Finding Fortune: How Do Institutional Investors Pick Asset Managers?☆

Gregory W. Brown[1], Oleg Gredil[2], Preetesh Kantak[3]

Abstract

We propose and test a framework of private information acquisition and decision timing for asset allocators hiring outside investment managers. Using unique data on due diligence interactions between an allocator and 860 hedge funds, we find that the production of private information complements (substitutes) public information at the intensive (extensive) margin. Our allocator strategically chooses the precision at which to acquire private signals, reducing due diligence time by 58% and improving outcomes. Selected funds outperform unselected funds by 9.0% over 20 months. This outperformance is related to the allocator learning about fund return-to-scale constraints and manager skill before other investors.

Keywords: Asset Management, Decreasing Return to Scale, Due Diligence, Hedge Funds, Information Choice, Institutional Investors, Learning, Private Information

JEL Classification: D81, D83, G11, G23

☆Version Date: November 4, 2019

[1]University of North Carolina, Kenan-Flagler Business School, gregwbrown@unc.edu, (919) 962-9250

[2]Tulane University, Freeman School of Business, ogredil@tulane.edu, (504) 314-7567

[3]Indiana University, Kelley School of Business, pkantak@iu.edu, (812) 855-4806

Helpful comments and suggestions were provided by Aleksandar Andonov (discussant), George Aragon (discussant), Yasser Boualam, Irem Demirci (discussant), Diego Garcia, Eitan Goldman, Philip Howard, Jiekun Huang (discussant), Veronika Pool, Alessandro Previtero, David Smith (discussant), Irina Stefanescu (discussant), Zheng Sun (discussant), Noah Stoffman, and Charles Trzcinka. We would like to thank seminar participants at the 2020 American Finance Association Annual Meeting, 2019 European Finance Association Annual Meeting, the 2018 Wabash River Conference at Notre Dame, the 2018 FIRS Conference, 2016 FMA Annual Meeting, Indiana University–Bloomington, 2017 Institute for Private Capital Spring Symposium, 2017 Nova-BPI Corporate Finance Conference, 2017 Private Equity Research Consortium Roundtable, 2017 SUNY-Albany Symposium, the University of North Carolina–Chapel Hill, 2016 USC Ph.D. Conference. We are also indebted to Mahesh Gopalkrishna for his research assistance. All errors are our own.

1. Introduction

Recent empirical evidence suggests that the primary role of private information collected by consultants on external managers is to provide "hand-holding" services to investors (see, e.g., Jenkinson, Jones, and Martinez, 2015). This limited role, however, belies the effort that some more sophisticated institutional investors, such as larger university endowments, funds-of-funds, sovereign wealth funds, insurance companies, and foundations (henceforth allocators), expend in acquiring their own proprietary information on managers. While we know that an allocator's ultimate investment decisions can be plagued by agency conflicts (see, e.g., Andonov, Hochberg, and Rauh, 2018), relatively little is documented about the actual due diligence process. This paper investigates the tradeoffs an allocator faces during due diligence by proposing an endogenous learning framework, and testing this framework's implications using a unique dataset from a large allocator under conditions in which it is fully incentivized to act using the collected information.

We model a Bayesian updating process whereby the allocator decides optimally when and how much to learn about a manager's skill. The framework yields a number of testable predictions. First, the choice whether to, and the precision with which to, acquire a costly private signal depends in specific ways on the moments of publicly available information—that is, private information complements public information at the intensive margin. Second, both public and private information ratios determine manager selection—that is, private and public information are substitutes at the extensive margin. Third, if the cost of private information acquisition is convex in the amount of information gathered, the allocator finds it optimal to spread acquisition over multiple meetings. Fourth, selected funds with higher private information ratios outperform peers over a longer period of time. A possible channel through which this due diligence adds value is by providing more accurate or precise information on manager skill than is available through public information alone. This value diminishes over time as information on skill becomes evident to other allocators—decreasing returns to scale lead to lower excess returns as fund size grows.

We test these predictions using a proprietary dataset with over $15 billion dedicated to alternative investment strategies. We examine the allocator's due diligence with respect to 860 long-short equity hedge funds from 2005 to 2012. Our dataset captures significant details of allocator-fund interactions, including over 600 pitchbooks prepared by fund managers and 3,000 notes stored in the allocator's internal database. Our empirical proxies are constructed from information contained in these meeting notes (private information), and from merged performance and size information from the allocator and commercial databases (public information). We show that our sample is plausibly representative.

Our empirical results are consistent with the predictions of the framework. The allocator spaces its information acquisition strategically to minimize the cost and maximize the value of due diligence—typically, for selected and unselected funds, the allocator holds 3.7 and 2.2 meetings, respectively. This is despite the similar frequency of initial meetings. In addition, the amount of information acquired during meetings (the precision) is strongly related to past return moments of the fund: a one-standard-deviation-higher excess return results in an approximately 3.5% higher meeting precision (as measured by the length of the meeting note), while a one-standard-deviation-higher return precision (as measured by the standard deviation of rolling excess returns) reduces meeting precision by more than 6.0%. Furthermore, private information precision exhibits an option-like relationship with a manager's past performance. That is, the amount of information collected is concave in the absolute value of excess returns, dropping 22% when excess returns are one standard deviation away from the mean. This implies that uncertainty in publicly available information drives the allocator's scrutiny of a manager. Funds with extremely high (low) Sharpe ratios are less ambiguously good (bad), making private information collection less valuable. Ascertaining the skill of funds with returns closest to the average fund, however, is more difficult. This is when private information derives its greatest value.

We also consider how the allocator uses private information to select funds in its portfolio (i.e., the extensive margin). We find that both private and public information positively predict the conditional probability (hazard rate) of manager selection. A one-standard-deviation-higher quality private (public) signal increases the hazard rate of selection by approximately 200% (40%), while a private signal that is one standard deviation more precise increases the hazard rate by 58%.

We also document important time variation in the impact of public and private information on selection. Generally, public (private) information is more important for manager selection earlier (later) in the process. This is consistent with our interpretation of the role of private information at the intensive and extensive margins. Uninformative returns drive the need for private information in the selection decision, but due to the cost, information acquisition requires time. The opposite is true when returns are more informative of skill: the allocator is able to make decisions that are both faster and less reliant on private information.

We next examine whether the ability to make informed decisions sooner adds value to the allocator. Pástor, Stambaugh, and Taylor (2015), henceforth PST, show that mutual fund manager outperformance decays from inception and they link this decay to growth in industry AUM.[1] Using this basic intuition, we show that cumulative benchmark-adjusted excess returns are 9.0% higher for the average selected versus matched (unselected) fund. Furthermore, we find that the outperformance of a skilled manager dissipates as information on skill is revealed in higher realized returns, implying decreasing returns to scale (see, inter alia, Berk and Green, 2004; Pástor and Stambaugh, 2012). Interestingly, the period over which selected funds outperform (20 months) is of the same order of magnitude as the reduction in due diligence time from higher levels of private information (17 months). We also find that nearly 65% of the post selection outperformance is driven by funds for which the public signal is ambiguous but the private signal strong on the selection date. Complementing the work of Gerken, Starks, and Yates (2018), who study the role of learning about manager reputation through investments in previous funds, our results show the value of a more conscious and proactive acquisition of private information.

In our framework, which is an infinite-horizon extension of that of Hermalin and Weisbach (1998), the allocator has an option to sequentially acquire costly private information on manager quality after observing public information at each moment in time. The allocator's problem stems from information asymmetry present in principal-agent relationships, whereby allocators are un-

[1] Zhu (2018) shows that, with stronger instruments, PST-like test also finds a negative return-to-scale effect at the individual fund AUM level.

able to determine the manager type. While pooling can arise for a variety of reasons (see, e.g., Admati and Pfleiderer, 1997), the difficulty of identifying good managers using public information alone is well documented (see, e.g., Korteweg and Sorensen, 2017). Furthermore, waiting to learn a fund's type using only public information increases the probability that other asset allocators will identify the manager's skill, diminishing the opportunity to earn excess returns. Berk and Green (2004) and Pástor and Stambaugh (2012) show that this equilibrium exists if there is a declining return to scale, an idea supported by empirical evidence (see, inter alia, Chen, Hong, Huang, and Kubik, 2004; Pástor, Stambaugh, and Taylor, 2015). Our analysis also builds on recent work by Gârleanu and Pedersen (2018), who show that endogenous private information acquisition in a general equilibrium framework can result in outperformance of informed allocators, and research by Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016), who show that information acquisition constraints affect investment allocations.

We make several important contributions to the literature. Prior research shows that allocators hire and fire managers based on past excess returns and that they monitor managers better than retail investors (see Evans and Fahlenbrach, 2012; Goyal and Wahal, 2008, among others), but little is known about the relative importance of other decision drivers. While Jones and Martinez (2017) highlight how external research impacts flows, and Roussanov, Ruan, and Wei (2018) highlight the role of marketing in fund selection, our paper studies the impact of in-house research, allowing us to evaluate the interplay between public and private information in manager selection. To the best of our knowledge, our allocator also did not employ external advisers for due diligence on prospective managers. We argue that the context of hedge fund manager selection is attractive for tests of our hypotheses because there are strong incentives both for the allocator to act quickly in identifying a good manager and for inferior managers to pool with the broader manager population (see Agarwal, Mullally, and Naik, 2015, for a recent review of the hedge fund literature). We mitigate concerns about heterogeneity in risk by normalizing returns using the performance of peers (Jagannathan, Malakhov, and Novikov, 2010; Pástor, Stambaugh, and Taylor, 2015) and by focusing on the equity long-short strategy to limit unobserved heterogeneity in fund strategy and allocator preferences. Discussions with institutional investors suggest that our results represent a good approximation of their processes and generalize beyond a single allocator and hedge funds.

By establishing the link between (i) due diligence pace, (ii) AUM growth and (iii) excess returns in a large sample of hedge funds, we contribute to the debate about the decreasing return to scale (DRS) in the alternative investments space. While Yin (2016) shows that hedge funds contracts do not preclude DRS in theory, he finds little empirical support for DRS effects outside a few strategies (e.g., 'Global Macro'). Our results pertain to a broad group of long-short equity funds and explain why observable return-to-asset dynamics might be weak—sophisticated institutional investors bankroll good managers *before* their skills are reliably observed in returns. This appears to be an important and previously undocumented element of Berk and Green-like equilibria. Of course, our results do not rule out other channels for performance erosion (see, e.g., Hoberg, Kumar, and Prabhala, 2017; Adams, Hayunga, and Mansi, 2018; Gupta and Sachdeva, 2019).

Our analysis also contributes to the literature on the value of manager recommendations (see, e.g., Bergstresser, Chalmers, and Tufano, 2009; Gennaioli, Shleifer, and Vishny, 2015). Our results relate closely to those of Jenkinson, Jones, and Martinez (2015), who focus on the performance of consulting firm (i.e., external) recommendations. While they show a prominent role for private information acquisition (see Kaniel and Parham, 2017; Benamar, Foucault, and Vega, 2019, for further evidence), they also show a weakly negative post recommendation alpha. This implies that private information offers little pecuniary benefit to the end investor. This begs the question: why do we find the opposite? First, our analysis examines the allocator's in-house process directly. The source of recommendation in their database (external) may add principal-agent tensions not present in ours (internal) such as fiduciary boards or differing sophistication of the investment processes (Andonov, Hochberg, and Rauh, 2018). Second, we find that the "shelf-life" of value generated by private information is of the approximate order of the frequency (annual) of their data. The higher frequency (monthly) of our data may thus allow better identification of the connection between private information and performance.

The paper proceeds as follows. Section 2 describes our theoretical framework and develops our predictions. Section 3 describes our unique data. Section 4 reports the main empirical results and additional tests, and section 5 concludes.

2. Allocator as a Bayesian Learner

This section develops a partial equilibrium framework of our allocator as a Bayesian learner. To motivate the role of private information in manager selection, consider Figure 1, Panel A. The histogram compares abnormal return estimates during the due diligence period for funds that were selected by the allocator with a matched sample of those that were not. While the distributions are statistically different, their overlap is substantial. In Appendix A, we review anecdotal evidence based on interviews with our allocator's senior investment professionals that also suggest that the allocator finds it difficult to select funds using historical returns alone. The need to separate good from bad funds under uncertainty about a manager skill drives the allocator to acquire private information. The propositions underlying our predictions are provided in Appendix B.

2.1. Acquisition and Uses of Private Information

An allocator is looking for a fund that can generate positive excess returns (henceforth α), but the fund's α is unknown. It has two sources of information through which it can learn about the fund's skill: one is a passive and costless public signal ($x \sim N(\alpha, 1/r)$), and the other a choice variable and costly private signal ($y \sim N(\hat{\alpha}, 1/s)$). To quote our allocator: "We want to know the basic premise for how they make money, how [this] compares to historical results, and if they are realistic in their assessment of performance."

Thus, the allocator first updates its diffuse prior with the fund's past returns to obtain an expectation ($\hat{\alpha}$) and variance ($1/\tau$) of α. The allocator then has three choices: to (i) do nothing, (ii) select the fund, or (iii) meet with the fund, acquiring another signal about skill, y. This is represented as a choice set, $V = \max\{0, V_{select}, V_{meet}\}$, for every period. Additionally, were the allocator to meet with the fund, it chooses an ex-ante intensity with which to meet. We assume that the precision of signal collected, s, increases in that intensity. In our empirical tests, we proxy for precision using the number of words (i.e., amount of information) collected. In section 3 we empirically validate that the amount of information is linked to meeting precision by analyzing the content in the allocator-fund meeting notes.

The value of doing nothing is normalized to zero in the one period case. Assuming a profit maximizing, risk neutral allocator, the value of selection (V_{select}) is a linear function of a single period estimate of α (i.e., $V_{select}(\hat{\alpha}) = -K + A\hat{\alpha}$). K and A represent the present value of monitoring costs and a leverage coefficient capturing the persistence of α were the allocator to select the fund, respectively. We assume that $K << A$. If this were not the case a fund would rarely if ever provide enough pecuniary value to justify an investment. The expected value of meeting (V_{meet}) is a function of the decision to do nothing or select in the subsequent period and the ex-ante choice of meeting precision,

$$V_{meet} = \max_{s} -c(s) + \beta \int_{-\infty}^{\infty} \max\{0, -K + A \cdot \frac{\tau \cdot \hat{\alpha} + s \cdot y}{\tau + s}\} \cdot f(y|\hat{\alpha}, \tau, s)\, dy \;. \tag{1}$$

The expression $c(s)$ is the cost of meeting and has a fixed portion and variable portion, which is convex in s, such that $c'(s), c''(s) > 0$. This captures a basic intuition: spending too much time with a single fund takes away valuable time from researching other possible investment opportunities. To quote our allocator: "[Meetings] take a large amount of our time and resources. The more senior people focus almost all of their efforts on understanding the people, process, and philosophy." Finally, the conditional probability density function, $f(\cdot)$, describes the distribution of the private signal about the fund's conditional α:

$$f(y|\hat{\alpha}, \tau, s) = \sqrt{\frac{H}{2\pi}} \exp\left((H/2)(y - \hat{\alpha})^2\right) \;, \tag{2}$$

where $H = \frac{s \cdot \tau}{s + \tau}$, is precision of y given $\hat{\alpha}$, τ, and s.

The maximization within the indefinite integral captures the key intuition of the model: meetings add value in expectation because they truncate a portion of the distribution of bad fund types. In particular, the integral in (1) can be represented as

$$V_{meet} = \max_{s} -c(s) + \beta \int_{Y_c(s)}^{\infty} \left(-K + A \cdot \frac{\tau \cdot \hat{\alpha} + s \cdot y}{\tau + s}\right) \cdot f(y|\hat{\alpha}, \tau, s)\, dy, \tag{3}$$

where $Y_c(s) = \frac{K \cdot (\tau + s)}{A \cdot s} - \frac{\tau \cdot \hat{\alpha}}{s}$. This reveals that both the choice to meet and intensity with which to meet are inextricably linked to the benefit of private information. If expected returns are ambiguous enough, the allocator will extract a net benefit from collecting additional information on the fund. Conditional on having a meeting, the more ambiguous a fund's returns the more the allocator can justify greater information acquisition. Assuming that K/A is small, the benefit of private information is that the allocator can better identify "right-tail" or bad-type funds—that is, $Y_c(s)$ is an increasing function of s. Given this one-period model, we can characterize the relationship between the moments of past returns and the amount of information acquired—see Figure 2.

2.2. *Timing of Investment*

Our allocator, however, views information acquisition as a multi-period, rather than single period, process. To quote our allocator: "[We] prefer to have shorter meetings to digest what [we've] learned and to determine if there [will be] another meeting." We therefore extend the framework by writing the choice set (i.e., do nothing, acquire private information, or select) as a dynamic programming problem.

$$V(\hat{\alpha}) = max\{V_{public}, V_{select}(\hat{\alpha}), -c(s) + \beta E\left(V'(M(\hat{\alpha}))\right)\}, \tag{4}$$

where $V'(M(\hat{\alpha}))$ is the continuation value of having a meeting and $M(\hat{\alpha})$ is the pre-posterior expected value. Additionally, an important distinction between the multi- and single-period problem is the substitution of 0 with V_{public} for the "do nothing" choice. Given that public information in our setup is costless, V_{public} represents the allocator's expected value from informing her decision using only public information going forward. This highlights that our allocator will never outright reject a fund given a costless source of information. This extension leads to several predictions.

First, given the optimality condition in equation 3 within a multi-period context, it is optimal for the allocator to split total information acquisition over time. This allows the allocator to more effectively synthesize information and utilize its resources. Within the theoretical framework, this is due to the convexity of the cost function, $c(s)$. By splitting the total information required to make a decision over multiple meetings, the allocator can effectively reduce the total cost of due-diligence. Given that the amount of information acquired on a fund is the interaction of the number of meetings with the per meeting amount of information acquired, the next prediction follows from the dynamic extension.

Prediction 1: Both the number of meetings and the amount of private information collected per meeting

(i) decreases when fund returns are precise,
(ii) decreases when the cross section of returns is precise,
(iii) is concave in the absolute value of returns around the fund's mean, and
(iv) increases in returns, primarily when fund return precision is high.

The intuition contained in the infinite period setup also provides a natural empirical proxy for the ex-post signal level (or quality), y, for a particular meeting. If the signal is extremely strong (weak) the next meeting will occur after a fewer (greater) number of periods. This is because ex-post a high (low) signal versus the prior estimate of α would suggest that more (less) information must be acquired going forward. In section 3.4, we exploit both pieces of information–the amount of information collected at each meeting and the periodicity between meetings–to generate an information ratio of private information.

Second, given only the available public data at the beginning of due-diligence, the allocator will have an estimated stopping point for further private information acquisition. The allocator expects that after this point the benefits of acquiring additional information will not be greater than the cost of meeting. Additionally, before this point, there are three zones across $\hat{\alpha}$, representing each of the choices in the set. As we proceed forward in due-diligence the decision to select a fund will therefore be driven by the updating of both private and public information.

Prediction 2: The information ratios of both public and private information predict selection.

Since it is optimal for the allocator to spread due diligence over multiple meetings, private information takes time to acquire. This implies that the conditional probability of investment (hazard rate) is endogenously hump-shaped in time.

Prediction 3: The hazard of investment is concave in due diligence time, initially increasing, peaking, and then decreasing to zero. Furthermore, the degree of concavity is increasing in the private-information ratio.

Furthermore, the allocator focuses not only on the α generated by the fund in a single-period, but also its persistence. This captured by a leverage coefficient, A, in the continuation value.

Prediction 4: Selected funds' excess returns will not only be higher on average at selection date, but also persistently positive for longer.

Although beyond the scope of our model, the amount of time over which the fund generates alpha will likely depend on its return-to-scale relationship. The smaller a fund is versus its full-information equilibrium size–where zero alpha will be generated–the longer one would expect is its positive alpha to last (i.e., in a general equilibrium context, A would be an increasing function of this AUM spread).

3. Data and Descriptive Statistics

In this section we describe the database that we use to conduct our empirical tests. Our analysis considers a single large institutional investor (our allocator) and covers the period June 2005 to June 2012. During this time our allocator managed more than $15 billion in assets targeting smaller institutional investors as the clients with no other services but alternative assets. We have entered into nondisclosure agreements that prevent us from revealing identifying information about the allocator and the specific funds in our dataset. We discuss the implications of our sample construction for interpretation of the empirical tests in the end of this section.

3.1. General Description

Our dataset contains the history of interactions between the allocator and 860 long-short equity hedge funds. The long-short portfolio for which these hedge funds were candidates averaged over $2 billion under management during the sample period. Each interaction (henceforth, meeting) is characterized by a date, type (phone call, meeting at a conference, on-site visit, etc.), list of participants, and related documents. The documents typically include a pitchbook prepared by the fund and meeting reports (henceforth, notes) written by the allocator's employees. The database also contains fund characteristics such as the education and previous professional affiliations of fund manager employees. The contents of the pitchbook, meeting notes, and other proprietary fund characteristics are considered private information that is acquired by expending resources.

We consider return and AUM history to be public data because we are able to supplement and cross-verify more than 90% of both variables using a combined database of hedge funds from Morningstar, Barclay Hedge, eVestment, HFR, and Lipper-TASS databases from 1990 through 2017 (see, e.g., Agarwal, Daniel, and Naik, 2009; Agarwal, Mullally, and Naik, 2015, for a description). We match funds by first limiting prospective matches to return correlations of ≥ 0.95. We then compare fund and firm names by both a fuzzy matching algorithm and by-hand verification. When multiple funds are associated with a given firm name across these databases, monthly returns are assumed to be the equal-weighted average.

From the allocator, we also obtain a complete history of due diligence stages for each fund, as well as the amounts invested. Most importantly, the stage code 'selection to investment universe' allows us to decouple the selection process from the actual investment. This is unique to our database and critical for pursuing our research question because not all selected funds receive investments immediately. According to the allocator, this has nothing to do with the perceived quality of the fund, but it rather reflects the allocator's portfolio constraints. For example, the allocator may already be fully invested or at its desired risk levels, or it may be experiencing net outflow of capital under management. In these cases the allocator would select, but not invest in, the fund. Given the single allocator-fund focus of our model and the resulting predictions, the fund due diligence status allows us to focus on the manager-related private information acquisition process and its pecuniary value.

The allocator selected 214 funds over the 8-year sample period, of which 114 received investments. The allocator's inflows were on average 13% higher in months when the allocator invested, whereas inflows were on average 2% lower in months when the allocator selected funds but did not invest. In Table 1, we present the summary statistics for our public information variables; Table 2 reports differences in summary statistics between selected and unselected funds.

3.2. Meetings and Pitchbooks

The meeting notes provide a largely objective picture of the topics explored during the interaction. For many of the funds we examine there are more than five preinvestment decision notes in the database. These meeting notes, along with the dates and attendees, are stored by the allocator.

Each meeting is given an interaction (i.e., meeting type) and stage code. We aggregate meeting interaction codes into three categories: informal (conference or email), semiformal (call or face-to-face meeting), and formal (on-site meeting). The stage code provides a snapshot of where in the due diligence process the allocator is at that point—for instance, preliminary screening, first step, selection, or investment. In Table 3 we present summary statistics between selected and unselected funds using data from these notes—the quantity (measured by the number and type of meetings and the number of meeting words) and quality (measured by the period between meetings (periodicity)) of meetings are statistically different during the preselection period.

In Table 3, Panel B, we also report summary statistics for start-up funds. The process involved in due diligence for start-up funds—where a fund's public data is the least informative—illustrates in its most pure form how private information informs selection. Of the 860 funds, about 10% are start-ups. Of these start-ups, roughly 26% were selected. This is not statistically different from the selection rate for the rest of the sample. However, if we condition on whether there was a prior investment relationship with the manager (e.g., the fund is a spin-off from a fund in which the allocator has already invested), the probability of selection jumps to a statistically different 50%.

Figure 3, Panel A, depicts the information reported in Table 3. Selected funds have a larger proportion of formal meetings (versus semiformal or informal meetings) than do unselected funds, even by the second and third meetings. In other words, the formality of the meetings increases more rapidly for selected versus unselected funds during the due diligence process. While not significant for the first meeting, the difference in means of meeting type and periodicity between selected and unselected funds is statistically significant over the entire due diligence period.

Figure 3, Panel B, reports the average number of meetings per month during the first and last nine months of due diligence for both selected and unselected funds on a three-month rolling basis. The meetings are more frequent earlier in the due diligence period. For both selected and unselected funds, the allocator meets with the fund roughly 0.4 times per month for the first three months of due diligence. The frequency of meetings falls for both sets of funds as time progresses, but much less so for the selected funds—starting with the sixth month, the meeting frequencies differ statistically between the two groups of funds.

We also examine the contents of the notes and pitchbooks using the financial word lists of Loughran and McDonald (2011), focusing on positive, negative, and uncertain proportions (see e.g., Loughran and McDonald, 2011; García, 2013). Details of this analysis are provided in online appendix IA.1. This analysis yields statistical differences in content between selected and unselected funds, but it is difficult to judge what the counts convey. In the next section, we capture this context by utilizing a machine-learning algorithm.

3.3. Meeting Notes Content and Context

This section provides evidence that the meetings constitute proactive acquisitions of private information by our allocator. We accomplish this goal by objectively measuring the temporal evolution of topics highlighted in the anecdotal evidence presented in Appendix A and validating the private information empirical proxies developed in section 3.4. The online appendix IA.2 details the machine learning method (Latent Dirichlet Allocation, or LDA, and Bayesian inference) used to determine topics discussed in the meeting notes. The algorithm exploits topical heterogeneity in the documents to endogenously create lists (topics) of frequently co-occurring words in a corpus of text (see, e.g., Hanley and Hoberg, 2018, for a recent application of the LDA method in finance).

We apply three standard filters to our data. First, although pitchbooks are statistically similar, their more structured content helps the algorithm decipher major themes discussed in the meeting notes. To assist the algorithm, we split each pitchbook into four sections: employee backgrounds, investment process, risk management, and performance. We then standardize the language used in the corpus, which reduces the size of the vocabulary over which the LDA learns. First, we spell out commonly used contractions (e.g., *don't* becomes *do not*) and acronyms (e.g., *GDP* becomes *gross domestic product*). Second, we lemmatize inflections and derivationally related word forms to a common base (Bird, Klein, and Loper, 2009). Third, we remove commonly occurring "stop" words (e.g., prepositions). Finally, we filter out words that appear too frequently (in $\geq 50\%$ of documents) or not frequently enough (in < 15 documents) and remove documents with fewer than 3 content words to generate our modified corpus.

We run the LDA over 30 topics. This number was chosen in order to maximize an average-topic fit measure: the topic coherence score described in online appendix IA.2. The coherence score codifies the underlying concept of the LDA, which is to decipher the most probable words in a topic by analyzing how often they occur in the same document. Based on the words that the LDA allocates to each topic, we assign titles to topics. Topics with coherence scores below -2.5 and those to which titles are difficult to ascribe are considered noise. Table 4 lists the remaining 23 topics, their inferred topic titles, and their coherence scores (CS).

We analyze the time variation in topics along two dimensions. First, we compute the weighted-average meeting number for each topic,

$$\text{AvgMeet}_j = \sum_{n=1}^{\#\text{Firms}} \sum_{i=1}^{\text{MaxMeet}_n} i \cdot \left[\frac{\text{Wrds}_{j,i,n}}{\text{TotalWrds}_j} \right] . \tag{5}$$

For a given topic, each meeting number is given a weight equal to the proportion of LDA-allocated words in that topic and appearing within the note associated with that meeting.[2]

To demonstrate the temporal evolution, we next use information gleaned during our interviews with the allocator to place the topics from Table 4 into three categories: early, middle, and late. The allocator stated that each meeting typically covers a different set of topics; this allows it to retain the option to end due diligence early, strategically conserving constrained resources. Early topics were generally associated with conversations about employee and firm background or the launch details of the fund. Middle topics were largely related to process, performance, and organization. Late topics discussed philosophy—specifically the managers' thoughts on opportunities in particular areas of the world or industries. We compute the weighted average statistics for the three topic categories (Figure 4, Panel A). The early, middle, and late category topics are on average discussed during meetings 3.9, 4.5, and 4.8, respectively. These mean meeting numbers are significantly different from one another, as demonstrated by their non-overlapping 95% confidence intervals. The histograms of scaled frequency, however, show that topics still overlap significantly across

[2]Figure D.3 in internet appendix presents the weighted average and standard deviation for all topics. Dark (light) grey symbols represent topics with coherence scores above (below) -2.5. Topic numbers correspond to those in Table 4. Variation in average meeting number across topics suggests significant temporal variation in topics covered during due diligence and comports with our allocator's preference for multiple shorter (rather than single longer) meetings.

meetings. In particular, the distribution of topics seems to shift from early to late topics, while the middle topics are distributed evenly over meetings. Finally, late (early) meeting notes are on average longer (shorter), driving up the weighted average meeting number across all topics.

As reported in Table 3, the average number of meetings for selected and unselected funds is 3.72 and 2.49, respectively. Given the high average meeting number for the early, middle, and late topic categories and the relatively low average number of meetings with funds overall, one potential concern is that the notes are tainted by reverse causality—that is, they already reflect the decision to select. To explore this possibility we analyze the evolution of breadth of information collected by our allocator through the due diligence process. Conditional on the meeting number, if reverse causality was driving note content, we would expect there to be significant difference in the breadth of information collected between selected and unselected funds. This corroborates evidence from Figure 3 Panel A which shows that the allocator continues to hold costly "formal meetings" even with the funds it never ends up selecting.

To test for reverse causality, we compute a rolling Kullback-Leibler divergence measure for each fund-meeting:

$$KL_{i,j} = \sum_{n=1}^{\text{NumTopics}} p\left(x_{i,j,n}\right) \cdot \left(\log p\left(x_{i,j,n}\right) - \log q\left(x_{i,\bar{j},n}\right)\right) \quad , \tag{6}$$

where $p\left(x_{i,j,n}\right)$ is the LDA-estimated topic proportion for each meeting note j with fund i, and $q\left(x_{i,\bar{j},n}\right)$ is a reference probability. For meetings after the first, the reference is the rolling topic proportion over all meetings before the jth ($\equiv \bar{j}$). For the first meeting, the reference is the average distribution for all selected start-up funds. This measure captures the degree of dissimilarity between a given meeting's note and the stock of information collected by the allocator up to that point. Figure 4, Panel B, presents the average and standard error of this measure for the notes conditioned on meeting number. This analysis reveals that the breadth of content between selected and unselected funds is statistically very similar, indicating that reverse causality is unlikely. If the allocator is meeting with the fund, it still has information to collect in order to make their decision.[3]

[3] In addition, the high Kullback-Leibler measure indicates that the information collected (flow) is more dissimilar to the stock early in due diligence, but becomes increasingly similar as due diligence proceeds. This corroborates the

In other words, the textual data is inconsistent with a conjecture that the allocator's meetings reflect a preordained decision to select a given fund.

3.4. Private Information Metrics

The content analysis in sections 3.2 and 3.3 informs our choice of proxies for the private information signal and precision. For our measure of meeting i precision ($s_{i,t}$), we use the log number of words in each note. This is tied to what we observe in section 3.2 and to the motivation behind our theoretical framework: the intensity with which the allocator meets should be reflected in the quantity of information captured.

For our measure of meeting signal quality, we exploit the heterogeneity in meeting frequency between selected and unselected funds. First, each meeting is given the value of the negative of the log of the number of months to the next meeting (henceforth $per_{i,t}$). This captures two pieces of intuition. First, if a meeting $i+1$ happens shortly after meeting i, information collected at i was likely of high quality. Second, the timing conforms with that described by our allocator: it decides whether to have a meeting $i+1$ with the candidate fund based on the quality of meeting i. The signal quality, however, is not known until after the choice to have a meeting has been made. For months in which there are multiple meetings, both the quantity and quality measures are scaled by the number of meetings.

Both $s_{i,t}$ and $per_{i,t}$, however, contain a likely common factor. If uncertainty during a meeting is high, we would expect both the number of months to the next meeting and the length of the meeting notes to be high. A positive correlation could confound the economic significance of the regression results in section 4. We thus orthogonalize the measure via the following regression:

$$per_{i,t} = \beta \cdot s_{i,t} + \varepsilon_{i,t} \quad . \tag{7}$$

Given the large number of meetings with low word counts (see Figure D.1), we follow Rigobon and Stoker (2007) in assuming that our regressor, $s_{i,t}$, is randomly censored. We find a statistically

prediction of our theoretical framework that the allocator strategically spreads information acquisition over multiple meetings.

significant β coefficient of 0.47 (t-statistic of 9.08). Our measure of private information signal quality ($y_{i,t}$) is then the error ($\varepsilon_{i,t}$) from the regression.

For predictions 2, 3, and 4, our private information variables of interest are related to an amalgamation of past meeting signals ($\hat{\alpha}_t$) and precisions (s_t). We follow a normal-normal Bayesian updating framework to construct these variables. On the dates t when there are meetings these variables are updated according to

$$\hat{\alpha}_{t+1} = \frac{s_{i,t} \cdot y_{i,t} + \hat{\alpha}_t \cdot s_t}{s_t + s_{i,t}} \tag{8}$$

$$s_{t+1} = \log\left(\sum_{i=1}^{\text{\# Meets up to t}} \exp\left(s_{i,t}\right)\right). \tag{9}$$

s_{t+1} can be interpreted as a stock of information. The greater the number of meetings and the higher their intensity, the more precise the signal of a manager's skill. $\hat{\alpha}_t$ can be interpreted as a weighted average of signals from meetings.[4]

3.5. *A Representative Sample?*

It is important to know if our allocator is exceptional in some way that would bias our results or preclude them from generalizing to a broader set of allocators. To this end we conduct a comparative analysis. First, our allocator is not a performance outlier. The performance was slightly above the median during our sample period—the average percentile rank against similarly focused fund-of-funds in the HFR database during the sample period is 57% with a standard deviation of approximately 3%. Some, but not all of the difference between this rank and the 9% add-value of private information we estimate is from management and incentive fees–1% and 10%, respectively–for our allocator. Likewise, not all funds in the portfolio are new additions earning high excess returns.

Second, the hedge funds on which the allocator performed due diligence in our sample represents a significant fraction of the broader investable funds in the long-short universe. Based on

[4]Equations 8 and 9 confer the additional benefit of downplaying the large portion of notes with between 5 and 25 words. Per our allocator, meetings with such short notes merely maintain engagement with a manager of potential interest rather contribute to due diligence.

the union over the five datasets described in section 3.1, the allocator met with 860 out of 1,500 managers that had at least one fund in the equity long-short strategy space and at least $50 million in AUM. Table 5 reports statistics similar to that presented within Figure 1, but now comparing performance of managers at three decision boundaries: whether to (i) start due-diligence on the fund, (ii) select the fund as eligible for investment, and (iii) invest in a fund from the selected subset. The boundaries are defined according to whether the allocator made a positive (Yes) or negative (No) decision. Each fund where the allocator's decision was positive is matched to a group of 3 peer funds where the allocator's decision was negative. These peer funds are matched according to the Mahalanobis distance calculated using a fund's log(AUM), age, and past information ratio as of the calendar month of the positive decision. Tests are conducted at each boundary, comparing either the 24-month forward Jensen's Alpha or Sharpe ratio (24-month return over the risk-free rate divided by 24-month volatility) between funds with a positive decision and the three controls.

Comparing the difference in returns between firms at the first boundary (the choice to start due-diligence) suggests that the allocator tends to initialize engagement with slightly worse, not better, performing mangers; the 2-year average difference in monthly alpha is *negative* 20bps. This could be a function of two issues worth highlighting. As discussed in Agarwal, Mullally, and Naik (2015), there is the potential for backfill bias in commercial databases which is likely to result in a positively biased returns inferred from these databases. Also, we have no information other than reported returns and AUM on the funds the allocator did not meet. It is possible that many were successful funds no longer open for new investments. Interestingly, when adjusting returns for risk (i.e., the Sharpe ratio), the spread between fund returns at this boundary are no longer statistically significant. The difference in fund returns for selected versus unselected managers is a *positive* 32bps. Given the +60bps difference in returns between both boundaries it seems unlikely that luck-driven sorting of managers into the allocator's due diligence sample explains the difference in outcomes we find for the 860 funds within our main sample.

Finally, it is worth reiterating that our main variable of interest is not the actual investment, but the *selection* of a fund into the "cleared for investment"-set. While this distinction alleviates concerns about confounding effects of access rationing and capital constraints (discussed in section 3), there could be systematic differences between the two subcategories that muddle our interpretations. In Table 5, we therefore compare returns of selected funds that received actual investments with those that did not. The analysis reveals an economically small and statistically insignificant difference of 1bp. Thus, any issues restricting the allocator's ability to actually deploy capital to the fund do not meaningfully correlate with the allocator's due-diligence on a manager's skill. Interestingly, the difference between invested (38bps) and selected (16bps) funds across rows is rather large. This is likely driven by the capital constraints—the allocator is more likely to actually invest when her own investment capital is abundant, which likely corresponds to a favorable market environment and higher subsequent excess returns realized by hedge funds.[5] This result additionally supports using hedge fund benchmark-adjusted returns (rather than rolling alpha estimate) and matching selected to unselected funds by calendar month in our main empirical tests.

4. Empirical Tests

Our main results cover three areas: (i) the acquisition of private information (prediction 1), (ii) ways in which the moments of private information and public information influence the allocator's selection decisions (predictions 2 and 3), and (iii) the assessment of whether private information is informative of excess returns (prediction 4). We also provide evidence of the underlying mechanism through which private information generates value: by providing the allocator a sense of a fund's true returns to scale before other allocators can discover this information (see Berk and Green, 2004; Pástor and Stambaugh, 2012).

[5]Comparisons across rows in Table 5 should be made only with respect to the differences ('Yes' minus 'No'). For example, returns for 'Selected for Investment?'-'Yes' and 'Actually Invested'-'Yes' are not equal because their corresponding decisions were made under different market conditions and allocator portfolio constraints. Also, the performance of 'Actually Invested'-'No' funds is measured as of the investment date in the closest matching 'Actually Invested'-'Yes' funds. So the two do not add-up to 'Selected for Investment?'-'Yes'.

4.1. *Acquisition of Private Information*

Prediction 1 states that the amount of information acquired during a meeting is decreasing in the public information precision (r), decreasing in the prior's precision (τ_0), increasing linearly in the signal (x), and concave in the absolute value of the signal around its mean. To test this prediction we estimate the following regression for each meeting m with a given manager i at due diligence duration t,

$$s_{m,i,t} = \mu + \beta_x \cdot x_{i,t} + \beta_r \cdot r_{i,t} + \beta_\tau \cdot \tau_{0,i,t} + \beta_{x^2} \cdot x_{i,t}^2 + \varepsilon_{m,i,t}. \tag{10}$$

Our proxies for $x_{i,t}$ and $r_{i,t}$ are the past 24-month average and the precision (inverse variance) of peer-adjusted returns, respectively. A good proxy for $\tau_{0,i,t}$ is the inverse of the cross-sectional variance of fund returns for any given month. In Figure 2, Panel D, we present the time series of this measure from January 1, 2000 through June 30, 2012. The shaded regions correspond to the aftermath of the Asian financial crisis, the dot-com bust, the financial crisis, and the European sovereign debt crisis. To limit the confounding of cross-sectional ($\tau_{0,i,t}$) with time-series ($r_{i,t}$) effects, we use an indicator variable for the prior precision that is set equal to one (zero) during nonshaded (shaded) periods. In order to better understand magnitudes, all continuous independent variables are standardized by their mean and standard deviation. The dependent variable, $s_{m,i,t}$, on the other hand, is only de-meaned. As $s_{m,i,t}$ represents the log of the number of words in a note, regression coefficients represent the percentage change in private information acquisition from a one-standard-deviation change in the regressor.

We report our estimation results in Table 6. In columns 1 and 2, we see that private information collected is negatively (positively) driven by the precision (level) of a fund's idiosyncratic returns. In column 3 we add a signal squared term, revealing a negative coefficient. This demonstrates the concave relationship between the amount of private information acquired and the fund's Sharpe ratio. In column 4, we include the interaction term $x_{i,t} \cdot \tau_{0,i,t}$. We would expect the allocator to put higher weight on the signal informing their decision to meet if the prior precision is less diffuse. This intuition is consistent with the results, although the significance of the coefficient on the stand-alone variable is now insignificant. A Wald test of both variables equaling zero, however, is still strictly rejected (reported at the bottom of Table 6). In column 5, we add prior fund-affiliation dummies (i.e., an indicator whether the fund manager has previous affiliations with the allocator), meeting number fixed effects, and standard errors clustered by fund and obtain similar results.

4.2. Decision to Meet

Prediction 1 suggests that the allocator has two dimensions along which to acquire private information: intensity of each meeting (tested in section 4.1) and number of meetings. The choice between these two dimensions is determined by a tradeoff between the fixed cost of a meeting and the convex cost related to spending too much time with a manager. To fully test prediction 1, we also regress the total number of meetings for each fund onto the average of each of our public information proxies across the entire due diligence period,

$$MeetNo_i = \mu + \beta_x \cdot \overline{x_{i,t}} + \beta_r \cdot \overline{r_{i,t}} + \beta_\tau \cdot \overline{\tau_{0,i,t}} + \beta_{x^2} \cdot \overline{x_{i,t}^2} + \varepsilon_i. \tag{11}$$

Our proxies again for $x_{m,t}$ and $r_{m,t}$ are the past 24-month average and the precision (inverse variance) of peer-adjusted returns, respectively. We use an indicator variable for the prior precision that is set equal to one (zero) during nonshaded (shaded) periods in Figure 2, Panel D. We present the results of this regression in Table 7. All independent variables are first averaged within a firm over the due diligence period. These variables are then standardized by their cross-sectional mean and standard deviation. Columns 1 and 2 show that if both the prior-precision and public-information signals are high, the number of meetings is lower. However, in column 3 we see that the concave signal term is not significant; this may be a function of our averaging procedure. In columns 4 and 5 we add the interaction term $\overline{x_{i,t} \cdot \tau_{0,i,t}}$. This variable has a positive relationship with the number of meetings, implying that the allocator puts more weight on the public information signal when the market environment is relatively calm.

The results from regression 11 provide additional support for our private information proxies. The allocator benefits from splitting the acquisition of private information over multiple meetings. Given that the assessment of a fund's skill is updated after each meeting, the number of periods between two meetings (periodicity) should be informative of the quality of the information received during the prior meeting. This is codified in our proxies for the private information signal and precision (equations 8 and 9). Next we show that while private information is a complement to public information at the intensive margin (i.e., the quantity of information collected), they are substitutes at the extensive margin (i.e., the selection decision).

4.3. Decision to Invest

Prediction 2 implies that the probability to select a fund for investment is increasing in the public and private information ratios. We begin with the partial-likelihood hazard model of Cox (1972) in which the conditional probability of selection at time t, given that selection has not yet occurred, (the hazard rate) is modeled as[6]

$$\lambda(t|X_{i,t}) = \lambda_0(t)\, exp\left(\beta_x \cdot x_{i,t} + \beta_r \cdot r_{i,t} + \beta_\alpha \hat{\alpha}_{i,t} + \beta_s s_{i,t}\right). \tag{12}$$

Our proxies for $x_{i,t}$ and $r_{i,t}$ are the past 24-month average and the precision (inverse variance) of peer-adjusted returns, respectively. Our proxies for $\hat{\alpha}_{i,t}$ and $s_{i,t}$ are related to the meeting note length weighted periodicity between pre-selection meetings and total pre-selection meeting notes length (see section 3.4 for construction).

We present the results of this regression in Table 8, utilizing the sample of funds for which there is at least one note in excess of 25 words. All variables are measured as of the last month a fund is retained in the panel. For selected funds this date is the month of selection, T. For unselected funds this is the later of the last month for which the fund has a documented return or the database truncation date (December 2017). All regressors are standardized. The coefficients are reported as odds ratios, representing the multiplicative change in the hazard rate from a one-standard-deviation change in the variable.

It follows that the hazard rate of selection increases in the levels of both private ($\hat{\alpha}_{i,t}$) and public ($x_{i,t}$) information signals, but not in their precisions. Furthermore, the effect of private information is stronger both statistically and economically than the effect of public information

[6]See online appendices IA.3 and IA.4 for tests under OLS and accelerated hazard frameworks. The potential for nonrandom censoring is mitigated by appending returns data through 2017 from the public databases, of which more than 90% of our funds are a part.

in determining selection. Interestingly, the coefficient on the interaction of the private signal with private precision suggests a significantly negative effect which is not consistent with prediction 2. We thus investigate this prediction in a more flexible statistical environment, one which allows for time-varying fund characteristics and their interactions with a time-varying hazard rate.

4.3.1. Time-Varying Hazard of Selection

The Cox model's assumption that the baseline hazard rate (a function of time only) is separable from the covariate-related hazard limits our ability to properly test predictions 2 and 3. Consequently, we estimate a more general discrete-time hazard model (see, e.g., Demyanyk and Van Hemert, 2011) as follows:

$$\begin{aligned} \lambda(t|X_{i,t}) &= P(T = t_i | T \geq t_i) = \log\left(\frac{p}{1-p}\right) \\ &= \eta_i + \beta X_{i,t} + f(t), \end{aligned} \tag{13}$$

where η_i includes fixed effects and time-invariant information about the fund manager i.[7] The parametric function $f(t)$ captures the hazard's proportional time effects. We use a functional form that allows for a direct test of prediction 3: $f(t) = \beta_1 t + \beta_2 \log t$. The coefficient on the $\log t$ term captures the rate of increase in the conditional probability of selection as private information is gathered, while the coefficient on the linear t term captures the rate of decay in this conditional probability if the fund never meets the allocator's threshold for selection. $X_{i,t}$ is time varying, as both public and private information are updated each month.

We present estimation results in Table 9, Panel A. All continuous regressors other than those in $f(t)$ are standardized. The coefficients are presented as odds ratios. Column 1 focuses on the public signal and precision proxies. The results are similar to those of the Cox model. Columns 2 and 3 add log(AUM), which exhibits a significantly positive effect. Growth in assets could be viewed as an additional public information measure (see, e.g., Berk and Green, 2004). All public information coefficients are robust to the addition of market controls and year and strategy fixed

[7]Consistent with this approach, we right-censor observations conditional on the event (selection) occurring. For unselected funds, we include all fund-month observations as long as there is at least one meeting with a note exceeding 25 words.

effects.[8] Column 4 adds the private information proxies, including affiliated fund and college dummies.[9] Importantly, the coefficients on the private information precision turn significantly positive, while the significant negative relation estimated in the Cox model on the coefficient for the private information ratio (interaction of $\hat{\alpha}_{i,t}$ and $s_{i,t}$) vanishes. The combined effect of the private information variables is now unambiguously positive and economically larger than that of the public information variables (verifying prediction 2).

Finally, Table 9, Panel A also shows that the baseline hazard rate follows the hump- or concave-shaped profile predicted by our framework: there is a strong upward slope in the initial hazard rate (positive coefficient on *log(Duration)*) that tails of as due-diligence time increases (negative coefficient on *Duration*). For all models, the baseline hazard follows the profile implied by prediction 3. In Figure 5 we present the margin plot of $f(t)$ for empirical specification 5 (discussed below). In both panels the hump-shaped hazard is statistically different from zero, with the hazard rate starting low, peaking at around 20 months with a conditional probability of selection around 0.4%, and then falling as the due diligence time extends.

4.3.2. The Interaction of Public and Private Information with Time

To further analyze the hazard profile implied by prediction 3, we add the interactions of the parameterized baseline hazard function, $f(t)$, and our primary public and private information variable proxies. Coefficients for the main terms are presented in column 5 of Table 9, Panel A. The coefficients and t-statistics for the interaction terms are presented in Panel B.

The interactions with a fund's average excess return ($x_{i,t}$) follow a profile similar to that of the baseline hazard—public information is much more informative about selection early in due diligence and less so later. This follows the intuition of our framework: if excess returns are sufficiently high and precise, the selection decision is straightforward and the allocator chooses not

[8] Market variables control for variation in the overall information set, including market returns (contemporaneous and lagged), rolling volatilities of four Fama-French factors, and predicted capital flows to the allocator.

[9] The college affiliation dummy takes a value of 1 if the hedge fund manager and an employee at the allocator attended the same college. The network or affiliated fund dummy takes a value of 1 if the hedge fund is a spin-off from a previous investment of the allocator (see, e.g., Cohen, Frazzini, and Malloy, 2008; Chevalier and Ellison, 1999; Grinblatt, Keloharju, and Linnainmaa, 2012).

to acquire costly private information. Interestingly, the odds ratio on the noninteracted $x_{i,t}$ term is now significantly less than 1. There are two possible explanations for this. First, even when returns are extremely high, our allocator still takes time (e.g., to collect private information that verifies skill) before selecting the fund. Second, the allocator may see little value in selecting funds whose skill has already been revealed to the broader market. We explore possible reasons for this in section 4.4.

The interactions with $s_{i,t}$, on the other hand, project in the opposite direction as the baseline hazard: the importance of private information is higher later in the due diligence process. Some private information, specifically that collected from due diligence on affiliated managers, however, is very important early in due diligence. Given that our allocator faces a tradeoff between making decisions too soon (e.g., investing in unknown, low-quality funds) or not soon enough (e.g., losing an α-generating opportunity), we would also expect them to use information gleaned from due diligence on other, connected firms when making selection decisions. We note that under all specifications the stand-alone fund- and college-affiliation variables have no statistical significance. This verifies that, at least for our professional allocator, information more relevant to skill (versus that related merely to personal affiliations), are of primary importance in the selection decision.

As noted, in Figure 5 we present the margin plots of the hazard function using the full model from column 5 at +/- one standard deviation in the levels of private information signal (Panel A) and precision (Panel B). In light of the results in Table 9, it is not surprising that the higher the private information signal ($\hat{\alpha}_{i,t}$), the higher the peak hazard rate. Given the interaction effects of the hazard with private information precision ($s_{i,t}$), it appears that the acquisition of private information shortens the average due diligence time not only by increasing the peak hazard rate, but also by extending the time over which the fund is truly at risk of selection.

To quantify the effects of private information on timing, we compare the due diligence time of the average fund in our sample with that of a fund with a one-standard-deviation-higher private information signal and precision, and a parent-allocator affiliation. The average due diligence time of fund i is the weighted sum of the due diligence periods (per equation 13):

$$\mu_i = \sum_{t=0}^{\infty} w_{i,t} \cdot t, \tag{14}$$

where $w_{i,t}$ is the scaled probability of investment at any period t. The probability of investment is the hazard rate, $\lambda(t|X_{i,t})$, times the survival rate, which is defined as the cumulative probability that a hazard as not occurred up to time t, i.e., $\prod_{j=0}^{t-1}\left(1-\lambda\left(t|X_{i,j}\right)\right)$. Therefore,

$$w_{i,t} = \frac{\lambda(t|X_{i,t}) \cdot \prod_{j=1}^{t-1}\left(1-\lambda\left(t|X_{i,j}\right)\right)}{\sum_{i=1}^{\infty}\lambda(t|X_{i,t}) \cdot \prod_{j=1}^{t-1}\left(1-\lambda\left(t|X_{i,j}\right)\right)}. \tag{15}$$

We estimate the expected due diligence time to be approximately 46 months for the average fund and 29 months for the high private information fund, representing a 17-month drop in due diligence duration. As we will see in the next section, this corresponds closely to the average time period over which the allocator generates excess returns in a fund investment.

4.4. *Selecting Diamonds in the Rough*

As motivated in sections 2 and 3, the acquisition of private information requires a significant commitment of resources. It is thus natural to ask whether this information has lead our allocator to better investment decisions, and if so, to understand the economic mechanism through which this information derives its pecuniary benefits. As in Figure 1, Panel A, we create matched samples based on calendar time, log(AUM), age, and past rolling 24 month information ratio. But now, rather than comparing past returns we compare future returns of selected and unselected funds. Histograms of returns for the two sets of funds are presented in Figure 1, Panel B. The mean selected fund outperforms the mean unselected fund by nearly 0.23% per month over 18 months.

4.4.1. Do Selected Funds Outperform?

To quantify both the outperformance and decay, the preferred experiment is to compare funds with identical public, but different private, information at the selection date. We can then assess the pecuniary benefits of private information by comparing the post–due diligence return dynamics. We approximate this experiment by following the same procedure above to match each selected fund with the three most similar unselected funds on the selection date. Our analysis then assumes that investments are made in all four funds on the same date (the selection date of the selected fund). We first quantify the realized effect of selection by estimating a fixed effects regression of *log(post-selection duration)* and $\mathbb{I}_{sel} \times$ *log(post-selection duration)* onto our allocator's pooled true (selected) and hypothetical (unselected) fund returns.

$$x_{i,t}^{peer} = a_i + \beta_{dur} \cdot log(Duration) + \beta_{int} \cdot \mathbb{I}_{sel} \times log(Duration) + \varepsilon_{i,t} \tag{16}$$

It is important to note that $x_{i,t}^{peer}$ is now the 1-month (not 24-month rolling average) peer-adjusted excess return; this allows us to estimate a parameterized decay coefficient. The results are presented in column 1 of Table 10. The coefficients on *log(Duration)* imply that selected and unselected fund excess returns diverge in opposite directions from the time of selection. We test prediction 4 by checking whether fund returns at the selection date are, on average, positive (negative) for selected (unselected) funds. A measure of initial expected excess returns is the fixed effects (i.e., $a_i + \varepsilon_{i,t}$) from our de-meaning procedure (see Figure D.8 in online appendix). On average, selected and unselected funds generate 0.65% and -0.15% in excess returns per month on selection date, respectively. This implies that our allocator sources good returns not only by selecting future outperforming managers, but also by avoiding underperforming managers. Furthermore, this outperformance is persistent, only mean-reverting towards zero in the long run. Taken together, these results support prediction 4.

To quantify the effect of higher and longer lasting excess returns, we estimate the post selection time, T_N, at which returns of selected (S) and unselected (US) funds are on average indistinguishable from one another:

$$\begin{aligned} 0 &= r_{S,T_N}^{peer} - r_{US,T_N}^{peer} \\ &= a_S + (\beta_{dur} + \beta_{int}) \cdot \log(T_N) - a_{US} - \beta_{dur} \cdot \log(T_N), \end{aligned}$$

where a_S and a_{US} are the mean fixed effects on the selection date, and β_{dur} and β_{int} are estimated from our fixed-effects regression. We find T_N to be approximately 20 months, which is close to the average reduction in due diligence time provided by private information. We use this information to estimate the average cumulative excess returns by integrating under the curve from the time of selection to T_N,

$$\int_1^{T_N} [(a_S - a_{US}) + \beta_{int} \cdot \log(x)]dx.$$

We find that the cumulative excess return of selected (over unselected) funds is on average 9%.

The magnitude of outperformance is similar to that of younger mutual funds in PST, albeit over a shorter horizon. The underlying assumptions of this estimate are that all investments (i) carry an equal weight, (ii) have no capital constraints, i.e., the allocator invests in all selected funds, (ii) will be unwound at 20 months when selected and unselected funds are indistinguishable from one another, (iv) will be conducted versus a short in funds not invested in, and (v) will be made versus an appropriate benchmark return. These assumption are obviously violated in reality; our estimate, while specifically approximating the value-add of private information, will not be entirely captured by our allocator on behalf of their clients.[10] This does, however, point to a possible economic channel through which private information derives its large pecuniary benefits: the tradeoff between fund size and returns (à la Berk and Green, 2004; Pástor and Stambaugh, 2012). Private information of the sort collected by our allocator provides an additional source of information through which an allocator can identify high-skill managers. The exclusivity of the information allows the allocator to find these managers before others that are relying on public information alone. The fund will outperform other funds, but only in the intermediate term. Assuming decreasing returns to scale, information on skill will eventually filter into the public sphere, attracting AUM and driving outperformance towards zero.

4.4.2. Private Information and Performance Persistence

Given this narrative, we would expect that funds with higher selection-date (henceforth, T) private information ratios will have both higher initial levels of and faster decay in post-T excess expected returns. This prediction is an extension of our Bayesian learning framework. The acquisition of private information (higher private information ratio) implies a lower prior precision of public information. For the broader market (i.e., investors with zero or inferior private signal) future

[10]As noted in section 3.5, our allocator is not an outlier in their performance (see, e.g., Andonov, Hochberg, and Rauh, 2018, for an example of reasons for other sources of performance drag).

public information will weigh heavier on their subsequently updated estimate of α (see, e.g., Pástor, Taylor, and Veronesi, 2009). We would likewise expect the opposite for funds with higher time-T public information ratios (i.e., little-to-no initial excess expected returns, as well as decay in these returns). This dichotomy is driven entirely by the exclusivity of private, as opposed to public, information. To test this hypothesis we replace the selection indicator, $\mathbb{I}_{sel}$, with indicators splitting the range of public and private information ratios into terciles. This allows us to analyze how variation in public and private characteristics known at time T inform both the level and decay of excess returns in the post selection period.

We find, first, that decay is increasing in private information (Table 10, column 2): the coefficients for $\mathbb{I}^{mid}_{\alpha_T \times s_T}$ and $\mathbb{I}^{high}_{\alpha_T \times s_T}$ are statistically different from one another. Second, although there is no decay in excess returns for funds with moderate public information ratios (see the coefficient on $\mathbb{I}^{mid}_{x_T \times r_T}$), there appears to be a reversal of returns in funds with high public information ratios. This interpretation would be consistent only if the average investor tends to rely too heavily on the public signal (i.e., a high public information ratio induces the marketplace on average to overshoot), increasing the fund's assets under management beyond its zero excess return equilibrium. At the bottom of column 2, we compare the average fixed effects of funds with first-tercile public and third-tercile private information ratios (henceforth, undershot funds) with those of third-tercile public and first-tercile private information ratios (henceforth, overshot funds) at time T. The results confirm that undershot and overshot funds have statistically different peer-adjusted returns of 0.25% and -0.24% per month, respectively. The fact that the net return of 0.50% is five-eighths of the 0.80% in excess return from the endogenous selection decision lends credence to the strength of our private information proxy.

4.4.3. Return-to-Scale Channel

To better illustrate the possibly connection between lagged AUMs and excess returns (negative returns-to-scale), we compare their average post-selection time series in Figure 6. Figure 6 Panel A compares the cumulative growth in AUMs (adjusted for returns) for the average selected and matched unselected fund. Early in the sample AUM growth is fairly robust for both fund types.

The growth rates, however, substantially diverge around 12 months post-selection. Selected funds continue to growth for an additional 8 months, but exhibit flatter growth at the tail end the window prsented. Unselected funds on the other hand flatline and then fall in asset growth at the end. If the Berk and Green framework is the driving force behind our findings, anecdotally, we should expect that returns will follow the same general profile. That is, selected fund returns will be high early, but tailoff later post due diligence, and unselected fund returns will be flat early and fall later post due diligence. This intuition is corroborated in the cumulative excess returns presented in Figure 6 Panel B.

We formally test for this predictive relationship by adding fund $log(AUM)_{t-1}$ to regression 16. The hypothesis behind its inclusion is derived from a simple question: does the private information collected provide our allocator with a superior signal of the "distance" between a fund's current AUM and that of a proper full-information equilibrium? To be consistent with our prediction, the coefficient on $log(AUM)_{t-1}$ should be negative, capturing negative returns to scale, while β_{dur} and β_{int} should be statistically zero, as $log(AUM)_{t-1}$ subsumes all information regarding the timing of excess returns.

As highlighted in PST, estimating this fixed effects model may produce downward-biased coefficients because $x_{i,t}^{peer}$ and $\log(\text{AUM})_{t-1}$ have structural negative correlation (see Stambaugh, 1999). Following PST and Hjalmarsson (2010), we recursively forward-demean all variables in the regression. This removes the need to estimate a_i. We then use backwards-demeaned AUM as an instrument for forward-demeaned AUM in a 2SLS regression framework. To verify that the new procedure does not change the regression results, we estimate the original regression using the forward-demeaned variables in column 3 of Table 10. Results are consistent with the regular fixed effects approach. To show that AUM is by itself an important driver of fund excess returns, in column 4 we present results with only AUM as a regressor. $\log(AUM)_{t-1}$ is negatively correlated with excess returns. Given the log-scale, these results tell us that a roughly 2.5-fold increase in fund size decreases excess returns by more than 1.20% per month.

In Table 10, column 5, we combine regressors from columns 3 and 4. The coefficient on

$\mathbb{I}_{sel} \times log(Duration)$ is now statistically insignificant. While the coefficient on $log(Duration)$ is still significant, its sign is now switched and magnitude larger. A possible missing variable in the regression is industry-wide size or AUM; per PST, industry rather than fund size is the primary driver of mutual funds' negative returns-to-scale. Unfortunately, the long-short hedge fund industry size is notoriously difficult to estimate. There is, however, consensus that the industry grew over our sample (2005–2017). Given the sample homogeneity, $log(Duration)$ may be capturing this trend. Its addition further clarifies the effects from variation in $log(AUM)_{t-1}$: a 2.5-fold increase in fund size now decreases excess returns by 2.20% per month.[11]

Notably, our results do not imply that all allocators that conduct in-house due diligence (i.e., collect private information) when selecting managers outperform. Gârleanu and Pedersen (2018) show that one would expect only a subset of allocators to outperform in an inefficient (e.g., high search cost) asset manager market. In their model, investors in a larger and more sophisticated allocator, such as ours, benefits from their favorable search economies-of-scale. While other allocators may have different cost to acquiring private information, our analysis illustrates the mechanism through which these allocators may derive their edge and demonstrates the important role that private information plays in making the asset manager markets more efficient. Nonetheless, our results point at one plausible reason for why the decreasing return-to-scale effect might be difficult to detect empirically in the hedge fund space in particular (see, e.g., Yin, 2016)—private information-based investment decisions by sophisticated allocators weaken the link between the lead-lag of realized returns (i.e., public information) and AUM changes.

5. Conclusion

This paper studies the fund manager selection problem from the standpoint of professional asset allocators. We develop a simple framework for their process, informed by analysis of due diligence notes and pitchbooks from the interactions of a representative allocator with 860 hedge funds. Our

[11] We note that adding the intercept to the first stage regression as advised in Zhu (2018) does not meaningfully affect estimates in our case—e.g. the second-stage t-statistic on lagged AUM increases from 3.09 to 3.14 in column 5.

analysis shows that the information gathered by the allocator comports with its stated objectives: to utilize research to identify skilled fund managers efficiently and quickly. In addition, our setting differs from previous studies in that we examine in-house research rather than external consultant recommendations. Our data allows us to disentangle manager selection from portfolio constraints, accurately determine decision timing, and measure the quality and quantity of private information involved. The cost of this unusually detailed information is a focus on just one allocator as in Becht, Franks, Mayer, and Rossi (2008).

We find that private information (i) complements public information at the intensive margin insofar as past returns and their moments inform the degree to which the allocator collects private information, and (ii) significantly affects the timing of manager selection. A one-standard-deviation increase in our proxy of private information triples the probability of the fund being selected and leads to an almost 40% drop in the time taken for due diligence. This is about five times the effect of a one-standard-deviation increase in past excess returns.

We also find no evidence that reliance on private information, which is potentially prone to poor subjective judgments, degrades our allocator's performance. On the contrary, cumulative gross excess returns are 9.0% higher for selected managers. This excess return decomposes into substantial outperformance at selection and a persistence of positive alpha over approximately two years. Interestingly, the timing of two years corresponds closely to the reduction in due diligence time achieved with a high private information-ratio. We then link this outperformance to the negative returns-to-scale relationship studied in the literature, finding that our allocator uses private information to gain a better understanding of a fund's capacity. This allows it to take advantage of a potentially transitory disconnect between the current and hypothetical full-information fund size–i.e., after a fund's true skill is common, public knowledge.

References

Adams, John C, Darren K Hayunga, and Sattar Mansi, 2018, Returns to scale in active and passive management, *University of Texas at Arlington Working paper*.

Admati, Anat R, and Paul Pfleiderer, 1997, Does it all add up? benchmarks and the compensation of active portfolio managers, *The Journal of Business* 70, 323–350.

Agarwal, Vikas, Naveen D Daniel, and Narayan Y Naik, 2009, Role of managerial incentives and discretion in hedge fund performance, *The Journal of Finance* 64, 2221–2256.

Agarwal, Vikas, Kevin A Mullally, and Narayan Y Naik, 2015, The economics and finance of hedge funds: A review of the academic literature, *Foundations and Trends® in Finance* 10, 1–111.

Andonov, Aleksandar, Yael V Hochberg, and Joshua D Rauh, 2018, Political representation and governance: Evidence from the investment decisions of public pension funds, *The Journal of Finance* 73, 2041–2086.

Becht, Marco, Julian Franks, Colin Mayer, and Stefano Rossi, 2008, Returns to shareholder activism: Evidence from a clinical study of the hermes uk focus fund, *The Review of Financial Studies* 22, 3093–3129.

Benamar, Hedi, Thierry Foucault, and Clara Vega, 2019, Demand for information, macroeconomic uncertainty, and the response of us treasury securities to news, *HEC Working Paper*.

Bergstresser, Daniel, John MR Chalmers, and Peter Tufano, 2009, Assessing the costs and benefits of brokers in the mutual fund industry, *Review of Financial Studies* 22, 4129–4156.

Berk, Jonathan B, and Richard C Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Bird, Steven, Ewan Klein, and Edward Loper, 2009, *Natural Language Processing with Python* (O'Reilly Media Inc.).

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey D Kubik, 2004, Does fund size erode mutual fund performance? the role of liquidity and organization, *American Economic Review* 94, 1276–1302.

Chevalier, Judith, and Glenn Ellison, 1999, Career concerns of mutual fund managers, *Quarterly Journal of Economics* pp. 389–432.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, 2008, The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* 116, 951–979.

Demyanyk, Yuliya, and Otto Van Hemert, 2011, Understanding the subprime mortgage crisis, *Review of Financial Studies* 24, 1848–1880.

Evans, Richard B, and Rüdiger Fahlenbrach, 2012, Institutional investors and mutual fund governance: Evidence from retail–institutional fund twins, *The Review of Financial Studies* 25, 3530–3571.

García, Diego, 2013, Sentiment during recessions, *The Journal of Finance* 68, 1267–1300.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2018, Efficiently inefficient markets for assets and asset management, *The Journal of Finance* 73, 1663–1712.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2015, Money doctors, *The Journal of Finance* 70, 91–114.

Gerken, William Christopher, Laura T Starks, and Michael Yates, 2018, The importance of family: The role of mutual fund family reputation in investment decisions, *University of Kentucky Working Paper*.

Goyal, Amit, and Sunil Wahal, 2008, The selection and termination of investment management firms by plan sponsors, *The Journal of Finance* 63, 1805–1847.

Grinblatt, Mark, Matti Keloharju, and Juhani T Linnainmaa, 2012, IQ, trading behavior, and performance, *Journal of Financial Economics* 104, 339–362.

Gupta, Arpit, and Kunal Sachdeva, 2019, Skin or skim? inside investment and hedge fund performance, *National Bureau of Economic Research Working Paper*.

Hanley, Kathleen Weiss, and Gerard Hoberg, 2018, Dynamic interpretation of emerging risks in the financial sector, *Forthcoming, Review of Financial Studies*.

Hermalin, Benjamin E., and Michael S. Weisbach, 1998, Endogenously chosen boards of directors and their monitoring of the ceo, *American Economic Review* 88, 96–118.

Hjalmarsson, Erik, 2010, Predicting global stock returns, *The Journal of Financial and Quantitative Analysis* 45, 49–80.

Hoberg, Gerard, Nitin Kumar, and Nagpurnanand Prabhala, 2017, Mutual fund competition, managerial skill, and alpha persistence, *The Review of Financial Studies* 31, 1896–1929.

Jagannathan, Ravi, Alexey Malakhov, and Dmitry Novikov, 2010, Do hot hands exist among hedge fund managers? an empirical evaluation, *The Journal of Finance* 65, 217–255.

Jenkinson, Tim, Howard Jones, and Jose Vicente Martinez, 2015, Picking winners? investment consultants' recommendations of fund managers, *The Journal of Finance*.

Jones, Howard, and Jose Vicente Martinez, 2017, Institutional investor expectations, manager performance, and fund flows, *Journal of Financial and Quantitative Analysis* 52, 2755–2777.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Kaniel, Ron, and Robert Parham, 2017, Wsj category kings–the impact of media attention on consumer and mutual fund investment decisions, *Journal of Financial Economics* 123, 337–356.

Korteweg, Arthur, and Morten Sorensen, 2017, Skill and luck in private equity performance, *Journal of Financial Economics* 124, 535–562.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of Finance* 66, 35–65.

Pástor, Ľuboš, and Robert F Stambaugh, 2012, On the size of the active management industry, *Journal of Political Economy* 120, 740–781.

——— , and Lucian A Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.

Pástor, Ľuboš, Lucian A Taylor, and Pietro Veronesi, 2009, Entrepreneurial learning, the ipo decision, and the post-ipo drop in firm profitability, *The Review of Financial Studies* 22, 3005–3046.

Rigobon, Roberto, and Thomas Stoker, 2007, Estimation with censored regressors: Basic issues, *International Economic Review* 48, 1441–1467.

Roussanov, Nikolai, Hongxun Ruan, and Yanhao Wei, 2018, Marketing mutual funds, *National Bureau of Economic Research Working Paper*.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

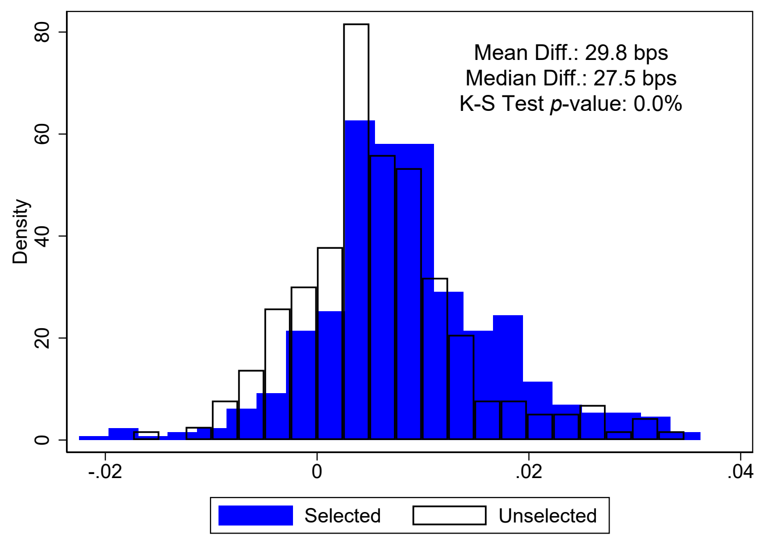
Yin, Chengdong, 2016, The optimal size of hedge funds: conflict between investors and fund managers, *The Journal of Finance* 71, 1857–1894.

Zhu, Min, 2018, Informative fund size, managerial skill, and investor rationality, *Journal of Financial Economics* 130, 114–134.

Figure 1. Selected Funds: Past and Future Alphas

This figure reports frequency distributions of 24-month rolling Jensen's α-estimates for funds that the allocator selected versus a matched group of peer funds that were never selected. In both panels, the peer funds are matched according to the Mahalanobis distance based on a fund's log(AUM), age, and past information ratio as of the calendar month on selection date. Statistics comparing the means and distributions (Kolmogorov-Smirnov tests) of selected and unselected samples are presented within each histogram. Robustness to matching scheme and to abnormal return definition are examined in the online appendix.

Panel A. Abnormal returns before selection



Panel B. Abnormal returns after selection

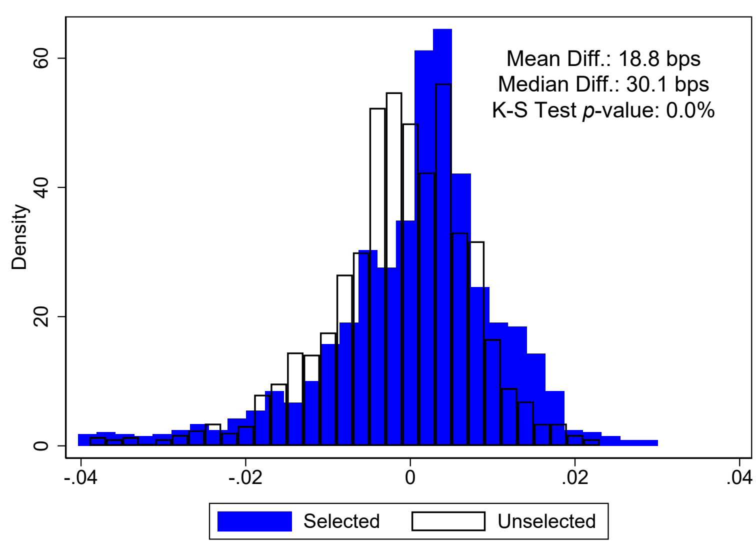
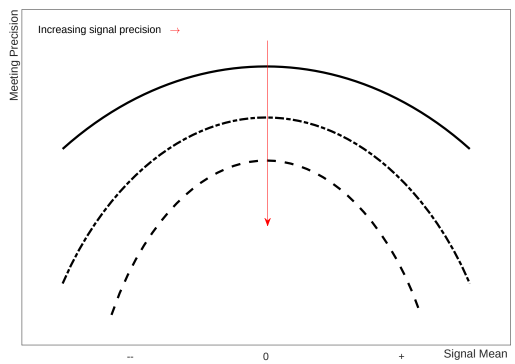


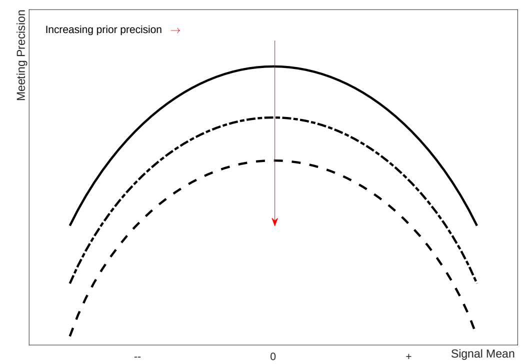
Figure 2. Model Comparative Statics

In Panel A– Panel C we plot the model's comparative statics. The cost function, $c(s)$, is parameterized as $c_0 + s^2$. Equation B.2 is solved on a fixed-point grid of parameter values for prior mean (μ), prior precision (τ_0), public signal level (x), and public signal precision (r). Panel A represents the effects of realized x on private signal precision (s); Panel B represents the effects of τ_0 on s; Panel C represents the effects of τ_0 on s assuming $\mu < 0$. Panel D is the cross sectional volatility of monthly idiosyncratic returns for the funds in our dataset, which illustrates our empirical proxy for τ_0.

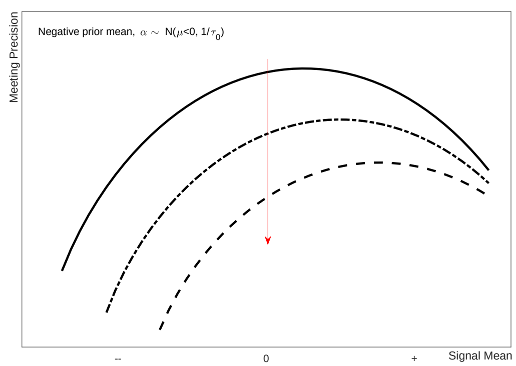
Panel A. s w.r.t. x



Panel B. s w.r.t. τ_0



Panel C. s w.r.t. τ_0 assuming $\mu < 0$



Panel D. Cross-sectional volatility of idiosyncratic fund returns

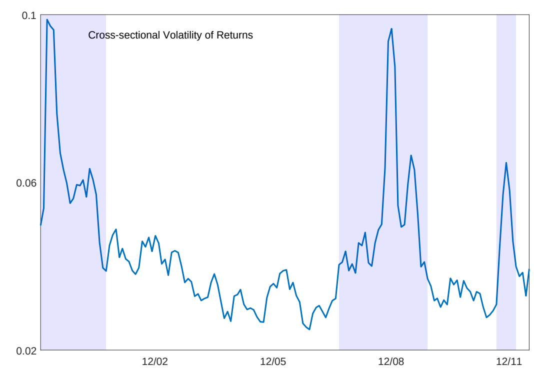
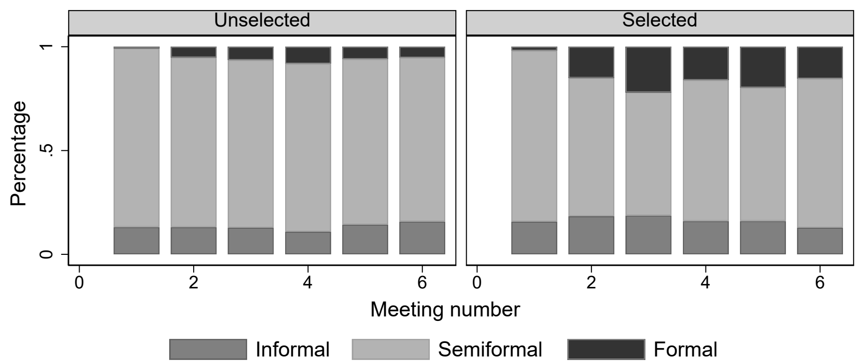


Figure 3. Fund Selection and Manager Meetings

This figure compares the meeting frequency and composition of selected and unselected funds during the due diligence process. To simplify exposition, we code meetings as informal ('conference' or 'email'), semiformal ('call' or 'face-to-face meeting'), and formal ('on-site visit'). Panel A shows the average change in composition of selected and unselected funds as the allocator holds progressively more meetings with the fund. Panel B reports the frequency of meeting during the first and last nine months of due diligence at three-month intervals. The x-axis is monthly, where 1 and -1 denote the first and last month of due diligence, respectively. We split the analysis into the first and last nine months in order to enhance the comparability of meeting frequencies across funds with different due diligence durations.

Panel A. Meeting composition



Panel B. Meeting frequency

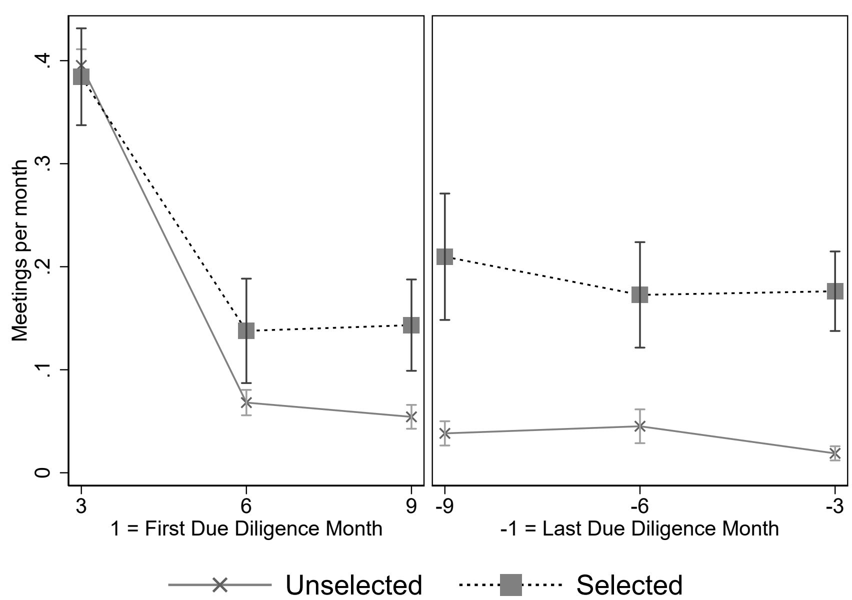
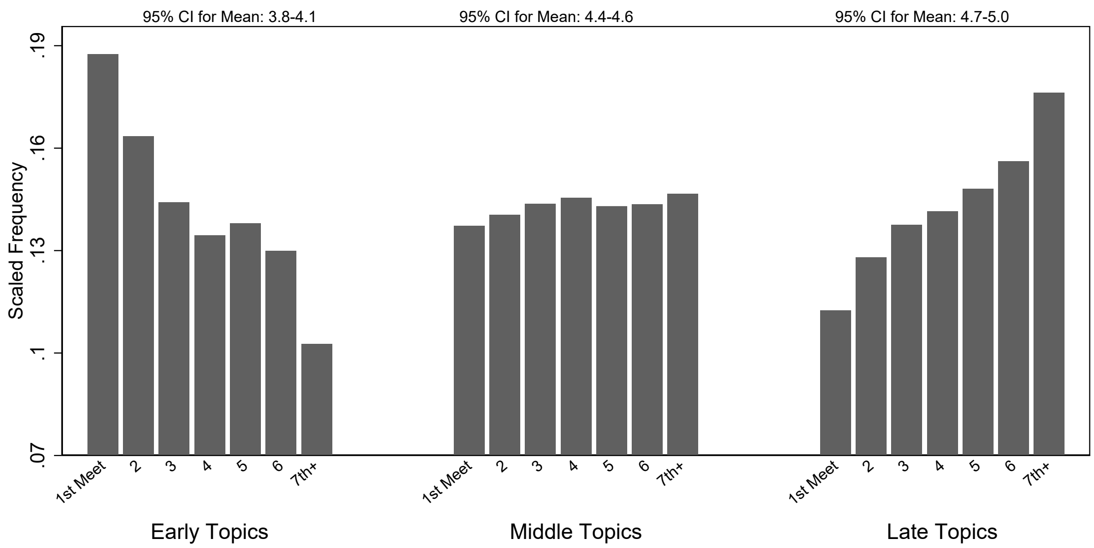


Figure 4. LDA Intuition

In Panel A we illustrate the relative time of discussion of the LDA-inferred topics from Table 4. We categorize the 23 discernible topics into three categories (early, middle, late) and estimate the weighted average meeting number and standard deviation for each. The 95% confidence interval statistics are presented at the top of the figure. The histograms plot the relative frequency of meeting category over each meeting. We first compute the relative mix, conditional on meeting number, of each topic category. These sum to one for each meeting across categories. Given that each category is allocated a different amount of attention on average, we then plot the scaled data such that the frequencies within each category sum to one. Panel B plots a rolling Kullback-Leibler measure of meeting-topic distributions in our preselection sample. For all meetings after the first, we use the rolling topic proportion over all previous meetings as our reference. For the first meeting, we use the topic distribution for all selected start-up funds as our reference.

Panel A. Meeting timing



Panel B. Meeting breadth

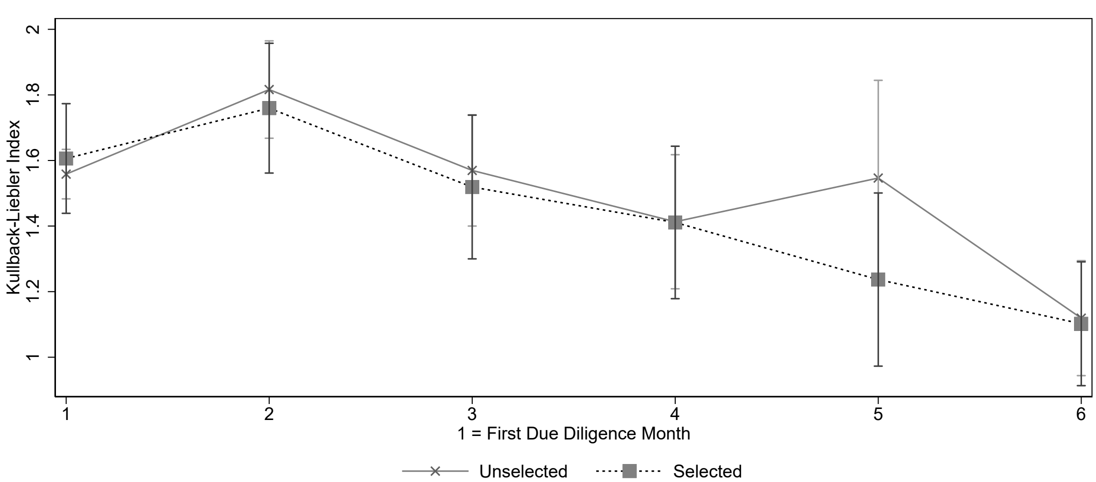
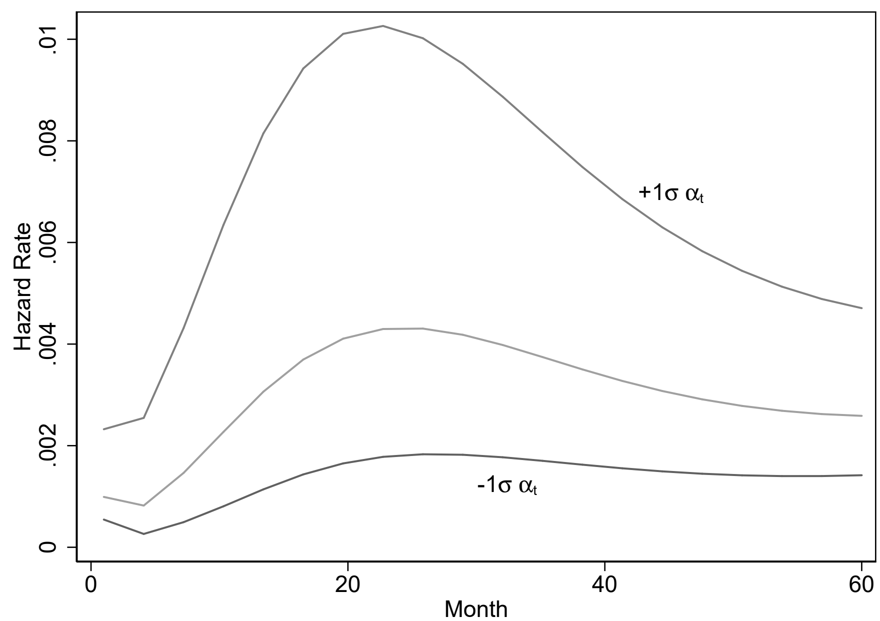


Figure 5. The Effects of Private Information Signal and Precision

This figure captures the marginal hazard rate given different values of the private information signal (+/- one standard deviation of $\hat{\alpha}_{i,t}$) in Panel A and private information precision (+/- one standard deviation of $s_{i,t}$) in Panel B. The marginals are computed using model 5 from Table 9.

Panel A. Signal level



Panel B. Signal precision

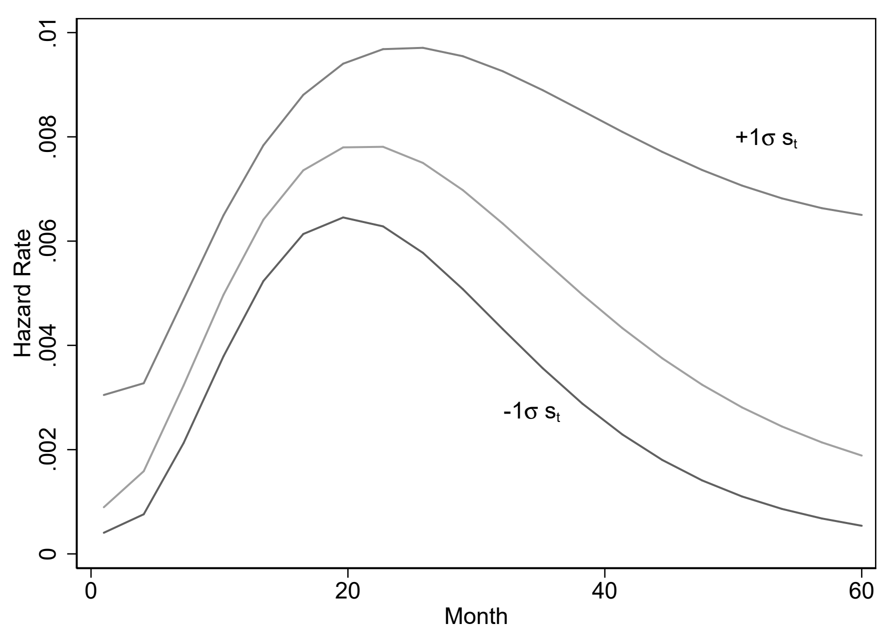
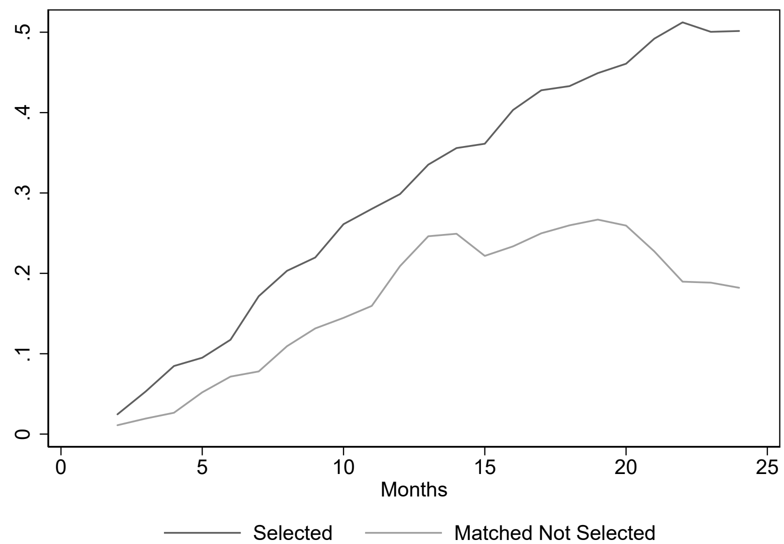


Figure 6. Post Selection Dynamics in AUM and Returns

This figure compares the cumulative changes in AUMs (gross of returns) and excess returns for up to 24-months after the selection date between selected funds and three unselected funds. The unselected funds were matched to the selected by calendar time, log(AUM), age, and past rolling 24 month information ratio as of the selection date. The figure helps illustrate the underlying economic mechanism quantified in regression model (16) and presented in columns (4) and (5) of table 10. The impact of returns are removed from the AUM growth rates, and therefore represents the flow of assets into the fund. See Appendix C for variable definitions.

Panel A. Cumulative AUM growth



Panel B. Cumulative Excess returns

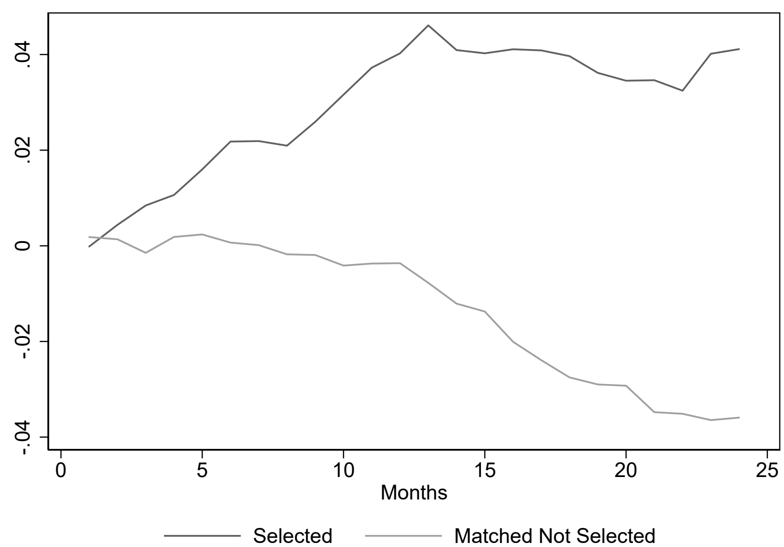


Table 1. Summary Statistics

This table reports summary statistics for public (Panel A), and private (Panel B) information measures that we use in our analysis. For Panel B, our primary source is a database of monthly returns and fund AUM estimates maintained by the allocator. We also match our fund names to funds in the combined databases of Morningstar, Barclay Hedge, eVestment, HFR, and Lipper-TASS. This data is used in our analysis when information is missing from our primary source or we require data for a longer period. For Panel B, meeting number is the maximum number of meetings our allocator had with a fund prior to selection. Informal and formal meetings are the number of informal ('conference' or 'email') or formal ('on-site visit') interactions, respectively, that each fund had over the course of due diligence. Periodicity is the average number of months between meetings. For unselected funds, we define the periodicity of the last meeting as the time between the meeting and the fund's dropping from the sample. This mechanically leads to a longer average periodicity. We therefore also report average periodicity ignoring the last meeting.

Panel A. Public information

	count	mean	sd	skewness	p5	p25	p50	p75	p95
Fund AUM (USD mln)	809	392.9	1014.2	6.48	7.09	37.8	113.2	317.7	1400
Fund age (months)	664	58.5	83.7	9.40	6	16.5	40.5	77.5	163.5
Raw return (R_t)	896	0.0074	0.015	-0.32	-0.017	-0.00073	0.0080	0.016	0.032
$\hat{E}_t(R - peers)$	896	0.0041	0.011	0.23	-0.012	-0.0018	0.0030	0.0090	0.024
Rolling alpha	893	0.0062	0.0097	1.29	-0.0063	0.0013	0.0052	0.010	0.024
Rolling beta	893	0.50	0.54	1.10	-0.13	0.17	0.40	0.76	1.51
Idiosync. bolatility	893	0.13	0.093	4.69	0.040	0.071	0.10	0.15	0.28
Idiosync. skewness	893	6.26	42.7	5.22	-12.6	-0.92	-0.022	1.42	41.9
$\hat{\sigma}_t(R - peers)$	896	0.11	0.072	3.87	0.043	0.069	0.096	0.14	0.23
$S\hat{kew}_t(R - peers)$	895	-0.039	0.75	-0.23	-1.33	-0.53	-0.034	0.49	1.15

Panel B. Private information

	count	mean	sd	skewness	p5	p25	p50	p75	p95
Meeting number (all)	895	2.81	2.16	2.17	1	1	2	4	7
Formal meetings	934	0.13	0.39	3.53	0	0	0	0	1
Informal meetings	934	0.30	0.65	2.99	0	0	0	0	1
Meetings/Notes:									
Words per document	647	370.6	178.7	0.96	138.3	239	341	478.5	708
Periodicity (months)	929	2.52	1.16	-0.047	0.69	1.62	2.43	3.56	4.23
Periodicity, ignore last	611	1.58	0.96	0.42	0	0.98	1.41	2.19	3.47
Pitchbooks:									
Words per document	523	3850.1	2877.3	2.73	805	2101	3194	4813	8602

Table 2. Univariate Tests: Returns and AUMs

This table reports difference tests on various public information measures 12 months from the beginning of due diligence between selected and unselected funds. Excess return, $E\widehat{(R-peer)}$, is the fund return minus a peer benchmark return. Our allocator flags long-short funds as global long-short, emerging market, market neutral, or relative value funds. Our peer benchmarks are thus the HFRI equity hedge, HFRI emerging market, HFRI equity market neutral, and HFRI relative value indices, respectively. The expected level and higher moments of excess return are then computed as a 12-month rolling average. Alpha is the 12-month rolling Jensen's α-estimate using the market return as the benchmark. The higher-order moments of its residuals are denoted as *idiosyncratic* below. We test the difference in means using t-tests and the Wilcoxon ranksum test.

	Selected			Unselected			P-values	
	N	Mean	Median	N	Mean	Median	T-test	Ranksum
Fund AUM (USD mln)	206	593	220	603	325	91	0.001	0.000
Fund age (months)	169	58.9	39.5	495	58.4	40.5	0.945	0.101
Raw return (R_t)	213	0.013	0.012	683	0.006	0.007	0.000	0.000
$\hat{E}_t(R-peers)$	213	0.007	0.005	683	0.003	0.003	0.000	0.000
Rolling alpha	211	0.010	0.008	682	0.005	0.004	0.000	0.000
Rolling beta	211	0.581	0.486	682	0.481	0.381	0.019	0.013
Idiosync. volatility	211	0.132	0.113	682	0.123	0.101	0.238	0.082
Idiosync. skewness	211	9.321	-0.018	682	5.317	-0.027	0.234	0.987
$\hat{\sigma}_t(R-peers)$	213	0.115	0.101	683	0.112	0.095	0.578	0.294
$\widehat{Skew}_t(R-peers)$	213	0.010	0.039	682	-0.054	-0.061	0.275	0.307
Sharpe ratio (months)	213	0.097	0.078	685	0.048	0.046	0.004	0.000
Information ratio (months)	213	0.051	0.056	682	0.023	0.029	0.065	0.000

Table 3. Univariate Tests: Manager Meetings and Textual Data

This table reports difference tests on estimates of private information between selected and unselected funds during the preselection period. Meeting number is the maximum number of interactions between the allocator and a fund during the due diligence period. Formal meetings are defined as 'on-site visit' and informal as 'conference calls' or 'email.' The table lists the fraction of meetings for each type of meeting. Periodicity is the average number of months between meetings. For unselected funds, we define the periodicity of the last meeting as the time between the meeting and the fund's dropping from the sample. We also compute average periodicity ignoring the last meeting. Panel A reports statistics computed over all selected and unselected funds. Panel B reports statistics only for start-up funds. Start-up funds are theoretically closest to funds for which selection decisions are made entirely on private information—funds that have little to no hard information and either are not yet operating or are funded entirely by employee-only money when due diligence commences.

Panel A. All funds

	Selected			Unselected			*P*-values	
	N	Mean	Median	*N*	Mean	Median	*T*-test	Ranksum
Meeting number (all)	227	3.72	3.00	668	2.49	2.00	0.000	0.000
Formal meetings only	227	0.31	0.00	707	0.07	0.00	0.000	0.000
Informal meetings	227	0.44	0.00	707	0.25	0.00	0.000	0.075
Meeting/Notes:								
Words per document	156	412	386	491	357	333	0.001	0.004
Periodicity(months)	227	1.50	1.45	702	2.85	2.85	0.000	0.000
Periodicity, ignore last	227	1.40	1.10	384	1.68	1.61	0.000	0.000
Pitchbooks:								
Words per document	134	4202.35	3230.00	389	3728.73	3180.00	0.100	0.213

Panel B. Start-ups Only

	Selected			Unselected			*P*-values	
	N	Mean	Median	*N*	Mean	Median	*T*-test	Ranksum
Meeting number (all)	72	3.71	3.00	217	2.29	2.00	0.000	0.000
Formal meetings only	72	0.35	0.00	240	0.04	0.00	0.000	0.000
Informal meetings	72	0.54	0.00	240	0.31	0.00	0.013	0.134
Meeting/Notes:								
Words per document	54	438	395	160	371	340	0.022	0.039
Periodicity(months)	72	1.51	1.47	239	2.84	2.78	0.000	0.000
Periodicity, ignore last	72	1.42	1.10	130	1.65	1.62	0.117	0.071
Pitchbooks:								
Words per document	45	3903.24	3173.00	100	3590.94	2981.00	0.525	0.755

Table 4. Top Topic Words From LDA

This table lists the top five words for each topic generated by the Latent Dirichlet Allocation algorithm applied to our corpus of text. Online appendix IA.2 provides details of the algorithm and our measure of fit. Section 3.3 describes the filters we apply to the data. We ascribe topic titles through visualization of the endogenously placed words. We then place these topics into categories (early, middle, late) corresponding to the time during due diligence each topic occurs as implied by our allocator during interviews. Below we list the topic number (from LDA) and the assigned topic name and category. Topic numbers not listed were considered "noise"– i.e., topics with low coherence scores (< -2.5) and/or difficult-to-assign titles.

Early Topics		Middle Topics		Late Topics	
4: Launch	25: Background	7: Process	18: Process	1: Medical	17: Latin America
Fund	Officer	Private	Value	Healthcare	Brazil
Launch	Chief	Public	Investment	Biotech	Banco
Small	Goldman-Sachs	Idea	Distribution	Drug	LatAm
Team	Advisor	Health	Opportunity	Medical	Mexico
Focus	Associate	Assets	Catalyst	Pharma	Currency
CS: -0.67	**CS**: -1.21	**CS**: -1.11	**CS**: -0.74	**CS**: -3.79	**CS**: -1.64
13: Background		9: Risk	22: Performance	3: Commodities	28: International
Degree		Analysis	Performance	Gold	Morgan-Stanley
Bachelor		Risk	Return	Russia	LatAm
Associate		Process	Index	Commodity	London
Founder		Fundamental	Inception	Africa	Emerging
Career		Research	Annualized	Coal	International
CS: -1.44		**CS**: -0.69	**CS**: -1.18	**CS**: -1.64	**CS**: -1.58
19: Background		11: Port. Mngt.	23: Process	6. E. Asia	
Tiger		Short	Growth	China	
Analyst		Position	Earnings	Asia	
Julian		Exposure	Price	Hong Kong	
Maverick		Long	Increase	Korea	
Kingdon		Portfolio	Inflation	Taiwan	
CS: -1.44		**CS**: -0.63	**CS**: -1.64	**CS**: -1.51	
21: Energy		12: Performance	26: Strategy	8: Real Estate	
Energy		Sharpe	Strategy	Bank	
Passport		Deviation	Equity	REIT	
Utility		Long	European	Credit	
Resources		Document	Trading	Debt	
Commodity		Return	Multi strategy	Mortgage	
CS: -1.4		**CS**: -2.39	**CS**: -0.83	**CS**: -0.91	
24. Background		16. Outlook	30. Organization	14. Technology	
University		Think	Investment	Technology	
Analyst		Like	Information	Internet	
Join		Look	Graduate	Apple	
Director		Today	Prime	Mobile	
Prior		Trade	Legal	Software	
CS: -0.85		**CS**: -1.07	**CS**: -0.85	**CS**: -1.49	

Table 5. Due Diligence Sample in Perspective

This table reports difference tests at a point in time (boundary) of three potential decisions: whether to (1) start due-diligence on the fund, (2) select the fund as eligible for investment—"Selected for Investment?", or (3) invest in a fund from the selected subset—"Actually Invested?". The boundaries are defined according to whether the allocator made a positive ("Yes") or negative ("No") decision at any point in time during our sample. Each fund where the allocator's decision was positive is matched to a group of 3 peer funds where the allocator's decision was negative. These peer funds are matched according to the Mahalanobis distance calculated using a fund's log(AUM), age, and past 24 month information ratio as of the calendar month of the positive decision. Tests are conducted at each boundary, comparing either the 24-month forward excess return and Sharpe ratio (24-month return over risk free divided by 24-month volatility) between funds with a positive decision and the three matched funds. P-values listed are for T-tests on mean and Ranksum Test for median value differences. For consistency, we condition the due diligence sample to have matched peers at the first boundary—the due diligence start date. The numbers for started due diligence (selected) [invested] at 675 (162) [77] are therefore lower than the 860 (214) [114] from the main analysis.

Panel A. Future Monthly Average Jensen's Alpha

	Mean				Median			
	Yes	No	Diff	P-value	Yes	No	Diff	P-value
Started Due-Diligence?	0.0012	0.0032	-0.0020	0.0069	0.0021	0.0028	-0.0008	0.0042
Selected for Investment?	0.0016	-0.0016	0.0032	0.0030	0.0027	-0.0011	0.0038	0.0000
Actually Invested?	0.0038	0.0037	0.0000	0.9797	0.0040	0.0036	0.0004	0.5007

Panel B. Future Monthly Sharpe Ratio

	Mean				Median			
	Yes	No	Diff	P-value	Yes	No	Diff	P-value
Started Due-Diligence?	0.1097	0.1582	-0.0485	0.1128	0.0858	0.0680	0.0178	0.1678
Selected for Investment?	0.1739	0.0492	0.1248	0.0155	0.1261	0.0231	0.1030	0.0122
Actually Invested?	0.0911	0.1334	-0.0423	0.2280	0.0900	0.1461	-0.0561	0.3268

Table 6. Intensity of Private Information Acquisition

This table reports the OLS regression estimates for equation 10 in which the dependent variable is the number of words in notes with length greater than 25 words. This cutoff is discussed in Section 3.4. Our proxy for x_t is the past 24-month peer-adjusted return ($\hat{E}_t\,(R-peer)$) and for r_t the past 24-month variance of peer-adjusted returns. $\tau_{0,t}$ is an indicator for the prior precision and takes the value of one during the non-shaded periods and zero during the shaded periods in Figure 2 Panel D. All RHS variables are standardized by their means and standard deviation. The LHS variable, $s_{i,t}$ is only demeaned. Since the variable reflects the log of the number of words in a note, the coefficients represent the percentage change in private information acquisition from a one-standard-deviation change in a given regressor. Reported t-statistics are robust to heteroskedasticity in models 1–4 and robust to fund-level clustering in model 5. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. See Appendix C for variable definitions.

	(1)	(2)	(3)	(4)	(5)
$\hat{E}_t(R-peers)$	0.040**	0.037*	0.064***	0.037	0.037
	[2.00]	[1.83]	[3.03]	[1.53]	[1.48]
r_t	−0.063***	−0.059***	−0.064***	−0.067***	−0.062***
	[−2.98]	[−2.80]	[−3.07]	[−3.18]	[−2.80]
$\tau_{0,t}$		−0.068**	−0.071**	−0.090***	−0.083**
		[−2.04]	[−2.15]	[−2.60]	[−2.42]
$\hat{E}_t(R-peers)^2$			−0.282***	−0.239**	−0.228**
			[−2.91]	[−2.51]	[−2.40]
$\hat{E}_t(R-peers)\times\tau_{0,t}$				0.060*	0.057
				[1.73]	[1.63]
Fund affiliation	No	No	No	Yes	Yes
Meeting # FE	No	No	No	No	Yes
Observations	1,356	1,356	1,356	1,356	1,356
R^2	0.0126	0.0153	0.0209	0.0343	0.0535
F-stat($\hat{E}(R-peers)$ variables)				5.48***	4.95***

Table 7. Quantity of Private Information Acquisition

This table reports the OLS regression estimates for equation (11), in which the dependent variable is the number of meetings. Our proxy for x_t is the past 24-month peer-adjusted return ($\hat{E}_t\,(R - peer)$) and for r_t the past 24-month variance of peer-adjusted returns. $\tau_{0,t}$ is an indicator for the prior precision and is one during the nonshaded periods and zero during the shaded periods in Figure 2 Panel D. RHS variables are then averaged over the full due diligence period for a given fund and standardized by the cross-sectional mean and standard deviation. Reported t-statistics are robust to heteroskedasticity. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. See Appendix C for variable definitions.

	(1)	(2)	(3)	(4)	(5)
$\hat{E}_t(R - peers)$	0.115*	0.096	0.119*	−0.021	−0.072
	[1.65]	[1.36]	[1.71]	[−0.15]	[−0.55]
r_t	−0.187**	−0.165**	−0.192**	−0.181**	−0.173**
	[−2.28]	[−1.97]	[−2.23]	[−2.11]	[−1.99]
$\tau_{0,t}$		−0.142**	−0.142**	−0.188**	−0.228***
		[−2.20]	[−2.18]	[−2.51]	[−3.11]
$\hat{E}_t(R - peers)^2$			−0.080	−0.042	−0.014
			[−0.94]	[−0.50]	[−0.18]
$\hat{E}_t(R - peers) \times \tau_{0,t}$				0.158	0.218**
				[1.43]	[2.01]
Fund affiliation	No	No	No	No	Yes
Observations	835	835	835	835	834
R^2	0.0120	0.0162	0.0169	0.0189	0.0588

Table 8. Time-Invariant Hazard of Selection

This table reports estimates of the Cox proportional hazard model for equation (12) on the cross section of funds that had a 24-month return history and at least one note with more than 25 words. The proxy for x_t, the public signal level, is the past 24-month peer-adjusted return ($\hat{E}_t\,(R - peer)$), and the proxy for r_t, the public signal precision, is the past 24-month variance of peer-adjusted returns. Both x_t and r_t are standardized. $\hat{\alpha}_{i,t}$ ($s_{i,t}$) is the text-based proxy for private signal level (precision). See Section 3.4 for construction details. For selected (unselected) funds, the proxies are computed as of the month preceding the selection date (the later of the last available sample or the truncation date, December 2017). The table reports odds ratios and t-statistics robust to heteroskedasticity across fund. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. See Appendix C for variable definitions.

	(1)	(2)	(3)	(4)	(5)
$\hat{E}_T\,(R - peers)$	1.688***	1.806**		1.444***	1.514*
	[4.76]	[2.45]		[4.13]	[1.65]
r_T	1.106	1.133		0.966	0.935
	[1.60]	[1.25]		[−0.57]	[−0.63]
$\hat{E}_T\,(R - peers) \times r_T$		0.931			0.986
		[−0.34]			[−0.05]
Private information:					
α_T			3.315***	3.302***	15.743***
			[12.32]	[12.02]	[4.76]
s_T			1.023	0.983	1.178
			[0.24]	[−0.17]	[1.31]
$\alpha_T \times s_T$					0.209***
					[−2.79]
Observations	802	801	805	797	796

Table 9. Time-Varying Hazard of Selection

This table reports estimates of the logistic discrete time hazard model. The sample includes all funds. Panel A reports selected odds ratios for five model specifications. Panel B reports odds ratios on the interaction terms for the specification in column 5 of Panel A. The proxy for x_t, the public signal level, is the past 24-month peer-adjusted return ($\hat{E}_t\,(R-peer)$), and the proxy for r_t, the public signal precision, is the past 24-month variance of peer-adjusted returns. Both x_t and r_t are standardized. $\hat{\alpha}_{i,t}$ ($s_{i,t}$) is the text-based proxy for private signal level (precision)—see Section 3.4 for details. *Duration* measures the number of months elapsed since the start of due diligence, as measured by the first meeting record in the database. Market variables include market returns (current and lags), rolling volatilities of Fama-French 4 factors, and predicted capital inflows to the allocator. For selected (unselected) funds, the proxies are computed for all months preceding the selection date (the later of the last available sample or the truncation date, December 2017). This table reports odds ratios and t-statistics clustered at the fund level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. See Appendix C for variable definitions.

Panel A. Baseline results

	(1)	(2)	(3)	(4)	(5)
Public information:					
$\hat{E}_t(R-peers)$	1.476***	1.501***	1.487***	1.470***	0.216***
	[4.38]	[4.69]	[4.57]	[4.31]	[−2.82]
r_t	0.936	0.843*	0.926	0.914	1.972
	[−0.61]	[−1.65]	[−0.60]	[−0.69]	[0.71]
$\hat{E}_t(R-peers)\times r_t$	1.023	1.023	1.031	1.016	1.013
	[0.93]	[0.89]	[1.33]	[0.43]	[0.26]
Log(AUM)		1.703***	1.909***	1.376***	1.398***
		[5.89]	[6.59]	[2.98]	[3.05]
Private information:					
Affiliated fund (D)				1.342	0.271
				[1.34]	[−0.94]
Affiliated college (D)				1.339	3.022
				[1.39]	[0.97]
α_t				2.479**	3.153
				[2.07]	[1.47]
s_t				1.585***	5.780**
				[3.32]	[2.17]
$\alpha_t \times s_t$				0.970	1.103
				[−0.07]	[0.21]
Due-diligence spell:					
log(*Duration*)	5.167***	4.657***	4.513***	4.089***	4.936**
	[5.46]	[4.76]	[4.64]	[4.05]	[2.39]
Duration	0.938***	0.942***	0.944***	0.951***	0.946
	[−4.62]	[−3.96]	[−3.78]	[−3.05]	[−1.63]
Market variables + Year FE	Yes	Yes	Yes	Yes	Yes
Strategy FE	No	No	Yes	Yes	Yes
Time interactions	No	No	No	No	Yes
Observations	30,487	27,812	27,812	27,208	27,208
F-stat(*added variables*)		34.66***	21.66***	118.13***	34.55***

Table 9. Time-Varying Hazard of Selection

(continued)

Panel B. Time interaction terms from specification 5

	Coefficient	T-Statistic	Description
$\alpha_t \times$ *Duration*	0.993	[−0.38]	
$\alpha_t \times$ Log(*Duration*)	0.940	[−0.19]	
$s_t \times$ *Duration*	1.052	[2.22]	More important later
$s_t \times$ Log(*Duration*)	0.423	[−2.04]	Diminished importance early
Affiliated fund (D) × *Duration*	0.915	[−2.50]	Diminished importance late
Affiliated fund (D) × Log(*Duration*)	3.652	[1.80]	More important early
Affiliated college (D) × *Duration*	1.018	[0.61]	
Affiliated college (D) × Log(*Duration*)	0.657	[−0.71]	
$\hat{E}_t(R - peers) \times$ *Duration*	0.958	[−2.73]	Diminished importance late
$\hat{E}_t(R - peers) \times$ Log(*Duration*)	2.795	[3.44]	More important early
$r_t \times$ *Duration*	1.008	[0.40]	
$r_t \times$ Log(*Duration*)	0.729	[−0.70]	

Table 10. Post-Selection Return Regressions

This table reports the results from estimating panel regression (16), which examines the post selection return of funds. The dependent variable is peer-adjusted excess returns of sets of four funds, with each set comprising a selected fund and three unselected funds matched by calendar time, log(AUM), age, and past rolling 24 month information ratio estimate as of the selection date. In columns 1 and 2 we run a fixed effects regression. In columns 3–5 we apply the recursive demeaned estimator described in Section 4.4, which instruments forward-demeaned quantities that involve AUM with their backward-demeaned values as in Pástor, Stambaugh, and Taylor (2015). α_T and s_T are our text-based proxies of private signal level and precision, respectively, for each fund. T is defined for each set of funds according to the selection date of the selected fund. *Duration* measures months elapsed after the end of due diligence under the assumption that each unselected fund was selected on the same date as the selected fund in its set. Reported t-statistics are robust to clustering at the fund level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. See Appendix C for variable definitions.

	(1)	(2)	(3)	(4)	(5)
log(Duration)	0.077^{**}	0.153^{**}	0.078^{**}		-0.400^{**}
	[2.22]	[2.32]	[2.23]		[−2.16]
$\mathbb{I}_{sel} \times \log(Duration)$	-0.276^{***}		-0.300^{***}		0.147
	[−4.77]		[−4.81]		[0.71]
$\mathbb{I}^{mid}_{\alpha_T \times s_T} \times \log(Duration)$		-0.241^{***}			
		[−3.17]			
$\mathbb{I}^{high}_{\alpha_T \times s_T} \times \log(Duration)$		-0.385^{***}			
		[−5.22]			
$\mathbb{I}^{mid}_{x_T \times r_T} \times \log(Duration)$		0.001			
		[0.01]			
$\mathbb{I}^{high}_{x_T \times r_T} \times \log(Duration)$		0.209^{***}			
		[2.76]			
$\log(\text{AUM})_{t-1}$				-1.248^{***}	-2.214^{***}
				[−3.40]	[−3.09]
Observations	39,953	32,193	39,953	39,297	39,297
(1): $\hat{E}(R-peers)\vert$ selected	0.667	0.245			
(2): $\hat{E}(R-peers)\vert$ unselected	-0.151	-0.241			
$Pr\{(1)=(2)\}$	0.000	0.000			

Appendix A. Anecdotal Evidence

In this appendix we review anecdotal evidence of the suggested due-diligence process based on interviews with our allocator's senior investment professionals and meeting notes from their interactions with an anonymized manager, fund *XYZ*. After seven meetings, the allocator selected fund *XYZ* for investment.

The allocator casts a wide net when sourcing funds in which it might invest. Initial contact with fund managers comes through two primary channels, the first of which is network relationships. As one senior manager at the allocator stated, "Most often an initial introduction is through people we know." Stressing the importance of a manager's network, the manager added, "In evaluating new managers, we want to know who they worked with and in what capacity." For example, *XYZ*'s chief investment officer (CIO) had been an analyst at a fund in which the allocator had previously invested. Prime brokers are the other common channel through which our allocator meets funds. Thanks to their business relationship with funds, prime brokers are able to provide dedicated capital introduction functions that directly reach out to asset allocators on behalf of funds, often through organized events and conferences.

For the allocator, the decision to initially meet with a fund is intentionally not algorithmic. There is no screening on fund size (AUM), returns, or track-record length. Instead, the allocator seeks to connect performance potential to "people, philosophy and process." As a senior manager at the allocator states, "We work to get past anecdotalism where managers make selective disclosures about trades and performances; it impresses us when a manager volunteers a discussion about a losing trade. This helps us understand the investment process and how the manager learns." This same individual notes, "There are a lot of subtleties in discussions around performance. We want to know the basic premise for how they make money, how their story compares to historical results, and if they are realistic in their assessment of performance."

The initial meeting between the allocator and fund is relatively short, usually lasting only 30 to 60 minutes, and occurs at a conference, via a video-conference call, or in a conference room at the allocator's office. This is in contrast to later meetings, which may occur at the hedge fund's office. After the initial meeting, a file is opened on the fund and includes any materials provided by the fund. In addition, an internal database entry is created that includes notes about the fund meeting by the employee who led the interview.

An almost universal first piece of private information about a fund comes in the form of the pitchbook (slides) presented in the first 15–20 minutes of the initial meeting. Pitchbooks tend to follow a standard format. The first few slides of fund *XYZ*'s pitchbook highlight historical milestones of the fund, organizational charts, and the backgrounds of the portfolio managers. The next ten slides discuss *XYZ*'s investment process: idea generation, portfolio construction, trade execution, and risk management. The general theme of this section is differentiation—what makes the fund's philosophy and process different and how this translates into an investment edge. The final section provides snapshots of the fund's portfolio (e.g., returns, and country and sector allocations).

Initial meetings with funds tend to focus heavily on the backgrounds of the fund managers and how their employees interact with one another. As the allocator's CIO states, "No one is born with pure investment talent; it usually takes deliberate practice under a good coach to become a good investor." Furthermore, the CIO "wants managers with confidence in their people and process. [They] appreciate the importance of how [various support functions] enhance the investment process." The allocator frowns heavily on managers that "exaggerate experience and do not give credit to the team or mentors." The allocator also does an in-depth analysis of the reasons a manager left his or her previous fund as a way to understand the fund's management style. For example, the first set of notes for fund *XYZ* reflects conversations about the reasons *XYZ* CIO thinks his previous fund was unsuccessful and what he would do differently: "[He] believes [that the previous fund] grew too big, too fast...and [that] the bulk of people that invested in [the previous fund] had an asset/liability mismatch, resulting in [their] inability to hold positions during crunch times."

According to the allocator, whether subsequent meetings are scheduled is determined largely by their perception of fund quality after the initial meeting. The topics of subsequent meetings shift from background to infrastructure, the economic specifics of the fund, and the philosophy behind the fund's investments. For example, the second meeting note for *XYZ* points out that "[the CIO] has put in about 1/3 of his personal net worth to fund operations for about two years. In his words, enough for him to care about, but not enough to lose sleep over." Additionally, "[the CIO] has the wealth and contacts to hire the right people and the [current] team seems impressive at first blush." These statements highlight the importance the allocator places on incentives when choosing a manager. Is the manager still hungry for success? Is there too much or too little personal skin in the game? And how do these incentives influence the operations of the fund?

As one partner at the allocator points out, as meetings progress "digging deeper into the key themes of people, philosophy, and process [is] essential." Philosophy covers investment (e.g., value versus growth, momentum versus mean reversion) and long-run themes (e.g., macroeconomic, sectoral, or position-specific issues) that inform a manager's portfolio. For fund *XYZ*, meetings 3 and 4 focus more on philosophy. Discussions include topics such as how investments are chosen for the portfolio, for example, "[The CIO] separates himself from [the previous fund] as more of a stock-picker versus one that would call markets," and "longs for [*XYZ*] need the proper balance-sheet and working capital for the business as it looks to shift from low to high margin business lines." In addition, current investment themes are discussed, for instance, *XYZ* sees their main long themes as "power generation in India with the country having a power deficit of 15%" and "consumer durables in China with the government pushing incentives to spend."

Process, on the other hand, covers how risk management is woven into allocations, and how the fund's institutional infrastructure is used in idea generation and thesis formation. Here, too, the allocator scrutinizes how process is reflected in past performance. As the allocator's CIO points out, "In evaluating the manager's process, we want to understand what types of risks they are comfortable with, how they define and measure risk, and how this is expressed in manager actions in a variety of market scenarios."

Appendix B. Model Analysis

At the inception of the due diligence process, the allocator assumes a prior belief about the distribution of $\alpha \sim N(0, 1/\tau_0)$. The allocator receives a passive, noisy signal $x \sim N(\alpha, 1/r)$ about the fund's α. Assuming the allocator is Bayesian, she will update his prior using this signal: $\hat{\alpha} \sim N\left(\frac{x \cdot r}{\tau_0 + r}, \tau_0 + r\right)$. We define τ as $\tau_0 + r$.

In addition, our allocator has an option to acquire private information. If she chooses to exercise that option she must decide the precision, s, with which to acquire it. This choice carries a cost function, $c(s)$, with standard properties, i.e., $c'(s)$, $c''(s) > 0$.

Normalizing the value of waiting for additional public information to zero, the allocator thus maximizes $V = \max\{0, V_{select}, V_{meet}\}$, where 0 is the value of doing nothing, V_{select} is the value of selecting the fund using $\hat{\alpha}$, and V_{meet} is the expected value of the fund given the option to meet. Henceforth we call V the choice set. Assuming V_{select} is a linear function of α, with K fixed cost and A leverage, estimated using public information only (i.e., $V_{select}(\hat{\alpha}) = -K + A\hat{\alpha}$), the meeting adds value in expectation because there exists the option either to reject the fund when $A \times \hat{\alpha}' < K$ in the next period or to select it.

Given this setup, the choice to acquire private information is given by

$$V_{meet} = \max_{\text{w.r.t. } s} -c(s) + \beta \int_{Y_c(s)}^{\infty} \left(-K + A \cdot \frac{\tau \cdot \hat{\alpha} + s \cdot y}{\tau + s}\right) f(y|\hat{\alpha}, \tau, s)\, dy, \tag{B.1}$$

where $Y_c = \frac{K \cdot (\tau + s)}{A \cdot s} - \frac{\tau \cdot \hat{\alpha}}{s}$ and $f(y|\hat{\alpha}, \tau, s) = \sqrt{\frac{H}{2\pi}} \exp\left((H/2)(y - \hat{\alpha})^2\right)$. H, $\frac{s \cdot \tau}{s + \tau}$, is the precision of y given $\hat{\alpha}$ and τ, and the optimal intensity of meeting. This single-period model leads to our first proposition.

Proposition 1: The optimal precision of private information signal acquisition, s^*, is
(i) decreasing in the public information signal precision, r_t;
(ii) decreasing in the prior precision, $\tau_{0,t}$; and
(iii) concave in the signal, x, around zero assuming $r_t > \tau_{0,t}$;
(iv) increasing in the signal, x, if the prior mean is less than zero, primarily when $\tau_{0,t}$ is high.

Proof: Using the properties of the standard normal distribution with a density $\phi(\cdot)$, we can solve for the optimal precision, s^*, and the value of meeting, $V_{meet}(s^*)$, by maximizing the above equation.

The first-order condition, assuming for simplicity that $A = 1$ and $K = 0$,

$$c'(s) = \frac{\beta\tau}{2(\tau+s)^2\sqrt{H}}\phi\left((Y_c - \hat{\alpha})\sqrt{H}\right), \tag{B.2}$$

yields comparative statics for a range of parameter values (see fig. 2) summarized in proposition 1.

We then extend the one-period model to infinite time dynamic programming problem,

$$V(\hat{\alpha}) = max\{V_{public}, V_{select}(\hat{\alpha}), -c(s) + \beta E\left(V'(M(\hat{\alpha}))\right)\}, \tag{B.3}$$

where $V'(M(\hat{\alpha}))$ is the continuation value of having a meeting and $M(\hat{\alpha})$ is the pre posterior expected value. V_{public} is the allocator's expected value from informing her decision using only public information going forward. This highlights that our allocator will never outright reject a fund, given that public information is costless.

Lemma 1: First, the expected value of collecting one more piece of private information is nonincreasing in time.

Proof: From the choice set we know that the expected change in value from one more piece of information is

$$-c\left(s_j^*\right) + \beta\hat{\alpha}_j\Phi\left(-\left(Y_{j,c}^* - \hat{\alpha}_j\right)\sqrt{H_j}\right) + \frac{\beta\sqrt{H_j}}{\tau_j}\phi\left(\left(Y_{j,c}^* - \hat{\alpha}_j\right)\sqrt{H_j}\right) - V_{select}\left(\hat{\alpha}_j\right). \tag{B.4}$$

As shown in Figure 2, as the precision of the prior increases the optimal precision at which the allocator acquires private information, s^*, falls. Since precision (both public and private) is an additive function, as $j \to \infty$, $s_j^* \to 0$. Taking the limit of the argument $\left(Y_{j,c}^* - \hat{\alpha}_j\right)\sqrt{H_j}$ as $s_j^* \downarrow$ shows that the cumulative distribution function and probability density function decrease monotonically to zero. Thus the benefit of an additional piece of private information monotonically falls.

Proposition 2: There exists an expected point in due diligence, $T(\mathbf{s}) \leq \infty$, which captures the stopping point for private information acquisition. As $T(\mathbf{s})$ is a function of precisions, and precisions are additive, assuming $\hat{\alpha} > 0$, the conditional probability of selection (hazard rate) is increasing in the precision of both public and private information.

Proof: From Lemma 1, taking $s_j^* \downarrow$ to zero we see that the additional piece of information monotonically falls to $-c(c_0) - V_{select}(\hat{\alpha}) < 0$. There will thus be a point in time, $T(\mathbf{s})$, after which the allocator will expect to no longer acquire more costly private information.

Proposition 3: For each time t there exists an $\underline{\alpha}_t$ and $\overline{\alpha}_t$ such that the allocator continues to acquire private information if $\underline{\alpha}_t \leq \hat{\alpha} \leq \overline{\alpha}_t$, selects the fund if $\widehat{\alpha} \geq \overline{\alpha}_t$, and expects to do nothing (i.e., continues to collect public information) if $\underline{\alpha}_t \geq \widehat{\alpha}$.

Proof: As max $\{0, V_{select}(\hat{\alpha})\}$ is a piecemeal convex function, from Jensen's inequality $E(V'(M(\hat{\alpha})))$ dominates with derivatives greater than zero but less than A $\forall\ \hat{\alpha}$. Thus, with the introduction of $c(s)$ and β, when $j < T(\mathbf{s})$ the continuation value meets max $\{0, V_{select}(\hat{\alpha})\}$ twice, at $\underline{\alpha}_j$ and $\overline{\alpha}_j$, respectively.

Proposition 4: Given $T(\mathbf{s}) > 1$ and an optimal aggregate precision of private information, the cumulative cost of private information acquisition can be decreased via multiple meetings if (i) the variable portion of meeting cost is sufficiently convex, (ii) the fixed portion of meeting cost is sufficiently low, and (iii) the loss of one-period income α is not too high.

Proof: Assuming $c(s)$ has a fixed and a variable portion, if the variable portion is sufficiently convex and/or the fixed portion sufficiently low, then the allocator can decrease information acquisition costs by simply spreading meetings over multiple periods. This assumes that the one-period alpha is relatively stable in time and that the opportunity cost is sufficiently low to miss one period of returns.

Appendix C. Key Variable Definitions

Variable Name	Description
$\hat{E}_t(R-peers) \equiv x_t$	24-month moving average of the fund return in excess of the matched HFRI strategy benchmark (equity hedge, emerging market, equity market neutral, relative value) *Tables 1, 6, 7, 8, 9. Sources: Allocator, Morningstar, Barclay Hedge, eVestment, HFR, TASS.*
τ_{0t}	is an indicator variable based on the cross-sectional variance of fund returns for any given month. Takes value of 1 the aftermath of the Asian financial crisis, the dot-com bust, the financial crisis, and the European sovereign debt crisis as indicated by the shaded regions in Figure 2, Panel D. *Tables 1, 6, 7, 8, 9. Sources: Allocator, Morningstar, Barclay Hedge, eVestment, HFR, TASS.*
r_t	the inverse of the 24-month moving variance estimate of the fund excess returns, where excess return is the difference between fund return and the matched HFRI strategy benchmark (equity hedge, emerging market, equity market neutral, relative value) *Tables 1, 6, 7, 8, 9. Sources: Allocator, Morningstar, Barclay Hedge, eVestment, HFR, TASS.*
s_t	Our fund-month level proxy of private signal precision, defined as $s_{t+1} = \log\left(\sum_{i=1}^{\#\text{ Meetings}\leq t} \exp\left(s_{i,t}\right)\right)$ where $s_{i,t}$ is the log number of words in the meeting note i recorded in the allocator's database. See sections 3.3 and 3.4 for filter details and discussion. *Tables 1, 6, 7, 8, 9. Sources: Allocator.*
α_t	Our fund-month level proxy of private signal quality, defined as $\hat{\alpha}_{t+1} = \frac{s_{i,t}\cdot y_{i,t}+\hat{\alpha}_t\cdot s_t}{s_t+s_{i,t}}$, where $s_{i,t}$ and s_t are as defined above and y_t is the the log of number of months until the next meeting orthogonalized with respect to the number of words in the previous meeting. See sections 3.3 and 3.4 for filter details and discussion. *Tables 1, 6, 7, 8, 9. Sources: Allocator.*
Duration	In table 9 is a count of months elapsed since the start of due diligence relatively to the first meeting record in the allocator database. In table 10 is a count of months elapsed after the end of due diligence under the assumption that each unselected fund was selected on the same date as the selected fund in its set, and for which the returns are available in the public databases *Sources: Sources: Allocator, Morningstar, Barclay Hedge, eVestment, HFR, TASS.*
Affiliated college, Affiliated fund	Indicator variables for whether the hedge fund is a spin-off from a previous investment by the allocator, or at least one senior employee (or partner) at the allocator attended the same college as the hedge fund senior employee (or partner). *Table 7, 9. Sources: Allocator employee bios and Meeting Notes, HF Pitchbooks*
$\mathbb{I}_{sel}$	Indicator variable for the fund-months after the date that the fund was designated as '*selected to investment universe*' in the allocator's database. *Tables 2, 3, 6, 7, 8, 9, 10. Sources: Allocator.*
$\mathbb{I}^{high}_{\alpha_T\times s_T}$, $\mathbb{I}^{mid}_{\alpha_T\times s_T}$ $[\mathbb{I}^{high}_{x_T\times r_T}, \mathbb{I}^{mid}_{x_T\times r_T}]$	The partition of $\mathbb{I}_{sel}$ dummy into terciles based on the level of private [public] signal and precision thereof, such that *high* indicates subset of selected funds in the top terciles by both the level, α_t $[x_t]$, and the precision, s_t $[r_t]$, of private [public] signal as of the selection date. *Tables 10. Sources: Allocator.*

Online Appendix. Finding Fortune: How Do Institutional Investors Pick Asset Managers?

In this appendix we provide additional analysis and robustness to the results reported in the main body of the paper. In section IA.1 we conduct textual analysis of pitchbooks and meeting notes using word counts similar to those in Loughran and McDonald (2011). In section IA.2, we provide technical details of the Latent Dirichlet Allocation estimation. In section IA.3, we present results and discuss problems of modeling the selection decision ignoring the baseline hazard (due diligence time) component. In section IA.4, we present results from modeling the time-to-decision using an accelerated hazard or OLS model, and demonstrate the bias generated by the always-at-risk assumption underlying this model. Finally, in section IA.5, we provide additional nonparametric statistical evidence of the α generated by selected versus unselected funds on selection date, and the subsequent decay of α as post selection time passes.

IA.1. Word Counts

In this section we present the results of a word count on the sample of 2,689 pre selection meeting notes using the financial word lists of Loughran and McDonald (2011), focusing on positive, negative, and uncertain proportions. In addition, we append the words CONSTRUED, HEDGE, HEDGING, LIQUIDITY, CASH, LEVERAGE, COMPLIANCE, and BETA to the uncertain word list. On average each note has 304 words, but we consider only notes with more than 25 words. Table D.1 lists the most commonly cited words in order of frequency from each of the 3 lists using this subsample of notes.

Both positive and negative words generally reflect discussions about portfolio themes. Positive words tended to be associated with descriptions of long positions, while negative words are associated with both past portfolio losses, short position descriptions, and discussions of how the fund managers learns from mistakes. Uncertain words appear in discussions about inconsistencies in pitches, lack of investment ideas, or decisiveness to deploy capital (asset hoarding). The proportion of positive and negative words are statistically larger for notes from selected funds on average. This along with the statistically confirmed longer average note size, imply that notes reflect the degree to which the allocator is scrutinizing the fund rather than the allocator's sentiment towards the fund.

Unlike meeting notes, pitchbooks are written by the hedge fund managers. Our sample includes pitchbooks from 677 funds. On average, there are 3,375 words per pitchbook. The pitchbooks are dedicated to a managers' experience, fund history, investment philosophy, current themes or positions, and risk management. We repeat the textual analysis done on the meeting notes with the text from the pitchbooks (see Table D.1). The key takeaway from this analysis is that while even simple word counts reflect differences in content between notes and pitchbooks, they seem to miss their context-specific information.

Figure D.1. Word counts in Meeting Notes

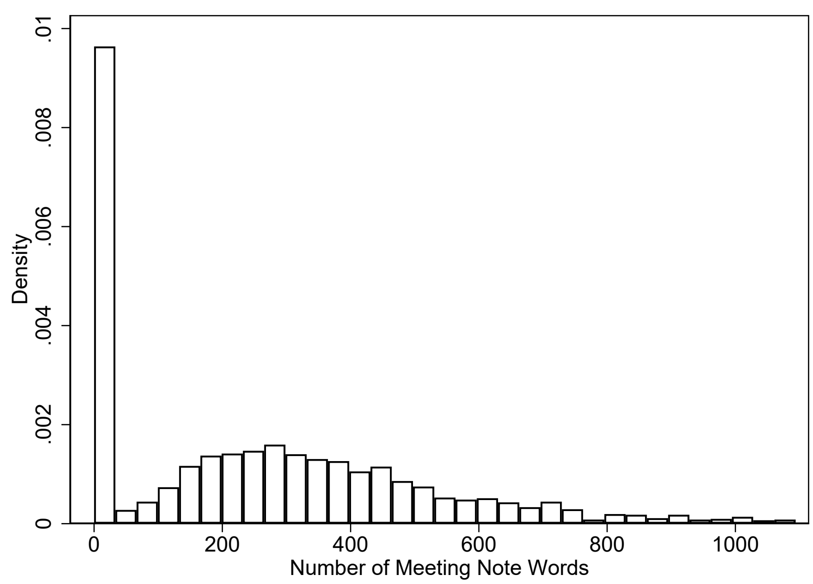


Table D.1. Top Words: Meeting Notes and Pitchbooks

This table reports the top words from the three words lists (positive, negative and uncertain) of Loughran, et al (2011) applied to our meeting notes and pitch books. The uncertain list is expanded with the additional words CONSTRUED, HEDGE, HEDGING, LIQUIDITY, CASH, LEVERAGE, COMPLIANCE, BETA. The two sources of private information are listed separately to show differences in their content.

Rank	Positive		Negative		Uncertain	
	Pbook	Notes	Pbook	Notes	Pbook	Notes
1	Opportunities	Good	Volatility	Volatility	Exposure	Exposure
2	Strong	Opportunities	Loss	Cut	Risk	Risk
3	Opportunity	Strong	Losses	Negative	Hedge	Liquidity
4	Attractive	Great	Conviction	Closed	Liquidity	Cash
5	Positive	Better	Negative	Against	Volatility	Hedge
6	Greater	Positive	Distressed	Bad	Cash	Could
7	Reward	Able	Deviation	Conviction	Leverage	Believes
8	Advantage	Opportunity	Opportunistic	Losses	Risks	Exposures
9	Leading	Attractive	Disclaimer	Late	Compliance	Volatility
10	Superior	Greater	Restructuring	Lost	Exposures	Beta
11	Achieved	Despite	Against	Difficult	Beta	Leverage
12	Good	Advantage	Poor	Hurt	Hedging	Believe
13	Achieve	Benefit	Volatile	Claims	Approximately	Risks
14	Gains	Excited	Stress	Loss	Could	Roughly
15	Successful	Reward	Closed	Crisis	Believe	Hedging
16	Better	Winners	Lack	Poor	Vary	Approximately
17	Highest	Gains	Lose	Distressed	Deviation	Might
18	Benefit	Strength	Construed	Wrong	Assumptions	Almost
19	Honors	Highest	Decline	Decline	Speculative	Seems
20	Success	Stable	Illiquid	Lose	Believed	Cautious
21	Able	Leading	Critical	Illiquid	Differ	Possible
22	Transparency	Constructive	Bankruptcy	Problem	Volatile	Probably
23	Profitable	Outperformed	Ill	Problems	Believes	Volatile
24	Gain	Improving	Crisis	Weak	Possible	Assuming
25	Profitability	Gain	Late	Concerns	Assumed	Depending
26	Great	Profitable	Disclosed	Restructuring	Depending	Uncertainty
27	Enhanced	Successful	Worst	Worst	Preliminary	Dependent
28	Stable	Confident	Weak	Lack	Anticipated	Compliance
29	Improving	Outperform	Inefficiencies	Forced	Assuming	Anticipates
30	Advantages	Happy	Declining	Volatile	Probability	Somewhat
31	Effective	Success	Conflicts	Concerned	Assume	Anticipate
32	Favorable	Improve	Exposed	Defensive	Assumes	Anticipated
33	Despite	Favorable	Claims	Slowing	Almost	Apparently
34	Strength	Easy	Bad	Correction	Nearly	Probability
35	Efficient	Winner	Limitations	Slowdown	Dependent	Sometimes
36	Excellent	Strengths	Force	Slow	Might	Occasionally
37	Outstanding	Improvement	Breakdown	Tightening	Hidden	Maybe
38	Outperformed	Transparency	Unlawful	Missed	Approximate	Perhaps
39	Improved	Optimistic	Difficult	Exposed	Variant	Possibility
40	Distinction	Rebound	Deteriorating	Recession	Anticipate	Risky

IA.2. *Latent Dirichlet Allocation Details*

The Latent Dirichlet Allocation's (LDA's) likelihood function is constructed assuming the following data generating process (dgp) (see Blei, Ng, and Jordan (2003) and Figure D.2):

1. M documents compose the corpus. The documents are assumed to be a random mixture of K-topics that are Dirichlet distributed (i.e., $\Theta \sim Dir$ (K-parameter vector α)). Each topic is assumed to be a mixture of a V-character vocabulary that are likewise Dirichlet distributed ($\beta \sim Dir$ (V-parameter vector η)).

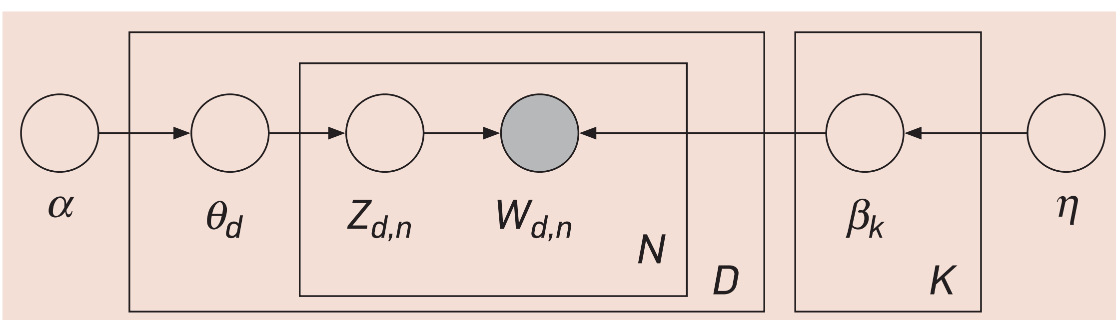
2. Each document, d, is of N_d words where $d \in \{1,...,M\}$. A random topic mixture, θ_d, is chosen from Θ.

3. For each word, n_d, in d a topic is then randomly selected from θ_d. This topic index is denoted z_{d,n_d}, where $n_d \in \{1,...,N_d\}$. Given that a Dirichlet is a distribution over the possible parameter vectors for a multinomial distribution $z_{d,n_d} \sim multinomial\,(\theta_d)$.

4. Finally for each topic–word index, z_{d,n_d}, a word, w_{d,n_d}, is selected according to the associated topic–vocabulary mixture, $\beta_{z_{d,n_d}}$. Thus, $w_{d,n_d} \sim multinomial\left(\beta_{z_{d,n_d}}\right)$.

Figure D.2. LDA Plate Notation

This figure represents the dgp assumed by LDA. Each document, d, of a M-document corpus is a composite of K-topics, θ_d. θ_d is chosen from a Dirichlet distribution with parameter α. Each word in the document is given a topic index, $z_{d,n}$, that is determined by θ_d. Each topic has a certain distribution over the vocabulary, β_k. Each word, which is the only observable, is randomly chosen according to this distribution. The distribution of all topics over the vocabulary is also a Dirichlet distribution with parameter η.



This setup implies, first, that the number of topics K is fixed and chosen before estimation. Second, the vocabulary is also fixed; e.g., there is no ability to see how the vocabulary changes through time. Third, while certain words appear with high probability together (how topics are deciphered), the algorithm still follows the "bag of words"-assumption. That is, the words appear together not because they are dependent on one another, but because they have been allocated to the same topic through any inference algorithm. The joint distribution is given by,

$$P(W,Z,\Theta,\beta|\alpha,\eta) = \prod_{k=1}^{K} P(\beta_k|\eta) \cdot \prod_{d=1}^{M} P(\theta_d|\alpha) \cdot \prod_{n_d=1}^{N_d} P\left(z_{d,n_d}|\theta_d\right) \cdot P\left(w_{d,n_d}|\beta_k\right)$$

Z, Θ and β are latent variables that we will infer using Bayesian Inference. Specifically, we are interested in inferring them given the corpus' words (observables) and priors, α and η.

$$P(Z,\Theta,\beta|W,\alpha,\eta) = \frac{P(W,Z,\Theta,\beta|\alpha,\eta)}{P(W|\alpha,\eta)}$$

As with many Bayesian inference problems the denominator of this RHS is intractable. This is because it requires the joint determination of Θ and β.

$$\begin{aligned} P(W|\alpha,\eta) &= \int_{\beta}\int_{\Theta}\sum_{z} P(W,Z,\Theta,\beta|\alpha,\eta)\, d\Theta d\beta \\ &= \int_{\beta} P(\beta|\eta)\int_{\Theta} P(\Theta|\alpha)\sum_{z} P(Z|\Theta)\, P(W|Z,\beta)\, d\Theta d\beta \end{aligned}$$

The second equation uses the assumption that the distributions of β and Θ are independent of one another. W on the other hand is determined by both latent variables. There are two primary methods with which the denominator is estimated: Gibbs sampling and Bayesian variational inference. We use a variational inference method developed by Hoffman, Blei and Bach (2010) and coded in Python by Radim Řehůřek and Petr Sojka (2010) for our main results.

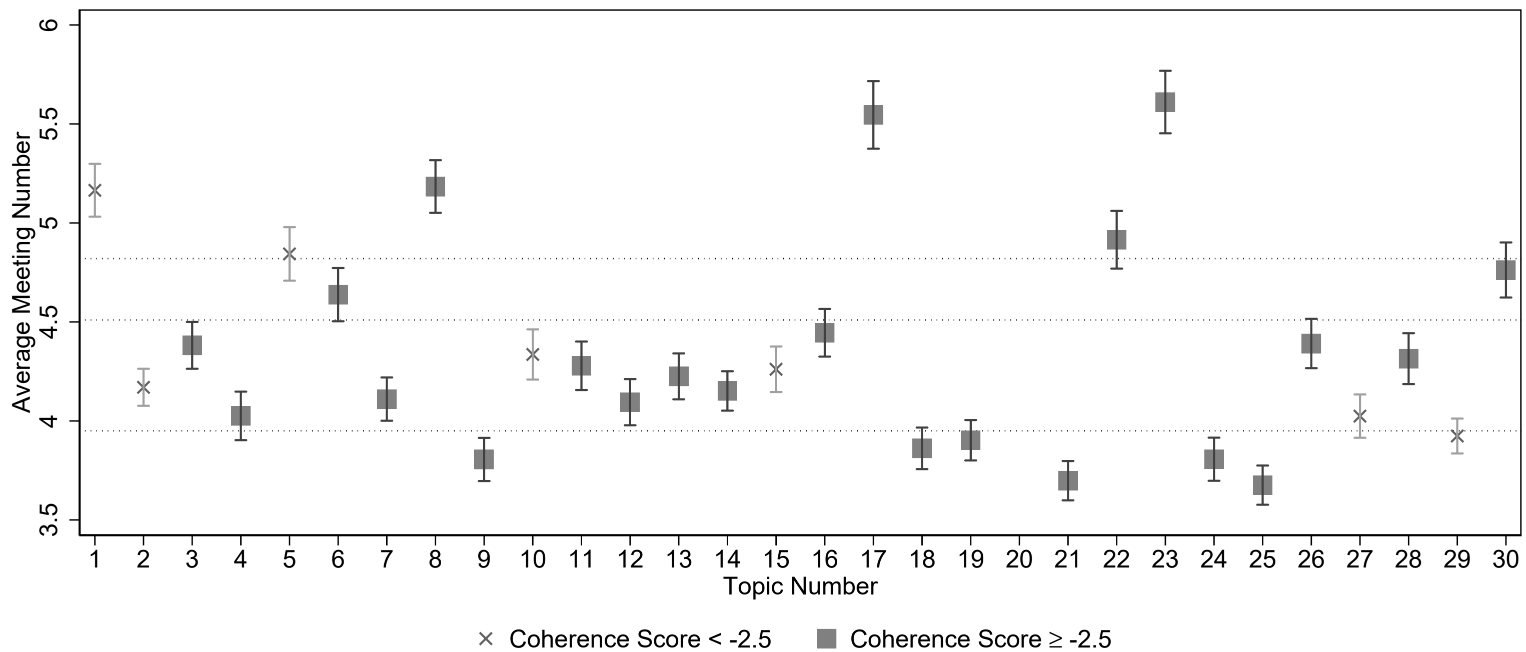
The optimal number of topics chosen was computed using the UMass coherence measure. The measure begins by computing, for each topic–word pair (v,i,j), a normalized number describing the number of documents ($D\left(v_i, v_j\right)$) over which both words appear.

$$score_{v,i,j} = log\frac{D\left(v_i, v_j\right)+\varepsilon}{D\left(v_j\right)}$$

This score is summed over the entire vocabulary for each topic, and then averaged over all topics for an aggregate model score. The model with the highest aggregate score (30 topics) was chosen for our baseline results. The algorithm tradesoff having too few topics such that the LDA is unable to separate words and maximize topic coherence with having too many topics such that the words in each topic rarely occur within the same document. In addition to looking at the aggregate score, we look at the individual topic scores, which are listed in Table 4. The cutoff between where a topic-header or topic-title goes from being easy to difficult to ascribe occurs around a topic coherence score (CS) of around -2.5.

Figure D.3. Cross sectional Timing of LDA Topics

This figure illustrates the timing of the LDA topics tabulated in Table 4 in a box plot. The average meeting number for each LDA-computed topic is estimated over our pre due diligence corpus. The weighted-average methodology is described in section 3.3. The horizontal lines uses the same methodology but averages over the allocator suggest timing of topics. These estimates are explored further in Figure 4, Panel A.



IA.3. *Decision to Invest, Ignoring Time*

To motivate the need to model the investment decision as a hazard function we run a regression on the selection decision for fund i at time t, ignoring due diligence time.

$$Selection_{i,t} = \mu + \beta_x \cdot x_{i,t} + \beta_r \cdot r_{i,t} + \beta_\alpha \hat{\alpha}_{i,t} + \beta_s s_{i,t} + \varepsilon_{i,t}, \tag{IA.1}$$

where $x_{i,t}$ and $r_{i,t}$ are defined as in section 4.1, and $\hat{\alpha}_{i,t}$ and $s_{i,t}$ are defined by equations 8 and 9. Our first set of results are on a purely cross sectional basis: each fund has only one observation. Results are presented in Table D.2. All regressors are standardized.

For Panel A, we use values for all funds in our sample at the end of due diligence (T). For funds that were selected, T is defined by their due diligence start date and selection date, whereas for funds unselected T is defined by the due diligence start date and either the end of the dataset (December 2017) or the fund's censored date (i.e., date fund drops from dataset). Overall, the decision to select seems to be driven by our public and private information proxies. One potential issue is our definition of T, which mixes together two decisions: (a) whether to select a fund and (b) conditional on selection, the speed at which the decision is made.

To address this, in Panel B, we follow the same procedure highlighted in the main paper, matching each selected fund to a maximum of 3 unselected funds based on their Mahalanobis distance using calendar time, log(AUM), age and the past information ratio. In this specification, T is defined by the due diligence start date and the selection date for both selected funds and their matched counterparts. Surprisingly, both the public information signal and precision now significant coefficients in the direction opposite of that predicted. In column 2 we add the information ratio. In columns 3, we look exclusively at private information; both the private information signal and precision are statistically significant and positive. In columns 4 and 5, we add to the private and public information proxies.

Finally, the choice to not select a fund, even if that fund is eventually selected, has significant informational content. This is especially true because we can never be fully sure of the validity of any matching algorithm. We thus take equation IA.1 to the full panel of data. Results are presented in Table D.3. All regressors are standardized and standard errors are robust to clustering by fund. The results are inline with those presented in Table D.2. One constant theme through these set of regressions, however, is the negative coefficient on the interaction of our private signal and precision (i.e., private information ratio). We believe this is due to model misspecification–i.e., we are not including the fact that acquiring private information is *time* intensive.

Table D.2. Probability of Selection Regressions

This table reports the results from regression IA.1 using only cross sectional data. Reported t-statistics are robust to heteroskedasticity. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. All Funds

	(1)	(2)	(3)	(4)	(5)
Public information:					
$\hat{E}_T(R-peers)$	0.075***	0.066***		0.054***	0.048**
	[4.82]	[2.81]		[4.04]	[2.46]
r_T	−0.014	−0.015		−0.023	−0.025
	[−0.65]	[−0.72]		[−1.33]	[−1.48]
$\hat{E}_T(R-peers)\times r_T$		0.011			0.007
		[0.49]			[0.37]
Private information:					
α_T			0.164***	0.158***	0.214***
			[12.45]	[11.97]	[3.17]
s_T			0.019	0.013	0.009
			[1.45]	[0.98]	[0.57]
$\alpha_T \times s_T$					−0.056
					[−0.86]
Observations	802	801	805	797	796
R^2	0.0404	0.0405	0.1904	0.2127	0.2136

Panel B. Matched Sample

	(1)	(2)	(3)	(4)	(5)
Public information:					
$\hat{E}_T(R-peers)$	0.005	−0.022		−0.024	−0.036**
	[0.27]	[−1.01]		[−1.63]	[−1.97]
r_T	−0.031*	−0.037**		−0.037**	−0.038**
	[−1.71]	[−2.02]		[−2.40]	[−2.52]
$\hat{E}_T(R-peers)\times r_T$		0.041***			0.022*
		[2.80]			[1.66]
Private information:					
α_T			0.254***	0.257***	0.371***
			[20.29]	[20.45]	[8.78]
s_T			0.047***	0.044***	0.050***
			[3.24]	[3.03]	[3.67]
$\alpha_T \times s_T$					−0.119***
					[−2.70]
Observations	766	766	733	733	733
R^2	0.0054	0.0107	0.3116	0.3182	0.3251

Table D.3. Probability of Selection Panel Regression

This table reports the results from regression IA.1 using the full panel (month–fund) data. Reported t-statistics are robust to clustering by fund. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Public information:					
$\hat{E}_t(R - peers)$	0.033***	0.031***		0.028***	0.025**
	[3.73]	[2.98]		[3.33]	[2.57]
r_t	−0.001	−0.002		−0.005	−0.007
	[−0.14]	[−0.24]		[−0.57]	[−0.83]
$\hat{E}_t(R - peers) \times r_t$		0.007			0.008
		[1.09]			[1.19]
Private information:					
α_t			0.068***	0.108***	0.104***
			[7.24]	[2.83]	[2.74]
s_t			0.006	0.004	−0.002
			[0.72]	[0.45]	[−0.24]
$\alpha_t \times s_t$				−0.040	−0.038
				[−1.06]	[−1.03]
Fund Affiliation	No	No	No	No	Yes
Observations	36,097	35,954	38,478	35,274	35,140
R^2	0.0118	0.0131	0.0529	0.0660	0.0767

IA.4. Accelerated Hazard Model

We start by assuming the hazard of selection follows an accelerated model. Our hazard event is a successful passage of the due diligence process by a fund. The spell begins at the first meeting, when a pitchbook is sent or presented. The end of a spell is either the date of selection to the investment universe or censoring of a fund. We define the due diligence period, T, as the difference in months between the end and beginning of spell. Our accelerated hazard specification for fund i is thus

$$\ln T_i = \mu + \beta_x \cdot x_{i,T} + \beta_r \cdot r_{i,T} + \beta_\alpha \hat{\alpha}_{i,T} + \beta_s s_{i,T} + \varepsilon_{i,T}. \tag{IA.2}$$

Given that the due diligence period is only defined relative to the first meeting and selection date, regression IA.2 requires us to collapse the data by fund such that each fund has a single observation. Regressors are defined as in section 4.3, but only on the end of due diligence date. Table D.4, Panel A reports our estimates for the full sample of funds. All variables are standardized.

Interestingly, none of the public information proxies load on the selection spell. In Panel B we run our regression specification on the selected matched sample of funds. Spell (T) is now defined by the due diligence start and end date of the selection for both the selected and their matched counterparts. Some results for this specification are inline with intuition: higher fund information ratios and higher private signals predict a shorter spell or due diligence time. We see, however, in columns 3–5 that private information precision and Sharpe ratio have positive coefficients.

We believe the source of bias in our coefficients is related to two assumptions underlying an accelerated hazard model. First, the model assumes that right censoring must be at random. Our data fails on this count due to our relatively short time series (June 2005–June 2012). Second, the model assumes that all funds in our regression are always at risk, i.e., the probability of investment is 100% as time approaches infinity. This is clearly violated; our allocator will never invest in a substantial portion of funds in our sample, even asymptotically. This requires a hazard model with a more flexible baseline (time only) hazard specification.

Table D.4. Accelerated Time Hazard Model

This table reports estimates of an accelerated hazard model of due diligence spell on public and private information. Our public information proxies are peer adjusted returns and precision on due diligence end date (T). For selected funds T is the selection date; for unselected funds it is date of censoring for Panel A and date of matched selected fund selection date. Variables are standardized. Panel A reports results for the full sample of funds. Panel B reports results for a matched sample. Reported *t*-statistics are robust to heteroskedasticity. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. All Funds

	(1)	(2)	(3)	(4)	(5)
$\hat{E}_T(R - peers)$	0.010	−0.009		0.027	0.011
	[0.30]	[−0.15]		[0.83]	[0.18]
r_T	0.019	0.019		0.024	0.031
	[0.48]	[0.45]		[0.67]	[0.81]
$\hat{E}_T(R - peers) \times r_T$		0.023			0.019
		[0.41]			[0.34]
Private information:					
α_T			−0.198***	−0.193***	−0.602***
			[−6.50]	[−6.29]	[−4.34]
s_T			0.067**	0.058**	0.082***
			[2.42]	[2.08]	[2.88]
$\alpha_T \times s_T$					0.413***
					[3.09]
Observations	802	801	805	797	796
R^2	0.0007	0.0011	0.0604	0.0579	0.0707

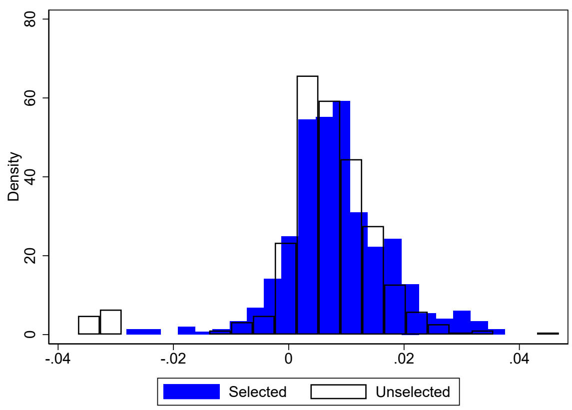
Panel B. Matched Sample

	(1)	(2)	(3)	(4)	(5)
$\hat{E}_T(R - peers)$	−0.029	0.046		0.042	0.065
	[−0.74]	[0.96]		[1.34]	[1.65]
r_T	−0.051	−0.035		0.011	0.008
	[−1.31]	[−0.90]		[0.35]	[0.29]
$\hat{E}_T(R - peers) \times r_T$		−0.116***			−0.048
		[−2.59]			[−1.32]
Private information:					
α_T			−0.596***	−0.601***	−1.508***
			[−15.81]	[−15.97]	[−12.51]
s_T			0.090***	0.090***	0.046
			[2.82]	[2.82]	[1.64]
$\alpha_T \times s_T$					0.943***
					[8.30]
Observations	766	766	752	752	752
R^2	0.0020	0.0088	0.3158	0.3171	0.3752

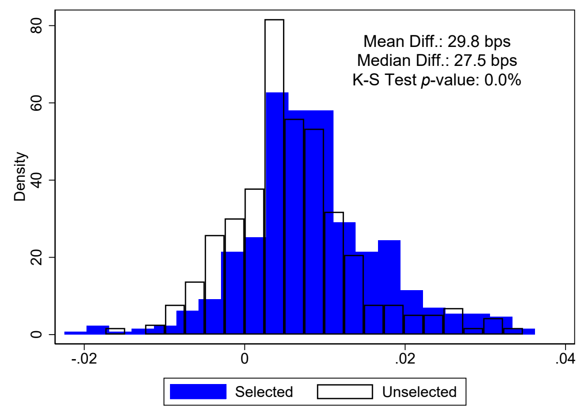
Figure D.4. Fund Selection and Past Alphas

This figure reports frequency distributions of a 24-month rolling Jensen's α for funds that the allocator selected for possible investments versus a matched group of peer funds that were never selected. In all panels, the peer funds are matched according to the Mahalanobis distance based on a fund's log(AUM), age and an additional variable, which is the information ratios for Panel A and Panel C and the due diligence month for Panel B and Panel D. In addition, for Panel A and Panel C, the control group for each fund is the 3 closest peers within a calendar month whereas for Panel B and Panel D it is a single peer within past performance tercile and calendar month. Panel A and Panel B pool monthly α estimates over 3 months before the allocator's decision date, Panel C and Panel D pool earlier months (up to 12).

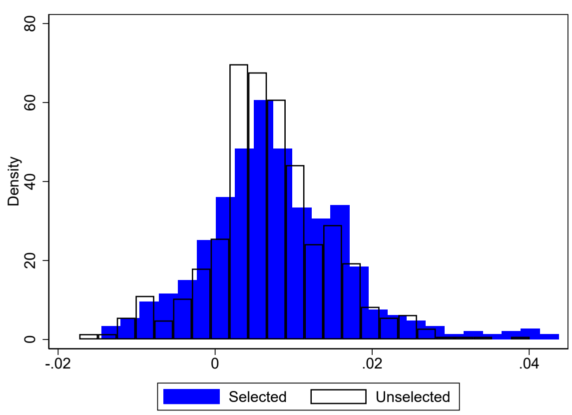
Panel A. Closest 3 peers by performance; 1–4 months before decision



Panel B. Closest peer by due diligence spell; 1–4 months before decision



Panel C. Closest 3 peers by performance; 5–12 months before decision



Panel D. Closest peer by due diligence spell; 5–12 months before decision

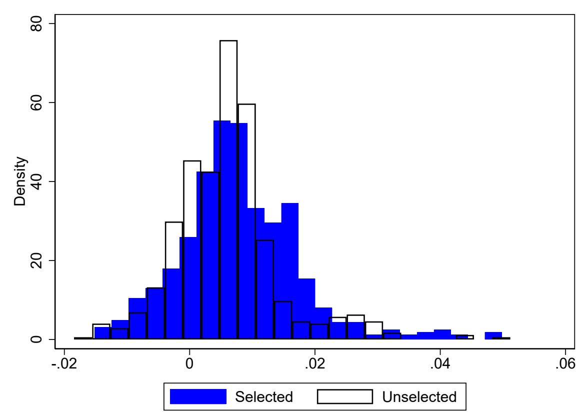
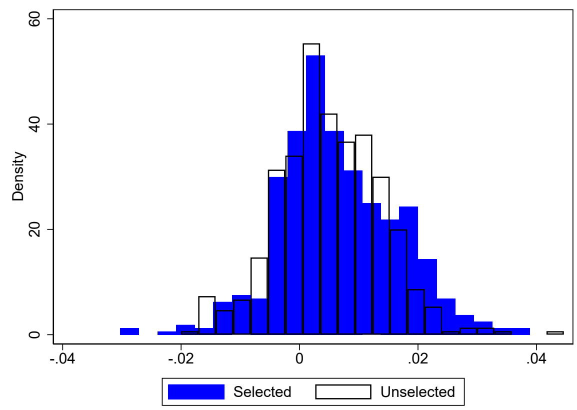


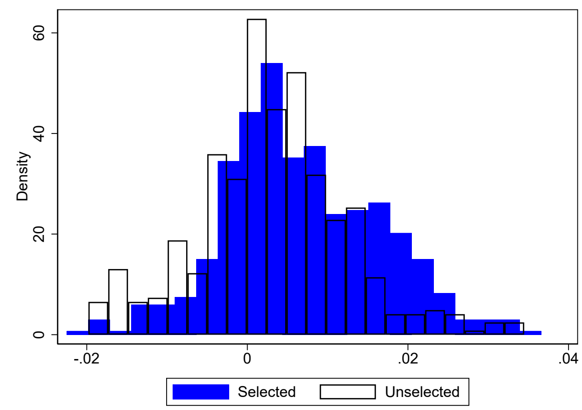
Figure D.5. Fund Selection and Past Information Ratios

The analysis represented in this figure is identical to that of Figure D.4, but using estimates of rolling excess return. The excess return is the fund return minus the peer benchmark returns. Our allocator flags each fund as either a global long-short, an emerging market specialist, a market neutral, or a relative value fund. Our peer benchmarks are thus the HFRI equity hedge (HFRIEHI), HFRI emerging market (HFRIEM), HFRI equity market neutral (HFRIEMNI) and HFRI relative value (HFRIRVA) indices, respectively. The expected excess returns computed as a 12-month rolling average of these excess returns.

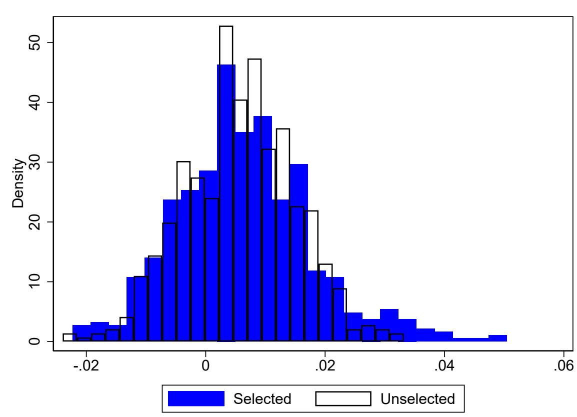
Panel A. Closest 3 peers by performance; 1–4 months before decision



Panel B. Closest peer by due diligence spell; 1–4 months before decision



Panel C. Closest 3 peers by performance; 5–12 months before decision



Panel D. Closest peer by due diligence spell; 5–12 months before decision

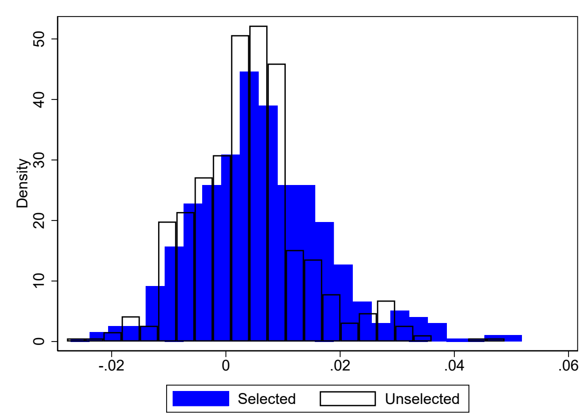
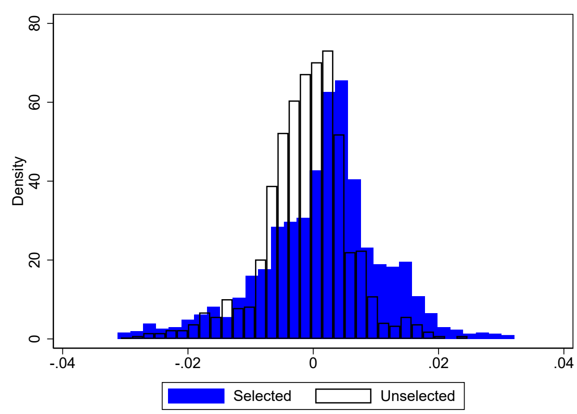


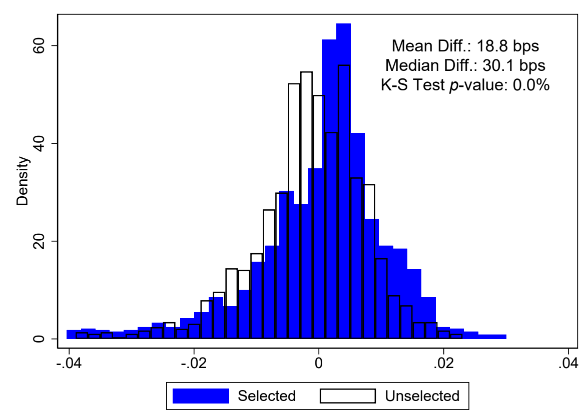
Figure D.6. Fund Selection and Future Alphas

This figure reports frequency distributions of a 24-month forward rolling Jensen's α for funds that the allocator selected for possible investments versus a matched group of peer funds that were never selected. As in Figure D.4, peer funds are matched according to the Mahalanobis distance based on a fund's log(AUM), age and an additional variable (*observables*), which is the information ratios for Panel A and Panel C and the due diligence month for Panel B and Panel D. For Panel A and Panel C, the control group for each fund is the 3 closest peers within a calendar month whereas for Panel B and Panel D it is a single peer within past performance tercile and calendar month. The histograms below represent the forward rolling alpha of these matched fund–months. That is, the *selected* group is matched after their selection date and αs are computed using future (at time t unobservable) returns. We do this to compare returns of selected versus unselected funds beyond the selection date. Panel A and Panel B pool monthly α estimates over the 12 months post decision, where as Panel C and Panel D pool estimates between 13 and 24 months post decision.

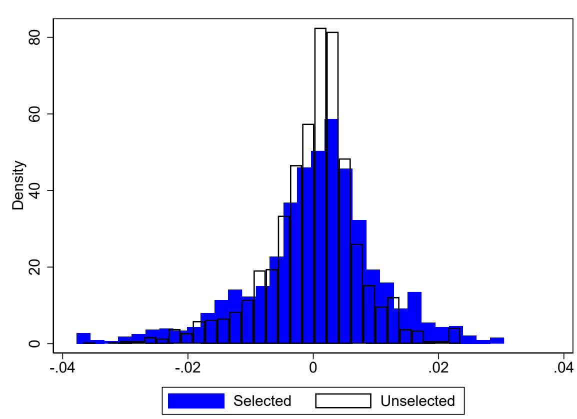
Panel A. Closest 3 peers by performance; 1–18 months after decision



Panel B. Closest peer by due diligence spell; 1–18 months after decision



Panel C. Closest 3 peers by performance; 19–36 months after decision



Panel D. Closest peer by due diligence spell; 19–36 months after decision

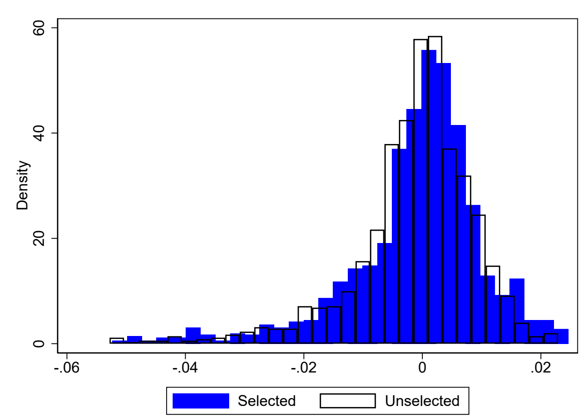
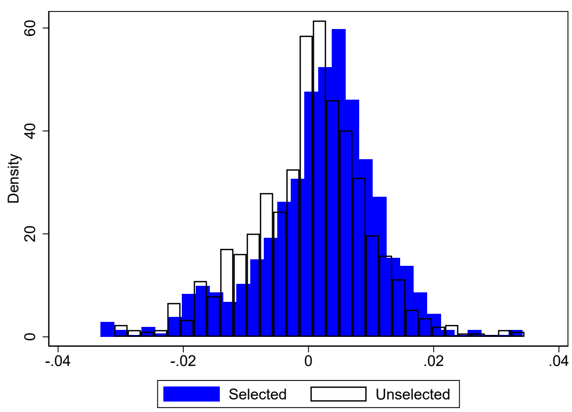


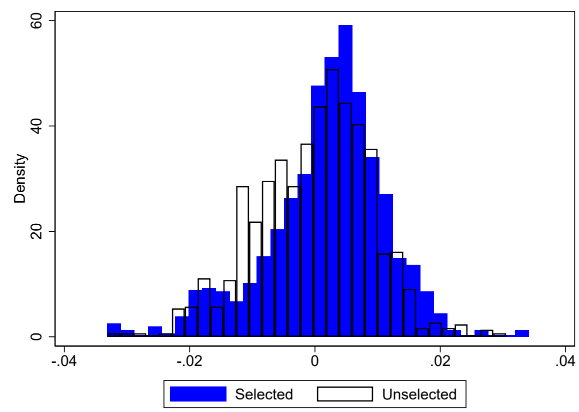
Figure D.7. Fund Selection and Future Excess Returns

The analysis represented in this figure is identical to that of Figure D.6, but using estimates of forward rolling excess return. The excess return is the fund return minus the peer benchmark returns. Our allocator flags each fund as either a global long-short, an emerging market specialist, a market neutral, or a relative value fund. Our peer benchmarks are thus the HFRI equity hedge (HFRIEHI), HFRI emerging market (HFRIEM), HFRI equity market neutral (HFRIEMNI) and HFRI relative value (HFRIRVA) indices, respectively. The realized excess returns computed as a forward 12-month rolling average of these excess returns. The matching is done as above.

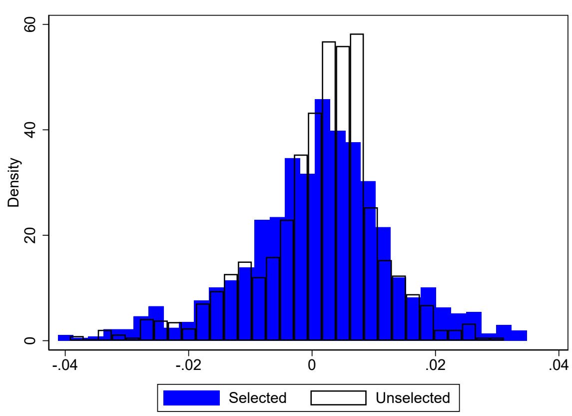
Panel A. Closest 3 peers by performance; 1–18 months after decision



Panel B. Closest peer by due diligence spell; 1–18 months after decision



Panel C. Closest 3 peers by performance; 19–36 months after decision



Panel D. Closest peer by due diligence spell; 19–36 months after decision

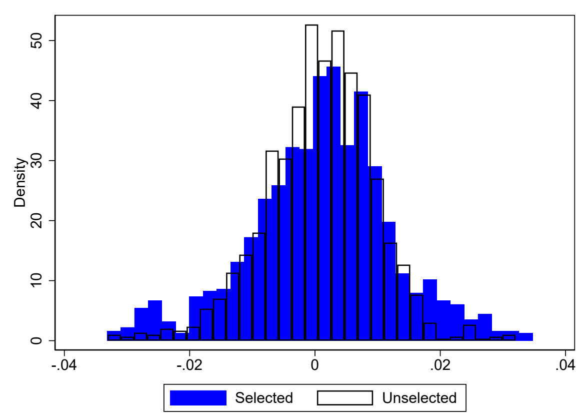


Figure D.8. Distribution of Skill Selected vs. Unselected Funds

For each selected fund, on selection date, 3 unselected funds are matched by calendar time, log(AUM), age and past 24 month rolling information ratio estimates. The fixed effect regression 16 is then run. This histogram is of the estimated fixed effects plus mean zero error from that regression. The distributions were then split on whether they were from an selected or matched, unselected fund.