Limited Partners versus Unlimited Machines

Reiner Braun[1], Borja Fernandez Tamayo[2], Florencio López-de-Silanes[3],

Ludovic Phalippou[4], and Natalia Sigrist[5]

We assemble a proprietary dataset of 395 PPMs to analyze private equity (PE) fund performance and fundraising success. We analyze PPMs' quantitative and qualitative information using econometric methods and machine learning techniques. PE fund performance is unrelated to standard fund characteristics, including prior performance, and measures of document readability. Measures of fundraising success, although correlated to most fund characteristics, are not related to future performance. Meanwhile, machine learning tools can use qualitative information to predict future fund performance: the performance spread between the funds within the top quartile and the lowest quartile of predicted probability of success is about 9% per annum. Our findings support the view that in opaque and non-standardized markets, sophisticated investors incorporate qualitative information in their asset manager selection, while the average investor does not.

Keywords: Private equity, performance predictability, natural language processing, machine learning

[1]Technische Universität München School of Management. Email at reiner.braun@tum.de
[2]Université Côte d'Azur SKEMA Business School. Email at borja.fernandeztamayo@skema.edu
[3]Université Côte d'Azur SKEMA Business School and NBER. Email at florencio.lopezdesilanes@skema.edu
[4]University of Oxford, Said Business School. Email at ludovic.phalippou@sbs.ox.ac.uk
[5]Unigestion, SA. Email at nsigrist@unigestion.com

1. Introduction

Private Equity (PE) markets have exploded in the past two decades. PE Assets under management (AUM) have multiplied more than tenfold since 2004 reaching over $10 trillion in 2022. The limited lifespan of these funds means that PE firms (also known as GPs) raise new funds every three years or so. The most sought-after PE firms complete fundraising in less than six months, while at the other end of the spectrum, fundraising can take two years. PE represents an increasing proportion of most institutional investor portfolios. These investors dominate the demand side of the market and spend considerable time and resources making PE investment decisions, probably as a result of the large commitment and multi-year nature of such investments.

While mutual fund flows have been well studied, the determinants of PE fundraising success and performance have received less attention, possibly due to the scarcity of information. In the mutual fund industry, fund flows have been associated with models based on the competitive provision of capital by investors, heterogeneous fund manager skills with diminishing returns to scale, and investor learning about manager skills based on past fund manager performance (Berk and Green (2004)). Since a large proportion of investment in public and private markets is made by the same institutions, similar forces may be driving these two markets. Yet, information asymmetry is a more pervasive feature of private markets. Unlike mutual fund databases, private equity datasets are thin. Moreover, PE fund managers have considerable degrees of freedom to frame their track records at the time of fundraising (Brown et al., (2019)). As a consequence, PE fund manager selection is particularly difficult.

Although access to *reliable* information is an issue in PE markets, GPs do provide a great deal of information to prospective investors (also known as LPs) at the time of fundraising. In this paper, we collect a unique dataset of close to 400 PE fundraising prospectuses and analyze the impact of quantitative data and qualitative information contained in these documents on fundraising success and performance using both econometric methods and machine learning algorithms for the first time.

We start by analyzing the impact of quantitative information extracted from private placement memorandums (PPMs). As in previous papers, we codify information on past performance, vintage year, fund size, and fund sequence. In addition, we also compute two common measures of readability proposed in related literature but never used in the PE context:

the number of words in relevant PPM sections and the total number of PPM pages (Loughran and McDonald (2014), Kim, Wang, and Zhang (2019)).

One could argue that investors use all of this information to learn about the perceived skills of PE firms and form their heterogeneous demand for the offered funds. Therefore, our first set of econometric tests (Section 3), analyzes fund-level determinants of fundraising success. To measure investors' demand, we compute fundraising success proxies used in previous work, such as the ratio of the realized fund size at the end of the fundraising period (ex-post) to the fund size targeted by the PE firm at the start of the fundraising effort (ex-ante), fund size growth, and the time to raise the fund. Our results confirm that investors seem to process quantitative information in their capital allocation decision. We find that PE firm reputation and past performance are significantly related to fundraising success. In contrast, we do not find consistent evidence that readability impacts fundraising success in private markets.

We then analyze whether the quantitative information apparently used by investors to allocate capital actually predicts fund success, i.e. whether investors really spotted fund manager skill. We find that fund performance is not explained neither by any of the aforementioned quantitative measures driving fundraising success nor by measures of fundraising success themselves. The result that past performance is not correlated to future performance is broadly consistent with previous work, (Hochberg, Ljunqvist, Vissing Jorgensen (2012), Harris, Jenkinson, and Kaplan (2020)). However, the finding that fundraising success is unrelated to future performance is novel and perhaps surprising as we could have anticipated that sought-after funds would perform better.[6]

In section 4, we use machine learning techniques to analyze whether qualitative information from the PPM can help investors. It is possible that in the absence of standardized sources to benchmark risk exposures and past performance in private markets, the qualitative information provided in PPMs is valuable in identifying PE fund manager skills. We focus our analysis on the qualitative information provided in the strategy section of the document in which GPs elaborate on market segments, strategy and planned activities.[7] This part of the document

[6] Although Cavagnaro et al. (2019) provide empirical evidence that some investors do persistently select skilled fund managers, our finding suggests that the average investor in private markets finds it difficult to learn about fund manager skills.

[7] The investment strategy section contrasts with the other sections of the PPM, which lawyers tend to largely copy-paste across PPMs. For details on the role played by lawyers in drafting a PPM please refer to: https://www.lexisnexis.com/lexis-practice-advisor/the-journal/b/lpa/posts/drafting-and-reviewing-the-key-documentation-for-a-private-equity-fund-and-its-offering. In addition to the mostly legal sections devoted to the

contains an average of 2,633 words and conveys considerable information about what the fund manager plans to do. The richness of this information contrasts with the more limited and generic set of quantitative fund characteristics and readability measures used in the previous section.

Since the information in the strategy section of the PPM is qualitative, our analysis uses the Term Frequency - Inverse Document Frequency (TF-IDF) approach, a well-established method in Natural Language Processing (NLP). This method produces a score indicating how characteristic a term is in each document.[8] In our context, each TF-IDF score is equal to the frequency with which a term occurs in the strategy section of a PPM divided by its average frequency across the strategy sections of the rest of the PPMs in the sample. The resulting matrices of TF-IDF scores of terms across PPMs, is then used as input in quantitative methods (Salton and Buckley, 1988). For example, the bigram (trigram) with the highest TF-IDF score in our sample is "portfolio company" ("due diligence process") which appears 4,065 (404) times in 92.6% (49.6%) of the 395 PPMs. An analysis of terms across time shows that the wording in PPMs has been remarkably stable over time, particularly since 2003.

OLS regressions are generally not successful techniques for either in-sample or out-of-sample predictions, especially in settings with a large number of terms (i.e. TF-IDF scores) and relatively few observations.[9] Therefore, we carry out our analysis using three of the most traditional machine learning techniques: Lasso Regression, Random Forest, and Gradient Boosting. Specifically, we train each algorithm on a training dataset equivalent to 80% of our total sample of funds raised between 1999 and 2013, and test its accuracy on the funds raised in 2014-2016.

To measure outperformance, we benchmark each fund's TVPI to the median TVPI of all other funds in the same geography, investment strategy, and vintage year in the Preqin database. Given the relatively small size of our sample, compared to other machine learning settings, we restrict ourselves to a binary indicator of outperformance. A second challenge for carrying out an out-of-sample exercise is the long horizon of PE investments. To do a standard

offer and its risks, PPMs also contain fund managers' background and selected case studies. Due to the extent of this paper, we analyse these sections in a separate paper.

[8] A term can be either a single word or a combination of adjacent words (i.e., bigrams or trigrams).

[9] Traditional econometric techniques are designed to estimate structural parameters and drawing causal inferences. In contrast, machine learning algorithms are designed to maximize out-of-sample predictive accuracy avoiding overfitting and not being constrained by parametric assumptions a linear structure, or a specific number of covariates.

out-of-sample test we would need to train the algorithm on the few funds raised before, say, 1985, and test the algorithm on funds raised between 2000 and 2005. Therefore, the trade-off between statistical power of a small sample and a look-ahead bias when training the algorithms needs to be managed. Since our main machine learning analysis exploits the full sample of 395 funds, we use performance information only known in June 2022. This choice ensures that all funds in the test sample are at least six years old, but the information was obviously unavailable at the start of our test period in 2014. We address the issue of a look-ahead bias in the robustness section.

Our main results show that the three machine learning algorithms are remarkably effective at predicting fund performance. When the algorithms are applied to our test sample of 72 funds from the 2014-2016 vintage years and using big- and trigrams, the different accuracy measures indicate that the algorithm is better than a random selector. Using all 323 funds raised between 1999 and 2013, the Lasso algorithm accurately classifies the outperformance of 43 of the 72 funds. Our models are better at predicting outperformance than underperformance. For robustness, we run an alternative analysis restricting the training sample to funds raised after 2002, when the language in the PPM becomes even more homogeneous. The models' predictive power generally improves: Gradient Boosting, for example, correctly classifies over 73% of the actual outperformers delivering TVPI returns higher than the median TVPI in their segment.

We perform a series of ancillary and robustness tests that put our machine learning results in the context of the previous analysis carried out with traditional econometric methods, and provide support for the robustness of the main findings of the paper. We start by showing that the qualitative information grasped by the algorithms does not simply mirror some quantitative fund characteristics. Our results show that the probabilities of fund success produced by the algorithms using qualitative information as input are uncorrelated with standard quantitative variables, such as reputation or past success. Similarly, the probabilities of success are also uncorrelated with fundraising success measures. This pattern of evidence seems to suggest that the average investor does not process qualitative information when selecting PE asset managers.

We also confirm that the power of the PPM's qualitative information translates into meaningful economic effects as the algorithms are particularly good at identifying outperformers. To illustrate this, we sort the funds in our test sample according to their predicted probability of success and look at their excess TVPI. The excess TVPI of the whole portfolio to be 0.16. If we now remove one fund at a time, starting with those with the lowest predictions, the median distance increases significantly: for the 25% of funds with the highest predicted probability of success, excess TVPI is 0.35. In our view, this finding is evidence that qualitative information is a valuable tool to learn about fund manager skills and maybe one of the reasons why some LPs are better at this exercise than others.

One could argue that it is unfair to compare machine learning-based results using qualitative information with results using quantitative information or fundraising success through linear regression models because the former models are able to capture non-linear complexities. For this reason, in the second part of Section 4, we use machine learning models training them with quantitative information and/or fundraising success. Our results show that the predictive power of these models for future fund success is substantially weaker than those algorithms using qualitative information.

Unlike traditional econometric models, machine learning algorithms do not provide a straightforward interpretation of the marginal influence of features (i.e. independent variables) on outcome variables. However, developments in the explainable AI (XAI) and interpretable ML have started to improve the interpretability of models (e.g., Lundberg and Lee (2017), Ribeiro, Singh, and Guestrin (2016) , and Vilone and Lungo (2020) for a review). In Section 4, we use some of the methods from this literature to gain insights into our algorithms, and to quantify the contribution of each feature in predicting future fund performance. Our analysis looking into the most relevant terms predicting success shows terms such as “investment criteria,” “developing portfolio,” or “growth opportunity” among these terms. This analysis gives us some confidence that our algorithms are picking up economically interpretable constructs.

In the final part of this Section 4, we address the previously raised issue of a look-ahead bias when training our algorithms. The additional tests in this last section show that our previous results are indeed robust. In particular, we carry out an exercise where the training data set consists of 184 funds raised before 2011 and the binary outperformance indicator is computed using performance information available as of December 2014. We use these algorithms to predict success for funds from vintage years 2015 and 2016 (using performance information as of June 2022 again). The algorithms achieve an out-of-sample accuracy of above 0.6 and correctly classify around 60% of the outperformers depending on the method.

Overall, our paper makes use of previously unexploited data on qualitative information provided to private market participants. This analysis allows us to provide the first empirical study assessing the impact the readability measures of fund manager disclosures and the application of textual analysis and machine learning in markets characterized by non-standardized disclosure and inherent information asymmetries.

Our results contribute to four different strands of the existing literature in finance. First, we contribute to the literature of fund flows and return predictability in private markets. As in Harris et al. (2020), we find that buyout fund performance is poorly predicted by traditional quantitative variables, such as the track record available at fundraising time. Our results are also in line with PE papers raising concerns about the reliability of returns reported to investors when raising a new fund (e.g., Barber and Yasuda (2017); Jenkinson et al., (2013)), and with other work arguing that prior returns are not an important issue for sophisticated investors (Brown et al., (2019); Robinson and Sensoy (2016)). We supplement this literature using new fundraising success measures and introducing the analysis of textual data. We show that the average investor finds it difficult to extract information about fund manager skills' heterogeneity using traditional quantitative measure, but that analyzing qualitative content can actually help. We therefore offer qualitative information as a potential explanation of why some investors seem to be persistently good at selecting funds (Cavagnaro et al., 2019).

Second, our paper expands the literature on document readability of disclosures in financial markets. Previous papers have examined the association between readability, fund flows an future firm performance in public markets (Li (2008); Loughran and McDonald (2014); Loughran and McDonald (2016)). To the best of our knowledge, we are the first to empirically analyze the informational value of document readability in private markets. We do not find readability proxies to be significantly correlated with fundraising success or fund performance in private markets. We posit that, as in security issuance (La Porta et al., (2006)), one potential explanation behind these results may be the lack of disclosure standardization.

Third, beyond the use of qualitative information in the form of readability and document presentation, our paper expands the nascent and rapidly growing literature using textual analysis in finance which has started to exploit public firms' 10K filings and earnings call transcripts. Hoberg and Phillips (2010) are attributed to be the first to use textual analysis to cluster firms according to product markets. Since then, these methods have been used for diverse objectives including categorizing corporate goals using letters to shareholders (Rajan et al., (2022)) and identifying risk factors raised by firms in their annual reports (Lopez-Lira (2023)).[10] In the area of private markets, the only other paper exploiting textual data thatwe are aware of is Biesinger et al. (2021), who apply this method to investigate value creation in LBO portfolio companies. We contribute to this literature showing the potential to extract meaningful patterns from textual data in private market disclosures. We find such qualitative information could be valuable in assessing or inferring PE manager skills beyond other quantitative information measures studied in previous papers (e.g., Kaplan and Schoar (2005)).

Finally, our paper is also connecting to the rising literature applying machine learning techniques identifying human biases and improving selection in both asset pricing (e.g., Bianchi et al. (2021); Ke, Kelly, and Xiu (2019); Gu et al. (2020)) and corporate finance (Li et al. (2020), Bubb and Catan (2020), Erel et al. (2021)). Erel et al. (2021), for example, use algorithms to predict director performance and also show that firms are more likely to nominate male candidates as directors than what a trained machine learning model would. Davenport (2022) and Lyonnet and Stern (2022) have used a similar method to compare investor choices in venture capital to algorithms' predictions. The thrust of our findings suggests that investors in private markets might be biased towards more established PE firms.

We organize the rest of the paper as follows. After this introduction, section 2 presents our new dataset of 395 PE funds. In section 3, we analyze the role of quantitative information and document readability in explaining fundraising success using traditional econometric methods. We also analyze whether a more successful fundraising campaign helps predict ultimate fund success in terms of returns. Section 4 expands the analysis presenting the main results of the paper including the study of qualitative information using textual analysis and machine learning techniques to predict PE fund success. We also provide some economic interpretation of the algorithms, and robustness tests to addressing potential biases and alternative explanations, and well as ancillary tests connecting quantitative and qualitative information. Section 5 concludes.

[10] For a extensive overview on textual analysis in several topics in accounting and finance see Loughran and McDonald (2016) and (2020).

2. Data

2.1. Data collection

When raising capital, PE firms send a large amount of information to potential investors. Most of this information is contained in Private Placement Memorandums (PPMs), which are long confidential documents averaging close to 80 pages (and about forty thousand words). There is no explicit or implicit industry standard for PE disclosures. Typically, a large part of the PPM content is mostly legal consisting of the terms of the security offered, the risks and the legal disclaimers protecting sellers from potential liabilities. Other common sections of the document include a description management biographies and selected investment examples. Importantly, PPMs contain a strategy section where management is heavily involved and which provides investors the GP's vision about the market, the strategy, and the business.

In terms of informational content, PPMs have been used as an important source of quantitative information in the literature. Hard information typically obtained from PPMs includes the PE firm's past performance, the number of funds the PE firm and its managers have raised before, the target amount to be raised and the fees associated with the offer (e.g., Reference 1, Reference 2, Reference 3). In our paper, we analyze quantitative information, but our main contribution is to go beyond hard data and extract qualitative information for the first time. To achieve this goal, we develop measures of the PPM readability and, most importantly, extract qualitative information from the strategy section of the PPM using machine learning techniques.

We source PPMs from a large global institutional investor based in Europe who is mostly focused on leveraged buy-out (LBO) funds. The proprietary database we assembled, using the investor's paper and electronic archives, consists of 941 PPMs received by the investor between 1999 and 2020. Panel A of Table 1 documents how we arrive at our final sample of 395 funds used throughout the rest of the paper. Appendix 1 provides detailed descriptions of all the variables used in the paper.

<Insert Table 1 here>

The data provider did not invest in more than 80% of the PPMs. In order to gather as much hard information as possible on these funds, we matched our sample with the Preqin dataset.[11] A total of 737 funds with a size above €5 million were identified in Preqin.

To ensure that we observe at least six years of fund performance (from 2022, the time of our data collection), we eliminate all funds raised after 2016. The following two filters aim to ensure similar style and geographic focus of the funds in the sample. We want to focus on buyout funds, so we eliminate 84 funds that include venture capital and other styles far from buyout. [12] We only keep the 503 funds that are classified as either buy-out, turnaround, growth capital, or balanced funds. The geographic focus filter only eliminates a couple of funds leaving us with 501 funds investing in either Europe, North America, or Asia. [13] Two funds in our proprietary sample do not have an "investment strategy and process" section in their PPMs, which we need to represent qualitative information. Finally, as the performance of mature funds is central for our analysis, we remove the 106 funds for which we do not find performance information in year six or later. The rest of the columns of Panel A provide statistics on the number of funds and their average and median size for the sample of fund in what we call the "Full Preqin" and the "Preqin Fund Performance" datasets. In terms of fund size, our final sample of 395 funds (median size of €355m) is comparable to the size of funds in the Preqin Fund Performance Sample (€284m), but significantly larger those funds in the the Full Preqin sample (€121m).

A feature of the Preqin dataset is that for a subset of funds, Preqin reports the fund characteristics (e.g., fund size, geographic focus, stage of target companies) but not the performance measures. The Preqin Fund Performance Sample refers to the subset of funds for which performance metrics are available. In addition, Preqin provides the detailed cash flow data behind the reported performance for a subset of these funds. For this sub-sample, we can calculate the time-series of performance and therefore compute the interim performance that we need in our analysis.

---

[11] Preqin is a commercial dataset that has now become widely used by academics (e.g., Cavagnaro et al., 2019). We need Preqin to obtain performance information for the funds in which our institutional investor data provider did not invest in.

[12] Using the Preqin classification, excluded funds belong to one of the following categories: Natural Resources, Special Situations, Secondaries, Distressed Debt, Co-Investment, Mezzanine, Infrastructure, Secondaries, Venture Debt, Fund of Funds, Real Estate, and Venture Capital.

[13] Using the Preqin classification, Asian countries include China, Australia, India, South Korea, and Turkey.

Panel B of Table 1 shows the sources of fund performance data that we use in our analysis for our final sample of 395 funds. Our data provider invested in a quarter of the 395 funds. For these 100 funds, we have the full time series of cash flows and NAVs. An additional data search using all available internal records (e.g., PPMs or subsequent fund presentations) allowed us to collect performance data for an additional 61 funds. Of the 234 funds for which our data provider did not have useful performance information, 34 funds were included in the Preqin Cash Flow dataset, and the remaining 200 funds were drawn from the Preqin Fund Performance sample. To ensure comparability across sources and currencies, we converted all return information into euros. Appendix 2 provides the details of the conversion procedure. We should metion that this exercise was quite involved as the Preqin data is not consistent in terms of currencies for the same funds across years.

2.2. Summary statistics: Performance, fundraising success and quantitative information

We construct three performance measures. We compute the Total Value to Paid-in (TVPI) as the ratio of the sum of all capital distributions plus the last reported NAV over the total amount of capital invested (including fees). Our second measure is the Internal Rate of Return (IRR), also comuted net of fees. Our final measure is an indicator variable we label Outperformer, which we compute as a binary indicator taking the value of one if the TVPI of the fund in question is greater than or equal to the median TVPI of the fund's peers in the Preqin dataset. We identify the Preqin peers of a fund as those funds with performance available in Preqin, having raised capital in the same year, and investing in the same type (e.g., Buyout).[14] Detailed definitions of these measures are provided in Appendix 1.

Panel A of Table 2 shows that the average performance for our final sample of 395 funds is equivalent to a net TVPI of 1.8x and a net IRR of 14.4%. These figures are comparable to those in other large scale buyout samples covering similar time windows (e.g., Cavagnaro et al., 2019). The fact that only half (51.7%) of the 395 funds in our final sample were outperformers, provides further confidence on the representativeness of our data.

[14] When we have the data to computere TVPI but we do not have an IRR (or the other way around) we infer the missing one using the following formula:

$$LN(TVPI) = 4 * LN(1 + IRR)$$

<Insert Table 2>

In panel B of Table 2, we provide additional summary statistics for the measures of fundraising success that we use in our study to proxy for investor demand. A well-established approach to measure fundraising success in the literature (see e.g., Hochberg, Ljungqvist, and Vissing-Jørgensen, 2014) is to calculate the oversubscription ratio, equivalent to the ratio of final to target fund size. The oversubscription ratio for our sample of funds has a mean (median) value of 1.05 (1.07). This means that the average fund in our sample is oversubscribed by 5%. This value is close to the average oversubscription of 1.014 (or 1.4%) documented in Hochberg, Ljungqvist, and Vissing-Jørgensen (2014).[15] The standard deviation of 0.31 indicates that we observe significant variation in fundraising success when using this measure. We use the oversubscription ratio to compute a binary indicator variable that takes a value of one if the so-called oversubscription ratio is greater than or equal to one, and zero otherwise. As panel B shows, 65% of the funds in our sample actually reached their target size.

Another way to construct a measure of investor demand in our context is to compute the increase in fund size of consecutive funds by the same PE fund manager, that is, the size increase from the previous fund to the current fund (see, e.g., Barber and Yasuda, 2017). Since revenue per PE firm partner has been show to strongly increase with fund scale (Metrick and Yasuda (2010)), it seems reasonable to assume that investor demand can be proxied by the change in fund size (divided by the number of years between the two successive funds). We compute fund size increase dividing the final size of the current fund by the size of the previous fund. This measure is similar to the measure of fund flows used in the mutual fund industry. To the extent that a PE firm is able to raise additional capital for a given investment strategy relative to its previous fundraising effort. Table 2 shows that the average (median) increase in fund size is quite substantial reaching 1.74 (1.47). This means that the average (median) follow-up fund by the same PE firm pursuing the same strategy is 74% (47%) larger than its predecessor.

In addition to the established measures of fundraising success we describe above, we also collected data on the duration of fundraising period. Our proxy is computed as the number of months between the date of the drafting of the PPM (i.e., when the PE firm starts marketing the fund) and the final closing date of the fundraising campaign. Caeteris paribus, higher

[15] A 1.014 average oversubscription ratio means that the average GP raises its target size plus 1.4% of the target size.

demand for a fund should translate in a shorter fundraising period. The average (median) fund in our sample spent 13.8 (12.0) months raising capital before closing the campaign. As with pervious success measures, we observe substantial differences across funds: the fund in the 75th percentile spent 19.3 months, while that in the 25th percentile raised its capital in only a little over half a year. The correlation matrix in Table A1 of the Appendix shows that all fundraising success measures we calculate are highly correlated with each other. This fact makes us more confident about the validity of our measure to proxy for the, ultimately unobservable investor demand for a fund at the time of fundraising.

Panel C of Table 2 provides the standard quantitative measures of fund characteristic used in the literature. We have 75 first-time funds as part of our final sample, but the median fund in our data represents the third generation of an investment strategy by a specific PE firm. The largest fund in our sample raised more than 17 billion euros. Indeed, our sample includes very large funds, as evidence by the significant difference between the mean and median fund. This difference is indicative of the fund size right-skewness typical of the PE industry. The fund characteristics in panel C have been widely used in previous work as proxies for PE firm reputation since Kaplan and Schoar (2005).

When making investment decisions across all asset classes, most investors of use past performance as a predictor for future success. In the context of PE, a large body of the literature has also studied the relationship between fundraising success and interim performance reported at the time of fundraising (e.g., Hochberg, Ljungqvist, Vissing-Jørgensen (2014); Barber and Yasuda (2017); Brown et al. (2019)). For this reason, we also collected the data on gross TVPI interim performance on the previous fund for 318 of the 320 second or higher fund generation in our sample. Following Lopez-de-Silanes et al. (2015), we compute gross TVPI as the ratio of the sum of the total cash received by the GP from the investment plus its current valuation (if not fully liquidated) over the total cash invested by the GP. The median (average) interim gross TVPI is 1.63 (1.52). Although these interim returns are largely based on reported net asset values, and hence can be somewhat managed by PE firms, the observed variation across funds is interesting. The standard deviation of a previous-fund gross TVPI is 0.62. One potential explanation behind this variation is the average four-year length between the closing of two subsequent funds in our sample. This length could translate is a large portion of the investments having realized distributions.

Inflated interim performance is a well-known issue PE. For this reason, investors might also consider the length of the PE term track record as an indicator of success. We were able to compute the gross TVPI of the previous two funds for a total of 228 of the 234 funds of third and higher generation in our sample. Since averaging the performance of two trailing funds may be conceptually difficult, we follow a different procedure to classify funds and obtain a measure of the performance of these funds. To get an idea of how well a previous fund performed, we compare its performance against competing funds of similar maturity at that point in time for each vintage year of the current fund. We then split the 228 funds into three groups based on their past track record. We code two three variables. The first one takes a value of one if both previous funds outperformed competing funds, and zero otherwise. Table 2 shows that 52 funds (23%) were marketed with such a track record of consistent past outperformance. The second dummy variable takes a value of one if to previous funds underperformed their peers, and zero otherwise. We have 68 funds (30%) that entered fundraising with this poor signal. The third dummy variable takes a value of one if both previous funds have a mixed track record, and zero otherwise. A total of 108 funds out of 228 funds in our sample (47%) exhibited a mixed track record at the time of fundraising of the current fund.

The final panel D of Table 2 departs from previous past performance measures used in the literature and introduces document readability measures in the area of PE for the first time. A large strand of the document readability literature relies on lexicas to determine complexity of language used in a text. However, we follow Loughran and McDonald (2014) who criticize lexica-based approaches, arguing that the complexity of words used seems to be more closely related to the complexity of the underlying business than to the actual readability of the document itself. For this reason, they use the size of 10-K filings and show that larger files are associated with high return volatility, earnings forecast errors, and earnings forecast dispersions. Although the setting of unregulated and non-standardized disclosure in private markets is somewhat different,[16] we adapt the idea of Loughran and McDonald' compute two simple readability proxies: the number of words in a PPM's strategy section and the number of pages of the PPM. The traditional hypothesis in this filed is that a larger numbers of words or pages translates into lower readability. The average PPM in our sample has a strategy section

[16] There is no document standard or regulating organization (like the SEC in public markets) publishing such documents in private markets. In fact, different PPMs are actually compiled (and provided) using different word processing software.

with 2775 words and the average PPM is 85 pages long. We use the natural logarithm of both variables in our subsequent tests.

2.3. Numerical Representation of Qualitative Information in PPMs

Although readability is a first step in measuring how qualitative information is presented to investors, in this section we go one step further and exploit methods of textual analysis to represent the content of qualitative information. Our goal is to represent numerically the text in the PPM so that it can be used for quantitative analysis. To characterize the qualitative information contained in each PPM, we rely on the term frequency- inverse data frequency (TF-IDF) approach, an established method in computational linguistics.

To implement the TF-IDF method in the strategy section of the PPM, we carry out the following steps. First, we stem words in the text corpse. Second, for the main analysis we present in the paper, we define a term as a combination of two or three adjacent words, respectively (called bigrams or trigrams).[17] Terms that appear in most documents, such as the bigram *portfolio companies*, are unlikely to help discriminating between funds. For this reason, the TF-IDF approach relies on a measure of originality of a term; it compares the number of times a term appears in a document with the number of documents the term appears in. We use the scaled version proposed by Pedregosa et al. (2011), which is a slightly modified version of the original measure proposed by Salton and Buckley (1988).[18] More specifically, the TF-IDF of a term *i* in document *j* is computed as:

$$TFIDF\ (i,j) = \frac{TF\ (i,j).\,(\ln(N) - \ln(N_i))}{\sqrt{(\ \sum(TF(i,j).\,(\ln(N) - \ln(N_i))^2}}$$

It is basically equivalent to the frequency of term i in document *j* ($TF(i,j)$), weighted by the ratio of the total number of documents in the corpse (*N*) to the number of documents

---

[17] In alternative exercises, we have defined a term as a single word instead, or used bigrams or trigrams alone or the three of them together.

[18] We use the TF-IDF vectorizer available in Scikit-Learn developed by Pedregosa et al. (2011). The authors of the paper modify the TF-IDF formula presented in the body of the text to produce more accurate results. They compute the natural logarithm to the IDF score to avoid high values for the score, preventing them from dominating the TF-IDF score. Furthermore, they normalize the TF-IDF score to make model training less sensitive to the scale of features:

$$TF - IDF\ (i,j) = \frac{TF\ (i,j)x\ LN(\ IDF(j))}{SQRT\ (Sum\ of\ Squares\ of\ the\ product\ of\ TF\ (i,j)x\ LN(\ IDF(j)))}$$

See Appendix Section 1 for implementation details.

containing the term i at least once (N_i). This means that the term frequency in each document is penalized by its frequency across documents. For example, if a term is mentioned in all the documents the weight attributed to it would be zero. The third step of the method is to obtain a TF-IDF score for each term and each strategy section of the PPM. A high TF-IDF score indicates that a term is particularly characteristic for a given document. For example, if a term appears twice in the focal document, and is not used in any of the other 394 documents, its TF-IDF would be equal to 3.60 in our case. A term that occurs 1,000 times altogether and appears in all of the 395 documents would have a score of 0.03

There are 37,001 unique terms across the 395 Strategy sections of the PPMs in our sample, and about half of the terms appear in more than one document. Table 3 shows the most common stemmed bigrams (Panel A) and trigrams (Panel B) that appear in the strategy sections of our PPMs. *Portfolio company* has the highest average TF-IDF score among all bigrams. It is mentioned at least once in more than 90% of the PPMs and, therefore, its presence is not particularly specific to certain documents. However, its high number of occurrences (4,065 times) and the high standard deviation of TF-IDF scores across PPMs indicate that it is used much more frequently in some documents. The highest standard deviation across these bigrams is displayed by *deal flow*. The trigram with the highest average TF-IDF score is *due diligence process,* while the trigram with the highest standard deviation is the term *buy build strategies*. Glancing through Table 3, we find terms associated with deal sourcing (e.g., *investment opportunity, proprietary deal flow*), value creation (e.g., *value creation, companies' management team, buy build strategy*), market segment (e.g., *investment criteria*), and decision making (e.g., *investment process*). Altogether, the list of terms shown in Table 3 suggests that the vectorization of this approach picks up concepts that appear meaningful in the PE context.

<Insert Table 3>

Our data set of PPMs includes documents issued over a period of 18 years (1999-2016). Therefore, one could argue that PE investment strategies may have evolved or changed over two decades. To analyze this possibility, Table 4 analyzes the evolution of the words used in the strategy sections of all the PPMs in our sample. To carry out this analysis, we compute the averages of all TF-IDF scores for each vintage year in our sample, and calculate the cosine similarities between all annual vectors.[19]

While one might have expected large changes, Table 4 shows that the wording used in the strategy section of the PPMs has remained remarkably stable, particularly since 2003. The cosine similarity among PPM term vectors after 2003 have values above 0.62. After 2006, when our sample becomes substantially larger, all cosine similarity scores are above 0.7, and most of them have values above 0.8. We believe that the stability of wording over time allays concerns regarding our approach to train algorithms using a body of historic PPMs to predict future fund success.

<Insert Table 4>

3. Fundraising Success

3.1. Standard Quantitative Determinants and Document Readability Measures

Following the previous literature on the determinants of fundraising success in Private Equity (Reference 1; Reference 2, Reference 3), we start this section examining the relationship between our fundraising success measures and the standard proxies for quantitative and qualitative information described in section 2. The goal is to analyze whether investors do indeed allocate capital according to PE firm reputation, past performance, and document readability.

<Insert Table 5>

Table 5 presents simple OLS regressions of two fundraising success for the full sample of 395 funds with available performance data and a PPM strategy section. In all regressions, we control for investment strategy style, region, and vintage year fixed effects. The first six columns present alternative specifications for our oversubscription ,which measures the ratio of actual fund size (ex-post) divided by targeted fund size (ex-ante). Columns 7 to 12 repeat the same specifications for our new proxy measuring the number of months it takes the PE firm to raise a fund. Appendix Table A2, provides the same analysis as in Table 5 for the two other

[19] To avoid that extremely uncommon or extremely common bigrams and trigrams influence the comparison, we filter out bigrams and trigrams that appear in less than 1% of all PPM Strategy sections (i.e., in 4 PPMs) and those that appears in more than 99% of all PPM Strategy sections (i.e., in 391 PPM).

fundraising success measures we compute for our sample: a binary indicator of whether the target fund size was reached, and the fund size increase of the most recent fund in comparison to the previous fund with the same investment focus. Given the high correlation across all four fundraising success measures, it is not surprising that most of the results in Appendix Table A2 are very similar to those in Table 5.[20]

The results in Table 5 are very similar for the two measures of fundraising success. There are three main findings that emerge from this table. First, fundraising success is positively correlated with PE firm reputation as proxied by the natural logarithms of fund sequence and fund size.[21] Since the correlation between these two measures is high (0.42 for our sample), the first regression model for each variable only includes fund sequence: the natural logarithm of fund sequence is positively and significantly correlated with fundraising success in models 1 and 7, respectively. The magnitude of the coefficients imply that a second-generation fund in our sample has a 4.4 percentage points higher ratio of actual vs. targeted fund size ratio compared to a first-time fund (0.063 x ln(2) = 0.044), and its fundraising period was about 1.5 months shorter than that of a first-time fund (Column 7). Fund size, our second proxy for PE firm reputation, is also significantly correlated with the oversubscription ratio and the length of fundraising. In our sample, a fund is 6.8 percentage points more oversubscribed (column 2), and needs 1.2 months less to raise capital than a fund half of its size (column 8).

The second set of findings pertain to the impact of current performance signals on fundraising success. As expected, we find that investors process current performance signals from the fund manager when making allocation decisions. The interim (i.e. at the time of fundraising) gross TVPI performance of the PE firm's previous fund, available for 318 funds, is positively associated with fundraising success. All else being equal, a fund manager with a one times higher TVPI has a 5.9%-points higher oversubscription (column 5), and a 2.3 months shorter fundraising period.

It is a well-established fact in PE that performance figures at the time of fundraising are not necessarily indicative of the ultimate performance of the fund (e.g., Brown et al., 2019).

[20] The only significant difference in results is that measures of PE firm reputation measures are not significant for our fund size increase variable. This may not be surprising given the way the variable is constructed though. The larger a PE firm's funds have become (as tenure and sequence increase), the harder it is to grow in size at a high rate.

[21] Since we could expect a decreasing marginal effect of fund sequence and fund size on fundraising success, we use the natural logarithm of these two measures in all regression specifications.

Therefore, some investors may also consider a longer track record as an alternative signal of future performance. To analyze this, we identify the gross TVPI returns of the two previous funds at the time of fundraising and compare them with the historical TVPIs of PE firms raising at the same time. Regressions 6 and 12 on 243 funds (with at least two trailing funds) show that PE managers' fundraising efforts are affected by past long(er)-term performance. Compared to fund managers with a mixed track record, funds in our sample that raised while both previous funds traded worse than the median (of previous funds at the time of fundraising) have a 11.5%-points lower ratio of actual to target fund size (Column 6). While the same effect is economically relevant in column 12 (1.9 months) it is not statistically significant. However, for fundraising time the positive effect of raising with two outperforming previous funds (2.5 months) is significant at the 10%-level.

The final finding of Table 5 and Appendix Table A2 is the lack of consistent evidence of a significant effect of document readability measures on fundraising success. Our readability measures are only significant for one of our four proxies of fundraising success (columns 9 and 10 of Table 5). Doubling the number of words in the investment strategy section of the PPM (total number of PPM pages) increases fundraising time by 1.4 months (2.1 months). Although readability measures have been shown to have an important impact in public market disclosure. It has been argued that in standardized public markets, complex language may b used to intentionally increase investors' information processing cost for negative price-relevant information. Our divergent set of findings here may be related to the non-standardized nature of disclosure in private markets. Additionally, it is also possible that the nature of the information analysis process carried out by sophisticated investors in private markets is also important. These investors typically carry out a thorough analysis of only a few funds per year. To the extent that they engage in substantial screening for each fund they analyze, the additional costs caused by low readability may not end up amounting to much for fund manager selection decisions.

Overall, the results in Table 5 and Appendix Table A2 provide a somewhat consistent pattern linking quantitative information to fundraising success in our sample. We interpret these results as suggesting that the average investor does consider several of these cues in her asset manager selection decisions. In contrast, traditional document readability proxies about the length of the strategy section and the PPM itself are not consistently connected to fundraising success.

3.2. **Fundraising Success and Fund Performance**

If investors successfully learn about differential PE fund manager ability to run funds, one could expect that their fund demand should be positively correlated with the fund's ultimate performance. However, private markets are characterized by a lack of regulation, non-standardized disclosures and significant information asymmetries between managers and investors (Robinson and Sensoy, 2013) which may interfere in the process. In the rest of this section, we analyze whether the ultimate net fund TVPI is correlated with the quantitative and qualitative fund characteristics that we found significant to predict fundraising success in the previous section, and, more directly, with fundraising success itself.

<Insert Table 6>

Table 6 presents the results of such an analysis using similar econometric specifications as in the previous section. The dependent variable for all specifications in the table is Net Fund TVPI and the set of regressors include the set of standard fund characteristics used in Table 5, (Columns 1-6) and the four proxies for fundraising success one at a time (Columns 7-10). In all regressions, we control for investment strategy, region, and vintage year fixed effects.

The summary of the findings in Table 6 is straightforward: the quantitative determinants of fundraising success and our proxies for fundraising success are poor predictors of future fund performance. Starting with the determinants of fundraising success (columns 1 to 6), only the number of pages in the PPM, a proxy for document complexity, is marginally negatively correlated with TVPI (specification 4). Fund sequence, fund size and past performance of previous funds have no predictive power. The finding that interim performance at the time of fundraising is unrelated to final fund performance may be surprising at first sight, as we have found that investors respond to this signal. However, the result is actually consistent with Harris, Jenkinson, Kaplan, and Stucke (2020), who find that when calculated at the time of fundraising, fund performance post-2000, does not predict the performance of the follow-on fund.

In the remaining columns of Table 6 (columns 7 to 10), we turn to examine if there is a the direct relationship between the four measures of fundraising success we used in the previous and future fund performance proxied by TVPI. All coefficients are statistically insignificant.

Although the coefficients of the proxies for fundraising success have the expected sign, i.e., are positively correlated with net TVPI, they do not show a statistically significant correlation with fund performance. The coefficient on the fundraising period is close to zero.[22]

The results in Table 6 provide a direct test to reject the hypothesis that investors are able to predict PE fund returns. This is evidence thus is broadly consistent with Harris, Jenkinson, Kaplan, and Stucke (2018), who find that, on average, investors have difficulty identifying successful buyout funds ex-ante.

If traditional quantitative fund characteristics and soft readability factors turn out to be poor predictors of future fund performance, the next logical question is if other type of information can help investors make better investment decisions in private markets? In the next section, we turn to explore the potential of natural language processing (NLP) techniques combined with traditional machine learning algorithms to process the *qualitative content* of financial disclosure in private markets to help investors in fund manager selection.

4. Predicting Fund Performance with Qualitative Information

4.1. Training Machine Learning Algorithms

Our PE setting is characterized by a very large number of terms (i.e. TF-IDF scores) and relatively few fund PPMs. For this reason, we chose a machine learning approach and apply three of the most traditional techniques in this field: Lasso regression, Random Forest, and Gradient Boosting. These approaches differ from traditional regression methods in several respects, including their ability to capture non-linearity and complexity in the data.

The general design of a machine learning exercise is to use a large part of the total sample, e.g. 80% of funds, to train the algorithms (in-sample). These models are then applied to the remaining sample (i.e., 20%) to test how well the algorithms predict patterns (out-of-sample). To compute the core results of the paper in this section, we train each of the three algorithms (i.e., Lasso, Random Forest, and Gradient Boosting) on a training dataset consisting of 323

[22] In unreported tests, we report the regressions shown in Table 5 but using net IRR or TVPI outperformance as alternative measures of fund performance. The results draw the same picture: there is no relevant explanatory variable in our set of proxies.

funds raised before 2014 (82% of our total sample). As we describe in section 2.3., we use a the TF-IDF scores of bigrams and trigrams as features for the training.

The objective of the training is to classify whether or not, a specific fund outperforms in terms of TVPI the rest of the funds from the same geography, investment strategy and vintage year in the Preqin database. We classify a fund as predicted to outperform if the predicted probability is greater than 50%. We use five-times repeated cross-fold cross-validation for each of our models. This is a re-sampling procedure to evaluate machine learning model and to estimate the unknown tuning parameters. Cross-validation is the most widely used method for estimating prediction error. The advantage of using repeated cross-validation, instead of only cross-validation, is that it reduces the variance of the cross-validation estimator. Since the size of our training sample is relatively small for prediction purposes, we set K=10 (Hastie, Tibshirani, and Friedman, 2009).[23] For each fund in the test set, our method produces an outperformance probability, which we call “predicted probability” from now on.

We compute an weshow the algorithms’ out-of-sample statistical performance using the two standard in the machine learning literature. Balance Accuracy represents the average number of correct predictions . It is calculated as follows:

$$Balanced\ Accuracy = \left[\frac{TP}{TP+FP} + \frac{TN}{TN+FN}\right]/2$$

Balanced Accuracy is the average of the ratio of true positives (TP) over the sum of true positives and false positives (FP), and the ratio of true negatives (TN) over the sum of true negatives and false negatives (FN).

The second standard metric in this field is the area under the receiver operating characteristic curve (ROC AUC). This metric is equivalent to the probability that the model will rank a random positive observation higher than a random negative observation. A model whose predictions are 100% wrong will have turn an AUC of 0, while a model whose predictions are 100% correct will turn an AUC of 1. Machine learning algorithms that outperform the 0.5 thresholds, are equivalent to having a higher predictive power than pure randomness.

[23] Appendix Table A3 depicts the implementation process of cross-validation for an example of cross-validation with K=5 (i.e., five-fold cross validation).

Table 7 presents the core machine learning results of the paper. The table has two panels. In Panel A, we train the algorithms using the funds whose vintage is from 1999 to 2003 and the out-of-sample test is carried out on fund of vintage years 2014 to 2016. Since in Section 2.3 of the paper we showed that the wording of the PPMs is much more similar from 2003 onwards, we carry out a second exercise in Panel B where we train the algorithms using the funds whose vintage is from 2003 to 2013 only. For each panel we show the AUC and the Balance Accuracy for the in-sample and the out-of sample using the three different algorithms of Lasso, Random Forest and Gradient Boosting.

<Insert Table 7>

If we focus in the results of Panel A, we find that the lowest overall value is the balanced accuracy for Lasso with 0.554. Meanwhile, the highest value is an AUC of 0.595 for Gradient Boosting. Looking at Panel B, and in line with this observation of more similar language starting in 2003, we find that models trained on the restricted sample of the 304 funds raised as of 2003 improve predictive power. Panel B shows that five out of six out-of-sample accuracy measures are higher than 0.6. The only exception is the balanced accuracy for the Random Forest algorithm. Meanwhile, the highest number in this panel is the AUC for the Lasso algorithm.

The two statistical accuracy metrics shown in Table 7 are indicative of the high predictive power of our methods, but they do not give us an idea of the predictive power for outperformers and underperformers separately. Ideally, we would like to also get an indication of whether our models are performing better at correctly classifying outperformers or underperformers.

In Table 8, we compute the Confusion matrices of the models presented in Table 7. The table shows that our algorithms are much better at correctly classifying outperformers. Panel A shows that this is particularly striking for Gradient Boosting model. Exploiting our full training set of funds from vintage 1999 to 2013, the models correctly classify 25 out of 37 outperforming funds (68%).

<Insert Table 8>

Panel B shows the same pattern of results. The better predictive power of outperformers becomes more apparent for Gradient Boosting when we train algorithms on the more homogenous PPM training sample that starts in 2003. The results show that the model correctly classifies 27 out of 37 (73%) outperforming funds.

The table also shows that the average TVPIs are higher for the subsample of funds that the algorithms classify as outperformers than for the funds that are predicted to underperform. In Panel A, the unweighted difference in mean (median) TVPIs for the Lasso algorithm is 0.38 (0.18). This result is not driven by increased downside risk. All 25th percentile values are higher for the predicted outperformer samples.

4.2. Determinants of Predictions

The results presented so far in this section show that qualitative information picked up by algorithms seems be correlated with future fund performance. The finding means that qualitative information can provide informational value beyond what quantitative information can do.

In Table 9 we carry out an additional exercise to put in context the results on quantitative and qualitative information that we presented in the previous two sections. This table shows OLS regressions of the probability of success predicted by Gradient Boosting on the quantitative and readability measures used in Table 5 and our fundraising proxies for the sample of 376 funds raised between 2003 and 2016.

The various specifications in the table show that the coefficients on quantitative cues and document complexity measures are practically all statistically insignificant and virtually undistinguishable from zero. The only exception in the table is the marginally statistically significant correlation with funds raised having had two previous underperforming funds (column 6). Compared to funds with mixed track record at fundraising, the average predicted probability of success for the former funds is 12.4 percentage points higher. These results suggest that the algorithms are picking up something that is uncorrelated to what the standard quantitative measures used in the literature are picking up.

In the last four columns of Table 9, we also test the association between the predicted probability of success obtained from Gradient Boosting and our fundraising success proxies. As the table shows, all coefficients are insignificant. These findings indicate that the average investor does not seem to incorporate the same qualitative information when selecting PE asset managers. These results are virtually identical when we use the probability of success obtained from Lasso regression or Random Forest. These results are included in Appendix Tables A5 and A6

4.3. **Economic Relevance of Predictions**

Machine learning predictions may also translate into meaningful and relevant economic outcomes. One potential way to show this with our data, is to rank the funds in our test sample (i.e., fund vintages 2014 to 2016) by their predicted outperformance probability and then calculate the TVPI distance of each of these 72 funds to the median TVPI of all funds in the Preqin dataset raised in the same year and with the same investment type and geography. We should note that this is the same information used to calculate our binary indicator of outperformance to train the algorithms.

Figure 1 depicts the calculations detailed above. The median distance across all 72 funds is equal to 0.17x. The lines in the figure are then calculated with the following procedure: we remove the fund with the lowest predicted probability of success and recalculate the median distance and r epeat this process moving up the ranks of predicted outperformance. The black lines in Figure 1 plot the median distances generated by this exercise for the three different algorithms: Lasso (Panel A), Random Forest (Panel B) and Gradient Boosting (Panel C). These algorithms are obtained using the training sample of 314 funds from 2003 to 2013. The figure shows that the median distance doubles to 0.35x when the 75% of funds with the lowest predicted outperformance probabilities of outperformance are discarded.

<Insert Figure 1>

We carry out a final exercise to try to bring together the results using traditional econometric methods and machine learning methods and compare the models trained on qualitative information fairly. We start by carrying out the exercise of training the machine learning models using the quantitative variables used in Table 5 as the features of the training (i.e., instead of the TF-IDF scores) and then predicting outperformance. We repeat the above procedure but now only using the predictions produced by these models to generate the blue line in Figure 1. The figure shows that the actual fund performance, measured by median TVPI distance, is also correlated with quantitative information when machine learning is used. This is probably picking up complexity and non-linearity. However, the figure also shows that, for the most part, the correlation is weaker than the one we observe when using TF-IDF scores to predict outperformance. This is in line with our previous finding that quantitative information is a not a good predictor of future success in linear models (as was shown in the regression setup in Section 3.2).

In addition to the above exercise, we also trained the algorithms using the TF-IDF scores as features, but this time predicting fundraising success instead of outperformance. We use time to raise a fund in month to measure fundraising success in this exercise. This exercise gives us models that detect qualitative information incorporated in the average investors fund investment decision. The red line in Figure 1 shows the results of this exercise. It depicts the median TVPI distance of fund portfolios ranked by their predicted probability of successfully raising a fund. The figure makes evident that the correlations of such predictions are weaker than those of the TF-IDF models trained with our outperformance outcome variable, as well as those of machine learning-based predictions exploiting quantitative information. Altogether, we believe that these finding suggest that qualitative information can be exploited to learn about fund manager skill in private markets.

In Table 10, we formally compare the relevance of qualitative information trained using our binary performance indicator to that of quantitative information and qualitative information picked up by the average investor. Table 10 presents regressions where the dependent variables are: (i) the net TVPI (columns 1 to 3); and (ii) the median net TVPI distance (columns 4 to 6). The table shows that the predicted success probabilities obtained from training Gradient Boosting on outperformance are the only ones that are statistically significantly correlated with ultimate fund performance. The predicted probabilities obtained from training Gradient Boosting using quantitative information or those obtained using qualitative information on fundraising success are both statistically insignificant. Our findings in Table 10 are aso confirmed using the Lasso ant the Random Forest models (see Appendix).

<Insert Table 10>

Figure 2 seems to confirm that our algorithms seem to be picking up meaningful terms that represent meaningful concepts. In this figure, we plot the 30 terms with the highest Shapley Additive exPlanations (SHAP) values when applying Gradient Boosting on the training sample of funds raised between 2003 and 2013. This method of analysis, developed by Lundberg and Lee (2017), is frequently used in the machine learning literature. The approach is based on the coalitional game, where the SHAP values measure the contribution of a feature to the prediction. An intuitive interpretation of a SHAP value is the difference between the prediction with and without using the feature. As the figure shows, the SHAP value of a term can have a strong (red) or weak (blue) positive or negative influence on whether a fund is predicted to outperform or not. Since machine learning algorithms capture combinations of terms, a term may have a positive impact on model performance in some combinations (or in the context of some other terms), and a negative impact in other combinations. In Figure 2, for example, the term *economies of scale* has a very strong negative impact in some contexts and documents, and a weak positive impact in other contexts.

<Insert *Figure 2* here>

Overall, we believe that the terms that appear in Figure 2 are consistent with previous literature in PE. For example, winning deals with an attractive *entry valuation* is one of the typical value creation levers in the PE ownership model (Acharya et al. (2013)). Similarly, the ability to take a private company to the public market (*initial public*) is often seen as a "successful" exit route (Jenkinson and Sousa, 2015). A reason behind the positive SHAP value of this term could be associated with PE fund managers who can report about a track record of successful past IPOs signaling experience usefull for successfully managing such a process in the future. Another example of a term in Figure 2 is the network of industry experts (*network relationship*) which is key to ensure an active and high-quality deal flow as shown in the venture capital industry (Hochberg et al., 2007).

Again, we would like to emphasize that machine learning algorithms use non-linear interactions among multiple word combinations in order to make their predictions. For this reason, we cannot state whether a fund will perform well (or poorly) because its PPM investment strategy section includes a specific combination of words. This is where the strength of machine learning algorithms lies. The ability to make sense of complex, non-linear relationships among various features makes machine learning algorithms a suitable tool for identifying patterns humans find hard, if not impossible, to detect.

<Insert *Figure 3* here>

In Figure 3 we provide an example from our data that illustrates how different terms contribute to the probability of success predicted by the Gradient Boosting algorithm in a random PPM. In the figure, combinations of words are ranked in decreasing order according to their contribution to the final predicted probability of outperformance *f(x)*, which equals 0.848 in this example. The *E[f(X)]* indicates the average mean predicted probability of Gradient Boosting in the training sample using funds raised between 2003 and 2013. In the x-axis, we represent the values of the probability of success. The length of each bar represents the Shapley value of each feature. The biagrams *gener superior, cash flows, companies generating, investment professionals, superior returns, track records,* and *growth profitability* are all positively correlated with the probability of success. On the other hand, the biagrams *Oper finance* and *hand approach* are all negatively correlated with the predicted probability of success.

4.4. **Robustness**

The core machine learning analysis presented in Table 7 kept in the sample as many funds as possible in order to maximize our sample size. However, as we argued before, this choice means that we have implicitly accepted a lookahead bias: in order to determine outperformance, we trained the algorithms using performance information not available at the time of predicting fund success in the training sample.

As a robustness check, we now carry out an exercise that sacrifices sample size but mitigates lookahead bias concerns. Our exercise retrains the algorithms using as a training sample that gives funds a minimum life of six years. Table 11 reports in- and out-of-sample statistical performance of the three algorithms we apply to predict fund success. In Panel A, the algorithms are trained on 141 funds raised between 1999 and 2007 using information available as of Dec 31st 2013. The test sample consists of 72 funds from vintage years between 2014 and 2016. Outperformance in the test sample is computed using TVPIs as of Jun 30th 2022 as before in Table 7. In Panel B, we train the algorithms using the 122 funds raised between 1999 and 2007 using information available as of Dec 31st 2013.

Table 11 presents these results and allows comparisons with the full sample results presented in Table 7. Both panels show similar results with Random Forest and Gradient Boostin performing at high levels. As in Table 7, we have reported results using bigrams and trigrams. However, algorithms trained using words and bigrams as inputs perform better with smaller samples (we can supply these results upon request).

5. Conclusion

Our study relies on the use of novel qualitative information provided to PE investors to predict future fund performance. Applying NLP (aka textual analysis) and machine learning techniques, we provide the first empirical analysis on readability and qualitative information in private markets. Our approach may be particularly useful if we consider that these markets are characterized by non-standardized disclosures and inherent information asymmetries.

We star out paper carrying out a traditional econometric analysis of the determinants of fund raising success using classic proxies of PE firm reputation and interim and past performance. Our results echo previous findings showing a positive association of most of these variables associated with success in PE raising capital. We then complement this analysis looking at qualitative information in the form of readability measures using the same econometric approach. Although these measures have been shown to matter in public market disclosure, we do not find they are associated with fundraising success in PE, possibly due to the non-standardized nature of disclosure in private markets or the thorough analysis carried out by sophisticated investors for a few funds each year.

Since investors seem to take into consideration quantitative factors PE in their PE capital allocation decisions, the next natural question we analyze is if these same quantitative factors are good predictors of future fund performance. Our results show the oposite: traditional quantitative factors and readability proxies, are poor predictors of future performance. In addition, we do not find our proxies of fundraising success at the beginning of a fund’s life to be actually be correlated with ultimate fund performance. Aggregate market demand does not seem to be a reliable signal of PE fund manager ability. \

The above pattern of findings seems difficult to reconcile with Cavagnaro et al. (2019) who show that some investors seem to make persistently good fund manager choices. One could argue that these investors may be exploiting different set of signals to learn about differential PE manager ability. In this paper, we test if qualitative information provided by the fund managers as they lay out investment strategy may be useful to extract such a set of signals. Our paper is the first to use Natural Language Process (NLP) techniques in textual analysis, and machine learning algorithms to analyzing whether this kind of qualitative information explains investor heterogeneity in predicting returns. Our findings show that approaches to exploit the qualitative information disclosed to investors in PPMs has important predictive power for ultimate fund success. Our study provides evidence of the value of applying these new technologies to process qualitative information in private markets.

The application of new techniques in the analysis of private markets in finance is just starting. Upcoming studies could address some of the limitations of our current work. The most important one is clearly realated to our sample size, which does not allow us to conduct a real backtest only using performance information that is really available at the time of training the algorithms. We hope our study shows the potential of such methods and motivates owners of proprietary qualitative data to cooperate with researchers, and put together larger samples that can be analyzed with even more realistic assumptions. There are certainly other areas of potential improvement as the disciplines of textual analysis and machine learning are growing fast and will provide even more powerful methods to be applied in the context of private capital markets.

We believe our findings have important implications and real world applications for investors in private markets. Our results suggest that there are signals of differential ability buried in qualitative information which can be exploited with the new methods introduced in this study.

Bibliography

Barber, B. M., & Yasuda, A. (2017). Interim fund performance and fundraising in private equity. Journal of Financial Economics, 124(1), 172-194.

Berk, J. B., & Green, R. C. (2004). Mutual fund flows and performance in rational markets. Journal of political economy, 112(6), 1269-1295.

Biddle, G. C., Hilary, G., & Verdi, R. S. (2009). How does financial reporting quality relate to investment efficiency?. Journal of accounting and economics, 48(2-3), 112-131.

Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. The Review of Financial Studies, 34(2), 1046-1089.

Biesinger, M., Bircan, C., & Ljungqvist, A. (2021). Value creation in private equity.

Brown, G. W., Gredil, O. R., & Kaplan, S. N. (2019). Do private equity funds manipulate reported returns?. Journal of Financial Economics, 132(2), 267-297.

Bubb, R., & Catan, E. M. (2022). The party structure of mutual funds. The Review of Financial Studies, 35(6), 2839-2878.

Castellaneta, F., Gottschalg, O., Kacperczyk, A., & Wright, M. (2021). Experience as Dr. Jekyll and Mr. Hyde: performance outcome delays in the private equity context. Journal of Management Studies.

Cavagnaro, D. R., Sensoy, B. A., Wang, Y., & Weisbach, M. S. (2019). Measuring institutional investors' skill at making private equity investments. The Journal of Finance, 74(6), 3089-3134.

Davenport, D. (2022). Predictably Bad Investments: Evidence from Venture Capitalists. Available at SSRN 4135861.

Dikolli, S. S., Mayew, W. J., & Nanda, D. (2014). CEO tenure and the performance-turnover relation. Review of accounting studies, 19(1), 281-327.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223-2273.

Erel, I., Stern, L. H., Tan, C., & Weisbach, M. S. (2021). Selecting directors using machine learning. The Review of Financial Studies, 34(7), 3226-3264.

Gompers, P., & Lerner, J. (1998). Venture capital distributions: Short-run and long-run reactions. The Journal of Finance, 53(6), 2161-2183.

Groh, A. P., & Gottschalg, O. (2011). The effect of leverage on the cost of capital of US buyouts. Journal of Banking & Finance, 35(8), 2099-2110.

Gunning, R. (1952). Technique of clear writing.

Harris, R. S., Jenkinson, T., Kaplan, S. N., & Stucke, R. (2018). Financial intermediation in private equity: How well do funds of funds perform?. Journal of Financial Economics, 129(2), 287-305.

Harris, R. S., Jenkinson, T., Kaplan, S. N., & Stucke, R. (2020). Has persistence persisted in private equity? Evidence from buyout and venture capital funds (No. w28109). National Bureau of Economic Research.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, andHober prediction. Springer Science & Business Media.

Hochberg, Y. V., Ljungqvist, A., & Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. The Journal of Finance, 62(1), 251-301.

Hochberg, Y. V., Ljungqvist, A., & Vissing-Jørgensen, A. (2014). Informational holdup and performance persistence in venture capital. The Review of Financial Studies, 27(1), 102-152.

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. The Review of Financial Studies, 23(10), 3773-3811.

Jenkinson, T., & Sousa, M. (2015). What determines the exit decision for leveraged buyouts?. Journal of Banking & Finance, 59, 399-408.

Jenkinson, T., Landsman, W. R., Rountree, B. R., & Soonawalla, K. (2020). Private equity net asset values and future cash flows. The Accounting Review, 95(1), 191-210.

Jensen, M. C. (1989). Active investors, LBOs, and the privatization of bankruptcy. Journal of applied corporate finance, 2(1), 35-44.

Kaplan, S. N., & Schoar, A. (2005). Private equity performance: Returns, persistence, and capital flows. The journal of finance, 60(4), 1791-1823.

Ke, Z. T., Kelly, B. T., & Xiu, D. (2019). Predicting returns with text data (No. w26186). National Bureau of Economic Research.

Kim, C., Wang, K., & Zhang, L. (2019). Readability of 10-K reports and stock price crash risk. Contemporary accounting research, 36(2), 1184-1216.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. Journal of Accounting and economics, 45(2-3), 221-247.

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. The Review of Financial Studies, 34(7), 3265-3315.

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2006) What Works in Securities Laws? Journal of Finance, **LXI**, No. 1, February, pp. 1-32.

Lopez-de-Silanes, F., Phalippou, L., & Gottschalg, O. (2015). Giants at the gate: Investment returns and diseconomies of scale in private equity. Journal of Financial and Quantitative Analysis, 50(3), 377-411.

Lopez-Lira, A. (2023). Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns. Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. the Journal of Finance, 69(4), 1643-1671.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 54(4), 1187-1230.

Loughran, T., & McDonald, B. (2020). Textual analysis in finance. Annual Review of Financial Economics, 12, 357-375.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Lyonnet, V., & Stern, L. H. (2022). Venture Capital (Mis) allocation in the Age of AI. Fisher College of Business Working Paper, (2022-03), 002.

Metrick, A., & Yasuda, A. (2010). The economics of private equity funds. The Review of Financial Studies, 23(6), 2303-2341.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. The Accounting Review, 85(6), 2107-2143.

Pan, Y., Wang, T. Y., & Weisbach, M. S. (2015). Learning about CEO ability and stock return volatility. The Review of Financial Studies, 28(6), 1623-1666.

Phalippou, L. (2011). Why is the evidence on private equity performance so confusing?. Available at SSRN 1864503.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

Rajan, R. G., Ramella, P., & Zingales, L. (2022). What Purpose Do Corporations Purport? Evidence from Letters to Shareholders. Evidence from Letters to Shareholders (February 1, 2022).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. arXiv preprint arXiv:1606.05386.

Robinson, D. T., & Sensoy, B. A. (2013). Do private equity fund managers earn their fees? Compensation, ownership, and cash flow performance. The Review of Financial Studies, 26(11), 2760-2797.

Robinson, D. T., & Sensoy, B. A. (2016). Cyclicality, performance measurement, and cash flow liquidity in private equity. Journal of Financial Economics, 122(3), 521-543.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information processing & management, 24(5), 513-523.

Vilone, G., & Longo, L. (2020). Explainable artificial intelligence: a systematic review. arXiv preprint arXiv:2006.00093.

Table 1 Panel A: Sample Construction and comparison with the Prequin dataset. Panel A shows fund performance variables. *TVPI* is the total value-to-paid net of fees and carried interest at fund maturity or, if still not liquidated, as of June 2022. *TVPI > Benchmark* is a binary indicator that adopts a value of one if the TVPI is higher than the median of funds in Preqin of the same investment strategy and vintage, and zero otherwise. *Net IRR* is the internal rate of return net of fees and carried interest.

	Our sample			**Preqin Sample**			**Preqin Fund Performance Sample**		
Panel A: Data Filters	**# Funds (%)**	**Average Fund Size (EURm)**	**Median Fund Size (EURm)**	**# Funds (%)**	**Average Fund Size (EURm)**	**Median Fund Size (EURm)**	**# Funds (%)**	**Average Fund Size (EURm)**	**Median Fund Size (EURm)**
Initial sample assembled from archives of data provider	941	836.91	291.21	37,798	242.11	54.38	9,272	560.72	176.61
Initial sample of funds trying to raise capital between 1999 and 2020 with fund size>5 and information available in Preqin	737 78%	836.91	291.21	28,309 75%	265.30	66.7	8,580 93%	572.68	183.54
Funds raised in 2016 or earlier (2/1)*100	589 80%	791.25	282.47	18,728 66%	237.89	66.01	6,743 79%	472.51	165.79
Buyout, Growth Capital, Turnaround, and Balanced funds (3/2)*100	505 86%	879.83	322.98	6,682 36%	402.34	117.59	2,799 42%	726.29	260.48
Funds investing in Europe, Asia, or North America (4/3)*100	503 100%	878.76	322.98	6,147 92%	420.62	121.14	2,601 93%	757.48	271.19
Funds with PPM including Investment Strategy section (5/4)*100	501 100%	881.03	322.98	6,147 100%	420.62	121.14	2,601 100%	757.48	271.19
Funds with Total Value to Paid-In (TVPI) and/or Internal Rate of Return (IRR) for more than six years after their vintage year (6/5)*100	395 79%	1025.34	354.9	6,147 100%	420.62	121.14	2,350 90%	810.73	283.87

Table 1 Panel B: Performance Information Sources

Panel B exhibits our fundraising success variables. *Fund target reached* is a binary variable taking a value of one if the fund reached its target size specified in the PPM at the beginning of the fundraising process, and zero otherwise. We obtain *Oversubscription* by dividing target size by actual fund size. *Months to fundraise* is the number of months that have passed between first and final closing of a fund. *Increase in fund size* is the ratio of current fund size divided by the size of the previous fund with the same investment strategy by the same PE firm.

	# Funds (%)	**Fund Size (EURm)**	
		Average	**Median**
Fund Cash Flows from Internal Sources	100 (25%)	1316	507
Summary Fund Performance from Internal Sources	61 (15%)	260	219
Fund Cash Flows from Preqin Database	34 (9%)	1826	1047
Preqin Fund Performance Sample	200 (51%)	976	319
All	395 (100%)	1025	354

Table 2: Descriptive Statistics

This table presents summary statistics for fund-level variables for the final sample of funds used in the paper. Panel A shows fund performance variables. Panel B exhibits our fundraising success variables. Panel C displays variables frequently used to explain private equity returns. Panel D shows measures of document readability. Detailed definitions of all variables are in the Appendix.

		Obs	Mean	SD	p25	Median	p75
Panel A. Performance							
(1)	TVPI Current Fund	395	1.80	0.70	1.39	1.72	2.12
(2)	TVPI > Benchmark (%)	395	51.65				
(3)	Net IRR Current Fund	395	14.36	13.01	7.67	14.19	21.09
Panel B. Fundraising success							
(4)	Oversubscription	395	1.05	0.31	0.87	1.07	1.25
(5)	Fund Target Reached	395	0.65				
(6)	Months in Fundraising	395	13.82	9.21	6.16	12	19.31
(7)	Increase in Fund Size	319	1.74	1.08	1.13	1.47	2.02
Panel C. Fund Characteristics							
(8)	Fund Sequence	395	3.29	2.02	2	3	4
(9)	Fund Size	395	1025.34	1913.11	176.85	354.90	815.16
(10)	Gross TVPI Previous Fund	318	1.63	0.62	1.23	1.52	1.90
(11)	Both Previous Funds Low Gross TVPI	228	0.30	0.46	0	0	1
(12)	Both Previous Funds High Gross TVPI	228	0.23	0.42	0	0	0
Panel D. Readability							
(13)	Strategy words	395	2774.96	1281.49	1829	2644	3724
(14)	PPM Pages	395	84.73	36.84	62	78	101

Table 3. Most Common stemmed bigrams and trigrams

In Panel A the table shows the 15 most common stemmed bigrams (i.e., combinations of two adjacent stemmed words) in our sample. For each stemmed bigram, the table reports the number of observations, the percentage of documents containing the stemmed bigram, and the average and standard deviation of the TF-IDF score. Panel B displays the 15 most common trigrams (combinations of three adjacent stemmed words). Detailed definitions of all variables are provided in the Appendix.

Stemmed terms	Nobs	% PPM	Avr. TF-IDF (in %)	Std. Dev. TF-IDF (in %)
Panel A: bigrams				
portfolio company	4,065	92.41%	1.52%	0.69%
manag team	2,752	92.15%	1.29%	0.57%
invest opportune	1,601	87.85%	1.08%	0.56%
due dilig	2,075	87.09%	1.18%	0.62%
privat equity	1,422	84.56%	1.00%	0.61%
invest strategi	878	82.28%	0.84%	0.55%
valu creation	1,423	72.91%	0.99%	0.76%
target compani	869	69.11%	0.83%	0.70%
compani manag	653	68.35%	0.75%	0.66%
invest process	669	66.84%	0.77%	0.66%
deal flow	818	65.82%	0.89%	0.81%
long term	529	60.51%	0.67%	0.64%
invest compani	457	60.51%	0.68%	0.68%
cash flow	569	58.48%	0.69%	0.70%
track record	460	55.19%	0.65%	0.67%
Panel B: trigrams				
due dilig process	404	49.62%	0.50%	0.58%
portfolio compani manag	293	39.24%	0.46%	0.66%
privat equiti firm	212	33.16%	0.36%	0.56%
manag portfolio compani	186	32.66%	0.38%	0.61%
compani manag team	204	32.41%	0.37%	0.59%
privat equiti invest	190	28.10%	0.34%	0.60%
attract invest opportun	166	27.59%	0.31%	0.55%
proprietari deal flow	158	26.08%	0.33%	0.60%
valu portfolio compani	129	23.54%	0.28%	0.56%
privat equiti fund	129	22.53%	0.30%	0.61%
decis make process	128	21.01%	0.27%	0.57%
buy build strategi	168	20.76%	0.33%	0.72%
potenti invest opportun	109	19.75%	0.24%	0.52%
fund portfolio compani	120	19.49%	0.27%	0.60%
privat equiti investor	105	18.73%	0.25%	0.55%

Table 4. Similarity of bigrams and trigrams across years

The table shows the cosine similarity score of vectors representing the annual average frequency of stemmed terms across years for all 396 funds in our sample. Stemmed terms are restricted to appear in at least 1 percent of all documents (i.e., 4 PPMs) and in less than 99 percent of all documents (i.e., 391 PPMs). The number of PPMs in each vintage year is shown in the second row of the table.

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Obs.	*3*	*6*	*5*	*5*	*11*	*13*	*22*	*39*	*49*	*39*	*25*	*20*	*27*	*27*	*32*	*25*	*28*	*19*
1999																		
2000	0.37																	
2001	0.48	0.40																
2002	0.51	0.42	0.35															
2003	0.45	0.61	0.46	0.48														
2004	0.46	0.52	0.48	0.45	0.63													
2005	0.46	0.58	0.50	0.51	0.71	0.66												
2006	0.54	0.65	0.56	0.57	0.77	0.74	0.80											
2007	0.52	0.64	0.53	0.57	0.78	0.76	0.81	0.88										
2008	0.51	0.61	0.51	0.56	0.74	0.70	0.80	0.88	0.87									
2009	0.49	0.64	0.49	0.58	0.75	0.68	0.78	0.86	0.88	0.85								
2010	0.52	0.55	0.55	0.49	0.68	0.67	0.73	0.81	0.77	0.78	0.73							
2011	0.52	0.59	0.51	0.57	0.74	0.68	0.77	0.84	0.86	0.82	0.82	0.74						
2012	0.49	0.60	0.50	0.54	0.74	0.72	0.77	0.87	0.88	0.87	0.85	0.75	0.83					
2013	0.50	0.59	0.50	0.55	0.72	0.73	0.77	0.86	0.88	0.88	0.84	0.75	0.83	0.88				
2014	0.46	0.54	0.47	0.52	0.71	0.66	0.77	0.83	0.85	0.86	0.81	0.73	0.81	0.85	0.86			
2015	0.52	0.60	0.52	0.56	0.75	0.72	0.79	0.88	0.88	0.86	0.85	0.80	0.85	0.89	0.88	0.85		
2016	0.51	0.57	0.50	0.54	0.74	0.68	0.76	0.84	0.87	0.85	0.84	0.76	0.85	0.86	0.87	0.83	0.87	

Table 5: Fundraising Success

This table presents results from OLS regressions in which the dependent variables are measures of fundraising success. In columns 1-6 we use *oversubscribed* which is the ratio of target size divided by actual fund size. *Month in fundraising* is the dependent variable in columns 7-12. We control for investment strategy, region, and vintage year fixed effects in all regressions. Standard errors are in parentheses and clustered at the PE firm level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Detailed definitions of all variables are provided in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Oversubscription						**Month in Fundraising**					
VARIABLES												
Fund Sequence (ln)	0.063**	-0.020	-0.023	-0.019	-0.033	-0.011	-2.146**	-1.242	-1.081	-1.319	-2.338*	-3.205**
	(0.025)	(0.026)	(0.026)	(0.026)	(0.034)	(0.038)	(0.896)	(0.950)	(0.927)	(0.932)	(1.289)	(1.525)
Fund Size (ln)		0.098***	0.100***	0.103***	0.082***	0.085***		-1.071**	-1.131**	-1.368***	-0.952**	-1.028*
		(0.014)	(0.015)	(0.015)	(0.016)	(0.019)		(0.469)	(0.473)	(0.452)	(0.463)	(0.525)
Number of Strategy words (ln)			-0.041						2.055**			
			(0.030)						(0.907)			
Number of PPM Pages (ln)				-0.046						3.046**		
				(0.044)						(1.338)		
Gross TVPI Previous Fund					0.059**						-2.310***	
					(0.030)						(0.797)	
Both Previous Funds Low Gross TVPI						-0.115***						1.930
						(0.039)						(1.217)
Both Previous Funds High Gross TVPI						0.036						-2.517*
						(0.041)						(1.444)
Observations	395	395	395	395	318	243	395	395	395	395	318	243
R-squared	0.228	0.319	0.324	0.321	0.291	0.358	0.126	0.138	0.151	0.149	0.169	0.233

Table 6: Total Value to Paid-in (TVPI) and Fundraising

This table presents results from OLS regressions in which the dependent variable is the total value to paid-in (TVPI) net of fees and carried interest. We control for investment strategy, region, and vintage year effects in all regressions. Standard errors are in parentheses and clustered at the PE firm level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Detailed definitions of all variables are provided in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES						**TVPI**				
Fund Sequence (ln)	0.031	0.037	0.036	0.041	-0.022	-0.099				
	(0.060)	(0.071)	(0.071)	(0.069)	(0.088)	(0.120)				
Fund Size (ln)		-0.007	-0.007	0.010	0.024	0.026				
		(0.034)	(0.034)	(0.037)	(0.034)	(0.038)				
Number of Strategy words (ln)			-0.014							
			(0.073)							
Number of PPM Pages (ln)				-0.177*						
				(0.105)						
Gross TVPI Previous Fund					-0.041					
					(0.064)					
Both Previous Funds Low Gross TVPI						0.042				
						(0.122)				
Both Previous Funds High Gross TVPI						-0.034				
						(0.094)				
Months in Fundraising							-0.002			
							(0.004)			
Oversubscribed								0.011		
								(0.110)		
Fund Target Reached									0.039	
									(0.075)	
Increase in Fund Size										0.036
										(0.039)
Observations	395	395	395	395	318	243	395	395	395	319
R-squared	0.106	0.106	0.106	0.112	0.094	0.134	0.105	0.105	0.106	0.095

Table 7. Statistical performance of machine learning algorithms

This table reports in- and out-of-sample statistical performance of the three algorithms we apply to predict fund success. The test sample consists of 72 funds from vintage years 2014 to 2016. Outperformance is computed using TVPIs as of Jun 30th 2022. We show Area Under the Receiver Operating Characteristic curve (AUC ROC) and Balanced Accuracy as measures of our models' ability to correctly classify outperforming funds. In Panel A the algorithms are trained on 323 funds raised between 1999 and 2013. Panel B shows results for models trained on trained on 304 funds raised between 2003 and 2012. Detailed definitions of all variables and algorithms are provided in the Appendix.

	Lasso	**Random Forest**	**Gradient Boosting**
Panel A: Training on 323 funds raised between 1999-2013 with info as of June 2022; test 2014-2016			
A.1 In-sample Fit:			
Area Under Curve (AUC)	0.565	0.603	0.611
Balanced Accuracy	0.544	0.570	0.585
A.2 Pseudo Out-of-sample:			
Area Under Curve (AUC)	0.604	0.568	0.596
Balanced Accuracy	0.597	0.585	0.538
Panel B: Training on 304 funds raised between 2003-2013 with info as of June 2022; test 2014-2016			
B.1 In-sample Fit:			
Area Under Curve (AUC)	0.584	0.615	0.
Balanced Accuracy	0.572	0.587	0.574
B.2 Pseudo Out-of-sample:			
Area Under Curve (AUC)	0.625	0.636	0.643
Balanced Accuracy	0.555	0.553	0.636

Table 8. Confusion Matrices

The confusion matrices show the number of correctly and incorrectly classified 37 outperformer funds and the number of correctly and incorrectly classified 35 underperformer funds in the test sample (out-of-sample, i.e. funds raised between 2014 and 2016). In Panel A, predictions are trained on 323 funds raised between 1999 and 2013, in Panel B trained on 304 funds raised between 2003 and 2013. All variables and algorithms are defined in the Appendix.

Panel A: Training on 323 funds raised between 1999-2013 with info as of June 2022; test 2014-2016

		Predicted Values:					
		Lasso		**Random Forest**		**Gradient Boosting**	
Actual Values:	**Total**	TVPI ≥ Benchmark	TVPI < Benchmark	TVPI ≥ Benchmark	TVPI < Benchmark	TVPI ≥ Benchmark	TVPI < Benchmark
TVPI ≥ Benchmark	37 51%	22 59%	15 41%	20 54%	17 46%	25 68%	12 32%
TVPI < Benchmark	35 49%	14 40%	21 60%	15 43%	20 57%	21 60%	14 40%
Total	72	36 50%	36 50%	35 49%	37 51%	46 64%	26 36%
TVPI							
Mean		2.05	1.67	2.01	1.71	1.93	1.74
p75		2.33	1.96	2.26	1.97	2.18	1.98
p50		1.91	1.73	1.91	1.74	1.78	1.76
p25		1.69	1.47	1.67	1.56	1.67	1.54

Panel B: Training: 2003-2013, Test: 2014-2016

		Predicted Values:					
		Lasso		**Random Forest**		**Gradient Boosting**	
Actual Values:	**Total**	TVPI ≥ Benchmark	TVPI < Benchmark	TVPI ≥ Benchmark	TVPI < Benchmark	TVPI ≥ Benchmark	TVPI < Benchmark
TVPI ≥ Benchmark	37 51%	21 57%	16 43%	24 65%	13 35%	27 73%	10 27%
TVPI < Benchmark	35 49%	16 46%	19 54%	19 54%	16 46%	16 46%	19 54%
Total	72	37 51%	35 49%	43 60%	29 40%	43 60%	29 40%
TVPI							
Mean		2.03	1.68	1.92	1.78	2.00	1.66
p75		2.26	1.97	2.20	1.98	2.25	1.92
p50		1.87	1.73	1.80	1.73	1.91	1.71
p25		1.69	1.44	1.69	1.52	1.69	1.44

Table 9. The determinants of predicted probability of success (Gradient Boosting)

This table presents results from OLS regressions on the sample of 376 funds raised between 2003 and 2016. The dependent variable is the Probability of Success predicted by Gradient Boosting. We control for investment strategy, region, and vintage year effects in all regressions. Standard errors are in parentheses and clustered at the PE firm level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Definitions provided in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES					**Probability of Success**					
Fund Sequence (ln)	-0.019	-0.038	-0.038	-0.039	0.006	-0.127				
	(0.035)	(0.040)	(0.040)	(0.040)	(0.056)	(0.080)				
Fund Size (ln)		0.022	0.022	0.017	0.030	0.027				
		(0.022)	(0.022)	(0.024)	(0.024)	(0.025)				
Number of Strategy words (ln)			-0.000							
			(0.043)							
Number of PPM Pages (ln)				0.049						
				(0.068)						
Gross TVPI Previous Fund					-0.019					
					(0.038)					
Both Previous Funds Low Gross TVPI						0.142**				
						(0.057)				
Both Previous Funds High Gross TVPI						0.028				
						(0.072)				
Months in Fundraising							0.002			
							(0.002)			
Oversubscribed								-0.088		
								(0.076)		
Fund Target Reached									-0.056	
									(0.047)	
Increase in Fund Size										0.003
										(0.024)
Observations	376	376	376	376	309	223	376	376	376	310
R-squared	0.033	0.036	0.036	0.037	0.068	0.141	0.033	0.036	0.036	0.060

Table 10. TVPI, Median TVPI Distance, and out-of-sample probability of success (Gradient Boosting)

This table presents results from OLS regressions in which the dependent variable is the Total Value to Paid-In (TVPI) net of fees and carried interest (columns 1-3) and the Median TVPI Distance (columns 4-6). The Median TVPI Distance is calculated as a fund's TVPI minus the median TVPI of the funds from the same vintage year and investment strategy in the Preqin database. All specifications in the table control for investment strategy, region, and vintage year effects. Standard errors are shown in parentheses and clustered at the PE firm level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively. Detailed definitions of all variables are provided in the Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES		TVPI			Median TVPI Distance	
Quantitative Info	0.308			0.297		
	(0.276)			(0.279)		
Qualitative Info (Fundraising Months)		0.005			0.022	
		(0.147)			(0.148)	
Qualitative Info (Outperformance)			0.406***			0.417***
			(0.138)			(0.138)
Observations	67	67	67	67	67	67
R-squared	0.238	0.209	0.300	0.216	0.208	0.428

Table 11. Statistical performance of the machine learning algorithms: Robustness

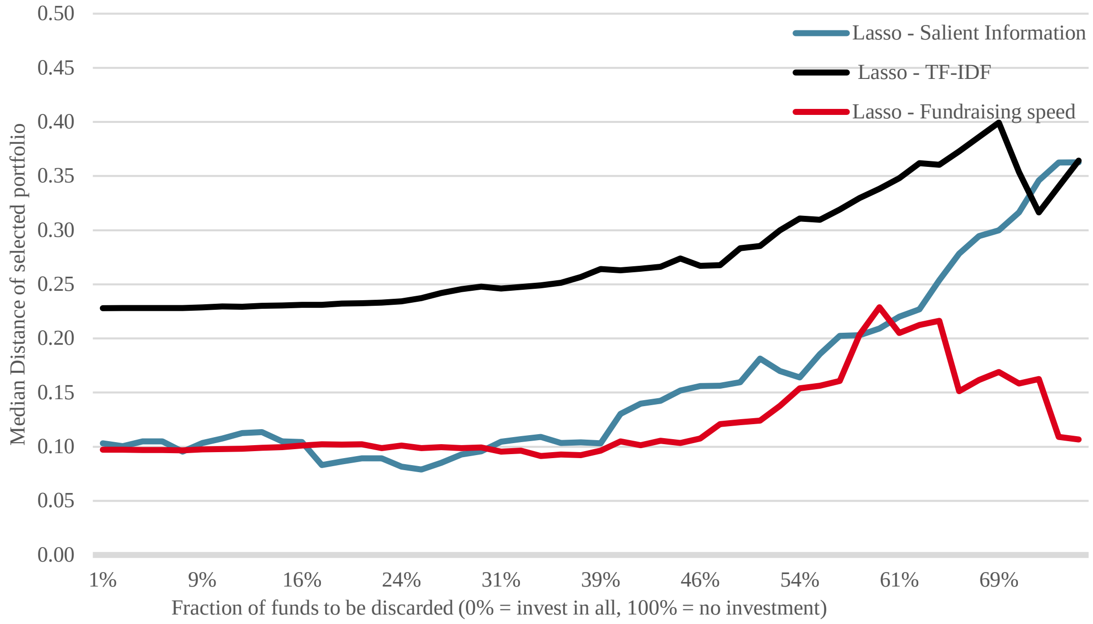
This table reports in- and out-of-sample statistical performance of the three algorithms we apply to predict fund success. In Panel A, the algorithms are trained on 141 funds raised between 1999 and 2007 using information available as of Dec 31st 2013. The test sample consists of 72 funds from vintage years between 2014 and 2016. Outperformance in the test sample is computed using TVPIs as of Jun 30th 2022. We show Area Under the Receiver Operating Characteristic curve (AUC ROC) and Balanced Accuracy as measures of our models' ability to correctly classify outperforming funds. In Panel A, we show statistical performance when training the algorithms on 141 funds raised between 1999 and 2007. In Panel B, we use 122 funds raised between 2003 and 2007 and using information available as of Dec 31st 2013. Detailed definitions of all variables and algorithms are provided in the Appendix.

	Lasso	**Random Forest**	**Gradient Boosting**
Panel A: Training on 141 funds raised between 1999-2007 with info as of Dec 2013; test 2014-2016			
A.1 In-sample Fit:			
Area Under Curve (AUC)	0.622	0.586	0.579
Balanced Accuracy	0.605	0.561	0.565
A.2 Pure Out-of-sample:			
Area Under Curve (AUC)	0.463	0.560	0.583
Balanced Accuracy	0.488	0.560	0.538
Panel B: Training on 122 funds raised between 2003-2007 with info as of Dec 2013; test 2014-2016			
B.1 In-sample Fit:			
Area Under Curve (AUC)	0.600	0.580	0.640
Balanced Accuracy	0.567	0.512	0.616
B.2 Pure Out-of-sample:			
Area Under Curve (AUC)	0.507	0.549	0.655
Balanced Accuracy	0.502	0.534	0.597

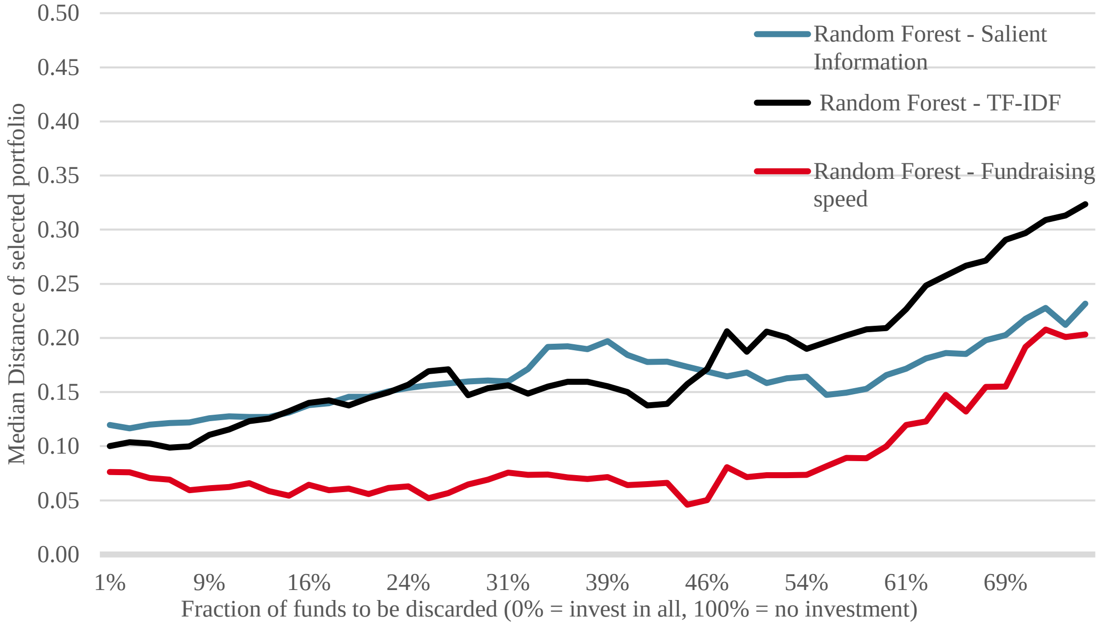
Figure 1. Economic significance of qualitative information

This figure depicts the average median distance of the selected portfolios of 72 out-of-sample funds from out test sample (i.e., raised between 2014 and 2016). Algorithms are trained using funds from 2003-2013. We first sort funds by their predicted probability of success (from low to high). We then calculate the median distance of all funds (i.e. with 0% funds removed). Median distance is calculated as the difference between the Total-Value to Paid-In (TVPI) and the median TVPI of funds raised in the same year and with the same investment type. We then calculate the median distance of the remaining portfolio of funds after having removed the fund with the next lowest predicted probability of success. We repeat this procedure removing one additional fund each time. The figure plots median distances for algorithms trained using quantitative information as features and outperformance as outcome variable, qualitative information (TF-IDF scores) and outperformance, and qualitative information (TF-IDF scores) and fundraising success measured by months in fundraising, respectively. Panel A displays predicted probabilities produced by Lasso regressions, Panel B for Random Forest algorithms and Panel C for Gradient Boosting. All variables and algorithms are defined in the Appendix.

Panel A: Lasso



Panel B: Random Forest



Panel C: Gradient Boosting

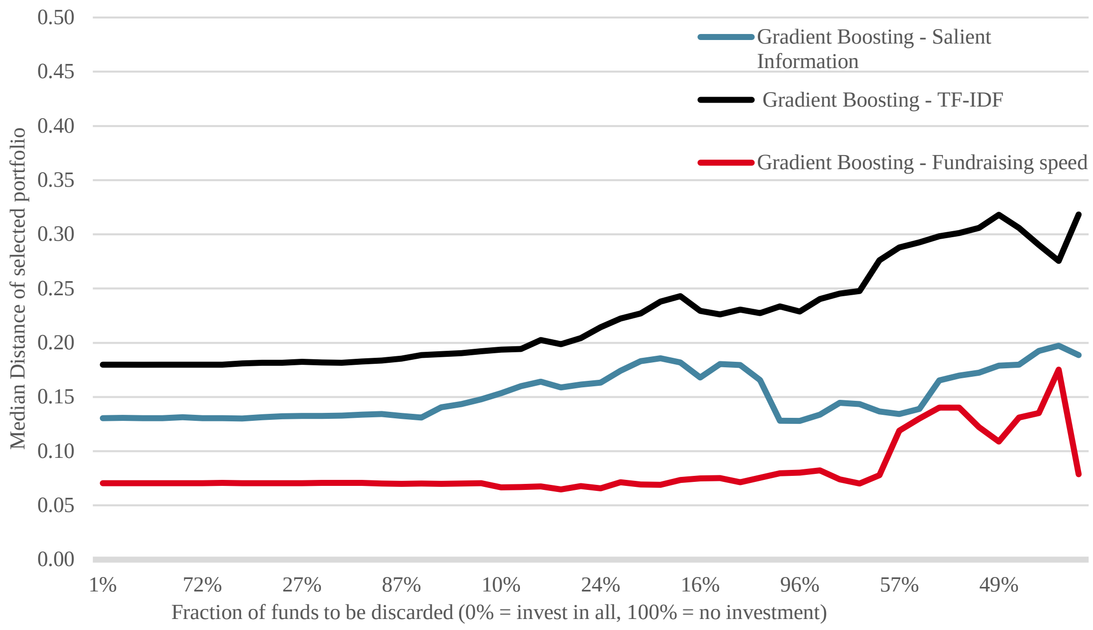


Figure 2. Most relevant combinations of words to make predictions (Gradient Boosting)

The figure presents the SHAP values for the top-30 characteristics in terms of variable importance in GP quality. We use the *Gradient Boosting* algorithm trained on 314 funds raised between 2003 and 2013 to make the predictions. Combinations of words are ranked in decreasing order according to their importance. The Shapley value determines the position on the x-axis, and the feature determines the y-axis. Points represent observations. The color represents the value of the feature from low (blue) to high (red).

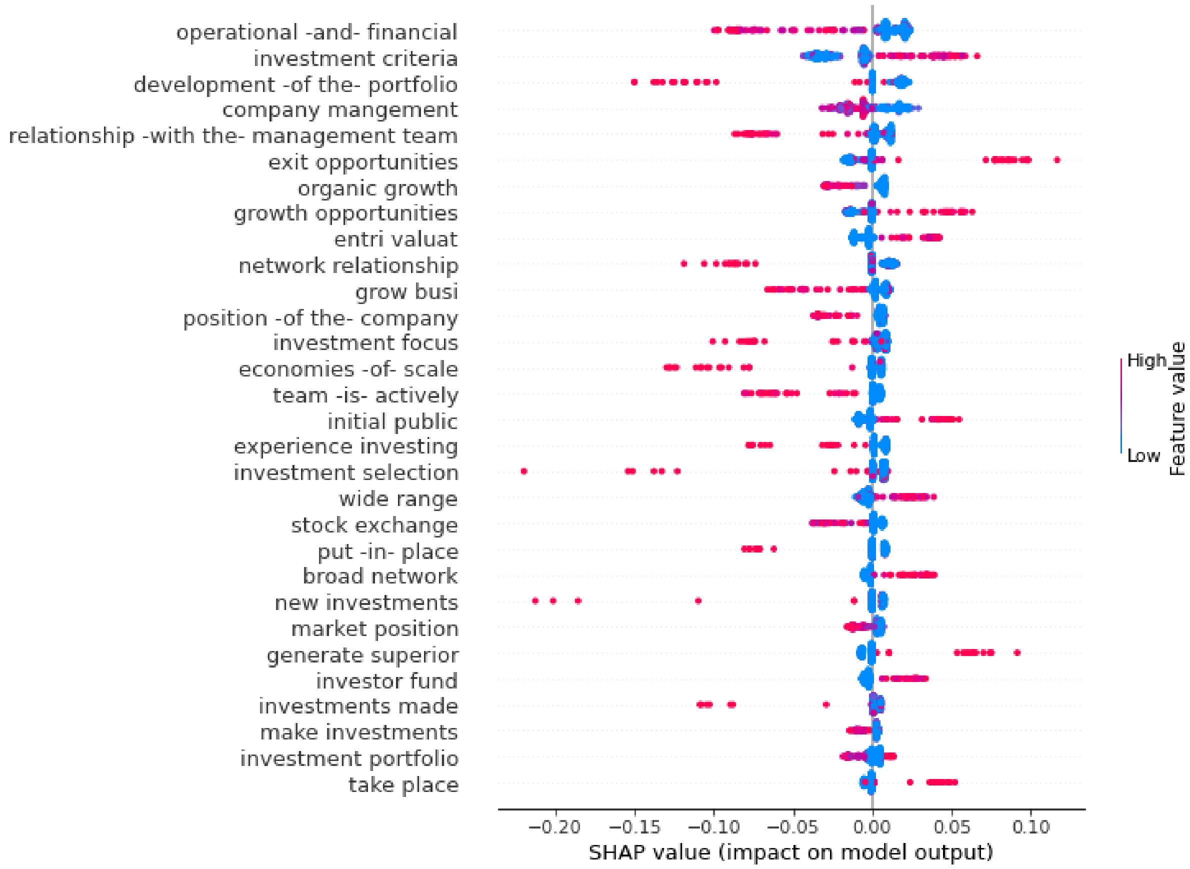


Figure 3. Shapley values at play in an example (Gradient Boosting)

This figure depicts how features contribute to the final prediction of a random example in our dataset. Combinations of words are ranked in decreasing order according to their contribution to the final predicted probability *f(x)*. *E[f(X)]* indicates the average mean predicted probability of Gradient Boosting in the training sample (i.e., funds raised between 2003 and 2013). The x-axis represents the values of the probability of success. The length of the bar represents the Shapley value of each feature.