Cherry-Picking Excellence: A Data-Driven Approach to Choosing the Best Funds and GPs in PE

José C. F. Abreu Neto* Eduardo Marinho† Richard Saito‡

Abstract

This paper employs high-dimensional econometric models to predict whether a private equity (PE) fund will achieve performance success, using only information available to prospective investors at the time of fundraising. In a novel approach, we incorporate identifiers that help isolate general partners (GPs) with a higher likelihood of launching top-performing funds. For venture capital (VC) funds, our models achieve an average accuracy of approximately 72% in predicting funds with abnormal returns. These results suggest that data-driven approaches can improve investment decisions, reduce due diligence efforts, and decrease reliance on subjective judgment.

Keywords: Machine Learning, Private Equity, Venture Capital, Performance, Limited Partners

*FGV - EAESP, Fundação Getúlio Vargas - Escola de Administração de Empresas de São Paulo, Av. 9 de Julho, 2029 Bela Vista, São Paulo - SP 01313-902, Brazil. *Email address*: jose.neto.7@fgv.edu.br

†PUC-Rio, Pontifícia Universidade Católica do Rio de Janeiro, Department of of Economics, Rua Marquês de São Vicente, 225 Gávea, Rio de Janeiro - RJ 22451-900, Brazil. *Email address*: edu.o.marinho@econ.puc-rio.br

‡FGV - EAESP, Fundação Getúlio Vargas - Escola de Administração de Empresas de São Paulo, Av. 9 de Julho, 2029 Bela Vista, São Paulo - SP 01313-902, Brazil. *Email address*: richard.saito@fgv.br

1 Introduction

From 2020 to mid-2023, private equity assets under management (AUM) increased from \$6.1 trillion to approximately \$8.2 trillion, reflecting an annual growth rate of around 8%–10%, driven by a rise in investors' portfolio allocations (McKinsey&Company, 2024). As private equity gains more relevance, limited partners (LPs) are spending more time and effort screening for the best general partners (GPs)[1]. Da Rin and Phalippou (2017) emphasizes the critical role of due diligence in private equity (PE), while David Swensen (Swensen, 2009), former CIO of the Yale Endowment, notes, "*Successful private equity investing requires identifying and affiliating with superior partners*".

In this paper, we aim to improve investors' decision-making by utilizing various econometric tools, with a focus on machine learning (ML) techniques, to predict whether a PE fund will achieve performance success, using only the information available to prospective investors at the time of fundraising. To enhance predictive power, we incorporate novel explanatory variables that reveal key drivers of fund performance, offering deeper insights into the factors that shape PE returns. These findings have the potential to enhance returns, reduce due diligence time, and decrease reliance on human judgment.

With the growing availability of detailed datasets on GPs and funds, we can incorporate more comprehensive information in performance analysis. However, traditional discrete choice econometric models struggle when the number of variables exceeds the sample size. To overcome this, we use ML methods capable of handling high-dimensional data. In this context, we evaluate six models to identify which one best predicts the performance success of PE funds. These models include traditional approaches like the Linear Probability Model (LPM), Logit, and Probit, as well as ML techniques such as Lasso, Ridge, and Random Forest (RF). This comparative analysis helps us identify the most robust method for forecasting private equity outcomes.

We measure whether a fund will have performance success or not, using two key metrics: net internal rate of return (Net IRR) and implied Public Market Equivalent (implied PME)[2]. Successful (or Abnormal) performance is defined as funds with a implied PME greater than 1 or a Net IRR

[1] A recent survey by Private Equity International reports that 69% of LPs dedicate most of their time to due diligence in search of top GPs - `https://www.privateequityinternational.com/lps-put-emphasis-on-due-diligence-as-they-get-selective/`

[2] The Public Market Equivalent (PME) is a metric introduced by Kaplan and Schoar (2005)) to evaluate private equity performance relative to public market benchmarks. It compares the present value of cash flows from a PE fund with the value of investing those same cash flows in a public index. In our analysis, we employ an implied PME, and the methodology underlying its construction is presented in Section 2.

falling within the top tercile, which serve as our dependent variables in the analysis.

For our explanatory variables, we capture six critical aspects from PE literature that are linked to performance: GP identification, fund characteristics, PE industry, market environment, macroeconomic conditions, and PE theory. Including GP identification is essential[3], as Hochberg, Ljungqvist, and Lu (2007) demonstrate that certain GPs consistently outperform their peers. In a novel approach, we include specific identifiers for each GP, enabling us to pinpoint those with stronger explanatory power for fund performance. Fund-specific factors, particularly fund size, show a concave relationship with performance (Kaplan & Schoar, 2005), and sectors such as technology and healthcare are known for superior returns. Additionally, papers by Gompers, Kaplan, and Mukharlyamov (2016); Gompers and Lerner (2000); Gompers, Lerner, Blair, and Hellmann (1998); Harris, Jenkinson, and Kaplan (2014); Kaplan and Schoar (2005); Ljungqvist, Richardson, and Wolfenzon (2020); Robinson and Sensoy (2016) emphasize the importance of PE industry, market environment, and macro conditions, with variables like GDP growth and interest rates providing crucial insights into PE returns. Lastly, PE theory variables relate to how GPs organize funding (Lerner & Schoar, 2004; Maurin, Robinson, & Strömberg, 2023), where the characteristics of LPs play a crucial role in securing capital across different market states, allowing GPs to capture attractive investment opportunities.

We also separate our sample into different groups that have underlying differences that might impact performance. To account for differences in risk-return profiles and investment strategies, we separate BO and VC funds. BO funds typically target mature companies with stable cash flows, while VC funds focus on early-stage firms with higher risks and greater potential upsides (Gompers & Lerner, 1999; Kaplan & Stromberg, 2009). Additionally, we divide our sample into first-time and sequential funds, aligning with the well-known phenomenon in PE called *fund performance persistence*—where prior fund performance predicts future returns for follow-on funds from the same GP (Harris, Jenkinson, Kaplan, & Stucke, 2023; Kaplan & Schoar, 2005). This persistence, explained by various theories (Hochberg, Ljungqvist, & Vissing-Jørgensen, 2014; Lerner & Schoar, 2004; Maurin et al., 2023), highlights the importance of analyzing sequential funds in comparison to first-time funds to better understand performance.

We distinguish our paper from the prior literature in several key ways. First, we are the first to introduce a GP identifier into the regressions, allowing us to pinpoint specific GPs that are more

[3]The GP identification variables are not used in traditional discrete choice econometric models because the number of variables exceeds the sample size. Only the machine learning methods applied here can handle this maximization problem.

likely to succeed, an insight that can improve the screening processes of LPs. Additionally, our analysis separates Venture Capital (VC) and Buyout (BO) funds, as well as first-time and follow-on funds. While Kruglikov and Forthun (2022) also employ machine learning, they group all fund types together. In our view, this distinction is important, as numerous empirical and theoretical studies highlight significant differences between these fund types, making it more appropriate to compare assets with similar characteristics. Furthermore, Morales and Tiozzo (2019) analyzes the probability that a fund's portfolio company will go public, inferring superior returns for those funds. In contrast, we use direct fund and GP information to predict performance without relying on indirect success measures such as IPOs. Additionally, Fernández Tamayo, Braun, Lopez-de Silanes, Phalippou, and Sigrist (2023) employs textual analysis of fund prospectuses to capture performance, while we use other financial data and identification strategies.

The dataset for this study is compiled from multiple sources based on the type of information required. PE data is gathered from Preqin©, a specialized platform for alternative investments, while public market and macroeconomic data are sourced from FRED (Federal Reserve Economic Data) and other specialized providers. The sample period covers the years 2000 to 2017, consisting of 1,267 BO funds and 868 VC funds, totaling 2,135 unique funds.

Our results from applying the six models across different sample group combinations[4], are derived using cross-validation. We train each model with 75% of the total data and test its predictive capacity on the remaining 25%, repeating this process 100 times per model and sample separation. To ensure robustness, we randomly select the training and testing samples with replacement. Additionally, to mitigate survivorship bias—since Preqin's dataset is self-reported and includes over 70% of funds with a PME greater than one—we ensure a balanced distribution by structuring the training and testing datasets to include 50% of funds with PME values above and below one. We apply a similar approach to the IRR threshold, creating a more realistic environment for analysis.

In analyzing the results from our models, we find that ML techniques generally outperformed classical econometric discrete choice models, particularly for VC funds. For ML models, we observe that the predictive accuracy is largely driven by sequential VC funds, achieving an average accuracy of approximately 72% in predicting funds with abnormal returns based on our two performance metrics (Net IRR and Implied PME). In contrast, for first-time VC funds, the accuracy

[4]We categorize our sample into six distinct groups based on fund type—Buyout or Venture Capital—and their classification as either first-time, sequence, or all funds: Buyout all, Buyout first-time funds, Buyout sequence funds, VC all, VC first-time funds, and VC sequence funds.

drops to around 59.8%. For buyout funds, the models perform less effectively, especially with first-time funds, where the results are nearly equivalent to random chance. When considering additional metrics such as precision, recall, specificity, and the F2 score, the same pattern emerges, with VC funds—particularly those in sequence—demonstrating stronger model performance across the board.

These prediction results align with Fernández Tamayo et al. (2023), where they correctly classify 75% of the outperformer funds in the top tercile using prospectus information. However, textual analysis relies heavily on the quality and completeness of prospectus documents, which are not always as readily available as the data used in this paper. Over time, GPs may standardize their approach in response to these findings, reducing the predictability and effectiveness of such analyses. In contrast, our investigation strategy uses structured, objective data, enabling more accurate and consistent predictions across various fund categories.

ML models have significant potential for uncovering insights in finance, but their "black box" nature often limits transparency. However, by employing variable selection techniques, we successfully identify key financial indicators driving our model predictions. Our analysis shows that variables related to GPs, macroeconomic factors, and fund information consistently hold more relevance across both VC and BO funds, regardless of the ML model or performance threshold. Among these, GPs stand out as the most influential predictor. This highlights the critical role of a thorough GP screening process in determining fund performance. Using ML tools, investors can make data-driven decisions and efficiently identify high-potential GPs.

We then examine the top 10 most influential variables across all simulations and observe the dominance of GP-related variables, particularly those linked to specific GP ID numbers. Additionally, macroeconomic variables such as CPIAUCSL (Consumer Price Index for All Urban Consumers), FEDFUNDS (Federal Funds Rate), HOUST (Housing Starts), and PERMIT (New Private Housing Permits) frequently emerge as key drivers. Interestingly, these variables are often interconnected: FEDFUNDS is typically adjusted in response to inflation (CPIAUCSL), which in turn influences housing-related indicators like HOUST and PERMIT.

Analyzing the beta direction for CPIAUCSL, our most relevant macroeconomic variable, reveals a negative correlation, indicating that higher inflation during the fundraising process is linked to lower fund performance. Inflation has the potential to increase operational costs, complicating efforts to improve efficiency—an essential objective in PE (Kaplan & Stromberg, 2009). Additionally, higher inflation often triggers rising interest rates, slowing economic activity and raising the cost of capital, which may depress future valuations. If GPs plan to exit in such conditions, performance might suffer. While these are just a few reasons for lower performance during investments in inflationary periods, the findings highlight the critical role of market timing in PE. This aligns with literature on timing strategies (Robinson & Sensoy, 2016), with inflation emerging as a key factor for analysis.

In addition, for BO funds, "median first-day return" frequently appears as a key predictive variable across several models. This metric, often used to assess the initial performance of public offerings, seems to serve as a proxy for positive market sentiment (Ritter, 1984), indicating favorable conditions in the broader financial environment. In our Lasso and Ridge models, its positive coefficients indicate higher returns for BO funds during fundraising periods marked by positive market sentiment. This aligns with Ljungqvist et al. (2020), who shows that buyout funds tend to earn higher returns when investment opportunities improve, competition for deal flow eases, and credit market conditions loosen. These findings highlight the critical role of timing and sentiment in BO investments, emphasizing the need to carefully evaluate external market conditions.

Our contributions are threefold. First, we apply machine learning tools to predict performance alpha by separating VC and BO funds, including a further division into first-time and sequence funds. This unique approach reflects the distinct underlying characteristics of these funds, which impact performance, and our results demonstrate that this separation enhances predictive power. Second, we include GP identifiers in our ML models, offering a new strategy for investors to identify specific GPs. Finally, we address the effects of unbalanced samples, which can introduce survivorship bias typical in PE databases.

The paper is organized as follows. Section II presents the dataset used in the analysis, Section III details the methodology, Sections IV and V address our results, and VI concludes the paper.

2 Data

The dataset was compiled from diverse sources, depending on the specific type of information required. For the PE data, we used Preqin©, a specialized platform for alternative investments, while public market and macroeconomic data were sourced from FRED and other specialized providers. The sample period spans from 2000 to 2017.

Specifically for PE, Preqin was used to collect information on funds, performance, LPs, and GPs, referencing data from 2022. Funds post-2017 were excluded to allow a five-year maturation period for performance data, aligning with the approach of Harris et al. (2023). From table 1, the sample includes 1,267 BO funds and 868 VC funds, totaling 2,135 unique funds. All VC subdivisions available in Preqin were included, while no subdivisions existed for BO funds in Preqin's database. In Table 2, we present the global regions and industries in which the funds operate. Most funds focus on North America and Europe, consistent with the higher deal values in these regions, which naturally attract more sophisticated investors (Bain&Company, 2024). Additionally, for BO funds, diversified industries dominate, while for VC funds, healthcare and IT emerge as the top two targeted sectors.

We use two measures of fund performance: Net IRR (net internal rate of return) and implied PME (public market equivalent). PME, as defined by (Kaplan & Schoar, 2005), is the ratio of the sum of discounted fund distributions to the discounted capital invested (with the S&P 500's total return as the discount rate). A PME greater than one indicates that the fund outperformed the S&P 500. However, Preqin does not report PME for all funds. To address this, we followed the approach of Sensoy, Wang, and Weisbach (2014) to derive an implied PME. Essentially, this strategy involves using the regression coefficients from Harris et al. (2014) to calculate PMEs from Net IRR, hence the term "implied PME".

Additionally, the performance metrics used in our models were not the original variables. Instead, we modeled these metrics as binary variables: invest or not invest. For implied PME, this adjustment is straightforward since an implied PME above 1 represents a good investment, while below 1 indicates otherwise. For Net IRR, we classified a fund as 1 only if it was in the top 33% of the entire sample, and as 0 otherwise.

The sample of funds was also separated into first-time funds and sequential funds. First-time funds are those classified by Preqin as 1 in their *fund number series* classification. In other words, these are the first-time funds of a specific series by a given GP, but not necessarily the GP's first overall fund. GPs might shift their focus to different regions or industries, positioning themselves in new ventures and taking on unfamiliar risks. This lack of prior experience can result in limited information for both GPs and LPs. We aim to capture this initial exposure through the *fund number series* classification provided by Preqin.

For sequential funds, this refers to subsequent funds launched by the same GP. To accurately separate the sequential funds, we followed the methodology described by Harris et al. (2023). Our initial step involved organizing the database by fund name, GP, vintage year, and sequence number. To ensure consistency, we excluded funds that did not align with their subdivision class or primary region of operation, such as those shifting focus from Europe to Asia. We also excluded funds with different characteristics, like annex or side funds. After sorting, we reviewed the sequence numbers to identify only adjacent pairs for analysis.

To examine aspects of private equity theory and empirical findings, we incorporate variables grounded in network theory. Following the methodology of Abreu Neto and Saito (2024), who analyze LP-GP interactions, they find that persistent LPs—also referred to as 1st Quartile LPs—those with strong, consistent reinvestments with GPs, serve as reliable indicators of GP quality/performance. This aligns with the theories proposed by Hochberg et al. (2014); Lerner and Schoar (2004); Maurin et al. (2023), which highlight a consistent match between successful GPs and specific LPs through follow-on funds, resulting in superior returns for these LPs. Additionally, we apply centrality measures at the GP level as a proxy for GP demand by LPs, akin to the oversubscription phenomenon identified by Kaplan and Schoar (2005), which is associated with higher returns. Similarly, Abreu Neto and Saito (2024) find that GPs with high centrality in a bipartite network of LPs and GPs tend to outperform their peers.

In addition to PE data, our analysis incorporated macroeconomic and public market information. For macroeconomic data, we utilized FRED and the Baker, Bloom, and Davis website, though the latter was used exclusively for a specific Uncertainty Index. Public market data was gathered from various sources, including Eikon, FRED, the Kenneth French website, and Jay R. Ritter's IPO Data website. Table 3 provides a detailed overview of the sources and corresponding website links for each variable.

3 Methodology

The central idea of this article is to use various econometric tools, including new machine learning techniques capable of handling big data, to predict a fund's potential success during the fundraising process. We aim to demonstrate, ex ante to the fund launch, the likelihood of a fund achieving abnormal performance based on different characteristic groups. Additionally, we seek to identify the main factors that influence PE returns. Abnormal performance is measured in two ways: PME above 1 or IRR above the first-tercile threshold (our dependent variables). To predict these binary outcomes, we employ discrete choice models.

The set of predictors (our independent variables or factors) is detailed in Table 3, grouped into six categories: GPs, fund information, PE theory and centrality measures, PE industry, macroeconomic factors, and market conditions/environment. These categories include 55, 58, and 55 covariates for VC first funds, VC sequence funds, and the full VC sample, respectively, and 54, 57, and 54 covariates for BO first funds, BO sequence funds, and the full BO sample, respectively. In VC, there is an additional strategy variable provided by Preqin, which is absent in BO data. Notably, categorical variables are transformed into dummy variables in the econometric models, significantly increasing the number of variables on the right-hand side of the equations. Due to the small sample size and the cross-validation approach used for population stratification, GP variables were excluded from traditional discrete choice econometric models, as the number of variables exceeded the sample size. Only the machine learning methods applied here are capable of handling this maximization issue to estimate the beta coefficients.

We conduct a horse race between six models to determine which has the best accuracy in predicting the binary events mentioned above. These models include traditional approaches like the Linear Probability Model (LPM), Logit, and Probit, as well as machine learning techniques such as Lasso, Ridge, and Random Forest (RF).

There are distinct idiosyncrasies between BO and VC funds, prompting us to conduct separate experiments for each fund type. BO funds typically target mature companies with stable cash flows, while VC funds focus on early-stage firms with higher risks and greater potential upsides (Gompers & Lerner, 1999; Kaplan & Stromberg, 2009). Additionally, we divide the experiment into three categories: first-time funds, sequential funds, and all funds within each type. The rationale is that during the fundraising process for sequential funds, more information is available, and in a market characterized by information asymmetry, this should make prediction easier—which is indeed the case. Moreover, performance persistence, a well-known phenomenon in PE where prior fund performance predicts future returns for follow-on funds from the same GP (Harris et al., 2023; Kaplan & Schoar, 2005), further justifies separating the analysis for first-time and sequential funds. This persistence is explained by several theories (Hochberg et al., 2014; Lerner & Schoar, 2004; Maurin et al., 2023).

For each group of funds, we implement cross-validation by training the model on 75% of the total data and testing its predictive capacity on the remaining 25%. This process is repeated 100 times for each model and each experiment, and we collect the average results. The training and testing samples are selected randomly, with replacement after each iteration of the 100 times.

There is ongoing discussion about the presence of survivorship bias in PE market data, as it is self-reported and privately held. Indeed, over 70% of the funds in our data show PME greater than one. To create a more realistic environment and mitigate potential survivorship bias, we ensure that both the training and testing samples contain an equal 50/50 split of funds with PME greater than and below one, as well as for the IRR threshold.

In summary, we categorize the funds into six classes (Buyout full, Buyout first funds, Buyout sequence funds, VC full, VC first funds, VC sequence funds) and aim to predict two binary events (implied PME and IRR threshold). We compare the performance of six models (LPM, Logit, Probit, Lasso, Ridge, and Random Forest) using bagging with 100 cross-validation iterations for each experiment, totaling 7,200 prediction exercises. Figure 3 below illustrates our methodology.

3.1 Models

We use six models to predict binary events (0 for below threshold and 1 for above threshold). The first model assumes a simpler relationship between covariates and the dependent variable, with a linear marginal effect. Probit and Logit models introduce non-linear functions, aiming for a better fit in binary outcomes. Ridge and Lasso models apply regularization, shrinking coefficients to reduce overfitting. Lastly, the Random Forest (RF) model captures complex, non-linear relationships using an ensemble of decision trees in an unsupervised learning framework.

3.1.1 LPM - Linear Probability Model

LPM applies linear regression to binary outcomes, estimating the probability that the dependent variable equals 1. It assumes a constant marginal effect of covariates on the event's probability, providing a straightforward approach but with potential limitations, such as predicted probabilities outside the [0,1] range.

$$P(y=1|x) = \beta_0 + \sum_{i}^{N} \beta_i x_i \tag{1}$$

3.1.2 Logit

The Logit model captures non-linear relationships by applying a logistic function, ensuring predicted probabilities remain between 0 and 1.

$$P(y=1|x) = \frac{e^{x'\beta}}{1+e^{x'\beta}} \tag{2}$$

3.1.3 Probit

Similar to Logit, the Probit model uses a normal cumulative distribution function (CDF) to model binary outcomes.

$$\Phi(x'\beta) = \int_{-\infty}^{x'\beta} \phi(z)dz \tag{3}$$

3.1.4 Random Forest (RF)

The Random Forest model builds multiple regression trees and using the principles of Bagging together, where each tree is trained on a random subset of the data. The model generates predictions by averaging the outputs of these trees, enhancing accuracy and reducing overfitting. The regression tree can be described as follows:

$$f(x) = E[P(y=1|x)] = \sum_{m=1}^{M} W_m I(x \in R_m) = \sum_{m=1}^{M} W_m \phi(x; V_m) \tag{4}$$

R_m represents the tree regions m, W_m the average answer of P(y=1|x) to this region, and V_m is the variable choice to be divided and the threshold. While x is the set of predictors, comprising approximately 55 variables across five groups, as detailed in Table 3.

A detailed explanation of how the algorithm determines the number of regions M and selects variables from each bootstrap sample is provided in Medeiros, Vasconcelos, Veiga, and Zilberman

(2021), Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. This can be found in the appendix of the Journal of Business and Economic Statistics, Section B8.

3.1.5 Lasso

Lasso introduces L1 regularization, shrinking coefficients and selecting variables by setting irrelevant ones to zero (Bühlmann & Van De Geer, 2011). This model is useful in high-dimensional settings to reduce overfitting and improve interpretability.

$$\hat{\mu}(\lambda), \hat{\beta}(\lambda) = argmin_{\mu,\beta}(\frac{1}{n}\sum_{i=1}^{n}\rho_{\mu,\beta}(X_i, Y_i) + \lambda\|\beta\|_1) \tag{5}$$

Being $Y_i|X_i = x\ Bernoulli(\pi(x)), \rho(f, y) = ln(1 + exp(-(2y - 1)f) = ln(1 + exp(-\ yf))$ $y = 2y - 1 \in -1, 1, ln(\frac{\pi(x)}{1-\pi(x)}) = \pi + \sum_{j=1}^{p}\beta_j x^{(j)} = f_{\mu,\beta}(x)$.

3.1.6 Ridge

Ridge regression applies L2 regularization, shrinking coefficients without eliminating variables, making it effective in managing multicollinearity (Bühlmann & Van De Geer, 2011).

$$\hat{\mu}(\lambda), \hat{\beta}(\lambda) = argmin_{\mu,\beta}(\frac{1}{n}\sum_{i=1}^{n}\rho_{\mu,\beta}(X_i, Y_i) + \lambda\beta^2) \tag{6}$$

Being $Y_i|X_i = x\ Bernoulli(\pi(x)), \rho(f, y) = ln(1 + exp(-(2y - 1)f) = ln(1 + exp(-\ yf))$ $y = 2y - 1 \in -1, 1, ln(\frac{\pi(x)}{1-\pi(x)}) = \pi + \sum_{j=1}^{p}\beta_j x^{(j)} = f_{\mu,\beta}(x)$.

4 Results

To evaluate the results of our prediction models, we use confusion matrices. These matrices display how well the model's predictions align with actual outcomes by categorizing the results into four groups: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). From this breakdown, we can derive key performance metrics that help assess the accuracy and effectiveness of each model (Powers, 2020).

We present our results by first separating Buyout and Venture Capital funds. Within each type of fund, we further divide the analysis into 1st-time, sequence and all funds. For each model, the results show that ML techniques consistently outperform classical discrete choice econometric models in terms of accuracy—the ability to correctly predict ones and zeros, indicating whether funds exceed or fall below the proposed threshold.

4.1 VC results

Panel A of Table 4 presents the accuracy results for VC funds. The first noticeable pattern is that ML techniques consistently outperform traditional discrete choice models across all cases. For first-time funds using the PME measure, the results from traditional models are worse than random chance. The Lasso model emerges as the top performer in most experiments, only being surpassed by the Ridge model when predicting PME for first-time funds.

It's also important to note that the population of VC funds is smaller than BO funds in all our experiments. However, despite the smaller sample size, the prediction results for the VC funds are significantly stronger.

When analyzing the confusion matrix in table 5, the results show a high degree of consistency, with balanced predictions of ones and zeros. Both Type I and Type II errors[5] are closely aligned, which is a promising outcome if these models are intended for use as decision-making tools by LPs at the investment stage. Achieving 75% accuracy in predicting a fund's relative success during the fundraising process would be highly beneficial in helping LPs decide whether to invest.

[5] A Type I Error or False Positive (FP) in our confusion matrix occurs when the model incorrectly predicts a positive outcome (e.g., abnormal fund performance) when the actual outcome is negative (e.g., the fund underperforms). On the other hand, a Type II Error or False Negative (FN) happens when the model incorrectly predicts a negative outcome (e.g., underperformance) when the actual outcome is positive (e.g., the fund outperforms).

The superior performance of ML models, particularly in predicting the success of VC funds, is evident. Despite potential complexity—such as having more independent variables than sample size due to the inclusion of GP dummies (as discussed in section 3)—the models used in this study remain interpretable. In the next section, we explore the most significant variables for predicting performance, providing further theoretical insights by dissecting the results.

4.2 Buyout results

Panel B of Table 4 presents the models' accuracy in predicting, out of sample, whether a BO fund will achieve abnormal performance. It can be observed that the models perform poorly, especially with first-time funds, with results nearly equivalent to random guessing. Specifically, for the PME measure, the task is to identify which funds underperformed (PME less than one), despite over 90% of funds having a PME greater than one.

The results improve slightly for sequence funds, with Lasso and LPM achieving 64% accuracy. This aligns with the theoretical expectation that, over time, some information becomes more accessible in markets characterized by asymmetric information.

In table 6, which provides the confusion matrices, the results are relatively balanced. The Lasso model, using the IRR threshold, performs better than the others in predicting both ones and zeros, supporting its higher accuracy.

Overall, the BO results, particularly with the Lasso model for sequence funds, provide useful insights for LP decision-makers at the time of investment. However, the more promising results are found in the VC funds, which will be discussed in the next section.

5 Results - Variable Selection

Machine learning models have significant potential for uncovering insights in finance, but their "black box" nature often limits transparency. In this section, however, by employing variable selection techniques, we were able to identify key financial indicators that drive our model predictions, thereby shedding light on the complex workings of machine learning models in financial applications.

In Tables 7 and 8, we present the main descriptive statistics for each category of variables used in our ML models. In Lasso and Ridge models, regressions capture the strength and direction of the linear relationship between variables and outcomes through beta coefficients (β), while RF evaluates variable importance based on its contribution to predictive accuracy in non-linear interactions.

To identify key drivers of performance across different modeling approaches, we aim to compare the absolute values of the beta coefficients for Lasso and Ridge, alongside the variable importance scores from RF, as detailed in Tables 9 and 10. Additionally, we refer to the outputs from ML models as "feature significance measures" because they indicate the relative importance or contribution of each feature (or variable) in making predictions.

Our results reveal that the categories of GPs, Macro factors, and Fund Information are consistently more relevant than others for both VC and BO across different ML models and performance thresholds, with GPs being by far the most influential. This highlights the importance of a thorough screening process when selecting GPs, as they emerge as a key predictive variable for fund performance.

When examining the top 10 most influential variables across all simulations, we find that GPs, macro, and fund information consistently emerge as the primary groups. This pattern holds true for both BO and VC funds. The dominance of GP-related variables, particularly those identified by GP ID numbers, is evident. Additionally, CPIAUCSL frequently emerges as a key macroeconomic variable, showing a negative beta direction. This shows the critical impact of inflation when investing in VC and BO funds. Interestingly, for BO funds, "median first day return" typically appears as a key predictive variable in several models. This variable, along with other PE industry variables identified in RF models, likely serves as a proxy for positive market sentiment, highlighting the importance of timing when participating in BO funds, particularly for first-time funds.

5.1 VC Analysis

5.1.1 Categories/Group Analysis - feature significance measures

In Table 9, we present key statistics for the absolute beta coefficients from Ridge and Lasso, along with variable importance scores from Random Forest (RF), applied to VC funds. These models are used to predict PME (Public Market Equivalent) and Net IRR (Internal Rate of Return) thresholds across different fund samples, including the Full Sample, First-Time Funds, and Sequence. For each sample, the variables used in the predictions are categorized into different groups: Fund Information, GPs, Macro, Market Conditions, PE Industry, and PE Theory & Centrality Measures.

When assessing the main categories, we find that the top three are GPs, Macro, and Fund Information. However, across various metrics—such as the sum of absolute values, averages, medians, and maximums— it is clear that GPs consistently have the most significant impact on both PME and Net IRR models. The GPs category demonstrates the highest sum of absolute values and strong averages, suggesting a considerable and consistent influence on performance metrics across various samples, including first-time and sequence funds. For example, in the full sample, the GPs group shows sum values of 117.47 (Ridge) and 133.46 (Ridge) for PME and Net IRR models, respectively, which far exceeds those of other groups. This indicates that the performance of GPs is crucial across different fund samples, making it the primary focus for decision-making.

The Macro category also plays a important role, especially when considering the maximum values observed. In certain cases, such as with sequence funds, Macro factors achieve a maximum value of 7.23 (PME Ridge) and 7.38 (Net IRR Ridge), indicating that while they may not always have the highest average impact, they can exert substantial influence under specific conditions/simulations. This variability suggests that macroeconomic variables should be closely monitored, particularly in environments with significant economic fluctuations. The consistency in the direction of action between PME and Net IRR models for Macro factors further strengthens this conclusion.

Fund Information, although slightly less impactful than GPs and Macro, still provides essential insights, particularly in terms of median values, which suggest a steady influence across models. For example, in first-time funds, the median values of Fund Information are relatively consistent across both PME and Net IRR models, reinforcing its role as a interesting source of information.

In summary, GPs are the most influential group in both PME and Net IRR models, across both first-time and sequence funds, with high sums and averages indicating significant impact on performance. Macro factors also play a important role, particularly under certain conditions, as suggested by their maximum values. Fund Information is slightly less influential but still valuable, particularly for its consistency across models. Other groups, such as Market Conditions, PE Industry, and PE Theory & Centrality Measures, are less relevant due to their lower sums and averages. Strategic focus should be on GPs and Macro, with Fund Information offering additional support, especially for first-time funds.

5.1.2 Top 10 most influential variables in VC funds

Table 13 shows that the top 10 variables with the highest feature significance measures in VC funds from our ML models. The analysis reveals that GP-related and Macro variables are the primary groups influencing performance across different models.

The dominance of GP related variables appears prominently in all models, performance threshold and fund sample levels. In the Ridge and Lasso models, the most significant variables often exhibit a negative direction, indicating potential challenges that specific GPs or their decisions might impose on fund performance.

In addition to GP-related variables, macroeconomic indicators such as CPIAUCSL and RPI in the Ridge and Lasso models highlight the sensitivity of private equity performance to broader economic conditions. The negative direction of their betas suggests that higher inflation correlates with lower fund performance, indicating that VC funds may struggle to maintain returns in inflationary environments, particularly at the time the fund begins its operations.

The RF model highlights FEDFUNDS (Federal Funds Rate) along with HOUST (Housing Starts) and PERMIT (New Private Housing Permits) as key drivers of private equity performance. Meanwhile, the Lasso and Ridge models emphasize CPIAUCSL (Consumer Price Index). These variables are interconnected: FEDFUNDS is often adjusted in response to changes in inflation (CPIAUCSL), and this, in turn, affects housing-related indicators like HOUST and PERMIT.

The inclusion of HOUST and PERMIT in the RF model suggests that the model captures broader economic conditions, where interest rates influence both inflation and the housing market. This reinforces the idea that macroeconomic factors—particularly those linked to inflation, monetary policy, and housing market conditions—are critical in shaping both relative (PME) and absolute (Net IRR) returns in private equity. The RF model's focus on these variables complements the direct emphasis on inflation seen in the Lasso and Ridge models, providing a more comprehensive view of how economic conditions impact fund performance.

5.2 BO Analysis

5.2.1 Categories/Group Analysis - feature significance measures

In our analysis of BO fund performance (table 10), similar to VC funds, we also found that the key categories are GPs, Macro factors and Fund Information.

GPs consistently emerge as the most influential category across all samples and thresholds. For example, in the Full Sample under the PME threshold, the Ridge model shows a sum of absolute betas of 81.08 with an average of 0.18, median of 0.13, and a maximum value of 0.85. This strong influence is mirrored under the Net IRR threshold, where the Ridge model again highlights the GPs' dominance with a sum of 107.90 and an average of 0.24. These findings suggest that GP characteristics are crucial across different performance measures and stages of fund development, indicating that PE firms should prioritize enhancing the capabilities and track records of their GPs.

Fund Information also proves to be significant, particularly under the Net IRR threshold, where the Sequence Sample shows a sum of 4.77 with an average of 0.14 and a median of 0.11 in the Ridge model. This category, however, exhibits less influence compared to GPs, especially in the First Fund sample, where the sum of absolute betas is only 1.37 in the Ridge model. Despite this, the consistent importance of Fund Information across all models suggests that maintaining comprehensive and transparent fund-specific data remains critical, particularly for sequential funds, where this information appears to have a more pronounced impact on performance.

Macro factors show varying levels of influence depending on the sample and threshold. In the Sequence Sample under the PME threshold, the Ridge model highlights a significant impact with a sum of 4.88 and an average of 0.24. This impact is slightly reduced under the Net IRR threshold but remains notable, suggesting that macroeconomic conditions are particularly relevant for sequential funds. The variability in the influence of Macro factors indicates that while they are important, their relevance may fluctuate depending on the specific context of the fund, requiring PE firms to stay attuned to broader economic trends.

When comparing the thresholds (PME and Net IRR), both generally indicate the same direction of action, particularly highlighting the dominance of GPs and the importance of Fund Information. However, the PME threshold appears to provide slightly more consistent results across different samples, suggesting it might be a more reliable performance measure in this context. Additionally, between the First Fund and Sequence Fund samples, there is a clearer direction and more robust conclusions drawn from the Sequence Sample, likely due to the accumulated experience and data available from multiple fund cycles.

5.2.2 Top 10 most influential variables in BO funds

The table 14 reveals that GP-related and Macro variables are also, just like for VC funds, important categories influencing performance across different models. Specially for our linear models Ridge and Lasso macro variable CPIAUCSL was the most important.

However, different from VCs funds, in the linear models ridge and lasso the group market conditions and envirnoment delivers interesting predictive power - being represented by "median first day return". This happens independent of the performance threshold, and fund sample.

The "median first day return" is a key predictive variable in Lasso and Ridge models as it effectively be capturing overall market sentiment and investor confidence during the time a fund is launched. This variable serves as a proxy for market conditions and timing (Ritter, 1984), and its positive direction shows that higher "median first day return" provides better performance prediction across thresholds and fund samples.

When analyzing the RF model, in the first-time funds sample, many of the variables identified by the RF model can be seen as capturing sentiments or conditions similar to those indicated by the "median first day return." These variables reflect investor confidence, market conditions, and sectoral or regional attractiveness, which are all factors that could drive both successful first-day returns in IPOs and the performance of first-time funds. While these variables may not be identical to the median first-day return, they can serve as proxies or are correlated with similar market sentiments, making them relevant in assessing the performance of first-time funds.

On the other hand, the lack of a similar pattern in sequential funds compared to first-time funds likely reflects the different nature and drivers of these funds. Sequence funds might benefit from an established GP track record and investor base, which shifts the focus away from short-term market sentiment (captured by the "median first day return") towards broader macroeconomic conditions and long-term strategic factors. The RF model, therefore, emphasizes variables that reflect these more stable and long-term influences, explaining why "median first day return" and similar sentiment-based variables are less prominent in the sequence funds sample.

6 Conclusion

We demonstrate that recent advances in machine learning methods, combined with the availability of new, rich datasets, make it possible to improve PE fund performance forecasts. Models such as Ridge, Lasso, and Random Forest generate more accurate predictions than traditional discrete choice econometric methods. These results leverages the value of ML and big data in private equity forecasting.

Our analysis shows that predictions for VC funds yield better outcome projections compared to BO funds, where Lasso models performed better than others. When we separate the sample into first-time and sequence funds, we find that the improved accuracy primarily comes from sequence funds rather than first-time funds.

Examining the variables selected by the ML models during the fundraising process, we identify inflation—measured through the CPI index—as a key driver of performance. Its negative coefficients indicate that inflationary periods challenge GPs to avoid overpaying and their portfolio companies in operating effectively to deliver superior performance. However, this finding promotes a deeper investigation in the reason behind the negative correlation between inflation and performance and creates a promising area for future research. Timing investments around inflation is crucial for generating alpha.

Although we apply ML techniques, the significance of GPs, as demonstrated by several other empirical studies on PE performance, remains consistent with findings using traditional econometric methods. In other words, this offers an alternative way to capture the fixed effects of GPs in a big data environment. Also, by identifying individual GPs throughout the simulations, we can pinpoint the top-performing GPs and show that this group plays a critical role in improving PE prediction accuracy.

References

Abreu Neto, J. C., & Saito, R. (2024). Unlocking the power of relationships: Limited partner networks and performance in private equity. *Available at SSRN 4754637*.

Bain&Company. (2024). Global private equity report 2024. *Global Private Equity Report 2024*, *1*(1).

Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Da Rin, M., & Phalippou, L. (2017). The importance of size in private equity: Evidence from a survey of limited partners. *Journal of Financial Intermediation*, *31*, 64–76.

Fernández Tamayo, B., Braun, R., Lopez-de Silanes, F., Phalippou, L., & Sigrist, N. (2023). Limited partners versus unlimited machines; artificial intelligence and the performance of private equity funds. *Artificial Intelligence and the Performance of Private Equity Funds (December 15, 2023)*.

Gompers, P., Kaplan, S. N., & Mukharlyamov, V. (2016). What do private equity firms say they do? *Journal of Financial Economics*, *121*(3), 449–476.

Gompers, P., & Lerner, J. (1999). *What drives venture capital fundraising?* National bureau of economic research Cambridge, Mass., USA.

Gompers, P., & Lerner, J. (2000). Money chasing deals? the impact of fund inflows on private equity valuations. *Journal of financial economics*, *55*(2), 281–325.

Gompers, P., Lerner, J., Blair, M. M., & Hellmann, T. (1998). What drives venture capital fundraising? *Brookings Papers on Economic Activity. Microeconomics*, *1998*, 149–204. Retrieved 2025-01-19, from http://www.jstor.org/stable/2534802

Harris, R. S., Jenkinson, T., & Kaplan, S. N. (2014). Private equity performance: What do we know? *The Journal of Finance*, *69*(5), 1851–1882.

Harris, R. S., Jenkinson, T., Kaplan, S. N., & Stucke, R. (2023). Has persistence persisted in private equity? evidence from buyout and venture capital funds. *Journal of Corporate Finance*, 102361.

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.

Kaplan, S. N., & Schoar, A. (2005). Private equity performance: Returns, persistence, and capital flows. *The Journal of Finance*, *60*(4), 1791–1823.

Kaplan, S. N., & Stromberg, P. (2009). Leveraged buyouts and private equity. *Journal of economic perspectives*, *23*(1), 121–46.

Kruglikov, N., & Forthun, A. (2022). *Predicting private equity fund performance with machine learning* (Unpublished master's thesis). Norwegian School of Economics.

Lerner, J., & Schoar, A. (2004). The illiquidity puzzle: theory and evidence from private equity. *Journal of Financial Economics*, *72*(1), 3–40.

Ljungqvist, A., Richardson, M., & Wolfenzon, D. (2020). The investment behavior of buyout funds: Theory and evidence. *Financial Management*, *49*(1), 3–32.

Maurin, V., Robinson, D. T., & Strömberg, P. (2023). A theory of liquidity in private equity. *Management science*, *69*(10), 5740–5771.

McKinsey&Company. (2024). Private markets in a slower era. *McKinsey Global Private Markets Review 2024*, *1*(1).

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119.

Morales, M. H., & Tiozzo, V. (2019). *Machine learning and survival prediction models for decision support in the private equity market* (Unpublished doctoral dissertation). Politecnico di Torino.

Powers, D. M. (2020). Evaluation: from precision, recall and f-measure to roc, informedness,

markedness and correlation. *arXiv preprint arXiv:2010.16061*.
Ritter, J. R. (1984). The" hot issue" market of 1980. *Journal of business*, 215–240.
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.
Sensoy, B. A., Wang, Y., & Weisbach, M. S. (2014). Limited partner performance and the maturing of the private equity industry. *Journal of Financial Economics*, *112*(3), 320–343.
Swensen, D. F. (2009). *Pioneering portfolio management: An unconventional approach to institutional investment, fully revised and updated*. Simon and Schuster.

Table 1

The table provides descriptive statistics for a sample of 1,267 venture capital and 868 buyout funds with vintage years between 2000 and 2017, categorized by the type of funds used (full sample, first-time funds, and sequential funds). Net IRR performance data is sourced from Preqin, while the Implied PME was calculated using regression coefficients reported by Harris et al. (2014), following a similar approach to Sensoy et al. (2014). Vintage year refers to the year the fund begins its operations. Size indicates the fund's volume in terms of invested capital. The overall fund sequence number represents the sequential fund reference number for a specific GP.

Fund Characteristics	Full Sample				First-time Funds				Sequence Funds			
	N	Mean	Median	SD	N	Mean	Median	SD	N	Mean	Median	SD
Buyouts funds												
Net IRR (%)	1267	17,24	15,80	14,34	177	17,72	16,00	14,49	841	17,57	16,00	13,37
Implied PME	1267	1,44	1,40	0,38	177	1,46	1,42	0,40	841	1,45	1,41	0,36
Vintage Year	1267	2009	2008	5	177	2006	2006	4	841	2010	2010	5
Size (millions of dollars)	1267	1.515,9	602,0	2.487,7	177	435,8	265,0	559,5	841	1.951,8	803,2	2.876,6
Overall fund sequence number	1267	5	3	6	177	2	1	4	841	6	4	6
Venture Capital funds												
Net IRR (%)	868	12,18	9,65	31,82	170	11,77	10,15	18,11	479	11,37	9,94	17,09
Implied PME	868	1,37	1,28	1,12	170	1,35	1,31	0,64	479	1,34	1,29	0,60
Vintage Year	868	2008	2007	5	170	2007	2007	5	479	2008	2008	5
Size (millions of dollars)	868	301,3	206,0	358,4	170	134,5	89,5	167,4	479	371,5	275,4	387,3
Overall fund sequence number	868	5	4	6	170	3	1	4	479	5	4	4

Table 2

The table presents characteristics of a sample of 1,267 venture capital and 868 buyout funds raised between 2000 and 2017, focusing on their primary investment regions and core industries. The statistics are provided for the full sample, as well as for first-time and sequential funds. The region and core industry classifications were provided by Preqin.

Funds Region Focus and Core Industries	Buyout Funds			Venture Capital Funds		
	Full Sample	First-time Fund Sample	Sequence Fund Sample	Full Sample	First-time Fund Sample	Sequence Fund Sample
Funds Primary Region Focus						
Africa	5	1	1	5	2	2
Americas	25	4	13	8	5	2
Asia	65	11	39	81	28	31
Australasia	37	6	22	9	7	1
Europe	383	52	248	110	35	49
Middle East & Israel	12	3	8	27	5	14
North America	740	100	510	623	87	377
Diversified Multi-Regional	-	-	-	5	1	3
Total	*1.267*	*177*	*841*	*868*	*170*	*479*
Funds Core Industries						
Business Services	18	2	15	5	2	3
Business Services, Consumer Discretionary	1	1	-	-	-	-
Business Services, Information Technology	2	-	1	-	-	-
Consumer Discretionary	79	14	51	14	5	7
Consumer Discretionary, Financial & Insurance Services	1	-	1	-	-	-
Consumer Discretionary, Raw Materials & Natural Resources	1	-	-	-	-	-
Diversified	924	120	626	179	38	82
Energy & Utilities	12	1	7	26	5	14
Energy & Utilities, Real Estate	2	-	2	-	-	-
Financial & Insurance Services	14	5	6	9	3	3
Healthcare	32	7	19	197	44	121
Healthcare, Information Technology	6	-	4	65	5	41
Industrials	73	13	42	1	1	
Information Technology	54	6	35	247	50	136
Information Technology, Telecoms & Media	6	4	2	66	4	40
Others	5	1	2	13	2	6
Raw Materials & Natural Resources	3	-	2	5	1	3
Real Estate	9	1	8	-	-	-
Telecoms & Media	25	2	18	41	10	23
Total	*1.267*	*177*	*841*	*868*	*170*	*479*

Table 3: Explanatory Variable description and other information

This table presents all 54 explanatory variables utilized in our prediction models. We have categorized each variable into six distinct groups based on the original exposure: GPs, Fund Information, PE Theory & Centrality Measures, PE Industry, Macro and Market Conditions/Market Environment. (1) Only for VC funds, as Preqin does not detail the strategy for BO funds. The VC strategies are: Early Stage: Seed, Early Stage: Start-up, Early Stage, Venture (General), Expansion / Late Stage. (2) We separate the regions as: North America, Europe, Asia, Australasia, Middle East & Israel, Americas. (3) Funds core industries are given/classified by Preqin. (4) For every vintage year, we found eigenvector centrality measure for GPs using a past 5yr rolling window. (5) For each fund, we calculated the average eigenvector centrality of all LPs invested in that fund. (6) For each fund, we determined the proportion of LPs classified in the 1st quartile (with the highest centrality measures) relative to the total number of LPs in the fund. (7) Time variable, where we have funds vintage year minus 1980 (base year). (8) `https://www.policyuncertainty.com/` (9) `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` (10) `https://site.warrington.ufl.edu/ritter/ipo-data/`

Explanatory Variable	**Group Classification**	**Type of Variables**	**Source**
GPs Firm Identification (ID)	GPs	Categorical	Preqin
Fund Strategy [1]	Fund information	Categorical	Preqin
Fund Primary Region of Focus [2]	Fund information	Categorical	Preqin
Fund Core Industries [3]	Fund information	Categorical	Preqin
GPs Headquarter Location/Country	Fund information	Categorical	Preqin
Fund Size (U$MM)	Fund information	Numerical	Preqin
Fund Number Overall in GP	Fund information	Numerical	Preqin
Eigenvector ex-ante Centrality measure GP [4]	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Average LPs Eigenvector Centrality per fund [5]	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Proportion of 1Q LPs (ex-ante) per fund [6]	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Quantity of LPs per fund	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Number of LPs reinvested per fund	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Number of 1Q LPs reinvested per fund	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Fund Size (U$MM) previous fund from same GP	PE Theory & Centrality Measures	Numerical	Preqin & Abreu Saito (2024)
Industry Dry Powder	PE Industry	Numerical	Preqin
Industry Fund Raising Volume	PE Industry	Numerical	Preqin
Industry Fund Raising Number	PE Industry	Numerical	Preqin
Industry Year [7]	PE Industry	Numerical	Preqin
University Michigan - Consumer Sentiment (UMCSENTx)	Macro	Numerical	FRED
Uncertainty Index (Uncertain)	Macro	Numerical	Baker, Bloom and Davis website [8]
Housing Starts: Total New Privately Owned (HOUST)	Macro	Numerical	FRED
New Private Housing Permits (PERMIT)	Macro	Numerical	FRED
Efective Federal Funds Rate (FEDFUNDS)	Macro	Numerical	FRED
1-Year Treasury Const. Minus Fed Funds (T1YFFM)	Macro	Numerical	FRED
5-Year Treasury Const. Minus Fed Funds (T5YFFM)	Macro	Numerical	FRED
10-Year Treasury Const. Minus Fed Funds (T10YFFM)	Macro	Numerical	FRED
Moody's Seasoned Aaa Corp. Bond Minus Federal Funds Rate (AAAFFM)	Macro	Numerical	FRED
Moody's Seasoned Baa Corp. Bond Minus Federal Funds Rate (BAAFFM)	Macro	Numerical	FRED
Help-Wanted Index for United States (HWI)	Macro	Numerical	FRED
Ratio of Help Wanted/No. Unemployed (HWI Ratio)	Macro	Numerical	FRED
Unemployment rate (UNRATE)	Macro	Numerical	FRED
Initial Jobless Claims (CLAIMSx)	Macro	Numerical	FRED
Industrial Production (INDPRO)	Macro	Numerical	FRED
Retail Price Index (RPI)	Macro	Numerical	FRED
Capacity Utilization: Manufacturing (CUMFNS)	Macro	Numerical	FRED
Crude Oil, spliced WTI and Cushing Price (OILPRICEx)	Macro	Numerical	FRED
Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)	Macro	Numerical	FRED
US GDP (last year)	Macro	Numerical	FRED
S&P 500 returns	Market Conditions/Environment	Numerical	FRED
S&P Dividend Yield	Market Conditions/Environment	Numerical	FRED
S&P Price-Earnings Ratio	Market Conditions/Environment	Numerical	FRED
Cleveland Fed's VIX (VIXCLSx)	Market Conditions/Environment	Numerical	FRED
Risk Premium: Market Return Minus Risk-Free Rate (Mkt_RF)	Market Conditions/Environment	Numerical	Kenneth R. French website [9]
Nasdaq returns	Market Conditions/Environment	Numerical	Eikon
Number of IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter
Average First day Return Proceeds Weighted	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Median First day Return	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Money left on the Table	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Aggregate proceeds	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Venture Capital (VC) IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Buyout IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Tech IPOs	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Venture Capital (VC) backed proceeds	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]
Tech proceeds	Market Conditions/Environment	Numerical	IPO Data website - Jay R. Ritter [10]

Table 4: Performance Prediction Accuracy - VC and BO funds

The table shows the accuracy of the models for performance prediction. Panel A presents the results for VC funds, considering different sample types (all funds, first-time funds, and sequence funds) and performance thresholds (PME and Net IRR). Panel B shows the results for BO funds. Each result represents the average of 100 cross-validation procedures. The model parameters are trained using 75% of the total available sample, selected randomly while maintaining the original ratio of ones (funds with PME greater than 1 or above the IRR threshold) and zeros (funds with PME below 1 or below the IRR threshold). The sample is forced into a 50/50 split between ones and zeros, meaning part of the data is disregarded. After training, the model is tested out-of-sample on the remaining 25%, also structured with a 50/50 split between ones and zeros.

Panel A: Venture Capital Funds		RF	Lasso	Ridge	LPM	Probit	Logit
All Funds	PME Threshold	0,647	0,704	0,690	0,679	0,668	0,668
	IRR Threshold	0,680	0,729	0,716	0,712	0,689	0,689
First-time Funds	PME Threshold	0,553	0,537	0,566	0,488	0,493	0,493
	IRR Threshold	0,618	0,658	0,657	0,526	0,522	0,522
Sequence Funds	PME Threshold	0,677	0,735	0,718	0,661	0,632	0,632
	IRR Threshold	0,677	0,748	0,746	0,708	0,679	0,679
Panel B: Buyout Funds							
All Funds	PME Threshold	0,596	0,570	0,624	0,600	0,590	0,590
	IRR Threshold	0,579	0,639	0,603	0,636	0,632	0,632
First-time Funds	PME Threshold	0,526	0,519	0,518	0,520	0,523	0,523
	IRR Threshold	0,496	0,508	0,511	0,498	0,500	0,500
Sequence Funds	PME Threshold	0,579	0,530	0,611	0,554	0,548	0,548
	IRR Threshold	0,576	0,640	0,622	0,636	0,629	0,629

Table 5: Confusion Matrices - Venture Capital Funds

The table presents the average results of each confusion matrix from the 100 cross-validation procedures explained for Venture Capital funds across different performance thresholds.Each matrix presents the true positive (TP - bottom right corner), true negative (TN - top left corner), false positive (FP - top right corner), and false negative (FN - bottom left corner) rates. Panels A, B, and C consider different sets of funds included in the analysis. Each panel presents the results based on varying fund types, such as all funds, first-time funds, or sequence funds. Details on accuracy, precision, recall, specificity, and F-score can be found in Powers (2020).

Venture Capital Funds *Panel A: All Funds*	RF		Lasso		Ridge	
PME Threshold	23,6	13,9	28,9	8,7	27,0	10,5
	12,6	24,9	13,5	24,0	12,8	24,7
Accuracy	64,7%		70,4%		69,0%	
Precision	64,2%		73,5%		70,2%	
Recall	66,4%		63,9%		66,0%	
Specificity	63,0%		76,9%		72,0%	
F-score	65,3%		68,4%		68,0%	
IRR Threshold	25,9	13,5	28,2	11,3	28,0	11,4
	11,7	27,7	10,1	29,3	11,0	28,5
Accuracy	68,0%		72,9%		71,6%	
Precision	67,3%		72,2%		71,4%	
Recall	70,2%		74,3%		72,2%	
Specificity	65,8%		71,5%		71,0%	
F-score	68,7%		73,3%		71,8%	

Panel B: First Funds	RF		Lasso		Ridge	
PME Threshold	4,1	2,4	2,9	3,7	3,7	2,9
	3,4	3,1	2,4	4,2	2,8	3,7
Accuracy	55,3%		53,8%		56,7%	
Precision	56,3%		53,2%		56,6%	
Recall	47,4%		63,9%		56,9%	
Specificity	63,1%		43,7%		56,4%	
F-score	51,5%		58,1%		56,8%	
IRR Threshold	5,2	2,9	4,7	3,3	5,1	3,0
	3,3	4,7	2,2	5,8	2,5	5,5
Accuracy	61,8%		65,8%		65,7%	
Precision	62,5%		63,9%		64,9%	
Recall	59,3%		72,4%		68,3%	
Specificity	64,4%		59,1%		63,1%	
F-score	60,8%		67,9%		66,5%	

Panel C: Sequence Funds	RF		Lasso		Ridge	
PME Threshold	21,9	11,2	25,9	7,1	24,4	8,6
	10,2	22,8	10,4	22,6	10,1	22,9
Accuracy	67,7%		73,5%		71,8%	
Precision	67,2%		76,1%		72,8%	
Recall	69,1%		68,6%		69,5%	
Specificity	66,2%		78,5%		74,1%	
F-score	68,1%		72,2%		71,1%	
IRR Threshold	22,7	13,3	25,6	10,4	25,7	10,3
	10,0	26,0	7,7	28,3	8,0	28,1
Accuracy	67,7%		74,8%		74,6%	
Precision	66,2%		73,1%		73,1%	
Recall	72,3%		78,5%		77,9%	
Specificity	63,1%		71,2%		71,3%	
F-score	69,1%		75,7%		75,4%	

Table 6: Confusion Matrices - Buyout Funds

The table presents the average results of each confusion matrix from the 100 cross-validation procedures explained for Buyout funds. Details on accuracy, precision, recall, specificity, and F-score score can be found in Powers (2020).

Buyout Funds *Panel A: All Funds*	RF		Lasso		Ridge	
PME Threshold	14,5	9,5	14,2	9,8	15,5	8,5
	9,9	14,1	10,8	13,2	9,5	14,5
Accuracy	59,6%		57,0%		62,4%	
Precision	59,7%		57,3%		62,9%	
Recall	58,6%		55,0%		60,3%	
Specificity	60,5%		59,1%		64,4%	
F-score	59,2%		56,1%		61,6%	
IRR Threshold	58,6	49,4	67,3	40,7	62,1	45,9
	41,6	66,4	37,4	70,6	39,8	68,2
Accuracy	57,9%		63,9%		60,3%	
Precision	57,3%		63,4%		59,8%	
Recall	61,5%		65,4%		63,2%	
Specificity	54,2%		62,3%		57,5%	
F-score	59,3%		64,4%		61,4%	

Panel B: First Funds	RF		Lasso		Ridge	
PME Threshold	2,2	1,8	1,4	2,6	1,8	2,2
	2,0	2,0	1,3	2,7	1,6	2,4
Accuracy	52,6%		51,9%		51,8%	
Precision	52,8%		51,4%		51,5%	
Recall	49,5%		67,8%		59,5%	
Specificity	55,8%		36,0%		44,0%	
F-score	51,1%		58,5%		55,2%	
IRR Threshold	8,1	7,4	3,3	12,2	4,2	11,3
	8,2	7,3	3,1	12,5	3,9	11,6
Accuracy	49,6%		50,8%		51,1%	
Precision	49,6%		50,5%		50,8%	
Recall	47,2%		80,3%		74,9%	
Specificity	52,0%		21,2%		27,4%	
F-score	48,3%		62,0%		60,5%	

Panel C: Sequence Funds	RF		Lasso		Ridge	
PME Threshold	8,3	5,8	6,2	7,8	8,1	5,9
	6,0	8,0	5,4	8,6	5,0	9,0
Accuracy	57,9%		53,0%		61,1%	
Precision	58,1%		52,6%		60,4%	
Recall	56,9%		61,5%		64,5%	
Specificity	58,9%		44,5%		57,8%	
F-score	57,5%		56,7%		62,4%	
IRR Threshold	39,9	34,1	44,8	29,3	43,9	30,1
	28,7	45,3	24,0	50,0	25,9	48,1
Accuracy	57,6%		64,0%		62,2%	
Precision	57,1%		63,1%		61,5%	
Recall	61,2%		67,6%		64,9%	
Specificity	53,9%		60,5%		59,4%	
F-score	59,1%		65,3%		63,2%	

Table 7: Summary of Feature Importance for Ridge, Lasso, and RF in VC Funds

The table presents the summary statistics of feature significance measures, including beta coefficients for Ridge and Lasso, and variable importance scores for RF, applied to VC funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

Betas Categories/Groups	Ridge						Lasso						RF					
	N	Avg	Med	Stdev	Min	Max	N	Avg	Med	Stdev	Min	Max	N	Avg	Med	Stdev	Min	Max
PME Threshold Performance																		
All funds sample																		
Fund information	35	0,01	- 0,00	0,17	- 0,35	0,49	35	- 0,01	-	0,14	- 0,67	0,31	6	0,00	0,00	0,00	0,00	0,00
GPs	373	- 0,02	0,09	0,36	- 0,94	0,63	373	0,01	-	0,12	- 0,83	0,38	1	0,08	0,08	-	0,08	0,08
Macro	20	- 0,39	0,00	1,26	- 5,51	0,32	20	- 0,04	-	0,14	- 0,59	0,05	20	0,01	0,00	0,00	0,00	0,01
Market Conditions - Market Environment	16	0,02	- 0,00	0,13	- 0,33	0,28	16	0,01	-	0,02	- 0,00	0,08	16	0,00	0,00	0,00	0,00	0,02
PE Industry	4	0,01	0,00	0,01	0,00	0,03	4	0,04	0,00	0,09	-	0,17	4	0,01	0,00	0,01	0,00	0,02
PE Theory & Centrality Measures	4	0,10	0,04	0,21	- 0,07	0,40	4	0,13	0,00	0,26	- 0,01	0,52	4	0,00	0,00	0,00	0,00	0,00
Total	*452*	*- 0,03*	*0,00*	*0,43*	*- 5,51*	*0,63*	*452*	*0,00*	*-*	*0,12*	*- 0,83*	*0,52*	*51*	*0,01*	*0,00*	*0,01*	*0,00*	*0,08*
First time funds sample																		
Fund information	35	- 0,01	- 0,02	0,10	- 0,21	0,13	35	0,00	-	0,03	- 0,04	0,09	6	0,02	0,02	0,01	0,01	0,03
GPs	153	0,00	0,07	0,14	- 0,31	0,20	153	0,00	-	0,02	- 0,10	0,10	1	0,36	0,36	-	0,36	0,36
Macro	20	- 0,12	- 0,00	0,44	- 1,94	0,17	20	- 0,07	-	0,24	- 1,05	0,08	20	0,03	0,03	0,01	0,02	0,07
Market Conditions - Market Environment	16	0,02	0,00	0,10	- 0,15	0,24	16	0,02	-	0,07	- 0,01	0,29	16	0,02	0,01	0,01	0,01	0,04
PE Industry	4	0,00	0,00	0,00	- 0,00	0,01	4	0,00	0,00	0,00	- 0,00	0,01	4	0,03	0,02	0,01	0,02	0,04
PE Theory & Centrality Measures	4	- 0,09	- 0,02	0,19	- 0,37	0,04	4	- 0,00	-	0,00	- 0,01	0,00	4	0,01	0,01	0,00	0,01	0,02
Total	*232*	*- 0,01*	*0,04*	*0,18*	*- 1,94*	*0,24*	*232*	*- 0,00*	*-*	*0,08*	*- 1,05*	*0,29*	*51*	*0,03*	*0,02*	*0,05*	*0,01*	*0,36*
Sequence funds sample																		
Fund information	33	0,00	- 0,01	0,18	- 0,59	0,33	33	- 0,03	-	0,21	- 1,19	0,12	6	0,01	0,01	0,00	0,00	0,01
GPs	259	- 0,00	0,12	0,37	- 1,06	0,67	259	0,01	-	0,19	- 1,41	0,55	1	0,13	0,13	-	0,13	0,13
Macro	20	- 0,58	- 0,00	1,71	- 7,23	0,36	20	- 0,14	- 0,00	0,52	- 2,28	0,01	20	0,01	0,01	0,01	0,01	0,03
Market Conditions - Market Environment	16	0,03	- 0,00	0,19	- 0,51	0,32	16	0,01	-	0,04	- 0,05	0,13	16	0,01	0,01	0,01	0,00	0,05
PE Industry	4	0,01	0,00	0,02	0,00	0,03	4	0,05	0,00	0,11	-	0,22	4	0,02	0,01	0,02	0,00	0,04
PE Theory & Centrality Measures	7	0,05	0,00	0,23	- 0,12	0,55	7	0,16	-	0,43	- 0,01	1,14	7	0,01	0,01	0,00	0,01	0,01
Total	*339*	*- 0,03*	*0,08*	*0,54*	*- 7,23*	*0,67*	*339*	*0,00*	*-*	*0,23*	*- 2,28*	*1,14*	*54*	*0,01*	*0,01*	*0,02*	*0,00*	*0,13*
Net IRR Threshold Performance																		
All funds sample																		
Fund information	35	0,02	0,03	0,19	- 0,63	0,47	35	- 0,03	0,00	0,23	- 1,32	0,13	6	0,00	0,00	0,00	0,00	0,00
GPs	373	- 0,03	- 0,14	0,42	- 0,96	1,39	373	- 0,00	-	0,26	- 0,80	1,74	1	0,06	0,06	.	0,06	0,06
Macro	20	- 0,31	- 0,00	1,19	- 5,29	0,29	20	- 0,04	-	0,11	- 0,49	0,00	20	0,01	0,00	0,00	0,00	0,02
Market Conditions - Market Environment	16	- 0,00	- 0,00	0,17	- 0,59	0,19	16	- 0,02	-	0,10	- 0,40	0,01	16	0,00	0,00	0,00	0,00	0,02
PE Industry	4	0,01	0,00	0,02	- 0,00	0,03	4	0,05	0,00	0,10	- 0,00	0,19	4	0,01	0,01	0,01	0,00	0,02
PE Theory & Centrality Measures	4	0,06	0,06	0,08	- 0,02	0,14	4	0,01	0,00	0,02	- 0,00	0,03	4	0,00	0,00	0,00	0,00	0,00
Total	*452*	*- 0,04*	*- 0,09*	*0,46*	*- 5,29*	*1,39*	*452*	*- 0,01*	*-*	*0,25*	*- 1,32*	*1,74*	*51*	*0,01*	*0,00*	*0,01*	*0,00*	*0,06*
First time funds sample																		
Fund information	35	- 0,00	- 0,01	0,10	- 0,22	0,30	35	0,00	-	0,02	- 0,05	0,10	6	0,01	0,01	0,00	0,01	0,02
GPs	153	0,00	- 0,09	0,17	- 0,23	0,41	153	0,00	-	0,02	- 0,05	0,10	1	0,30	0,30	.	0,30	0,30
Macro	20	- 0,07	- 0,00	0,44	- 1,78	0,69	20	- 0,01	-	0,09	- 0,37	0,10	20	0,02	0,02	0,01	0,01	0,06
Market Conditions - Market Environment	16	0,01	- 0,00	0,11	- 0,25	0,22	16	0,00	-	0,01	- 0,03	0,03	16	0,01	0,01	0,01	0,00	0,04
PE Industry	4	0,00	0,00	0,01	- 0,00	0,01	4	0,01	-	0,02	-	0,04	4	0,02	0,02	0,02	0,01	0,06
PE Theory & Centrality Measures	4	0,17	0,10	0,23	- 0,01	0,50	4	0,03	0,02	0,03	- 0,00	0,07	4	0,01	0,01	0,00	0,00	0,01
Total	*232*	*- 0,00*	*- 0,06*	*0,20*	*- 1,78*	*0,69*	*232*	*0,00*	*-*	*0,03*	*- 0,37*	*0,10*	*51*	*0,02*	*0,01*	*0,04*	*0,00*	*0,30*
Sequence funds sample																		
Fund information	33	0,07	0,09	0,24	- 0,83	0,46	33	- 0,06	-	0,28	- 1,55	0,10	6	0,00	0,00	0,00	0,00	0,01
GPs	259	- 0,03	- 0,14	0,36	- 0,87	1,17	259	- 0,01	-	0,16	- 0,68	0,95	1	0,12	0,12	.	0,12	0,12
Macro	20	- 0,53	- 0,00	1,70	- 7,38	0,12	20	- 0,20	-	0,69	- 2,97	0,00	20	0,01	0,01	0,01	0,01	0,04
Market Conditions - Market Environment	16	- 0,03	- 0,00	0,27	- 1,01	0,29	16	- 0,01	-	0,12	- 0,40	0,23	16	0,01	0,01	0,01	0,00	0,04
PE Industry	4	0,01	0,00	0,02	- 0,00	0,03	4	0,04	0,00	0,09	- 0,00	0,18	4	0,02	0,02	0,02	0,01	0,04
PE Theory & Centrality Measures	7	0,05	0,00	0,09	- 0,05	0,15	7	0,01	0,00	0,02	- 0,00	0,06	7	0,01	0,01	0,00	0,00	0,01
Total	*339*	*- 0,04*	*- 0,07*	*0,54*	*- 7,38*	*1,17*	*339*	*- 0,03*	*-*	*0,24*	*- 2,97*	*0,95*	*54*	*0,01*	*0,01*	*0,02*	*0,00*	*0,12*

Table 8: Summary of Feature Importance for Ridge, Lasso, and RF in BO Funds

The table presents the summary statistics of feature significance measures, including beta coefficients for Ridge and Lasso, and variable importance scores for RF, applied to BO funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

Betas Categories/Groups	Ridge						Lasso						RF					
	N	Avg	p50	Stdev	Min	Max	N	Avg	p50	Stdev	Min	Max	N	Avg	p50	Stdev	Min	Max
PME Threshold Performance																		
All funds sample																		
Fund information	35	- 0,00	0,01	0,24	- 0,53	0,51	35	- 0,02	0,00	0,22	- 0,65	0,59	5	0,01	0,01	0,01	0,01	0,02
GPs	452	0,00	0,09	0,24	- 0,85	0,34	452	- 0,00	0,01	0,06	- 0,71	0,13	1	0,19	0,19	.	0,19	0,19
Macro	20	- 0,06	0,00	0,59	- 2,39	0,88	20	- 0,05	-	0,35	- 1,53	0,20	20	0,02	0,02	0,00	0,02	0,02
Market Conditions - Market Environment	16	0,11	0,00	0,40	- 0,07	1,59	16	0,22	-	0,86	- 0,03	3,42	16	0,02	0,01	0,00	0,01	0,03
PE Industry	4	0,00	- 0,00	0,01	- 0,00	0,01	4	0,00	- 0,00	0,01	- 0,00	0,01	4	0,02	0,02	0,00	0,01	0,02
PE Theory & Centrality Measures	4	0,14	0,16	0,10	0,00	0,23	4	0,11	0,07	0,13	0,00	0,28	4	0,02	0,02	0,00	0,01	0,02
Total	*531*	*0,00*	*0,08*	*0,27*	*- 2,39*	*1,59*	*531*	*0,00*	*0,01*	*0,18*	*- 1,53*	*3,42*	*50*	*0,02*	*0,02*	*0,03*	*0,01*	*0,19*
First time funds sample																		
Fund information	28	- 0,00	0,00	0,05	- 0,10	0,08	28	0,00	-	0,03	- 0,08	0,13	5	0,04	0,04	0,02	0,01	0,06
GPs	171	0,00	0,01	0,03	- 0,15	0,03	171	0,00	-	0,01	- 0,11	0,04	1	1,00	1,00	.	1,00	1,00
Macro	20	0,07	0,00	0,24	- 0,07	1,07	20	0,13	-	0,54	- 0,03	2,41	20	0,04	0,03	0,02	0,02	0,11
Market Conditions - Market Environment	16	0,04	0,00	0,20	- 0,06	0,80	16	0,17	0,00	0,69	- 0,00	2,76	16	0,05	0,04	0,04	0,01	0,15
PE Industry	4	- 0,00	- 0,00	0,00	- 0,00	0,00	4	- 0,00	- 0,00	0,00	- 0,00	0,00	4	0,06	0,06	0,02	0,04	0,09
PE Theory & Centrality Measures	4	- 0,37	0,02	0,82	- 1,59	0,08	4	- 0,23	0,01	0,54	- 1,04	0,08	4	0,03	0,04	0,02	-	0,04
Total	*243*	*0,00*	*0,01*	*0,14*	*- 1,59*	*1,07*	*243*	*0,02*	*-*	*0,25*	*- 1,04*	*2,76*	*50*	*0,06*	*0,04*	*0,14*	*-*	*1,00*
Sequence funds sample																		
Fund information	33	- 0,02	0,00	0,16	- 0,35	0,31	33	- 0,00	-	0,07	- 0,26	0,19	5	0,02	0,02	0,01	0,01	0,03
GPs	373	0,00	0,04	0,13	- 0,45	0,16	373	- 0,00	-	0,02	- 0,13	0,04	1	0,32	0,32	.	0,32	0,32
Macro	20	- 0,12	0,00	0,62	- 2,16	0,81	20	- 0,04	-	0,18	- 0,68	0,18	20	0,03	0,02	0,01	0,02	0,04
Market Conditions - Market Environment	16	0,06	0,00	0,19	- 0,04	0,76	16	0,04	0,00	0,16	- 0,01	0,64	16	0,02	0,02	0,01	0,01	0,05
PE Industry	4	0,00	0,00	0,00	- 0,00	0,01	4	0,00	0,00	0,00	- 0,00	0,01	4	0,02	0,02	0,01	0,02	0,03
PE Theory & Centrality Measures	7	0,05	0,00	0,10	- 0,07	0,18	7	0,02	0,00	0,05	- 0,04	0,12	7	0,03	0,03	0,00	0,02	0,03
Total	*453*	*- 0,00*	*0,03*	*0,18*	*- 2,16*	*0,81*	*453*	*- 0,00*	*-*	*0,05*	*- 0,68*	*0,64*	*53*	*0,03*	*0,02*	*0,04*	*0,01*	*0,32*
Net IRR Threshold Performance																		
All funds sample																		
Fund information	35	0,01	- 0,00	0,19	- 0,38	0,41	35	- 0,03	-	0,23	- 0,96	0,47	5	0,00	0,00	0,00	0,00	0,00
GPs	452	- 0,02	- 0,01	0,27	- 0,48	0,57	452	0,01	- 0,00	0,20	- 0,53	0,96	1	0,02	0,02	.	0,02	0,02
Macro	20	- 0,10	0,00	0,63	- 2,72	0,54	20	0,07	-	0,50	- 0,97	1,92	20	0,00	0,00	0,00	0,00	0,00
Market Conditions - Market Environment	16	0,06	- 0,00	0,29	- 0,08	1,16	16	0,19	-	0,81	- 0,13	3,22	16	0,00	0,00	0,00	0,00	0,00
PE Industry	4	0,00	0,00	0,00	- 0,00	0,01	4	0,00	-	0,00	- 0,00	0,00	4	0,00	0,00	0,00	0,00	0,00
PE Theory & Centrality Measures	4	- 0,05	- 0,06	0,05	- 0,10	- 0,00	4	- 0,04	- 0,04	0,04	- 0,07	- 0,00	4	0,00	0,00	0,00	0,00	0,00
Total	*531*	*- 0,02*	*- 0,00*	*0,29*	*- 2,72*	*1,16*	*531*	*0,01*	*- 0,00*	*0,26*	*- 0,97*	*3,22*	*50*	*0,00*	*0,00*	*0,00*	*0,00*	*0,02*
First time funds sample																		
Fund information	28	- 0,00	- 0,00	0,07	- 0,13	0,12	28	0,00	-	0,05	- 0,13	0,20	5	0,02	0,02	0,01	0,01	0,03
GPs	171	- 0,00	- 0,05	0,09	- 0,13	0,17	171	0,00	-	0,01	- 0,02	0,06	1	0,69	0,69	.	0,69	0,69
Macro	20	0,02	0,00	0,14	- 0,22	0,57	20	0,02	-	0,13	- 0,15	0,56	20	0,01	0,01	0,01	0,00	0,03
Market Conditions - Market Environment	16	0,01	0,00	0,06	- 0,05	0,23	16	0,01	-	0,04	- 0,00	0,14	16	0,01	0,01	0,01	0,00	0,02
PE Industry	4	- 0,00	- 0,00	0,00	- 0,00	- 0,00	4	- 0,00	- 0,00	0,00	- 0,00	-	4	0,04	0,04	0,00	0,04	0,04
PE Theory & Centrality Measures	4	- 0,03	- 0,04	0,03	- 0,06	- 0,00	4	- 0,01	- 0,01	0,01	- 0,03	-	4	0,03	0,03	0,01	0,01	0,04
Total	*243*	*0,00*	*- 0,03*	*0,09*	*- 0,22*	*0,57*	*243*	*0,00*	*-*	*0,04*	*- 0,15*	*0,56*	*50*	*0,03*	*0,01*	*0,10*	*0,00*	*0,69*
Sequence funds sample																		
Fund information	33	0,00	-	0,20	- 0,45	0,39	33	- 0,03	-	0,25	- 1,16	0,34	5	0,00	0,00	0,00	0,00	0,00
GPs	373	- 0,01	- 0,06	0,26	- 0,56	0,61	373	- 0,00	- 0,00	0,15	- 0,52	0,80	1	0,04	0,04	.	0,04	0,04
Macro	20	- 0,15	0,00	0,74	- 3,27	0,29	20	0,01	-	0,23	- 0,66	0,73	20	0,00	0,00	0,00	0,00	0,01
Market Conditions - Market Environment	16	0,09	- 0,00	0,41	- 0,12	1,61	16	0,23	-	0,98	- 0,19	3,89	15	0,00	0,00	0,00	0,00	0,00
PE Industry	4	0,00	0,00	0,01	0,00	0,01	4	0,01	0,00	0,01	-	0,03	4	0,00	0,00	0,00	0,00	0,00
PE Theory & Centrality Measures	7	- 0,01	- 0,01	0,02	- 0,05	0,01	7	- 0,01	- 0,00	0,01	- 0,02	-	7	0,00	0,00	0,00	0,00	0,00
Total	*453*	*- 0,01*	*- 0,00*	*0,30*	*- 3,27*	*1,61*	*453*	*0,00*	*-*	*0,24*	*- 1,16*	*3,89*	*52*	*0,00*	*0,00*	*0,00*	*0,00*	*0,04*

Table 9: Summary of Absolute Betas and Variable Importance for Ridge, Lasso, and RF in VC Funds

The table provides key statistics of the absolute betas for Ridge and Lasso, along with variable importance scores for Random Forest, applied to VC funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

	Ridge						Lasso						Randon Forest (RF)					
Betas Categories/Groups	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.
PME Threshold Performance																		
All funds Sample																		
Fund information	35	3,87	0,111	0,060	0,49	0,00004	35	1,89	0,054	0,007	0,67	-	6	0,01	0,002	0,002	0,004	0,001
GPs	373	117,47	0,315	0,279	0,94	-	373	21,57	0,058	0,004	0,83	-	1	0,08	0,076	0,076	0,076	0,076
Macro	20	8,56	0,428	0,010	5,51	0,00012	20	0,92	0,046	0,000	0,59	-	20	0,12	0,006	0,004	0,015	0,003
Market Conditions - Market Environment	16	1,08	0,068	0,002	0,33	0,00000	16	0,10	0,006	0,000	0,08	-	16	0,05	0,003	0,002	0,018	0,001
PE Industry	4	0,03	0,007	0,001	0,03	0,00017	4	0,17	0,043	0,000	0,17	-	4	0,03	0,008	0,004	0,020	0,002
PE Theory & Centrality Measures	4	0,54	0,135	0,069	0,40	0,00032	4	0,53	0,132	0,006	0,52	-	4	0,01	0,003	0,003	0,004	0,002
Total	452	131,55	0,291	0,260	5,51	-	452	25,18	0,056	0,003	0,83	-	51	0,30	0,006	0,003	0,076	0,001
First Fund Sample																		
Fund information	35	2,72	0,078	0,063	0,21	0,00003	35	0,42	0,012	0,001	0,09	-	6	0,12	0,020	0,019	0,033	0,007
GPs	153	18,33	0,120	0,092	0,31	0,00705	153	1,03	0,007	-	0,10	-	1	0,36	0,359	0,359	0,359	0,359
Macro	20	2,97	0,149	0,015	1,94	0,00006	20	1,55	0,077	0,001	1,05	-	20	0,68	0,034	0,033	0,066	0,018
Market Conditions - Market Environment	16	0,76	0,048	0,001	0,24	0,00000	16	0,39	0,025	0,000	0,29	-	16	0,27	0,017	0,013	0,039	0,009
PE Industry	4	0,01	0,002	0,001	0,01	0,00002	4	0,01	0,002	0,000	0,01	-	4	0,10	0,025	0,021	0,042	0,017
PE Theory & Centrality Measures	4	0,45	0,113	0,037	0,37	0,00817	4	0,01	0,002	0,001	0,01	-	4	0,04	0,010	0,009	0,015	0,007
Total	232	25,24	0,109	0,084	1,94	0,00000	232	3,40	0,015	0,000	1,05	-	51	1,57	0,031	0,021	0,359	0,007
Sequence funds Sample																		
Fund information	33	4,20	0,127	0,090	0,59	-	33	1,9	0,057	0,006	1,19	-	6	0,03	0,006	0,005	0,012	0,001
GPs	259	79,66	0,308	0,266	1,06	-	259	20,5	0,079	0,005	1,41	-	1	0,13	0,126	0,126	0,126	0,126
Macro	20	12,47	0,624	0,008	7,23	0,00017	20	2,8	0,142	0,000	2,28	-	20	0,26	0,013	0,009	0,030	0,005
Market Conditions - Market Environment	16	1,52	0,095	0,004	0,51	0,00000	16	0,2	0,015	0,000	0,13	-	16	0,14	0,009	0,006	0,047	0,003
PE Industry	4	0,04	0,009	0,001	0,03	0,00022	4	0,22	0,054	0,000	0,22	-	4	0,08	0,019	0,014	0,044	0,005
PE Theory & Centrality Measures	7	0,94	0,135	0,084	0,55	0,00012	7	1,17	0,167	0,003	1,14	-	7	0,06	0,008	0,008	0,011	0,005
Total	339	98,83	0,292	0,238	7,23	-	339	26,84	0,079	0,004	2,28	-	54	0,69	0,013	0,008	0,126	0,001
Net IRR Threshold Performance																		
All funds Sample																		
Fund information	35	4,49	0,128	0,096	0,63	0,00002	35	2,12	0,061	0,007	1,32	-	6	0,01	0,002	0,002	0,003	0,001
GPs	373	133,46	0,358	0,328	1,39	-	373	43,94	0,118	0,001	1,74	-	1	0,06	0,055	0,055	0,055	0,055
Macro	20	7,12	0,356	0,015	5,29	0,00022	20	0,79	0,040	0,000	0,49	-	20	0,11	0,006	0,004	0,017	0,003
Market Conditions - Market Environment	16	1,19	0,074	0,002	0,59	0,00000	16	0,41	0,025	0,000	0,40	-	16	0,05	0,003	0,002	0,016	0,001
PE Industry	4	0,03	0,009	0,001	0,03	0,00023	4	0,19	0,048	0,000	0,19	-	4	0,03	0,008	0,008	0,017	0,002
PE Theory & Centrality Measures	4	0,29	0,073	0,074	0,14	0,00456	4	0,04	0,011	0,007	0,03	0,00	4	0,01	0,002	0,002	0,003	0,002
Total	452	146,6	0,324	0,284	5,29	-	452	47,5	0,105	0,001	1,74	-	51	0,27	0,005	0,003	0,055	0,001
First Fund Sample																		
Fund information	35	2,65	0,076	0,057	0,30	0,00009	35	0,37	0,011	0,001	0,10	-	6	0,09	0,014	0,014	0,018	0,009
GPs	153	24,08	0,157	0,135	0,41	0,06308	153	1,50	0,010	-	0,10	-	1	0,30	0,302	0,302	0,302	0,302
Macro	20	3,36	0,168	0,011	1,78	0,00012	20	0,64	0,032	0,000	0,37	-	20	0,49	0,024	0,021	0,055	0,011
Market Conditions - Market Environment	16	0,91	0,057	0,002	0,25	0,00000	16	0,08	0,005	0,000	0,03	-	16	0,19	0,012	0,008	0,042	0,005
PE Industry	4	0,01	0,003	0,001	0,01	0,00006	4	0,04	0,010	-	0,04	-	4	0,10	0,024	0,016	0,056	0,009
PE Theory & Centrality Measures	4	0,71	0,177	0,100	0,50	0,00648	4	0,11	0,028	0,021	0,07	-	4	0,02	0,006	0,005	0,008	0,005
Total	232	31,7	0,137	0,117	1,78	0,00000	232	2,7	0,012	-	0,37	-	51	1,2	0,023	0,014	0,302	0,005
Sequence funds Sample																		
Fund information	33	5,84	0,177	0,119	0,83	-	33	2,58	0,078	0,008	1,55	-	6	0,03	0,005	0,005	0,009	0,002
GPs	259	79,32	0,306	0,290	1,17	-	259	17,58	0,068	0,000	0,95	-	1	0,12	0,120	0,120	0,120	0,120
Macro	20	11,17	0,559	0,015	7,38	0,00027	20	4,03	0,201	0,000	2,97	-	20	0,27	0,013	0,010	0,040	0,006
Market Conditions - Market Environment	16	1,68	0,105	0,002	1,01	0,00000	16	0,63	0,040	-	0,40	-	16	0,13	0,008	0,006	0,042	0,004
PE Industry	4	0,04	0,009	0,002	0,03	0,00023	4	0,18	0,044	0,000	0,18	-	4	0,08	0,021	0,019	0,041	0,007
PE Theory & Centrality Measures	7	0,51	0,074	0,053	0,15	0,00011	7	0,10	0,015	0,005	0,06	0,00	7	0,04	0,006	0,006	0,007	0,003
Total	339	98,6	0,291	0,226	7,38	-	339	25,1	0,074	0,000	2,97	-	54	0,7	0,012	0,007	0,120	0,002

Table 10: Summary of Absolute Betas and Variable Importance for Ridge, Lasso, and RF in BO Funds

The table provides key statistics of the absolute betas for Ridge and Lasso, along with variable importance scores for Random Forest, applied to BO funds. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds. The absolute beta statistics are further categorized by beta types.

Betas Categories/Groups	Ridge						Lasso						Randon Forest (RF)					
	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.	N	Sum Abs. Values	Avg.	Median	Max.	Min.
PME Threshold Performance																		
All funds Sample																		
Fund information	35	5,87	0,17	0,10	0,53	-	35	3,65	0,10	0,01	0,65	-	5	0,07	0,01	0,01	0,02	0,007
GPs	452	81,08	0,18	0,13	0,85	-	452	14,22	0,03	0,02	0,71	-	1	0,19	0,19	0,19	0,19	0,193
Macro	20	4,39	0,22	0,03	2,39	0,00000	20	2,13	0,11	0,00	1,53	-	20	0,39	0,02	0,02	0,02	0,016
Market Conditions - Market Environment	16	1,94	0,12	0,00	1,59	0,00000	16	3,52	0,22	0,00	3,42	-	16	0,25	0,02	0,01	0,03	0,010
PE Industry	4	0,01	0,00	0,00	0,01	0,00002	4	0,01	0,00	0,00	0,01	0,00	4	0,07	0,02	0,02	0,02	0,014
PE Theory & Centrality Measures	4	0,56	0,14	0,16	0,23	0,00278	4	0,42	0,11	0,07	0,28	0,00	4	0,07	0,02	0,02	0,02	0,013
Total	531	93,86	0,18	0,12	2,39	-	531	23,95	0,05	0,01	3,42	-	50	1,04	0,02	0,02	0,19	0,007
First Fund Sample																		
Fund information	28	1,06	0,04	0,04	0,10	-	28	0,42	0,01	0,00	0,13	-	5	0,21	0,04	0,04	0,06	0,013
GPs	171	3,26	0,02	0,01	0,15	-	171	0,82	0,00	-	0,11	-	1	1,00	1,00	1,00	1,00	1,000
Macro	20	1,51	0,08	0,00	1,07	0,00001	20	2,65	0,13	0,00	2,41	-	20	0,71	0,04	0,03	0,11	0,016
Market Conditions - Market Environment	16	0,96	0,06	0,00	0,80	0,00000	16	2,78	0,17	0,00	2,76	-	16	0,81	0,05	0,04	0,15	0,010
PE Industry	4	0,00	0,00	0,00	0,00	0,00011	4	0,00	0,00	0,00	0,00	-	4	0,25	0,06	0,06	0,09	0,037
PE Theory & Centrality Measures	4	1,71	0,43	0,06	1,59	0,00293	4	1,15	0,29	0,05	1,04	0,00	4	0,11	0,03	0,04	0,04	-
Total	243	8,49	0,03	0,01	1,59	-	243	7,82	0,03	-	2,76	-	50	3,08	0,06	0,04	1,00	-
Sequence funds Sample																		
Fund information	33	3,87	0,12	0,08	0,35	-	33	1,13	0,03	0,00	0,26	-	5	0,10	0,02	0,02	0,03	0,013
GPs	373	30,27	0,08	0,05	0,45	-	373	2,72	0,01	0,00	0,13	-	1	0,32	0,32	0,32	0,32	0,318
Macro	20	4,88	0,24	0,02	2,16	0,00001	20	1,34	0,07	0,00	0,68	-	20	0,50	0,03	0,02	0,04	0,017
Market Conditions - Market Environment	16	1,10	0,07	0,00	0,76	0,00000	16	0,72	0,05	0,00	0,64	-	16	0,30	0,02	0,02	0,05	0,011
PE Industry	4	0,01	0,00	0,00	0,01	0,00009	4	0,01	0,00	0,00	0,01	0,00	4	0,09	0,02	0,02	0,03	0,015
PE Theory & Centrality Measures	7	0,58	0,08	0,07	0,18	0,00003	7	0,25	0,04	0,02	0,12	0,00	7	0,18	0,03	0,03	0,03	0,020
Total	453	40,72	0,09	0,05	2,16	-	453	6,17	0,01	0,00	0,68	-	53	1,48	0,03	0,02	0,32	0,011
Net IRR Threshold Performance																		
All funds Sample																		
Fund information	35	4,88	0,14	0,12	0,41	-	35	3,36	0,10	0,01	0,96	-	5	0,01	0,00	0,00	0,00	0,001
GPs	452	107,90	0,24	0,23	0,57	-	452	52,36	0,12	0,05	0,96	-	1	0,02	0,02	0,02	0,02	0,021
Macro	20	4,04	0,20	0,03	2,72	0,00003	20	3,33	0,17	0,00	1,92	-	20	0,05	0,00	0,00	0,00	0,002
Market Conditions - Market Environment	16	1,39	0,09	0,00	1,16	0,00000	16	3,41	0,21	0,00	3,22	-	16	0,03	0,00	0,00	0,00	0,001
PE Industry	4	0,01	0,00	0,00	0,01	0,00002	4	0,01	0,00	0,00	0,00	-	4	0,01	0,00	0,00	0,00	0,002
PE Theory & Centrality Measures	4	0,21	0,05	0,06	0,10	0,00109	4	0,15	0,04	0,04	0,07	0,00	4	0,01	0,00	0,00	0,00	0,002
Total	531	118,43	0,22	0,21	2,72	-	531	62,62	0,12	0,04	3,22	-	50	0,12	0,00	0,00	0,02	0,001
First Fund Sample																		
Fund information	28	1,37	0,05	0,03	0,13	-	28	0,54	0,02	0,00	0,20	-	5	0,08	0,02	0,02	0,03	0,006
GPs	171	14,24	0,08	0,08	0,17	-	171	0,93	0,01	0,00	0,06	-	1	0,69	0,69	0,69	0,69	0,693
Macro	20	1,01	0,05	0,00	0,57	0,00000	20	0,73	0,04	-	0,56	-	20	0,27	0,01	0,01	0,03	0,005
Market Conditions - Market Environment	16	0,35	0,02	0,00	0,23	0,00000	16	0,14	0,01	-	0,14	-	16	0,17	0,01	0,01	0,02	0,001
PE Industry	4	0,00	0,00	0,00	0,00	0,00012	4	0,00	0,00	0,00	0,00	-	4	0,17	0,04	0,04	0,04	0,041
PE Theory & Centrality Measures	4	0,13	0,03	0,04	0,06	0,00049	4	0,05	0,01	0,01	0,03	-	4	0,10	0,03	0,03	0,04	0,011
Total	243	17,09	0,07	0,07	0,57	-	243	2,40	0,01	0,00	0,56	-	50	1,49	0,03	0,01	0,69	0,001
Sequence funds Sample																		
Fund information	33	4,77	0,14	0,11	0,45	-	33	3,46	0,10	0,01	1,16	-	5	0,01	0,00	0,00	0,00	0,001
GPs	373	85,73	0,23	0,23	0,61	-	373	28,30	0,08	0,02	0,80	-	1	0,04	0,04	0,04	0,04	0,036
Macro	20	4,32	0,22	0,03	3,27	0,00004	20	1,88	0,09	0,00	0,73	-	20	0,09	0,00	0,00	0,01	0,003
Market Conditions - Market Environment	16	1,93	0,12	0,00	1,61	0,00000	16	4,13	0,26	0,00	3,89	-	15	0,04	0,00	0,00	0,00	0,002
PE Industry	4	0,01	0,00	0,00	0,01	0,00000	4	0,03	0,01	0,00	0,03	-	4	0,01	0,00	0,00	0,00	0,003
PE Theory & Centrality Measures	7	0,12	0,02	0,01	0,05	0,00000	7	0,04	0,01	0,00	0,02	-	7	0,03	0,00	0,00	0,00	0,003
Total	453	96,88	0,21	0,20	3,27	-	453	37,84	0,08	0,02	3,89	-	52	0,22	0,00	0,00	0,04	0,001

Table 11: Category Importance for Ridge, Lasso, and RF in VC Funds

The table presents the proportion of absolute betas for Ridge and Lasso, along with the proportion of variable importance scores for Random Forest, applied to VC funds across each category/group. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds.

PME Threshold				**IRR Treshold**			
All Funds	**Ridge**	**Lasso**	**RF**	**All Funds**	**Ridge**	**Lasso**	**RF**
GPs	89,3%	85,6%	25,1%	GPs	91,0%	92,5%	20,3%
Fund information	2,9%	7,5%	4,5%	Fund information	3,1%	4,5%	4,9%
PE Theory & Centrality Measures	0,4%	2,1%	4,4%	PE Theory & Centrality Measures	0,2%	0,1%	3,4%
PE Industry	0,0%	0,7%	10,2%	PE Industry	0,0%	0,4%	12,5%
Macro	6,5%	3,6%	38,2%	Macro	4,9%	1,7%	40,8%
Market Conditions - Market Environment	0,8%	0,4%	17,7%	Market Conditions - Market Environment	0,8%	0,9%	18,1%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
1st Funds	**Ridge**	**Lasso**	**RF**	**1st Funds**	**Ridge**	**Lasso**	**RF**
GPs	72,6%	30,1%	22,8%	GPs	75,9%	54,6%	25,4%
Fund information	10,8%	12,3%	7,6%	Fund information	8,4%	13,5%	7,2%
PE Theory & Centrality Measures	1,8%	0,2%	2,6%	PE Theory & Centrality Measures	2,2%	4,1%	2,0%
PE Industry	0,0%	0,2%	6,4%	PE Industry	0,0%	1,5%	8,2%
Macro	11,8%	45,5%	43,1%	Macro	10,6%	23,5%	41,0%
Market Conditions - Market Environment	3,0%	11,6%	17,5%	Market Conditions - Market Environment	2,9%	3,0%	16,2%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
Sequence	**Ridge**	**Lasso**	**RF**	**Sequence**	**Ridge**	**Lasso**	**RF**
GPs	80,6%	76,3%	18,2%	GPs	80,5%	70,0%	17,9%
Fund information	4,3%	7,0%	4,9%	Fund information	5,9%	10,3%	4,4%
PE Theory & Centrality Measures	1,0%	4,3%	8,1%	PE Theory & Centrality Measures	0,5%	0,4%	5,9%
PE Industry	0,0%	0,8%	10,9%	PE Industry	0,0%	0,7%	12,6%
Macro	12,6%	10,6%	37,1%	Macro	11,3%	16,0%	39,9%
Market Conditions - Market Environment	1,5%	0,9%	20,8%	Market Conditions - Market Environment	1,7%	2,5%	19,3%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*

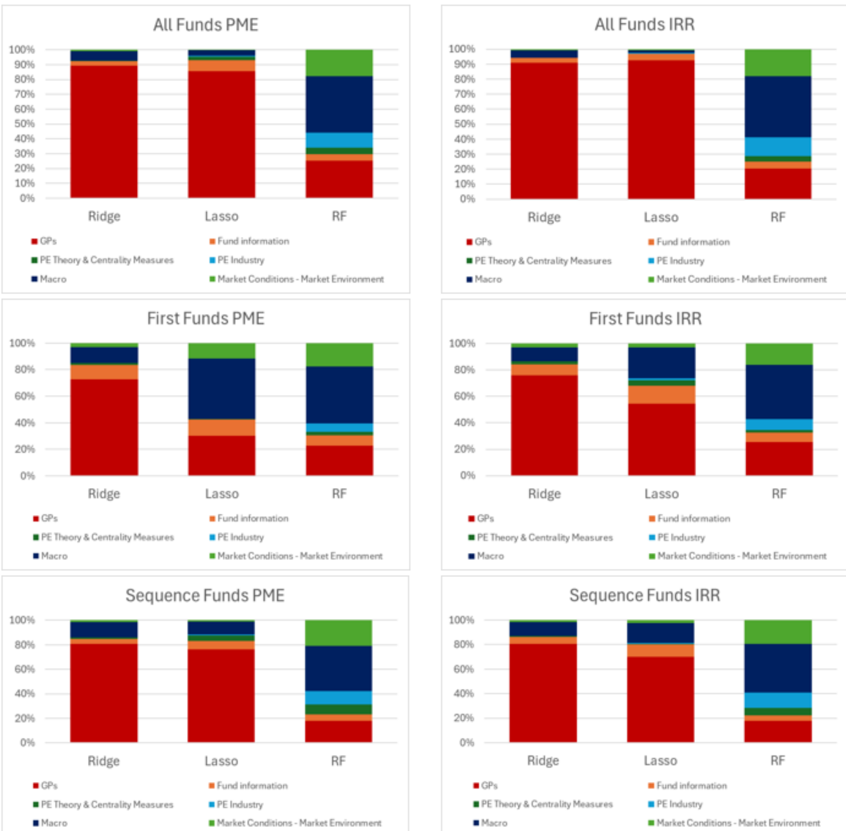

Figure 1: Proportion of Category Importance for Ridge, Lasso, and RF in VC Funds

Table 12: Category Importance for Ridge, Lasso, and RF in BO Funds

The table presents the proportion of absolute betas for Ridge and Lasso, along with the proportion of variable importance scores for Random Forest, applied to BO funds across each category/group. These statistics are calculated across two performance thresholds: PME and Net IRR. The analysis considers the entire sample of funds, as well as subsets including only first-time funds and sequential funds.

PME Threshold				IRR Treshold			
All Funds	**Ridge**	**Lasso**	**RF**	**All funds**	**Ridge**	**Lasso**	**RF**
GPs	86,4%	59,4%	18,5%	GPs	91,1%	83,6%	17,2%
Fund information	6,3%	15,2%	6,3%	Fund information	4,1%	5,4%	6,8%
PE Theory & Centrality Measures	0,6%	1,8%	7,0%	PE Theory & Centrality Measures	0,2%	0,2%	7,4%
PE Industry	0,0%	0,1%	6,7%	PE Industry	0,0%	0,0%	7,1%
Macro	4,7%	8,9%	37,2%	Macro	3,4%	5,3%	40,0%
Market Conditions - Market Environment	2,1%	14,7%	24,3%	Market Conditions - Market Environment	1,2%	5,4%	21,6%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
1st Funds	**Ridge**	**Lasso**	**RF**	**1st Funds**	**Ridge**	**Lasso**	**RF**
GPs	38,3%	10,5%	32,4%	GPs	83,3%	38,9%	46,6%
Fund information	12,5%	5,4%	6,8%	Fund information	8,0%	22,4%	5,5%
PE Theory & Centrality Measures	20,1%	14,7%	3,7%	PE Theory & Centrality Measures	0,8%	2,1%	6,8%
PE Industry	0,0%	0,0%	8,1%	PE Industry	0,0%	0,0%	11,4%
Macro	17,7%	33,9%	23,0%	Macro	5,9%	30,6%	18,5%
Market Conditions - Market Environment	11,3%	35,5%	26,1%	Market Conditions - Market Environment	2,0%	6,0%	11,2%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*
Sequence	**Ridge**	**Lasso**	**RF**	**Sequence**	**Ridge**	**Lasso**	**RF**
GPs	74,3%	44,2%	21,4%	GPs	88,5%	74,8%	16,1%
Fund information	9,5%	18,3%	6,5%	Fund information	4,9%	9,1%	6,0%
PE Theory & Centrality Measures	1,4%	4,0%	12,1%	PE Theory & Centrality Measures	0,1%	0,1%	11,8%
PE Industry	0,0%	0,1%	6,1%	PE Industry	0,0%	0,1%	6,5%
Macro	12,0%	21,7%	33,7%	Macro	4,5%	5,0%	39,9%
Market Conditions - Market Environment	2,7%	11,7%	20,3%	Market Conditions - Market Environment	2,0%	10,9%	19,8%
Total	*100%*	*100%*	*100%*	*Total*	*100%*	*100%*	*100%*

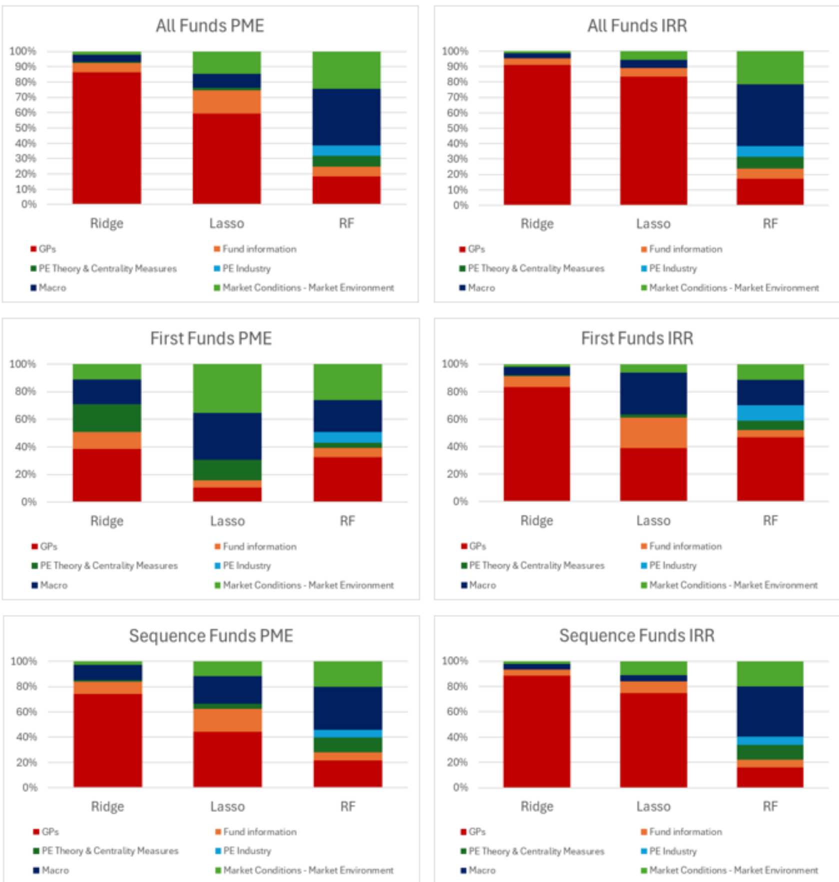

Figure 2: Proportion of Category Importance for Ridge, Lasso, and RF in BO Funds

Table 13: Top 10 variables - Venture Capital funds

The table presents the top 10 variables with the highest absolute betas, considering all ML models (Lasso, Ridge, and RF), performance thresholds (PME and Net IRR), and samples of VC funds (full sample, first-time, and sequential funds). For each variable, we identified its group classification as outlined in Table 3, the beta direction (positive or negative), and the absolute beta value. Note that for RF models, beta direction is not applicable, so we have marked it as "NA" in the respective column. The variables are ranked from the largest (#1) to the 10th largest absolute beta value. In the variables column, you can match the names with Table 3 for more details. Variables associated with a number represent the ID of a specific GP in our dataset. Due to confidentiality, the names of the GPs are not disclosed.

PME Threshold Performance											
Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Full sample				*Full sample*				*Full sample*			
Macro	CPIAUCSL	-	5,511	GPs	102485	-	0,827	GPs	Firm(GP) ID	NA	0,076
Macro	RPI	-	1,336	Fund information	Energy & Utilities	-	0,667	PE Industry	Industry years	NA	0,020
GPs	2290	-	0,939	Macro	CPIAUCSL	-	0,590	Market Cond. & Environ.	S&P Dividend Yield	NA	0,018
Macro	US GDP (last year)	-	0,921	GPs	53913	-	0,583	Macro	FEDFUNDS	NA	0,015
GPs	102485	-	0,817	GPs	72331	-	0,544	Macro	HOUST	NA	0,014
GPs	53913	-	0,816	PE Theory & Cent. Meas.	Eigenv ex-ante GP	+	0,516	Macro	PERMIT	NA	0,012
GPs	49310	-	0,809	GPs	178323	-	0,469	Macro	OILPRICEx	NA	0,010
GPs	72331	-	0,800	GPs	2290	-	0,423	Macro	INDPRO	NA	0,007
GPs	4561	-	0,800	GPs	6712	+	0,383	Macro	UNRATE	NA	0,007
GPs	62315	-	0,788	GPs	247	+	0,372	PE Industry	Fund Raising Number	NA	0,006
First time fund sample				*First time fund sample*				*First time fund sample*			
Macro	CPIAUCSL	-	1,940	Macro	CPIAUCSL	-	1,049	GPs	Firm(GP) ID	NA	0,359
PE Theory & Cent. Meas.	Eigenv ex-ante GP	-	0,371	Market Cond. & Environ.	Mkt_RF	+	0,286	Macro	CLAIMSx	NA	0,066
Macro	RPI	-	0,361	Macro	CLAIMSx	-	0,258	Macro	FEDFUNDS	NA	0,056
GPs	62315	-	0,309	GPs	3798	+	0,100	Macro	CPIAUCSL	NA	0,046
GPs	18446	-	0,304	GPs	62315	-	0,099	PE Industry	Industry years	NA	0,042
GPs	102485	-	0,287	Macro	HWIURATIO	-	0,099	Market Cond. & Environ.	S&P Dividend Yield	NA	0,039
GPs	838	-	0,271	Market Cond. & Environ.	S&P 500 returns	+	0,098	Macro	Uncertain	NA	0,039
GPs	768	-	0,270	Fund information	Asia	+	0,094	Market Cond. & Environ.	Mkt_RF	NA	0,036
GPs	2616	-	0,265	Macro	UNRATE	+	0,080	Macro	CUMFNS	NA	0,036
GPs	7809	-	0,250	Fund information	LOCATION_Asia	+	0,073	Macro	UNRATE	NA	0,036
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	7,234	Macro	CPIAUCSL	-	2,282	GPs	Firm(GP) ID	NA	0,126
Macro	RPI	-	2,263	GPs	178323	-	1,406	Market Cond. & Environ.	S&P Dividend Yield	NA	0,047
Macro	US GDP (last year)	-	2,135	Fund information	Energy & Utilities	-	1,189	PE Industry	Industry years	NA	0,044
GPs	178323	-	1,059	PE Theory & Cent. Meas.	Eigenv ex-ante GP	+	1,144	Macro	HOUST	NA	0,030
GPs	9398	-	0,971	GPs	14527	-	1,076	Macro	PERMIT	NA	0,028
GPs	14527	-	0,953	GPs	238	-	0,695	Macro	FEDFUNDS	NA	0,028
GPs	9443	-	0,881	GPs	5462	+	0,549	Macro	OILPRICEx	NA	0,025
GPs	1697	-	0,837	GPs	6479	-	0,530	Macro	INDPRO	NA	0,020
GPs	7473	-	0,799	GPs	9398	-	0,518	PE Industry	Fund Raising Number	NA	0,019
GPs	1986	-	0,768	GPs	553	-	0,507	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,016

Net IRR Threshold Performance											
Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Full sample				*Full sample*				*Full sample*			
Macro	CPIAUCSL	-	5,289	GPs	2280	+	1,737	GPs	Firm(GP) ID	NA	0,055
GPs	2280	+	1,391	Fund information	Energy & Utilities	-	1,317	Macro	HOUST	NA	0,017
GPs	1390	+	1,243	GPs	1390	+	1,251	PE Industry	Industry years	NA	0,017
GPs	8595	+	1,173	GPs	140	+	1,170	Macro	PERMIT	NA	0,016
GPs	48370	+	0,990	GPs	48370	+	1,085	Market Cond. & Environ.	S&P Dividend Yield	NA	0,016
GPs	7379	+	0,971	GPs	381	+	1,055	Macro	FEDFUNDS	NA	0,012
GPs	9403	-	0,957	GPs	8846	+	1,034	PE Industry	Fund Raising Number	NA	0,011
GPs	381	+	0,950	GPs	2214	+	0,977	Macro	CUMFNS	NA	0,006
GPs	11944	+	0,925	GPs	11944	+	0,957	Macro	Uncertain	NA	0,005
GPs	2214	+	0,917	GPs	7379	+	0,918	Macro	OILPRICEx	NA	0,005
First time fund sample				*First time fund sample*				*First time fund sample*			
Macro	CPIAUCSL	-	1,779	Macro	CPIAUCSL	-	0,369	GPs	Firm(GP) ID	NA	0,302
Macro	RPI	+	0,689	Macro	RPI	+	0,102	PE Industry	Industry years	NA	0,056
PE Theory & Cent. Meas.	Eigenv ex-ante GP	+	0,501	Fund information	Early Stage	+	0,099	Macro	PERMIT	NA	0,055
GPs	1390	+	0,408	GPs	9225	+	0,099	Macro	HOUST	NA	0,054
GPs	140	+	0,407	GPs	1390	+	0,091	Market Cond. & Environ.	S&P Dividend Yield	NA	0,042
GPs	3795	+	0,353	GPs	7379	+	0,089	Macro	FEDFUNDS	NA	0,042
GPs	7379	+	0,351	GPs	65	+	0,087	Macro	CUMFNS	NA	0,035
GPs	11944	+	0,342	GPs	140	+	0,073	Macro	CPIAUCSL	NA	0,030
GPs	65	+	0,338	GPs	11944	+	0,072	Macro	Uncertain	NA	0,029
GPs	8710	+	0,328	GPs	15991	+	0,072	Market Cond. & Environ.	S&P 500 returns	NA	0,025
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	7,376	Macro	CPIAUCSL	-	2,967	GPs	Firm(GP) ID	NA	0,120
Macro	RPI	-	2,287	Fund information	Energy & Utilities	-	1,551	Market Cond. & Environ.	S&P Dividend Yield	NA	0,042
GPs	8595	+	1,168	Macro	RPI	-	1,037	PE Industry	Industry years	NA	0,041
Market Cond. & Environ.	Med. 1st day Return	-	1,012	GPs	8846	+	0,946	Macro	PERMIT	NA	0,040
GPs	247	+	0,909	GPs	8595	+	0,822	Macro	HOUST	NA	0,040
Macro	US GDP (last year)	-	0,907	GPs	140	+	0,727	Macro	FEDFUNDS	NA	0,029
GPs	664	+	0,896	GPs	247	+	0,700	PE Industry	Fund Raising Number	NA	0,027
GPs	2017	-	0,865	GPs	2017	-	0,677	Macro	Uncertain	NA	0,014
Fund information	Energy & Utilities	-	0,831	GPs	1449	+	0,533	Macro	OILPRICEx	NA	0,012
GPs	140	+	0,741	GPs	8792	+	0,524	Macro	CPIAUCSL	NA	0,012

Table 14: Top 10 variables - Buyout funds

The table presents the top 10 variables with the highest absolute betas, considering all ML models (Lasso, Ridge, and RF), performance thresholds (PME and Net IRR), and samples of BO funds (full sample, first-time, and sequential funds). For each variable, we identified its group classification as outlined in Table 3, the beta direction (positive or negative), and the absolute beta value. Note that for RF models, beta direction is not applicable, so we have marked it as "NA" in the respective column. The variables are ranked from the largest (#1) to the 10th largest absolute beta value. In the variables column, you can match the names with Table 3 for more details. Variables associated with a number represent the ID of a specific GP in our dataset. Due to confidentiality, the names of the GPs are not disclosed.

PME Threshold Performance

Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Full sample				*Full sample*				*Full sample*			
Macro	CPIAUCSL	-	2,386	Market Cond. & Environ.	Median First day Return	+	3,423	GPs	Firm(GP) ID	NA	0,193
Market Cond. & Environ.	Median First day Return	+	1,595	Macro	CPIAUCSL	-	1,529	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,029
Macro	US GDP (last year)	+	0,883	GPs	1390	-	0,708	Macro	T5YFFM	NA	0,025
GPs	1390	-	0,852	Fund information	Africa	-	0,655	Macro	AAAFFM	NA	0,022
GPs	43451	-	0,728	Fund information	Americas	-	0,639	Macro	HWIURATIO	NA	0,022
GPs	581	-	0,720	Fund information	Australasia	+	0,588	Macro	T10YFFM	NA	0,022
GPs	12067	-	0,693	Fund information	Information Technology	+	0,524	Macro	OILPRICEx	NA	0,021
GPs	429	-	0,658	GPs	8837	-	0,341	Fund information	Fund Size (U$MM)	NA	0,021
GPs	782	-	0,651	GPs	11936	-	0,314	PE Theory & Cent. Meas.	Quant_all_lps	NA	0,021
GPs	6629	-	0,647	Fund information	Latin America & Caribbean	-	0,290	Macro	CUMFNS	NA	0,021
First time fund sample				*First time fund sample*				*First time fund sample*			
PE Theory & Cent. Meas.	Eigenv_ex_ante_GP	-	1,592	Market Cond. & Environ.	Median First day Return	+	2,765	GPs	Firm(GP) ID	NA	1,000
Macro	RPI	+	1,072	Macro	RPI	+	2,410	Market Cond. & Environ.	Median First day Return	NA	0,153
Market Cond. & Environ.	Median First day Return	+	0,799	Fund information	LOCATION_Europe	+	0,130	Market Cond. & Environ.	Tech_proceeds	NA	0,112
GPs	1390	-	0,149	Macro	HWIURATIO	+	0,089	Macro	RPI	NA	0,107
GPs	6839	-	0,143	PE Theory & Cent. Meas.	Proportion 1Q LPs ex ante	+	0,084	Market Cond. & Environ.	VC backed proceeds	NA	0,100
GPs	895	-	0,136	Fund information	Europe	+	0,071	Market Cond. & Environ.	VC IPOs	NA	0,089
GPs	25072	-	0,127	Macro	HWI	+	0,059	PE Industry	Fund Raising Volume	NA	0,088
GPs	9389	-	0,123	Macro	CLAIMSx	+	0,040	PE Industry	Dry Powder	NA	0,069
Macro	US GDP (last year)	+	0,121	GPs	46046	+	0,038	Fund information	GP Headquarter Location	NA	0,064
GPs	8734	-	0,110	Fund information	Australasia	+	0,027	Market Cond. & Environ.	Aggregate proceeds	NA	0,057
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	2,159	Macro	CPIAUCSL	-	0,675	GPs	Firm(GP) ID	NA	0,318
Macro	RPI	-	1,491	Market Cond. & Environ.	Median First day Return	+	0,639	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,048
Macro	US GDP (last year)	+	0,812	Macro	RPI	-	0,341	Macro	T5YFFM	NA	0,041
Market Cond. & Environ.	Median First day Return	+	0,756	Fund information	Americas	-	0,262	Macro	T10YFFM	NA	0,038
GPs	7550	-	0,449	Fund information	Australasia	+	0,192	Macro	FEDFUNDS	NA	0,033
GPs	581	-	0,445	Macro	US GDP (last year)	+	0,183	PE Industry	Industry years	NA	0,033
GPs	8935	-	0,445	GPs	11936	-	0,132	Macro	AAAFFM	NA	0,032
GPs	11936	-	0,439	Fund information	Information Technology	+	0,124	Fund information	Fund Size (U$MM)	NA	0,031
GPs	429	-	0,432	PE Theory & Cent. Meas.	Proportion 1Q LPs ex ante	+	0,119	Macro	OILPRICEx	NA	0,031
GPs	386	-	0,409	GPs	8837	-	0,113	Macro	BAAFFM	NA	0,031

Net IRR Threshold Performance

Ridge				Lasso				Random Forest (RF)			
Group	Variable	Direction	Beta	Group	Variable	Direction	Beta	Group	Variable	Direction	Beta
Macro	CPIAUCSL	-	2,724	Market Cond. & Environ.	Median First day Return	+	3,221	GPs	Firm(GP) ID	NA	0,021
Market Cond. & Environ.	Median First day Return	+	1,164	Macro	HWIURATIO	+	1,919	Macro	HWIURATIO	NA	0,003
GPs	6674	+	0,569	Macro	CPIAUCSL	-	0,971	Macro	CUMFNS	NA	0,003
GPs	129	+	0,568	Fund information	Americas	-	0,963	Macro	OILPRICEx	NA	0,003
GPs	8723	+	0,566	GPs	438	+	0,956	Fund information	Fund Size (U$MM)	NA	0,003
GPs	939	+	0,562	GPs	4936	+	0,829	Macro	AAAFFM	NA	0,003
GPs	438	+	0,559	GPs	2317	+	0,820	Macro	BAAFFM	NA	0,003
Macro	RPI	+	0,544	GPs	14457	+	0,790	Macro	HOUST	NA	0,003
GPs	21861	+	0,535	GPs	6674	+	0,714	PE Theory & Cent. Meas.	Eigen fund exante 5yr	NA	0,003
GPs	2317	+	0,529	GPs	21861	+	0,702	Macro	PERMIT	NA	0,003
First time fund sample				*First time fund sample*				*First time fund sample*			
Macro	RPI	+	0,567	Macro	RPI	+	0,559	GPs	Firm(GP) ID	NA	0,693
Market Cond. & Environ.	Median First day Return	+	0,234	Fund information	Healthcare	+	0,196	PE Industry	Dry Powder	NA	0,043
Macro	CPIAUCSL	-	0,219	Macro	CPIAUCSL	-	0,153	PE Industry	Industry years	NA	0,043
GPs	12952	+	0,175	Market Cond. & Environ.	Median First day Return	+	0,141	PE Industry	Fund Raising Number	NA	0,042
GPs	6955	+	0,168	Fund information	Australasia	-	0,127	PE Industry	Fund Raising Volume	NA	0,041
GPs	14457	+	0,164	GPs	10177	+	0,061	PE Theory & Cent. Meas.	Eigenv_ex_ante_GP	NA	0,038
GPs	7995	+	0,156	GPs	46046	+	0,058	Macro	OILPRICEx	NA	0,033
GPs	14055	+	0,151	Fund information	IT, Telecoms & Media	-	0,055	PE Theory & Cent. Meas.	Eigen fund exante 5yr	NA	0,032
GPs	21861	+	0,148	GPs	12952	+	0,052	Fund information	Fund Number Overall in GP	NA	0,025
GPs	14052	+	0,147	Fund information	Industrials	+	0,041	Macro	AAAFFM	NA	0,025
Sequence funds sample				*Sequence funds sample*				*Sequence funds sample*			
Macro	CPIAUCSL	-	3,271	Market Cond. & Environ.	Median First day Return	+	3,891	GPs	Firm(GP) ID	NA	0,036
Market Cond. & Environ.	Median First day Return	+	1,611	Fund information	Americas	-	1,161	Macro	CUMFNS	NA	0,006
GPs	438	+	0,605	GPs	438	+	0,797	Macro	HWIURATIO	NA	0,006
GPs	548	+	0,605	Macro	HWIURATIO	+	0,732	Macro	Uncertain	NA	0,005
GPs	2317	+	0,601	GPs	6674	+	0,677	Macro	HOUST	NA	0,005
GPs	129	+	0,599	Macro	CPIAUCSL	-	0,664	Macro	AAAFFM	NA	0,005
GPs	7168	+	0,570	GPs	548	+	0,634	Fund information	Fund Size (U$MM)	NA	0,005
GPs	455	-	0,564	Fund information	LOCATION_Asia	-	0,626	Market Cond. & Environ.	S&P Price-Earnings Ratio	NA	0,005
GPs	718	+	0,561	GPs	4936	+	0,624	Macro	T10YFFM	NA	0,005
GPs	6674	+	0,545	GPs	2317	+	0,593	Macro	BAAFFM	NA	0,005

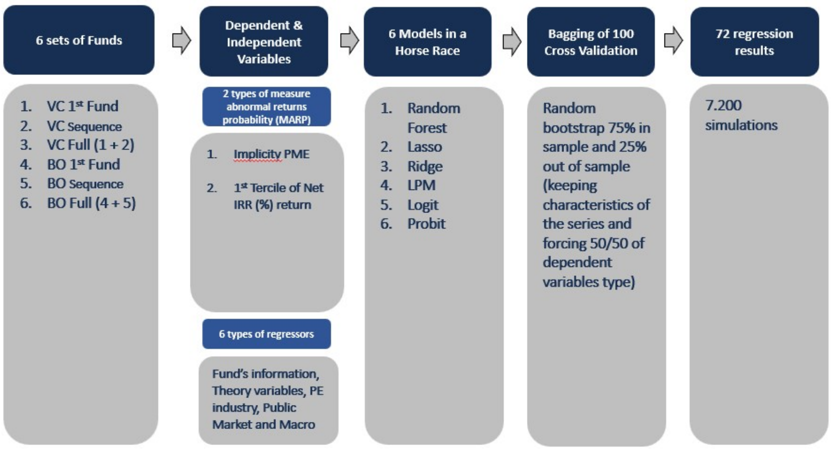

Figure 3: Methodology scheme.