fund skill measures: active returns and value added (long-term averages), constructed based on 24-month rolling regressions of excess fund returns on the FF6 factors.[14]

We observe a large and statistically significant difference in average fund size between the bottom and top terciles of *FinCon*, with funds in the top tercile being approximatively 11.8% ($e^{0.1113} - 1$) smaller than funds in the bottom tercile, significant at the 1% level. We also observe a positive relationship between *FinCon* and expense ratio, but with a much smaller economic magnitude ($0.0005/0.0117 = 4.27\%$ higher for the top tercile–significant at the 1% level); and a negative relationship between *FinCon* and turnover ratio ($0.0112/0.7542 = 1.49\%$ lower for the top tercile–significant at the 10% level) and age ($e^{0.0146} - 1 = 1.5\%$ younger for the top tercile–significant at the 5% level), both of very small economic magnitude. However, for fund skill, we do not observe any differences in mean long-term active returns or value added between

[13]Results are obtained by regressing each variable on two dummy variables indicating, respectively, the middle and top terciles of *FinCon*, and on a set of demeaned fund-level controls (all other characteristics except the LHS variable, as well as style controls based on the FF6 model—see discussion in section 4.2).

[14]The construction of value-added follows exactly the procedure of Berk and van Binsbergen (2015). Long-term averages are computed from January 2000 or from fund inception, whichever is later. All control variables included when estimating the relationship between fund skill and *FinCon* are averaged over the same time period as the active return/value-added; thus the reported differences in sample means for these variables indicate cross-sectional covariation. For the remaining fund characteristics, the reported values are averaged across the full panel of quarterly observations.

funds belonging to the top or bottom terciles of $FinCon$, suggesting that the potential value of prospectus content to investors is not to provide a direct signal of managerial ability.

3 A Model of Prospectus Choice

Given the large variation in financial content of mutual fund prospectuses documented in section 2.3, 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 with a continuum of investors and asset management funds.

Funds. Each fund is characterized by a fund mandate (or broad investment category), which identifies a specific risky factor that the fund might be exposed to, and a passive fund strategy, which identifies the fund's exposure to that factor.[15] We model the net return generated by fund i at time t as follows:

$$r_{i,t} = \alpha_i + \beta_i^k F_{k,t} + \varepsilon_{i,t} - \mathcal{C}(q_{i,t-1}) - f, \tag{2}$$

where $k \in \{1, 2, ..K\}$ denotes the fund mandate, F_k is the risk factor specific to the fund mandate,[16] β_i^k is the fund exposure to that factor, α_i is the skill of the fund manager (i.e., the additional return generated by actively managing the fund), and $\varepsilon_{i,t}$ is the fund's idiosyncratic risk. $\mathcal{C}(\cdot)$ is a mapping capturing returns to scale as a function of the fund assets q_i, and f denotes the (exogenous) management fees charged per dollar invested. We assume that $\varepsilon_{i,t}$ and $F_{k,t}$ are independent for any t and distributed as

$$\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \text{and} \quad F_{k,t} \sim \mathcal{N}(\mu_{F_k}, \sigma_{F_k}^2), \tag{3}$$

[15] While the term "strategy" in general also includes the fund's alpha-generating activities, in the context of the model we use the term simply to refer to the passive factor exposure.

[16] To keep the model as simple as possible, we assume that the systematic risk of each mandate is captured by a single factor. However, one might interpret that factor as a portfolio of multiple underlying factors or principal components.

and maintain, as in Berk and Green (2004), that $\mathcal{C}(q) = c_k \cdot q$, where the fund has decreasing returns to scale (DRS) if $c_k > 0$ and increasing returns to scale if $c_k < 0$. For simplicity, we assume that the returns to scale parameter c_k is mandate-specific and is common across all funds within the same mandate. We interpret the return that fund i generates by taking exposure to the risky factor, $\beta_i^k F_{k,t}$, as the return of the fund's *true benchmark*.

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 only about the fund mandate and, as a consequence, about the average exposure to the mandate-specific risk factor across all funds within that mandate. In our setting, therefore, a generic prospectus reveals the mandate k and the average factor exposure $b^k \equiv \mathbb{E}[\beta_i^k]$.

Detailed disclosures, instead, provides additional and more comprehensive information about the specifics of the fund strategy that are unique to the fund. In our setting, a detailed prospectus, besides revealing k and b^k as a generic prospectus does, also reveals how the fund exposure to the mandate-specific factor deviates from the average. If we let γ_i denote fund i's deviation from the average exposure within its mandate, we can write

$$\beta_i^k = b^k + \gamma_i. \tag{4}$$

It follows that, by providing information about the average mandate exposure and the specific fund deviation from it, a detailed prospectus perfectly reveals the fund strategy β_i^k.

Investors. Before reading a fund prospectus, fund investors do not know the mandate k and the fund strategy β_i^k, and are uncertain about the fund manager's skill α_i. By reading either a generic or a detailed prospectus, they learn the fund mandate and the average factor exposure within the mandate. The uncertainty about skill is the same regardless of prospectus type, captured by the prior distribution:

$$\alpha_i \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2). \tag{5}$$

However, the uncertainty about the fund strategy β_i^k depends on whether the prospectus is generic or detailed. If it is detailed, fund investors can perfectly infer γ_i, thus resolving any uncertainty about the fund specific deviations from the average mandate exposure. If, instead, the prospectus is generic, such deviations remain unknown and are captured by the prior distribution

$$\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2). \tag{6}$$

Intuitively, the distribution in (6) captures the cross-sectional variation in fund strategies within a mandate. When reading a generic prospectus, therefore, the investors are uncertain about how to evaluate the fund performance (i.e., about what return due to factor exposure to remove from the fund performance), which feeds back into their ability to estimate the manager's skill. As we shall see later, this is at the core of the equilibrium mechanism that drives the optimal choice of fund prospectuses.

Over time, investors observe the net performance of their fund, $r_{i,t}$, as well as the mandate-specific factor return $F_{k,t}$, and form posterior beliefs about α_i and β_i^k. We refer to the corresponding posterior means, denoted by $\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,t}^k$, as fund i's *perceived skill* and *perceived strategy*, respectively, where in our context *strategy* refers to the fund's factor exposure and is thus synonymous with its *benchmark*. We then define a fund's *benchmark misalignment* as the absolute value of the difference between the true and the perceived fund exposure to the risky factor, $\left|\beta_t^k - \hat{\beta}_{i,t}^k\right|$.

Additional assumptions. We make two additional assumptions that allow us to focus on the equilibrium relationship between fund strategy and fund prospectuses, while keeping the model parsimonious.

First, we assume as in Chung et al. (2012) that fund managers are uncertain about their investment skill when writing their fund prospectuses, and thus use the prior distribution of α_i in (5) 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^k. This assumption is supported by our finding in section 2.3 that the financial content of prospectuses is unrelated to the performance of the fund.

Second, while fund managers fully internalize the way investors learn and allocate their capital, we assume that investors are less sophisticated—specifically, that they only learn directly from the informational content of the prospectuses, 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. This assumption is broadly consistent with the literature showing that investors have lower levels of sophistication than fund managers (see discussion in the introduction).

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.

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 net active returns are driven to zero in expectation. However, since investors must learn about funds' benchmarks in our model, capital allocation depends on their *perceived* active returns, rather than active returns derived from true benchmarks.

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 $\mathbb{E}_t^{\mathcal{I}^p}[\hat{r}^A_{i,t+1}] = 0$. Perceived active returns, $\hat{r}^A_{i,t+1}$, are defined as

$$\hat{r}^A_{i,t+1} = (\alpha_i + \varepsilon_{i,t+1}) + (\beta_i^k - \hat{\beta}_{i,t}^k)F_{t+1} - c_k \cdot q_{i,t} - f. \tag{7}$$

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

$$q_{i,t}(p) = \frac{\hat{\alpha}_{i,t}(p) - f}{c_k}, \tag{8}$$

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, $q_{i,t} \cdot f$:

$$p_i^* = \underset{p \in \mathcal{P}}{\arg\max}\; \mathbb{E}_0^{\mathcal{M}_i}[v_i(q_{i,1}(p)f) + \delta \cdot v_i(q_{i,2}(p)f)], \tag{9}$$

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 average mandate exposure b^k, the means and variance of β_i^k, computed by investors $\mathcal{I}^g$, are

$$\mathbb{E}_1^{\mathcal{I}^g}[\beta_i^k] = b^k \quad \text{and} \quad \mathbb{V}\text{ar}_1^{\mathcal{I}^g}[\beta_i^k] = \sigma_\gamma^2. \tag{10}$$

Investors $\mathcal{I}^d$, instead, perfectly learn β_i^k. Therefore, the corresponding mean and variance are

$$\mathbb{E}_1^{\mathcal{I}^d}[\beta_i^k] = \beta_i^k \quad \text{and} \quad \mathbb{V}\text{ar}_1^{\mathcal{I}^d}[\beta_i^k] = 0. \tag{11}$$

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 $F_{k,2}$. Since the existing fund size $q_{i,1}$, the management fees f, and the DRS/scale parameter c_k are also known, the investors can back out the fund gross return $R_{i,2} \equiv r_{i,2} + c_k \cdot q_{i,1} + f = \alpha_i + \beta_i^k F_{k,2} + \varepsilon_{i,2}$. Using this information, they update their beliefs about managerial skill.[17]

[17] Investors also update their beliefs about fund strategy at $t = 2$ if the fund has adopted a generic prospectus. In this case, the posterior mean and variance of the fund strategy are equal to

$$\hat{\beta}_{i,2}^k(g) = (1 - \lambda_g^{\beta^k})b^k + \lambda_g^{\beta^k}\left(\frac{R_{i,2} - \bar{\alpha}}{F_{k,2}}\right), \quad \hat{\sigma}^2_{\beta_i^k,2}(g) = \sigma_\gamma^2(1 - \lambda_g^{\beta^k}), \quad \text{where} \quad \lambda_g^{\beta^k} = \frac{\sigma_\gamma^2 F_{k,2}^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\gamma^2 F_{k,2}^2}.$$

Since our economy ends at $t = 3$, these quantities are not relevant for the equilibrium analyzed here.

Given a prospectus of type $p \in \mathcal{P}$, the perceived managerial skill is equal to

$$\hat{\alpha}_{i,2}(p) \equiv \mathbb{E}_2^{\mathcal{I}^p}[\alpha_i | R_{i,2}, F_{k,2}] = (1 - \lambda_p^\alpha)\bar{\alpha} + \lambda_p^\alpha (R_{i,2} - \mathbb{E}_1^{\mathcal{I}^p}[\beta_i^k] F_{k,2}), \tag{12}$$

where $(R_{i,2} - \mathbb{E}_1^{\mathcal{I}^p}[\beta_i^k] F_{k,2})$ represents the *perceived active return* (PAR), and

$$\lambda_p^\alpha = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \mathbb{V}\mathrm{ar}_1^{\mathcal{I}^p}[\beta_i^k] F_{k,2}^2}. \tag{13}$$

Intuitively, the best forecast of a fund manger's skill is a weighted average of the investor's prior belief $\bar{\alpha}$ and the PAR, where the weight λ_p^α captures the "speed of learning" and is given by the ratio of the covariance of managerial skill with the PAR, and the variance of the latter. The uncertainty on the perceived managerial skill is given by the posterior variance $\mathbb{V}\mathrm{ar}_2^{\mathcal{I}^p}[\alpha_i | R_{i,2}, F_{k,2}] = \sigma_\alpha^2(1 - \lambda_p^\alpha)$.

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:

$$\begin{aligned} p_i^* &= \underset{p \in \mathcal{P}}{\arg\max} \;\; \mathbb{E}_0^{\mathcal{M}_i}[\hat{\alpha}_{i,2}(p)] - \frac{f}{2c_k} \mathbb{V}\mathrm{ar}_0^{\mathcal{M}_i}[\hat{\alpha}_{i,2}(p)] && (14) \\ &= \underset{p \in \mathcal{P}}{\arg\max} \;\; \bar{\alpha} + \gamma_i \mathbb{E}_0^{\mathcal{M}_i}[\lambda_g^\alpha F_{k,2}] \mathbb{1}_{\{p=g\}} - \frac{f}{2c_k} \mathbb{V}\mathrm{ar}_0^{\mathcal{M}_i} \left[\lambda_p^\alpha \left((\alpha_i - \bar{\alpha}) + \varepsilon_i + \gamma_i F_{k,2} \mathbb{1}_{\{p=g\}} \right) \right]. && (15) \end{aligned}$$

When $p = d$, evaluating the expression inside the arg max in (15) yields

$$\bar{\alpha} - \frac{f \sigma_\alpha^4}{2c_k(\sigma_\alpha^2 + \sigma_\varepsilon^2)}, \tag{16}$$

whereas, when $p = g$, it yields

$$\bar{\alpha} + \gamma_i \mathbb{E}_0^{\mathcal{M}_i}[\lambda_g^\alpha F_{k,2}] - \frac{f}{2c_k} \mathbb{V}\mathrm{ar}_0^{\mathcal{M}_i} \left[\lambda_g^\alpha \left((\alpha_i - \bar{\alpha}) + \varepsilon_i + \gamma_i F_{k,2} \right) \right], \tag{17}$$

where

$$\lambda_g^\alpha = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\gamma^2 F_{k,2}^2}. \tag{18}$$

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 (17) is larger (smaller) than (16), 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 ($\mu_{F_k} > 0$), or when the exposure is smaller than average ($\gamma_i < 0$) and the factor return is expected to be negative ($\mu_{F_k} < 0$). 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-averse manager's utility. This increase in uncertainty is particularly severe when the fund's factor exposure β_i^k is sufficiently far from the within-mandate cross-sectional average b^k. 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 (17) are not available in closed-form, we are able to establish several useful propositions (all proofs can be found in Appendix B).

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

$$m = \Phi(\overline{\gamma}_k/\sigma_\gamma) - \Phi(\underline{\gamma}_k/\sigma_\gamma), \tag{19}$$

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

Proposition 1 formally shows that the endogenous distribution of prospectus types 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 top-left panel in figure 3 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 ($\underline{\gamma}_k, \overline{\gamma}_k$) 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 $\gamma_i > 0$).[18]

Based on the equilibrium prospectus choice, the next propositions present our key model implications by highlighting cross-sectional differences in various equilibrium quantities across funds adopting different types of prospectuses.

Proposition 2 (**Fund Size**). *The cross-sectional average of expected size across funds adopting prospectus p within mandate k is equal to*

$$\frac{\bar{\alpha} - f}{c_k} + \frac{\sigma_\gamma}{c_k} \left(\frac{\phi(\underline{\gamma}_k/\sigma_\gamma) - \phi(\overline{\gamma}_k/\sigma_\gamma)}{m} \right) \mathbb{E}[\lambda_g^\alpha F_{k,2}] \mathbb{1}_{\{p=g\}}, \tag{20}$$

where $\phi(\cdot)$ is the probability density function of a standard normal distribution, and m is as in (19). It follows that the expected size of funds adopting a generic prospectus is on average larger than the expected size of funds adopting a detailed prospectus within the same mandate.

Proposition 2 characterizes equilibrium fund size as a function of prospectus choice. The key finding is that expected fund size for generic prospectuses is higher on average than for detailed prospectuses. This is because the average mistake made by investors, captured by $\sigma_\gamma \left(\phi(\underline{\gamma}_k/\sigma_\gamma) - \phi(\overline{\gamma}_k/\sigma_\gamma) \right) / m = \mathbb{E}[\gamma_i | \underline{\gamma}_k < \gamma_i < \overline{\gamma}_k]$ in (20), is positive when the factor return is expected to be positive, and negative when the factor return is expected to be negative. The intuition, illustrated in the top-middle panel in figure 3, is that managers who optimally adopt a generic prospectus tend to benefit from investors' mistakes, which increase as the specific component of factor loading (γ_i) becomes more positive. This benefit comes in the form of higher expected fund flows, which lead to higher AUM, all else equal.

Proposition 3 (**Benchmark Misalignment**). *The cross-sectional average of benchmark misalignment across funds adopting prospectus p within mandate k is equal to*

$$\sigma_\gamma \left(\frac{2\phi(0) - \phi(\underline{\gamma}_k/\sigma_\gamma) - \phi(\overline{\gamma}_k/\sigma_\gamma)}{m} \right) \mathbb{1}_{\{p=g\}}. \tag{21}$$

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

It follows that the benchmark misalignment of funds adopting a generic prospectus is on average larger than the benchmark misalignment of funds adopting a detailed prospectus within the same mandate.

The results in Proposition 3 are a direct consequence of the investors' learning from prospectuses. Since a detailed prospectus perfectly reveals the fund exposure to the mandate-specific risk factor, the average benchmark misalignment across funds adopting a detailed prospectus is zero. A generic prospectus, instead, reveals only the average exposure within the fund's mandate, implying that the perceived fund exposure might differ form the true one. As a consequence, the average benchmark misalignment across funds adopting a generic prospectus, captured by $\sigma_\gamma \left(2\phi(0) - \phi(\underline{\gamma}_k/\sigma_\gamma) - \phi(\overline{\gamma}_k/\sigma_\gamma)\right)/m = \mathbb{E}[|\gamma_i||\underline{\gamma}_k < \gamma_i < \overline{\gamma}_k]$ in (21), is positive. The top-right panel in figure 3 plots the benchmark misalignment against a fund's strategy γ_i, and shows that it is always positive for $\underline{\gamma}_k < \gamma_i < \overline{\gamma}_k$.

Proposition 4 (**Perceived Active Return**). *The cross-sectional average of the perceived active return variance across funds adopting prospectus p within mandate k is equal to*

$$\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_{F_k}^2\ \sigma_\gamma \left(\sigma_\gamma + \frac{\underline{\gamma}_k \cdot \phi(\underline{\gamma}_k/\sigma_\gamma) - \overline{\gamma}_k \cdot \phi(\overline{\gamma}_k/\sigma_\gamma)}{m}\right) \mathbb{1}_{\{p=g\}}. \tag{22}$$

It follows that the uncertainty about the perceived active returns of funds adopting a generic prospectus is on average larger than the uncertainty about the perceived active returns of funds adopting a detailed prospectus within the same mandate.

As illustrated in the bottom-left panel of figure 3, the mistakes that investors make in evaluating a manager's performance introduce uncertainty, which makes the PAR of a fund adopting a generic prospectus more volatile than the PAR of a fund adopting a detailed prospectus within the same mandate. Intuitively, by mistakenly under- or over-estimating the fund's exposure to the risky factor, the investors of a fund adopting a generic prospectus are unable to perfectly "remove" the true benchmark return $\beta_i^k F_{k,t}$ from the fund return. The effect of investor mistakes is captured by the expression $\sigma_\gamma \left(\sigma_\gamma + (\underline{\gamma}_k \phi(\underline{\gamma}_k/\sigma_\gamma) - \overline{\gamma}_k \phi(\overline{\gamma}_k/\sigma_\gamma))/m\right) = \mathbb{E}[\gamma_i^2|\underline{\gamma}_k < \gamma_i < \overline{\gamma}_k]$ in (22), which is related to the cross-sectional average benchmark misalignment among funds with generic prospectuses. As a consequence, the realizations of factor returns remain a driver of the fund's PAR, and therefore contribute to the PAR volatility.

Fund managers care about the PAR uncertainty because it affects the volatility of fund flows and in turn the uncertainty of their fees. However, fund flows (and hence fees) are also affected by the endogenous response of the investors to PAR uncertainty, which determines the equilibrium flow-performance sensitivity.

Proposition 5 (**Flow-Performance Sensitivity**). *Within a given mandate k, the flow-performance sensitivity,*

$$\frac{\partial flow_{i,2}}{\partial \hat{r}^A_{i,2}} = \frac{\lambda^\alpha_p}{\bar{\alpha} - f}, \tag{23}$$

is always larger for funds adopting a detailed prospectus, irrespective of the mandate-specific factor return $F_{k,2}$. Moreover, the difference in flow-performance sensitivity between funds adopting a detailed prospectus and funds adopting a generic prospectus is increasing in $|F_{k,2}|$.

The result in proposition 5 further highlights the core learning mechanism of our model. The PARs 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 PAR 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 bottom-middle panel of figure 3 illustrates this result.

On the one hand, higher PAR uncertainty increases fund flow uncertainty. On the other hand, a lower flow-performance sensitivity reduces uncertainty. Thus, the net effect of the choice of prospectus on the uncertainty of fund flows is ambiguous. 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 play 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, $\gamma_i F_{k,2}$, which tends to increase the 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^a_g, reflecting the slower learning rate of the investors who read generic prospectuses.

The bottom-right 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, which is equal to $\sigma_\alpha^2\big((\bar{\alpha}-f)\sqrt{\sigma_\alpha^2+\sigma_\varepsilon^2}\big)^{-1}$ for funds adopting a detailed prospectus, and

$$\frac{1}{m(\bar{\alpha}-f)}\int_{\underline{\gamma}_k/\sigma_\gamma}^{\overline{\gamma}_k/\sigma_\gamma}\sqrt{\mathbb{E}[(\lambda_g^\alpha)^2](\sigma_\alpha^2+\sigma_\varepsilon^2)+\mathbb{V}\mathrm{ar}[\lambda_g^\alpha F_{k,2}]\sigma_\gamma^2 z^2}\;\phi(z)dz \tag{24}$$

for funds adopting a generic prospectus, might be higher or lower for generic prospectuses, depending on the distribution of fund strategies.

4 Empirical Tests

4.1 Measurement

To test the predictions of the model, we require empirical measures/proxies for the following key concepts: (i) the investment mandate/category of each fund; (ii) the amount of detail in the fund's prospectus; (iii) the fund's true/objective benchmark, the investor-perceived benchmark, and the perceived active return; and (iv) the specialization of the fund's strategy.

4.1.1 Investment Mandates

We consider three different methods of grouping similar funds into investment mandates, based alternatively on fund holdings, fund returns, and PIS text.

Return-based mandates are constructed using the three-factor model of Fama and French (1993). Each quarter, funds are ranked separately according to their market, SMB, and HML betas, estimated from monthly rolling regressions over the past 24 months. We then split funds into terciles along each dimension, forming 27 groups of funds. For example, funds that independently fall into the bottom tercile of market beta, the middle tercile of SMB beta, and the bottom tercile of HML beta, are placed in the group labeled *FF3-121*. As a result of this methodology, the number of funds in each group will be potentially uneven. In practice, the smallest group is *FF3-123* with 2197 fund-quarter observations, and the largest is *FF3-333* with 5200 observations.

Holdings-based mandates are constructed similarly by ranking funds on the average market capitalization, book-to-market ratio, and past returns of the stocks in their portfolios. We call these "DGTW" mandates, after the Daniel et al. (1997) characteristics-based benchmarks. This methodology also results in 27, potentially uneven, groups. In practice, the smallest group is *DGTW-311* with 844 fund-quarter observations, and the largest is *DGTW-323* with 5658 observations. Note that Morningstar style-box is also based on a TNA-weighted avereage of stock holdings' characteristics. The 9 styles included in Morningstar's style box are nested by our methodology.

For text-based mandates we use the Strategy Peer Groups (SPGs) of Abis and Lines (2020), which are obtained using a machine learning algorithm to form clusters of funds with similar PIS descriptions. The methodology results in 17 distinct SPGs.[19] Abis and Lines (2020) show that funds belonging to the same SPG display significantly higher similarity in risk factor exposures, stock characteristics, returns, and portfolio holdings, and their measured characteristics are generally consistent with their stated objectives.

4.1.2 Prospectus Detail

Section 2.3 describes the construction of our financial content measure ($FinCon$). To better align with the model, we construct dummy variables for low, medium, and high $FinCon$, based on the terciles of its distribution. We associate *Low FinCon* with *generic* prospectuses, and *High FinCon* with *detailed* ones. In all of our empirical tests, we control for the alternative textual measures discussed in section 2.3 (length, complexity, uniqueness, and boilerplate content) to ensure that our results are driven explicitly by the amount of financial content in PIS descriptions, and not by correlated but conceptually irrelevant textual characteristics.

4.1.3 Benchmarks and Perceived Active Returns

For our purposes, a fund's *objective* benchmark is a composite portfolio constructed from its true exposure to each relevant factor, while its *perceived* benchmark is investors' "best guess" about the objective benchmark, given the information signals they have received from fund returns and

[19] For a full description of the algorithm (*k-means*), further details and validation tests, see Abis and Lines (2020). The peer groups are: *Large Cap*, *Mid Cap*, *Small Cap*, *Fundamental*, *Quantitative*, *Long Term*, *Intrinsic Value*, *Defensive*, *Dividends*, *New Products & Services*, *PE-Ratio*, *Competitive Advantage*, *Foreign (ADR)*, *Foreign (Emerging Markets)*, *Fixed Income*, *Derivatives*, and *Tax-Managed*. Note that, while all funds in our sample are equity funds—defined as holding at least 80% of their assets in common stock—their most distinguishing textual features may refer to the remaining 20% of assets; e.g. *Derivatives* and *Fixed Income*.

prospectuses. We construct an estimate of each fund's *objective benchmark* using the vector of factor loadings from a regression of monthly fund excess returns $r_{i,t}$ on monthly factor returns F_t:

$$r_{i,t} - r_{f,t} = \alpha_i + (\beta_i^o)' F_t + \varepsilon_{i,t}. \tag{25}$$

For each fund, the regression is run over the length of time it appears in our sample (i.e., from 2000 Q1 or the fund's inception date, whichever is later). Using the six-factor model of Carhart (1997) and Fama and French (2015), the estimated vector of coefficients $\hat{\beta}_i^o$ represents the objective benchmark of fund i.

In line with recent work by Berk and Van Binsbergen (2016) and Barber et al. (2016), we use fund flows to infer investors' *perceived* benchmarks.[20] The idea is to select, for each fund, the vector of loadings which best explains variations in 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:

$$flow_{i,t+1} \equiv \frac{TNA_{i,t+1} - TNA_{i,t}(1 + r_{i,t+1})}{TNA_{i,t}}. \tag{26}$$

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

$$flow_{i,t+1} = a_i + b_i r_{i,t} - b'_{F,i} F_t + \varepsilon_{i,t+1}, \tag{27}$$

and obtain the *perceived benchmark* of fund i as the vector $\hat{\beta}_i^p = \frac{1}{\hat{b}_i} \hat{b}_{F,i}$. It follows that fund i's perceived active returns (PAR) can be obtained as

$$\hat{r}_{i,t}^A = r_{i,t} - (\hat{\beta}_i^p)' F_t. \tag{28}$$

For PAR volatility, we compute a rolling 24-month standard deviation. In the subsequent quarterly regressions, we use the PAR volatility as of the last month in each quarter.

[20] Unlike these papers, we remain agnostic about the correct asset pricing model. If a particular model (e.g., the CAPM) is used by most investors, our framework will capture this; however, we also allow the relevant factors to vary by fund.

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

$$BenchMisalign_i = \sqrt{\sum_{k=1}^{K} \left(\frac{\left(\hat{\beta}_i^o(k) - \hat{\beta}_i^p(k)\right) - mean\left(\hat{\beta}_i^o(k) - \hat{\beta}_i^p(k)\right)}{std\left(\hat{\beta}_i^o(k) - \hat{\beta}_i^p(k)\right)} \right)^2}, \tag{29}$$

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 $mean(\cdot)$ and $std(\cdot)$ 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.

4.1.4 Fund Strategy Specialization

Our measure of strategy specialization captures the distance between the factor loadings of the fund in question and the average loadings of all funds in the same investment mandate. Mathematically, it is defined as the (square root of the) sum of squared normalized objective FF6 betas (as estimated in (25)), where the normalization takes place within mandates:

$$\hat{\beta}_{i,t}^{norm} = \frac{\hat{\beta}_{i,t}^o - mean(\hat{\beta}_{i,t}^o)}{std(\hat{\beta}_{i,t}^o)}, \tag{30}$$

where $mean(\hat{\beta}_{i,t}^o)$ and $std(\hat{\beta}_{i,t}^o)$ represent, respectively, the cross-sectional average and standard deviation of the vector of objective factor loadings across all funds belonging to the same mandate as fund i at time t. Betas are winsorized at the 1% level before normalization. Intuitively, subtracting the mean betas implements the desired comparison with the average fund in the mandate, while dividing by the standard deviation adjusts each beta to be of comparable scale. Our measure of specialization is then given by:

$$Specialization_{i,t} = \sqrt{\sum_{k=1}^{K} \left(\hat{\beta}_{i,t}^{norm}(k)\right)^2}, \tag{31}$$

where K is the number of factors (i.e., 6 for the FF6 model). We compute a version of this measure for each different mandate specification (see section 4.1.1).

4.2 Results

4.2.1 Prospectus Choice

The main equilibrium relationship predicted by the model, expressed in Proposition 1, is that managers face a stronger incentive to disclose detailed qualitative information when their fund strategy is more specialized—i.e., when their risk factor loading is more extreme relative to the average fund. To test this prediction, we run variants of the following regression:

$$HighFinCon_{i,t} = a + b \cdot Specialization_{i,t} + \gamma' X_{i,t-1} + FEs + \varepsilon_{i,t}, \tag{32}$$

where $HighFinCon_{i,t}$ is a dummy variable indicating that financial content (see section 2.3) in the PIS description of fund i at time t is in the top tercile of its empirical distribution. The regression omits observations where financial content is in the middle tercile of its distribution, allowing for a comparison between generic ($LowFinCon$) and detailed ($HighFinCon$) prospectuses. $Specialization_{i,t}$ is as defined in subsection 4.1.4. $X_{i,t-1}$ is a vector of fund level control variables: log fund size and age, turnover and expense ratios, and additional style controls constructed as the average loading of stocks held by the fund on the Fama and French (2015) and Carhart (1997) market, size, value, momentum, investment and profitability factors. In all subsequent regressions, we use the same set of fund-level controls, only omitting variables if 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 mandate-quarter fixed effects, denoted in (34) by FEs, using the three different mandates constructed in section 4.1.1. Standard errors are double-clustered at the quarter and fund level.

Table II displays the results of these regressions. All specifications indicate that higher specialization is associated with a significantly greater probablity of a fund having a detailed strategy description ($HighFinCon$), confirming the main prediction of the model.[21] The results are similar across the different fixed effect specifications, which correspond to "within FF3 mandates" (column 1), "within DGTW mandates" (column 2), and "within SPGs" (column 3) identification. A one-standard-deviation increase in specialization raises the probability of a fund having a detailed prospectus by 2.0% to 2.3%. Given the unconditional probability of 1/3 (by construction), the relative increase is 6-7%. The results are statistically significant at the 5% level for the "within SPGs" specification, and at the 1% level for the "within FF3 mandates" and "within DGTW mandates" specifications. Note that while our proxies for prospectus detail and strategy specialization

[21] In untabulated results, we confirm that this relationship is robust to using the continuous $FinCon$ measure instead of the top tercile dummy.

are based on sound theory, they are nonetheless likely to contain some measurement error, biasing coefficient estimates towards zero. Therefore, these estimates should be interpreted as a lower bound on the strength of the effect.

Adjusted R^2 is highest for the "within SPGs" specification (0.26), indicating that the text-based mandate fixed effects (unsurprisingly) captures the most baseline variation in financial content. Indeed, it is encouraging that the relationship between financial content and specialization persists when controlling for this baseline textual similarity, suggesting that the relationship is not mechanical.

4.2.2 Fund Size

The next prediction of the model, expressed in Proposition 2, is that funds with generic prospectuses should be allocated more capital by investors on average (within each investment mandate). We test this prediction by running the following regression:

$$ln(TNA_{i,t}) = a + b \cdot HighFinCon_{i,t} + c \cdot MedFinCon_{i,t} + d \cdot ln(TNA_{i,i_0}) + \gamma' X_{i,t-1} + FEs + \varepsilon_{i,t}, \tag{33}$$

where $HighFinCon_{i,t}$ is a dummy variable indicating that financial content (see section 2.3) in the PIS description of fund i at time t is in the top tercile of its empirical distribution, and $MedFinCon_{i,t}$ indicates the middle tercile. $ln(TNA_{i,t})$ is the natural logarithm of the size (total net assets) of fund i at time t, and $ln(TNA_{i,i_0})$ is the log of the fund's size at its inception date i_0. We log-transform the size variables due to their high skewness. $X_{i,t-1}$ is a vector of the standard fund-level control variables (see section 4.2.1). FEs represent mandate-quarter fixed effects, using the three different mandates constructed in section 4.1.1. Standard errors are double-clustered at the quarter and fund level.

Table III reports the results of estimating (33). Across all variations, the negative $\hat{b}$ coefficients indicate that fund size decreases when financial content is high (i.e., for detailed prospectuses), as predicted by the model. Across the different mandate specifications our estimates indicate that *High FinCon* funds are about 13% smaller overall relative to *Low FinCon* funds, and the results are significant at the 5% level. The regressions all control for initial size at fund inception, which captures potential (static) differences between funds in distribution channels and marketing.

4.2.3 Benchmark Misalignment

The first key aspect of the learning mechanism in our model is that investors should be able to identify the true benchmark of a fund more accurately if a fund's prospectus contains detailed information about its strategy (Proposition 3). 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 (29), running variants of the following regression:

$$BenchMisalign_{i,t} = a + b \cdot HighFinCon_{i,t} + c \cdot MedFinCon_{i,t} + \gamma' X_{i,t-1} + FEs + \varepsilon_{i,t}, \quad (34)$$

where $HighFinCon_{i,t}$ is a dummy variable indicating that financial content (see section 2.3) in the PIS description of fund i at time t is in the top tercile of its empirical distribution, and $MedFinCon_{i,t}$ indicates the middle tercile. $X_{i,t-1}$ are the usual fund-specific control variables (see section 4.2.1). FEs represent mandate-quarter fixed effects, using the three different mandates constructed in section 4.1.1. Standard errors are double-clustered at the quarter and fund level.

Table IV presents the results, with columns 1 to 3 showing coefficients for the different mandate-quarter fixed effects specifications. The estimated relationship is consistent with the model's learning mechanism. Funds with more detailed strategy descriptions (i.e., *High FinCon* funds) have lower benchmark misalignment, with coefficients of -0.20 across the different mandate specifications. In terms of magnitude, this corresponds to a decrease of about 16% relative to funds with generic descriptions. This finding indicates that investors learn about funds' true benchmarks by reading their prospectus strategy descriptions.

4.2.4 Perceived Active Return Volatility

The second aspect of the learning mechanism in our model is that investors' perceived active returns (PAR) should exhibit higher volatility for funds with generic prospectuses (Proposition 4). This occurs because investors do not observe specific signals about funds' factor loadings but rather only the average loading among peers. The increased PAR volatility then arises from the volatility of factor returns mistakenly considered to be part of the active returns. Using the rolling PAR volatility ($\sigma(\hat{r}^A_{i,t})$) computed as in section 4.1.3, we run the following regression:

$$\sigma(\hat{r}^A_{i,t}) = a + b \cdot HighFinCon_{i,t} + c \cdot MedFinCon_{i,t} + \gamma' X_{i,t-1} + FEs + \varepsilon_{i,t}. \quad (35)$$

Funds in the top and middle terciles of our financial content measure are indicated by the dummy variables $HighFinCon_{i,t}$ and $MedFinCon_{i,t}$, respectively. The vector $X_{i,t-1}$ contains the usual fund-level control variables (see section 4.2.1). FEs represent mandate-quarter fixed effects, using the three different mandates constructed in section 4.1.1. Standard errors are double-clustered at the quarter and fund level.

Table V reports the estimated coefficients. Again supporting the model's learning mechanism, we find that perceived active returns are less volatile for funds with detailed prospectuses. Across the different mandate specifications in columns 1 to 3, the coefficient of interest, $\hat{b}$, ranges from -0.956 to -0.984, indicating that the volatility is approximately 15% lower relative to generic prospectuses.

4.2.5 Flow-Performance Sensitivity

The final implication of our model mechanism is that investors with access to more detailed strategy descriptions should place more weight on their estimates of active return when allocating capital to funds (Proposition 5). In practical terms, funds' flow-performance sensitivity (FPS) should be increasing in the amount of financial content in their prospectuses. To test this prediction, we run the following regression:

$$\begin{aligned} Flow_{i,t+1} = a + b_0 \cdot \hat{r}^A_{i,t} + b_1 \cdot (\hat{r}^A_{i,t} \times MedFinCon_{i,t}) + b_2 \cdot (\hat{r}^A_{i,t} \times HighFinCon_{i,t}) + \\ \delta \cdot Flow_{i,t} + \xi \cdot (\hat{r}^A_{i,t})^2 + \gamma' X_{i,t} + FEs + \varepsilon_{i,t}, \end{aligned} \tag{36}$$

where $MedFinCon_{i,t}$ and $HighFinCon_{i,t}$ are dummy variables associated with the middle and top terciles of the financial content of fund i's prospectus in quarter t. $\hat{r}^A_{i,t}$ 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 4.1.3). $X_{i,t}$ are fund-specific control variables (see section 4.2.1). We also include past flows ($Flow_{i,t}$) and squared PAR ($(\hat{r}^A_{i,t})^2$), the latter to capture potential convexity in the flow-performance relationship. FEs represent mandate-quarter fixed effects, using the three different mandates constructed in section 4.1.1. Standard errors are double-clustered at the quarter and fund level. 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 form the bottom to the top tercile of its cross-sectional distribution.

Table VI presents the results, which are consistent across the different mandate specifications in columns 1 to 3, and continue to support the mechanism of our model. For funds with generic prospectuses, investors exhibit a positive but only marginally significant sensitivity to lagged PARs (coefficient of 0.006). This sensitivity more than triples for funds with detailed prospectuses: the coefficient on the interaction term between the top *FinCon* tercile and lagged PAR is 0.014, significant at the 5% level. This confirms our model's prediction that flow-performance sensitivity is stronger for funds with detailed prospectuses, and provides new insight into this important relationship. Indeed, our results suggest that investors' performance-chasing behavior is mediated by their ability to estimate the skill-based component of returns, specifically using non-return information.[22]

4.2.6 Age Effects

Although not an explicit prediction of the model, if our proposed learning mechanism is indeed behind the documented empirical findings, we would expect our results on the model's mechanism to be more pronounced for younger funds. Without a long history, investors cannot learn as much about the strategy of young funds from their returns, and in these cases the information communicated via prospectuses may be particularly helpful. This turns out to be the case for all three components of the learning mechanism: benchmark misalignment, perceived active return volatility, and flow-performance sensitivity.

First, as reported in columns 4-5 of table IV, we find that detailed prospectuses are associated with lower benchmark misalignment only for the subsample of young funds (coefficient of -0.28, significant at 1%). For older funds with longer performance histories, prospectus information does not appear to make a material difference to investors' inference about fund strategies. Second, the same age effects can be seen in our results for perceived active return volatility (columns 4-5 of table V). The coefficient on *High FinCon* for young funds (-1.348) is almost double that for old funds (-0.723), and the latter is not statistically significant. Third, strong age effects can also be found when analyzing flow-performance sensitivity (columns 4-5 of table VI). We find that detailed fund prospectuses are associated with a much higher FPS for younger funds (below-median age), with a coefficient of 0.021 compared to 0.011 for old funds (the latter only significant at the 10% level). This suggests that investors are less certain about their performance estimates for the older funds and so are less sensitive to perceived performance when allocating capital.

[22] We also note that, unlike in 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.

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 core learning mechanism of our model.

5 Conclusion

In this paper, we study the interplay between fund managers' incentives to disclose information through prospectuses and investors' learning from this information. We show both theoretically and empirically that greater disclosure provides a clearer signal about the fund's strategy to investors, and therefore about the appropriate performance benchmark. In response to this signal, investors perceive the fund's benchmark more accurately, learn faster about the manager's skill, and therefore exhibit a higher flow-performance sensitivity. Less detailed disclosures, on the other hand, cause investors to make more mistakes in detemining the correct benchmark for the fund, which leads to higher uncertainty in their perception of active returns (and thus in their capital allocation). Investor mistakes can also be beneficial for some fund managers, if they have above-average factor exposures while average factor returns are also positive.

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 the Principal Investment Strategy descriptions in fund prospectuses. Lastly, we show that investors' learning from prospectuses is greater for young funds, reinforcing the proposed mechanism.

Figure 1: Financial Content: Distribution

This figure displays the empirical distribution of our financial content measure ($FinCon$) in the quarterly panel of PIS descriptions, constructed as in equation (1), and winsorized at the 0.1% level.

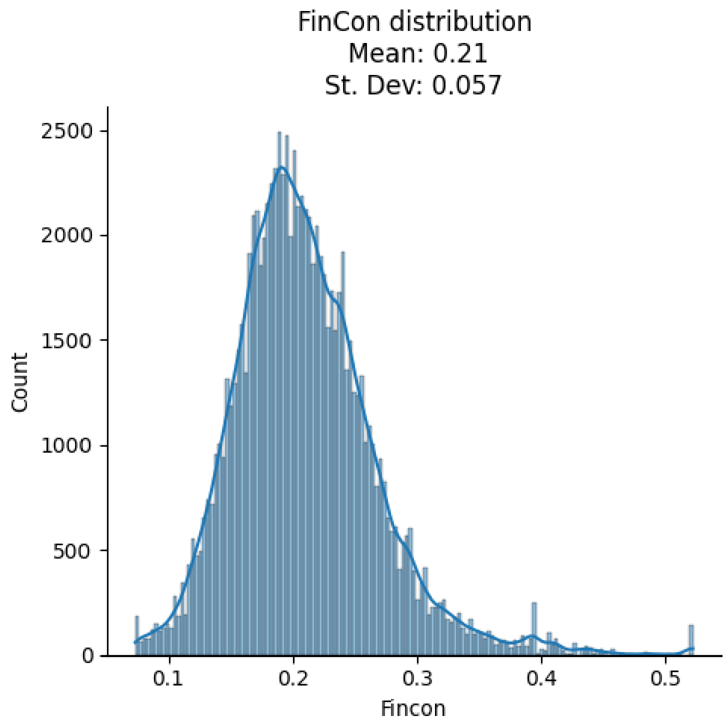

Figure 2: Financial Content: Correlations

This figure reports correlations between our financial content measure ($FinCon$) and other commonly-used textual statistics: length (word count), Kincaid et al. (1975) complexity, Kostovetsky and Warner (2020)'s uniqueness measure, and boilerplate content. The construction of these measures is described in section 4.1. In particular, we implement two versions of Kostovetsky and Warner (2020)'s measure: (i) using only the first 70 words, to match their sample of PIS excerpts from Morningstar, and (ii) using the full PIS text. All variables are winsorized at the 0.1% level.

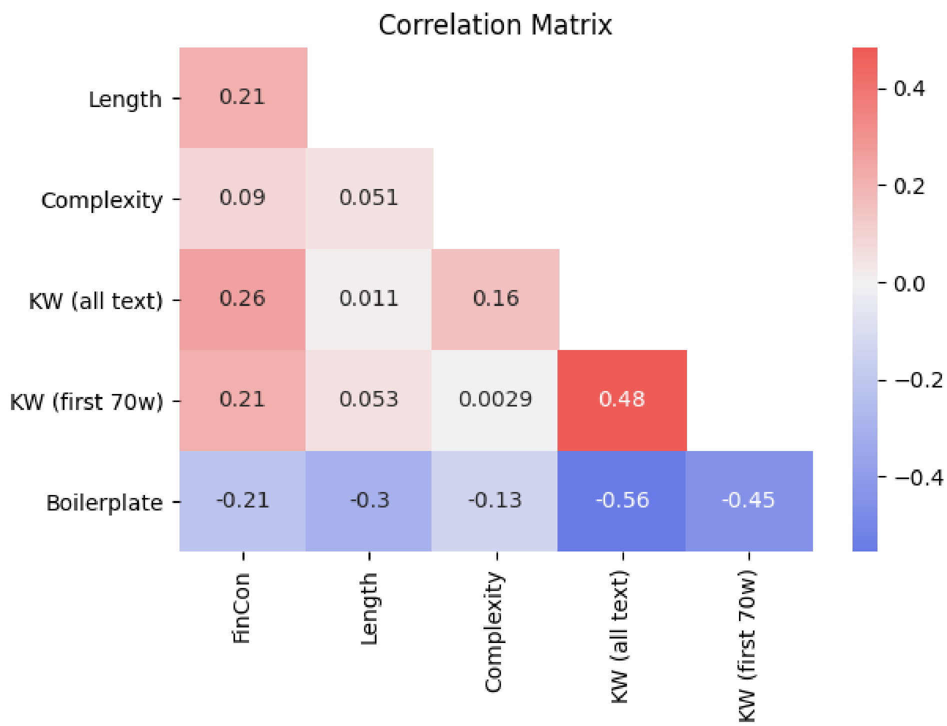

Figure 3: Equilibrium Within Investment Mandate

This figure plots the fund manager's expected utility $\mathbb{E}[v_i(q_{i,2} \cdot f)]$ (top-left panel), the expected fund size $\mathbb{E}[q_{i,2}]$ (top-middle panel), the benchmark misalignment $|\beta_i^k - \hat{\beta}_{i,t}^k(p)|$ (top-right panel), the perceived active return uncertainty $\mathbb{V}\text{ar}[R_{i,2} - \hat{\beta}_{i,t}^k(p) F_{k,2}]$ (bottom-left panel), and the fund flow uncertainty $\mathbb{V}\text{ar}[q_{i,2}/q_{i,1} - 1]$ (bottom-right panel), as a function of the fund strategy γ_i. The bottom-middle panel plots the flow-performance sensitivity $\lambda_p^\alpha / (\bar{\alpha} - f)$, as a function of the mandate-specific factor return $F_{k,2}$. Parameter values are: $\mu_{F_k} = 0.05$, $\sigma_{F_k} = 0.15$, $\bar{\alpha} = 0.02$, $\sigma_\alpha = \sigma_\varepsilon = \sigma_\gamma = 0.2$, $f = 0.01$, $c_k = 0.001$.

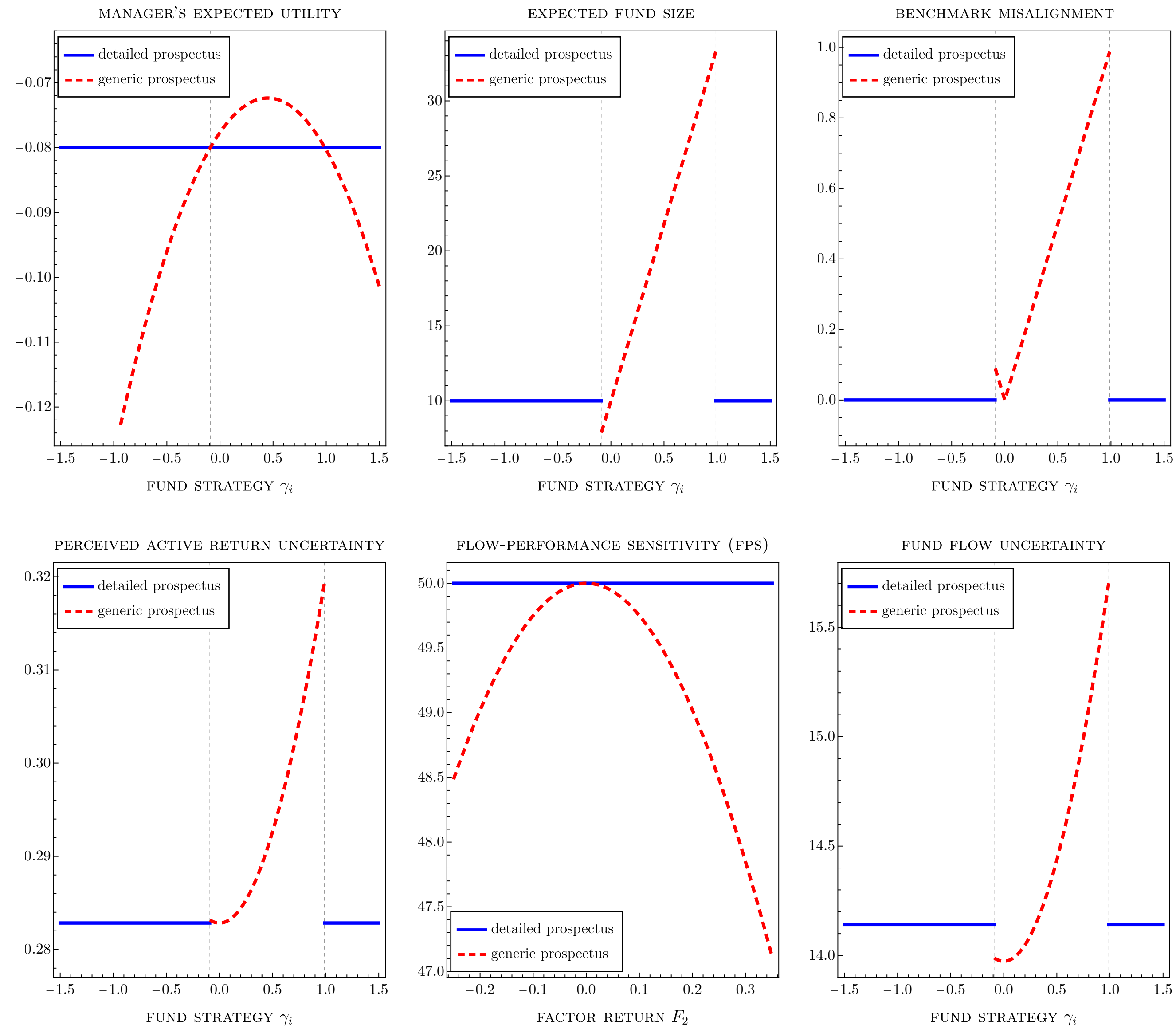

Table I: Financial Content and Fund Characteristics

This table reports differences in average characteristics between funds in the top and bottom terciles of our financial content measure. The results are obtained by regressing each characteristic on dummy variables for the middle and top terciles of $FinCon$, and on a set of demeaned fund-level controls (i.e., all characteristics other than the LHS variable, including holdings-based style controls derived from the FF6 model—see the discussion in section 4.2). The regressions with log total net assets (TNA), log age, expense ratio, and turnover ratio on the left-hand side are run in the quarterly panel. The regressions with skill measures on the left hand side are run cross-sectionally: active return (AR) and value added (VA) (along with all controls) are first averaged over the time-series for each fund, starting in 2000 Q1 or the fund's inception date, whichever is later. All variables are winsorized at the 0.1% level. Standard errors are clustered by fund and time, with the corresponding t-statistics reported in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Fund Characteristics				Fund Skill	
	Log(TNA)	Log(Age)	Exp Ratio	Turn Ratio	LT AR	LT VA
Bottom Tercile FinCon	5.9459	5.2641	0.0117	0.7542	-0.0002	0.0126
Top Tercile FinCon	5.8346	5.2495	0.0122	0.7429	-0.0002	0.0009
(Top - Bottom)	-0.1113***	-0.0146**	0.0005***	-0.0112*	-0.0001	-0.0117
	(-7.452)	(-2.854)	(14.1635)	(-1.7105)	(-0.7552)	(-1.5124)

Table II: Specialization and Prospectus Choice

This table documents the relationship between financial content and fund strategy specialization, estimated by equation 34. The dependent variable is a dummy variable indicating that an observation falls in the top tercile of the *FinCon* distribution (observations from the middle tercile of the distribution are excluded from the regression). Specialization is measured as the distance between the FF6 beta vector of each fund and the average, with each beta normalized within the investment mandate (details in section 4.1.4). For the mandates, we use three alternative specifications: in column 1, they are derived from similar exposure to the Fama and French (1993) factors; in column 2, they are derived from similar holdings characteristics, as in Daniel et al. (1997); and, in column 3, they are derived from the text-based Strategy Peer Groups (SPGs) of Abis and Lines (2020) (see section 4.1.1). All regressions contain fund-level control variables: size (log total net assets), log age, expense ratio, turnover ratio, and additional style controls (weighted averages of the FF6 betas of stocks held by the fund). Standard errors are clustered by fund and time, with the corresponding t-statistics reported in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Within FF3 (1)	Within DGTW (2)	Within SPG (3)
Specialization	0.021***	0.023***	0.020**
	(2.68)	(2.84)	(2.48)
Log(TNA)	-0.019***	-0.019***	-0.018***
	(-3.21)	(-3.14)	(-3.06)
Log(Age)	0.014	0.015	0.011
	(1.10)	(1.18)	(0.91)
Exp Ratio	0.047**	0.038	0.057***
	(2.11)	(1.66)	(2.72)
Turn Ratio	-0.000**	-0.000**	-0.000
	(-2.34)	(-2.33)	(-1.41)
Intercept	0.159*	0.155*	0.174**
	(1.83)	(1.73)	(2.09)
Style Controls	Y	Y	Y
Textual Controls	Y	Y	Y
Quarter FE	N	N	N
FF3-Quarter FE	Y	N	N
DGTW-Quarter FE	N	Y	N
SPG-Quarter FE	N	N	Y
AdjR2	0.17	0.17	0.26
Obs	57,048	57,144	57,621

Table III: Fund Size

This table documents the relationship between fund size (dependent variable) and financial content. Fund size is measured as the natural logarithm of total net assets, *Log(TNA)*. *Init Log(TNA)* is the initial size of the fund at inception. The dummy variables *Med FinCon* and *High FinCon* identify funds in the middle and top terciles of $FinCon$, respectively. All specifications feature quarter-mandate fixed effects, isolating the cross-sectional relationship within each investment mandate. For the mandates, we use three alternative specifications: in column 1, they are derived from similar exposure to the Fama and French (1993) factors; in column 2, they are derived from similar holdings characteristics, as in Daniel et al. (1997); and, in column 3, they are derived from the text-based Strategy Peer Groups (SPGs) of Abis and Lines (2020) (see section 4.1.1). All regressions contain fund-level control variables: size (log total net assets), log age, expense ratio, turnover ratio, and additional style controls (weighted averages of the FF6 betas of stocks held by the fund). Standard errors are clustered by fund and time, with the corresponding t-statistics reported in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Within FF3 (1)	Within DGTW (2)	Within SPG (3)
Med FinCon	-0.068	-0.070	-0.082*
	(-1.40)	(-1.44)	(-1.68)
High FinCon	-0.136**	-0.137**	-0.141**
	(-2.41)	(-2.43)	(-2.34)
Init Log(TNA)	0.470***	0.471***	0.467***
	(23.73)	(23.87)	(23.49)
Cuml PAR	0.012	0.011	0.012
	(1.52)	(1.42)	(1.50)
Log(Age)	0.881***	0.878***	0.858***
	(24.56)	(24.65)	(24.08)
Exp Ratio	-1.016***	-1.039***	-1.045***
	(-11.89)	(-12.15)	(-11.98)
Turn Ratio	-0.003***	-0.003***	-0.003***
	(-9.28)	(-9.16)	(-8.22)
Intercept	0.627**	0.600**	0.883***
	(2.46)	(2.38)	(3.35)
Style Controls	Y	Y	Y
Textual Controls	Y	Y	Y
Quarter FE	N	N	N
FF3-Quarter FE	Y	N	N
DGTW-Quarter FE	N	Y	N
SPG-Quarter FE	N	N	Y
AdjR2	0.44	0.45	0.44
Obs	86,896	87,024	87,729

Table IV: Benchmark Misalignment

This table estimates the relationship between benchmark misalignment (dependent variable) and financial content. Benchmark misalignment is measured as the Euclidean (normalized) distance between a fund's true/objective benchmark and the benchmark perceived by investors; the former is estimated using a full-sample linear factor regression for each fund, and the latter is inferred from fund flows (see section 4.2.3). The dummy variables *Med FinCon* and *High FinCon* identify funds in the middle and top terciles of $FinCon$, respectively. All specifications feature quarter-mandate fixed effects, isolating the cross-sectional relationship within each investment mandate. For the mandates, we use three alternative specifications: in column 1, they are derived from similar exposure to the Fama and French (1993) factors; in column 2, they are derived from similar holdings characteristics, as in Daniel et al. (1997); and, in column 3, they are derived from the text-based Strategy Peer Groups (SPGs) of Abis and Lines (2020) (see section 4.1.1). The regressions in columns 4-5 feature the same specification as in column 3, but are run separately for the sub-samples of below-median-age (young) and above-median-age (old) funds. All regressions contain fund-level control variables: size (log total net assets), log age, expense ratio, turnover ratio, and additional style controls (weighted averages of the FF6 betas of stocks held by the fund). Standard errors are clustered by fund and time, with the corresponding t-statistics reported in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Within FF3 (1)	Within DGTW (2)	Within SPG (3)	Young (4)	Old (5)
Med FinCon	-0.115	-0.118	-0.117	-0.280***	0.022
	(-1.56)	(-1.59)	(-1.60)	(-3.25)	(0.22)
High FinCon	-0.200**	-0.200**	-0.204**	-0.280***	-0.129
	(-2.39)	(-2.40)	(-2.51)	(-2.74)	(-1.21)
Log(TNA)	-0.066**	-0.068**	-0.064**	-0.104***	-0.030
	(-2.59)	(-2.64)	(-2.54)	(-3.47)	(-0.85)
Log(Age)	0.089	0.083	0.090	0.079	0.169
	(1.26)	(1.17)	(1.25)	(1.25)	(0.94)
Exp Ratio	-0.041	-0.018	-0.005	0.010	-0.014
	(-0.34)	(-0.14)	(-0.04)	(0.09)	(-0.07)
Turn Ratio	0.000	0.000	0.000	-0.000	0.001
	(0.26)	(0.17)	(0.40)	(-0.85)	(1.14)
Intercept	0.766*	0.744*	0.794*	1.288***	-0.043
	(1.89)	(1.82)	(1.90)	(2.80)	(-0.05)
Style Controls	Y	Y	Y	Y	Y
Textual Controls	Y	Y	Y	Y	Y
Quarter FE	N	N	N	N	N
FF3-Quarter FE	Y	N	N	N	N
DGTW-Quarter FE	N	Y	N	N	N
SPG-Quarter FE	N	N	Y	Y	Y
AdjR2	0.03	0.03	0.02	0.04	0.04
Obs	85,049	85,177	85,862	39,716	46,117

Table V: Perceived Active Return Volatility

This table documents the relationship between perceived active return volatility (dependent variable) and financial content. PAR volatility is computed as the rolling 24-month standard deviation of flow-implied active returns (see section 4.1.3). The dummy variables *Med FinCon* and *High FinCon* identify funds in the middle and top terciles of *FinCon*, respectively. All specifications feature quarter-mandate fixed effects, isolating the cross-sectional relationship within each investment mandate. For the mandates, we use three alternative specifications: in column 1, they are derived from similar exposure to the Fama and French (1993) factors; in column 2, they are derived from similar holdings characteristics, as in Daniel et al. (1997); and, in column 3, they are derived from the text-based Strategy Peer Groups (SPGs) of Abis and Lines (2020) (see section 4.1.1). The regressions in columns 4-5 feature the same specification as in column 3, but are run separately for the sub-samples of below-median-age (young) and above-median-age (old) funds. All regressions contain fund-level control variables: size (log total net assets), log age, expense ratio, turnover ratio, and additional style controls (weighted averages of the FF6 betas of stocks held by the fund). Standard errors are clustered by fund and time, with the corresponding t-statistics reported in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Within FF3	Within DGTW	Within SPG	Young	Old
	(1)	(2)	(3)	(4)	(5)
Med FinCon	-0.367	-0.359	-0.368	-1.296***	0.220
	(-1.00)	(-0.98)	(-1.00)	(-2.79)	(0.46)
High FinCon	-0.959**	-0.956**	-0.984**	-1.348**	-0.723
	(-2.35)	(-2.34)	(-2.46)	(-2.57)	(-1.43)
PAR	0.181	0.214	0.218	1.339	-0.449
	(0.21)	(0.25)	(0.25)	(1.20)	(-0.36)
Log(TNA)	-0.340***	-0.340***	-0.322**	-0.513***	-0.197
	(-2.69)	(-2.67)	(-2.53)	(-3.27)	(-1.17)
Log(Age)	0.454	0.431	0.427	0.615	0.693
	(1.20)	(1.13)	(1.12)	(1.43)	(0.88)
Exp Ratio	0.030	0.149	0.215	0.269	0.248
	(0.05)	(0.25)	(0.37)	(0.46)	(0.29)
Turn Ratio	0.003	0.003	0.004	0.001	0.007*
	(1.01)	(1.01)	(1.41)	(0.24)	(1.68)
Intercept	4.138*	3.826*	4.202*	5.956**	0.908
	(1.97)	(1.81)	(1.92)	(2.12)	(0.22)
Style Controls	Y	Y	Y	Y	Y
Textual Controls	Y	Y	Y	Y	Y
Quarter FE	N	N	N	N	N
FF3-Quarter FE	Y	N	N	N	N
DGTW-Quarter FE	N	Y	N	N	N
SPG-Quarter FE	N	N	Y	Y	Y
AdjR2	0.05	0.05	0.04	0.06	0.05
Obs	71,595	71,688	72,299	29,266	43,027

Table VI: Flow-Performance Sensitivity

This table reports the estimated sensitivity of fund flows to perceived active returns (PAR), and how this sensitivity varies for different levels of prospectus financial content ($FinCon$). PARs are computed by subtracting investors' perceived benchmark returns (inferred from flows) from funds' total returns (see section 4.1.3). The dummy variables *Med FinCon* and *High FinCon* identify funds in the middle and top terciles of $FinCon$, respectively. All specifications feature quarter-mandate fixed effects, isolating the cross-sectional relationship within each investment mandate. For the mandates, we use three alternative specifications: in column 1, they are derived from similar exposure to the Fama and French (1993) factors; in column 2, they are derived from similar holdings characteristics, as in Daniel et al. (1997); and, in column 3, they are derived from the text-based Strategy Peer Groups (SPGs) of Abis and Lines (2020) (see section 4.1.1). The regressions in columns 4-5 feature the same specification as in column 3, but are run separately for the sub-samples of below-median-age (young) and above-median-age (old) funds. All regressions contain fund-level control variables: size (log total net assets), log age, expense ratio, turnover ratio, and additional style controls (weighted averages of the FF6 betas of stocks held by the fund). Standard errors are clustered by fund and time, with the corresponding t-statistics reported in parenthesis. ***, **, * indicate significance at less than 1%, 5%, and 10%, respectively.

	Within FF3 (1)	Within DGTW (2)	Within SPG (3)	Young (4)	Old (5)
PAR	0.006*	0.006	0.006	0.006	0.006
	(1.71)	(1.55)	(1.59)	(1.18)	(1.23)
PAR × MedFinCon	0.009*	0.010**	0.010**	0.016**	0.006
	(1.88)	(2.15)	(2.17)	(2.04)	(1.03)
PAR × HighFinCon	0.014**	0.014**	0.014**	0.021**	0.011*
	(2.41)	(2.37)	(2.36)	(2.04)	(1.75)
Med FinCon	-0.001	-0.001	-0.001	-0.001	-0.002
	(-0.89)	(-1.15)	(-1.30)	(-0.67)	(-1.23)
High FinCon	-0.001	-0.001	-0.001	-0.003	-0.000
	(-1.08)	(-1.18)	(-1.09)	(-1.41)	(-0.14)
Flow(t)	0.415***	0.415***	0.417***	0.421***	0.399***
	(34.15)	(34.70)	(34.19)	(28.79)	(26.96)
PAR2	0.008***	0.007***	0.008***	0.008**	0.008***
	(3.92)	(3.73)	(3.92)	(2.24)	(3.39)
Log(TNA)	-0.002***	-0.002***	-0.002***	-0.003***	-0.001***
	(-6.48)	(-7.08)	(-6.45)	(-5.78)	(-3.04)
Log(Age)	-0.009***	-0.009***	-0.009***	-0.017***	-0.000
	(-10.81)	(-10.46)	(-10.73)	(-7.97)	(-0.40)
Exp Ratio	-0.007***	-0.007***	-0.007***	-0.008***	-0.005***
	(-4.39)	(-4.69)	(-4.12)	(-3.07)	(-2.89)
Turn Ratio	-0.000***	-0.000***	-0.000***	-0.000*	-0.000***
	(-4.40)	(-4.58)	(-3.51)	(-1.76)	(-4.31)
Intercept	0.075***	0.077***	0.074***	0.114***	0.017**
	(13.31)	(13.71)	(12.99)	(10.72)	(2.27)
Style Controls	Y	Y	Y	Y	Y
Textual Controls	Y	Y	Y	Y	Y
Quarter FE	N	N	N	N	N
FF3-Quarter FE	Y	N	N	N	N
DGTW-Quarter FE	N	Y	N	N	N
SPG-Quarter FE	N	N	Y	Y	Y
AdjR2	0.23	0.23	0.22	0.24	0.20
Obs	85,916	86,038	86,735	40,643	46,075

APPENDIX

A PIS Examples

To better understand our $FinCon$ measure, consider the examples below. For exposition purposes, they have been chosen to be between 100 and 200 words in length, that is relatively short (between about the 20^{th} and 60^{th} percentiles of PIS length). To emphasize the textual differences between high and low $FinCon$, the first example is chosen from the bottom 10% of the $FinCon$ distribution, while the others are chosen from the top 10%.

For each example we provide the full text of the PIS description, as well as a pre-processed version of that text. The latter is used in the construction of $FinCon$. Financial terms and short phrases are color-coded. The color gradation indicates the weight each term or short phrase is given in the computation of $FinCon$, with lower weights tending towards green and higher weights tending towards blue. Given the definition of $FinCon$ in (1), $weight = \frac{1}{(1+Coverage_{\omega})^4} \in [0, 1]$. Intuitively, the more commonly used a term is, the more its weight is shrunk towards zero, meaning that the term is treated more similarly to a non-financial term. Less common financial terms, instead, have weights closer to one, hence contributing more to a higher $FinCon$.

Common financial terms, which are frequently utilized by most funds and thus receive low weights, are part of the context of the majority of prospectuses (e.g., it is very common to mention the terms *fund* and *invest* in describing a fund's investment strategy). Therefore, what drives the difference between the $FinCon$ of generic and detailed prospectuses is the use of relatively less common financial terms. Accordingly, in the examples below, terms highlighted in green are present in the text of both the generic and the detailed prospectuses. Instead, terms highlighted in blue are mostly present in the text of detailed prospectuses.

Example 1 – Low financial content. Fund Family: The Advisors' Inner Circle; Fund: FMC Strategic Value Fund; Date: 2012-3.

Full text before pre-processing:

> The Fund invests primarily in common stocks of U.S. companies with small to medium market capitalizations (between $250 million and $5 billion) that FirstManhattan Co. ("FMC" or the "Adviser") believes are selling at a market price below their true value and offer the potential to increase in value. The Fund will generally invest in equity

securities of domestic companies, but may also invest in equity securities of foreign companies and American Depositary Receipts("ADRs"). The Adviser expects that the Fund's investments in foreign companies will normally represent less than 20% of the Fund's assets. In selecting investments for the Fund, the Adviser focuses on companies in industries and sectors about which the Adviser believes it has a substantial understanding. The Adviser also seeks to invest Fund assets in securities of companies where the Adviser has identified a catalyst which could have a significant positive impact on the market price of the company's stock. The Fund intends to buy and hold securities of companies for the long-term, and seeks to limit portfolio turnover. The Fund may sell a security, however, if the security achieves a designated price target or there is a fundamental change in a company's outlook.

Pre-processed text, with financial terms and short phrases colored according to their weight:

fund invest primarili common stock unit state compani small medium market cap million billion first manhattan fmc advis believ sell market price below true valu offer potenti increas valu fund gener invest equiti secur domest compani may also invest equiti secur foreign compani american depositari receipt adr advis expect fund invest foreign compani normal repres less fund asset select invest fund advis focus compani industri sector advis believ substantial understand advis also seek invest fund asset secur compani advis identifi catalyst signific posit impact market price compani stock fund intend buy hold secur compani long term seek limit portfolio turnov fund may sell secur howev if secur achiev design price target fundamental chang compani outlook

In example 1, most financial terms and short phrases are highlighted in green. These terms are very common and not particularly informative about the fund's specific strategy. There are very few terms highlighted in blue, with the highest weight being only 0.77. Since most financial terms in this prospectus are green-shaded (i.e., they have weights close to zero), the fund's $FinCon$ is only 13.63%. Indeed, this prospectus only provides a description of the fund's investment mandate (invest in smaller stocks, keep a value perspective), without including any details on the specific security-selection criteria or risk exposures of the fund.

Example 2 – High financial content. Fund Family: Perritt Funds, Inc; Fund: Perritt MicroCap Opportunities Fund; Date: 2013-03.

Full text before pre-processing:

The Fund normally invests at least 80% of its net assets (plus any borrowings for investment purposes) in the common stocks of United States companies with market capitalizations that are below $500 million at the time of initial purchase, which the Fund's investment adviser refers to as "micro-cap" companies. The Fund invests in both value-priced and aggressive growth stocks. Generally, the Fund's investment adviser seeks to invest in companies with the following attributes: (1) Have demonstrated above-average growth in revenues and/or earnings; (2) Possess relatively low levels of long-term debt; (3) Have a high percentage of their shares owned by company management; and (4) Possess modest price-to-sales ratios and price-to-earnings ratios that are below their long-term annual growth rate. At times, the Fund may invest in "special situations" such as companies that possess valuable patents, companies undergoing restructuring, and companies involved in large share repurchase programs. Although the Fund seeks long-term capital appreciation, stocks may be sold in the short-term for several reasons. These include: (1) a company's market capitalization grows beyond $1.5 billion; (2) a company's financial condition deteriorates to the point that the Fund's investment adviser believes that the company's long-term growth prospects may be impaired; (3) a company receives a purchase offer from another company; or (4) a company's price-to-sales ratio or price-to-earnings ratio expands to the point that the Fund's investment adviser believes the company's stock is significantly overvalued. Generally, the Fund's portfolio contains 150 to 200 stocks. The Fund is intended for investors who are willing to withstand the risk of short-term price fluctuations in exchange for potential long-term capital appreciation.

Pre-processed text, with financial terms and short phrases colored according to their weight:

fund normal invest least net asset plu borrow invest purpos common stock unit state compani market cap million time initi purchas fund invest advis refer micro cap compani fund invest valu price aggress growth stock gener fund invest advis seek invest compani follow attribut demonstr averag growth revenu earn possess rel low level long term debt high percentag share own compani manag possess modest price sale ratio price earn ratio long term annual growth rate time fund may invest special situat compani possess valuabl patent compani undergo restructur compani involv larg share repurchas program although fund seek long term capit appreci stock may sold short term sever reason includ compani market cap grow beyond billion compani financi condit deterior point fund invest advis believ compani long term growth prospect may impair compani receiv purchas offer anoth compani compani price sale ratio price earn

> ratio expand point fund invest advis believ compani stock significantli overvalu gener fund portfolio contain stock fund intend investor will withstand risk short term price fluctuat exchang potenti long term capit appreci

Example 2 contains many more financial terms but, most importantly, it contains disproportionally more financial terms that are not used by most of the other funds. This results in many more blue-shaded terms, with the highest weight being 0.99. Aggregating all the weights, and rescaling by the text's length, the fund's $FinCon$ is 31.76%. To ease the comparison between the two examples above, example 2 was chosen among funds within the same text-based mandate as the fund in example 1 (mandates are defined in Section 4.1.1). As in example 1, the prospectus in example 2 provides a generic description of the fund's investment focus (invest in smaller stocks, keep both value-priced and aggressive growth perspectives). In contrast to example 1, it includes more information about the fund's specific strategy, e.g., the criteria utilized in selecting securities and in deciding when to sell them.

Example 3 – High financial content. Fund Family: John Hancock Investment Trust; Fund: John Hancock Small Cap Core Fund; Date: 2016-03.

Full text before pre-processing:

> Under normal market conditions, the fund invests at least 80% of its net assets (plus any borrowings for investment purposes) in equity securities of small-capitalization companies. The fund considers small-capitalization companies to be those within the capitalization range of the Russell 2000 Index, with a maximum capitalization of $6.47 billion as of December 31, 2015. The fund generally will not invest in companies with market capitalizations of $5 billion or more. Equity securities include common and preferred stocks, rights, warrants, and depositary receipts. The manager emphasizes a fundamental, bottom-up approach to individual stock selection, looking for companies with durable, niche business models with the potential for high returns on capital and that the manager believes are undervalued. Companies are screened based on a number of factors, including balance sheet quality, profitability, liquidity, size, and risk profile. The fund intends to invest in a number of different sectors based on stock selection and sector weightings may vary significantly from its benchmark. The fund may focus its investments in a particular sector or sectors. The fund may invest up to 10% of its total assets in foreign securities including emerging-market securities and securities of non-U.S. companies traded on a U.S. exchange. The fund may invest in initial public

offerings (IPOs), real estate investment trusts (REITs) or other real estate-related equity securities, and certain exchange-traded funds (ETFs). The fund normally will invest 10% or less of its total assets in cash and cash equivalents, including repurchase agreements, money market securities, U.S. government securities, and other short-term investments. The fund may invest in derivatives to a limited extent. Derivatives may be used to reduce risk and/or obtain efficient market exposure, and may include futures contracts and foreign currency forward contracts.

Pre-processed text, with financial terms and short phrases colored according to their weight:

normal market condit fund invest least net asset plu borrow invest purpos equiti secur small cap compani fund consid small cap compani within cap rang russel maximum cap billion decemb fund gener invest compani market cap billion more equiti secur includ common prefer stock right warrant depositari receipt manag emphas fundament bottom approach individu stock select look compani durabl nich busi model potenti high return capit manag believ undervalu compani screen base number factor includ balanc sheet qualiti profit liquid size risk profil fund intend invest number differ sector base stock select sector weight may vari significantli benchmark fund may focu invest particular sector sector fund may invest total asset foreign secur includ emerg market secur secur non unit state compani trade unit state exchang fund may invest initi public offer ipo real estat invest trust reit real estat relat equiti secur certain exchang trade fund etf fund normal invest less total asset cash cash equival includ repurchas agreement money market secur unit state govern secur short term invest fund may invest deriv limit extent deriv may use reduc risk obtain effici market exposur may includ futur contract foreign currenc forward contract

Similarly to example 2, example 3 contains a greater proportion of less common financial terms than example 1, resulting in more blue-shaded terms, with the highest weight being 0.997. In this case, the fund's $FinCon$ is 31.21%. Differently from example 2, example 3 was chosen among funds within the same return-based and holdings-based mandate (as opposed to text-based mandate) as the fund in example 1. The prospectus in example 3 provides a greater amount of detail about the investment process, the investment criteria, as well as the range of assets and financial instruments utilized by the fund.

B Proofs

Lemma B.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 $Y \equiv (X | \underline{x} < X < \overline{x})$. It follows that Y has a truncated normal distribution with density equal to*

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

and first and second moments equal to

$$\mathbb{E}[Y] = \mu_X + \sigma_X \left(\frac{\phi(\underline{z}) - \phi(\overline{z})}{\Phi(\overline{z}) - \Phi(\underline{z})} \right), \tag{B.2}$$

$$\mathbb{V}\text{ar}[Y] = \sigma_X^2 \left(1 + \frac{\underline{z}\phi(\underline{z}) - \overline{z}\phi(\overline{z})}{\Phi(\overline{z}) - \Phi(\underline{z})} - \left(\frac{\phi(\underline{z}) - \phi(\overline{z})}{\Phi(\overline{z}) - \Phi(\underline{z})} \right)^2 \right), \tag{B.3}$$

where $\underline{z} \equiv (\underline{x} - \mu_X)/\sigma_X$, $\overline{z} \equiv (\overline{x} - \mu_X)/\sigma_X$, 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 (17) is larger than (16). Let $\Psi(\gamma_i)$ denote the difference between (17) and (16), multiplied by $2c_k/f$:

$$\Psi(\gamma_i) = \frac{2c_k}{f}\gamma_i \mathbb{E}[\lambda_g^\alpha F_{k,2}] - \mathbb{V}\text{ar}[\lambda_g^\alpha((\alpha_i - \bar{\alpha}) + \varepsilon_i + \gamma_i F_{k,2})] + \frac{\sigma_\alpha^4}{(\sigma_\alpha^2 + \sigma_\varepsilon^2)}, \tag{B.4}$$

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 (B.4) can be written as

$$\begin{aligned} \mathbb{V}\text{ar}[\lambda_g^\alpha(z_i + \gamma_i F_{k,2})] &= \gamma_i^2 \mathbb{V}\text{ar}[\lambda_g^\alpha F_{k,2}] + \mathbb{V}\text{ar}[\lambda_g^\alpha z_i] + 2\gamma_i \mathbb{C}\text{ov}[\lambda_g^\alpha z_i, \lambda_g^\alpha F_{k,2}] \\ &= \gamma_i^2 \mathbb{V}\text{ar}[\lambda_g^\alpha F_{k,2}] + \left(\sigma_\alpha^2 + \sigma_\varepsilon^2\right) \mathbb{E}[(\lambda_g^\alpha)^2], \end{aligned} \tag{B.5}$$

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

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\lambda_g^\alpha z_i] &= \mathbb{E}[(\lambda_g^\alpha)^2 (z_i)^2] - \mathbb{E}[\lambda_g^\alpha z_i]^2 \\ &= \mathbb{E}[(\lambda_g^\alpha)^2]\mathbb{E}[(z_i)^2] - (\mathbb{E}[\lambda_g^\alpha]\mathbb{E}[z_i])^2 \\ &= \left(\sigma_\alpha^2 + \sigma_\varepsilon^2\right) \mathbb{E}[(\lambda_g^\alpha)^2], \end{aligned} \tag{B.6}$$

and

$$\begin{aligned}\mathbb{C}\mathrm{ov}[\lambda_g^\alpha z_i, \lambda_g^\alpha F_{k,2}] &= \mathbb{E}[z_i (\lambda_g^\alpha)^2 F_{k,2}] - \mathbb{E}[\lambda_g^\alpha z_i]\mathbb{E}[\lambda_g^\alpha F_{k,2}] \\ &= \mathbb{E}[z_i]\mathbb{C}\mathrm{ov}[\lambda_g^\alpha, \lambda_g^\alpha F_{k,2}] \\ &= 0, \end{aligned} \tag{B.7}$$

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

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

with coefficients

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

$$a_1 = (c_k/f)\mathbb{E}[\lambda_g^\alpha F_{k,2}], \tag{B.10}$$

$$a_2 = -\mathbb{V}\mathrm{ar}[\lambda_g^\alpha F_{k,2}]. \tag{B.11}$$

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

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

Proof of Proposition 2. Given the equilibrium fund size in (8), the expected fund size associated with detailed and generic prospectuses are equal to

$$\mathbb{E}[q_{i,2}(d)] = \frac{\mathbb{E}[\hat{\alpha}_{i,2}(d)] - f}{c_k} = \frac{\bar{\alpha} - f}{c_k}, \tag{B.12}$$

$$\mathbb{E}[q_{i,2}(g)] = \frac{\mathbb{E}[\hat{\alpha}_{i,2}(g)] - f}{c_k} = \frac{\bar{\alpha} - f}{c_k} + \frac{1}{c_k}\mathbb{E}[\lambda_g^\alpha F_{k,2}]\gamma_i, \tag{B.13}$$

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

$$\frac{\bar{\alpha} - f}{c_k} + \frac{1}{c_k}\mathbb{E}[\lambda_g^\alpha F_{k,2}]\mathbb{E}[\gamma_i | \underline{\gamma}_k < \gamma_i < \overline{\gamma}_k], \tag{B.14}$$

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

$$\mathbb{E}[\gamma_i | \underline{\gamma}_k < \gamma_i < \overline{\gamma}_k] = \sigma_\gamma \frac{\phi(\underline{\gamma}_k/\sigma_\gamma) - \phi(\overline{\gamma}_k/\sigma_\gamma)}{\Phi(\overline{\gamma}_k/\sigma_\gamma) - \Phi(\underline{\gamma}_k/\sigma_\gamma)} \tag{B.15}$$

and (20) obtains.

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

Proof of Proposition 3. Let $M_{i,t}^{\beta^k}(p)$ denote the absolute value of the difference between the true and the perceived fund exposure to the mandate-specific risky factor, $\left|\beta_i^k - \hat{\beta}_{i,t}^k(p)\right|$, for a fund adopting prospectus of type p. Since after reading either prospectus types ($p = g$, or $p = d$), an investor comes to know the average exposure b^k across all funds belonging to the same mandate k,

$$M_{i,t}^{\beta^k}(p) = \left|\gamma_i - \mathbb{E}_t^{\mathcal{I}^p}[\gamma_i]\right|. \tag{B.16}$$

Since $\mathbb{E}_1^{\mathcal{I}^d}[\gamma_i] = \gamma_i$ and $\mathbb{E}_1^{\mathcal{I}^g}[\gamma_i] = 0$, it follows that the cross-sectional average of benchmark misalignment at $t = 1$ across funds adopting a detailed prospectus within mandate k is equal to

$$\mathbb{E}\left[M_{i,1}^{\beta^k}(d)\right] = 0, \tag{B.17}$$

whereas the cross-sectional average of benchmark misalignment at $t = 1$ across funds adopting a generic prospectus within the same mandate is equal to

$$\begin{aligned}
\mathbb{E}\left[M_{i,1}^{\beta^k}(g)\right] &= \mathbb{E}\left[|\gamma_i|\Big|\underline{\gamma}_k < \gamma_i < \overline{\gamma}_k\right] \\
&= \frac{1}{\int_{\underline{\gamma}_k}^{\overline{\gamma}_k} pdf_\gamma(\gamma_i)d\gamma_i} \int_{\underline{\gamma}_k}^{\overline{\gamma}_k} |\gamma_i| pdf_\gamma(\gamma_i)d\gamma_i \\
&= \frac{1}{\int_{\underline{\gamma}_k}^{\overline{\gamma}_k} pdf_\gamma(\gamma_i)d\gamma_i} \left(\int_0^{\overline{\gamma}_k} \gamma_i pdf_\gamma(\gamma_i)d\gamma_i + \int_{\underline{\gamma}_k}^0 -\gamma_i pdf_\gamma(\gamma_i)d\gamma_i\right) \\
&= \frac{1}{\int_{\underline{\gamma}_k}^{\overline{\gamma}_k} pdf_\gamma(\gamma_i)d\gamma_i} \left(\int_0^{\overline{\gamma}_k} pdf_\gamma(\gamma_i)d\gamma_i \cdot \mathbb{E}[\gamma_i|0 < \gamma_i < \overline{\gamma}_k] - \int_{\underline{\gamma}_k}^0 pdf_\gamma(\gamma_i)d\gamma_i \cdot \mathbb{E}[\gamma_i|\underline{\gamma}_k < \gamma_i < 0]\right) \\
&= \frac{1}{\int_{\underline{\gamma}_k/\sigma_\gamma}^{\overline{\gamma}_k/\sigma_\gamma} \phi(z)dz} \left(\int_0^{\overline{\gamma}_k/\sigma_\gamma} \phi(z)dz \cdot \mathbb{E}[\gamma_i|0 < \gamma_i < \overline{\gamma}_k] - \int_{\underline{\gamma}_k/\sigma_\gamma}^0 \phi(z)dz \cdot \mathbb{E}[\gamma_i|\underline{\gamma}_k < \gamma_i < 0]\right) \\
&= \frac{(\Phi(\overline{\gamma}_k/\sigma_\gamma) - 0.5) \cdot \mathbb{E}[\gamma_i|0 < \gamma_i < \overline{\gamma}_k] - (0.5 - \Phi(\underline{\gamma}_k/\sigma_\gamma)) \cdot \mathbb{E}[\gamma_i|\underline{\gamma}_k < \gamma_i < 0]}{\Phi(\overline{\gamma}_k/\sigma_\gamma) - \Phi(\underline{\gamma}_k/\sigma_\gamma)} \\
&= \sigma_\gamma \left(\frac{2\phi(0) - \phi(\underline{\gamma}_k/\sigma_\gamma) - \phi(\overline{\gamma}_k/\sigma_\gamma)}{\Phi(\overline{\gamma}_k/\sigma_\gamma) - \Phi(\underline{\gamma}_k/\sigma_\gamma)}\right) > 0,
\end{aligned} \tag{B.18}$$

where $pdf_\gamma(\gamma_i) = \frac{1}{\sigma_\gamma}\phi(\gamma_i/\sigma_\gamma)$ since $\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2)$, and the last equality follows from Lemma B.1. □

Proof of Proposition 4. The variance of fund i's perceived active return (PAR) is equal to

$$\mathbb{V}\mathrm{ar}\left[R_{i,2} - \beta_i^k F_{k,2}\right] = \sigma_\alpha^2 + \sigma_\varepsilon^2 \tag{B.19}$$

if the fund adopts a detailed prospectus, and equal to to

$$\mathbb{V}\mathrm{ar}\left[R_{i,2} - b^k F_{k,2}\right] = \sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_{F_k}^2 \gamma_i^2 \tag{B.20}$$

if it adopts a generic prospectus. It follows that the cross-sectional average of the PAR variance associated with detailed prospectuses is equal to $(\sigma_\alpha^2 + \sigma_\varepsilon^2)$, whereas the cross-sectional average of the PAR variance associated with generic prospectuses is equal to

$$\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_{F_k}^2 \mathbb{E}[\gamma_i^2 | \underline{\gamma}_k < \gamma_i < \overline{\gamma}_k] = \sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_{F_k}^2 \sigma_\gamma^2 \left(1 + \frac{(\underline{\gamma}_k/\sigma_\gamma)\phi(\underline{\gamma}_k/\sigma_\gamma) - (\overline{\gamma}_k/\sigma_\gamma)\phi(\overline{\gamma}_k/\sigma_\gamma)}{\Phi(\overline{\gamma}_k/\sigma_\gamma) - \Phi(\underline{\gamma}_k/\sigma_\gamma)}\right), \tag{B.21}$$

where the conditional expectation in the last term is obtained by Lemma B.1. □

Proof of Proposition 5. Fund flows at time 2, associated with fund prospectus p, $q_{i,2}(p)/q_{i,1} - 1$, are equal to

$$flow_{i,2}(p) = \frac{\hat{\alpha}_{i,2}(p) - \bar{\alpha}}{\bar{\alpha} - f}. \tag{B.22}$$

where, based on (12), $\hat{\alpha}_2(p) = (1 - \lambda_p^\alpha)\bar{\alpha} + \lambda_p^\alpha(\hat{r}_{i,2}^A + c_k \cdot q_{i,1} + f)$. 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

$$\frac{\lambda_p^\alpha}{\bar{\alpha} - f}. \tag{B.23}$$

Since $\lambda_d^\alpha \geqslant \lambda_g^\alpha$ for any realization of the factor return $F_{k,2}$, the flow-performance sensitivity is always larger for funds adopting detailed prospectuses. Moreover, since λ_g^α decreases with the absolute value of the factor return $|F_{k,2}|$ (while λ_d^α is independent of it), the difference in flow-performance sensitivity across funds adopting detailed and generic prospectuses within the same mandate increases in $|F_{k,2}|$. □

References

Abis, S. (2020). Man vs. machine: Quantitative and discretionary equity management. Working paper.

Abis, S. and A. Lines (2020). Do mutual funds keep their promises? 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.

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

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

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 (2022). A signaling theory of mutual fund activeness. 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.

Choi, J. J. and A. Z. Robertson (2020). What matters to individual investors? evidence from the horse's mouth. *The Journal of Finance 75*(4), 1965–2020.

Chung, J.-W., B. A. Sensoy, L. Stern, and M. S. Weisbach (2012). Pay for performance from future fund flows: the case of private equity. *Review of Financial Studies 25*(11), 3259–3304.

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.

Evans, R. B. and Y. Sun (2021). Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *The Review of Financial Studies 34*(1), 67–107.

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.

Frazzini, A. and O. A. Lamont (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of financial economics 88*(2), 299–322.

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.

Holden, S., D. Schrass, and M. Bogdan (2020). Characteristics of mutual fund investors, 2020. *Perspective 26*(9), 1–20.

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.

Kacperczyk, M., C. Sialm, and L. Zheng (2008). Unobserved actions of mutual funds. *Review of Financial Studies 21*(6), 2379–2416.

Kaniel, R., Z. Lin, M. Pelger, and S. Van Nieuwerburgh (2022). Machine-learning the skill of mutual fund managers. Technical report, National Bureau of Economic Research.

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 personal. *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.

Krakow, N. J. and T. Schäfer (2020). Mutual funds and risk disclosure: Information content of fund prospectuses. Working Paper 20-54.

Lang, M. and L. Stice-Lawrence (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics 60*(2-3), 110–135.

Pástor, L. and R. F. Stambaugh (2002). Investing in equity mutual funds. *Journal of Financial Economics 63*(3), 351–380.

Reid, B. K. and J. D. Rea (2003). Mutual fund distribution channels and distribution costs. *Perspective 9*(3).

Sensoy, B. A. (2009). Performance evaluation and self-designated benchmark indexes in the mutual fund industry. *Journal of Financial Economics 92*(1), 25–39.

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.

Zhu, Q. (2020). The missing new funds. *Management Science 66*(3), 1193–1204.