Computational Economics
https://doi.org/10.1007/s10614-024-10577-6

Determining Drivers of Private Equity Return with Computational Approaches

Prosper Lamothe-Fernández[1] · Eduardo García-Argüelles[1] · Sergio Manuel Fernández-Miguélez[2] · Omar Hassani-Zerrouk[3]

Accepted: 21 February 2024

Abstract
Private equity (PE) represents the acquisition of stakes in non-listed companies, often long-term, with the objective of improving the performance and value of the company to obtain significant benefits at time of disinvestment. PE has gained particular importance in the global financial system for delivering superior risk-adjusted returns. Knowing the PE return drivers has been of great interest among researchers and academics, and some studies have developed statistical models to determine PE return drivers. Still, the explanatory capacity of these models has certain limitations related to their precision levels and exclusive focus on groups of countries located in Europe and the EE.UU. Therefore, in the current literature, new models of analysis of the PE return drivers are demanded to provide a better fit in worldwide scenarios. This study contributes to the accuracy of the models that identify the PE return drivers using computational methods and a sample of 1606 PE funds with a geographical focus on the world's five regions. The results have provided a unique set of PE return drivers with a precision level above 86%. The conclusions obtained present important theoretical and practical implications, expanding knowledge about PE and financial forecasting from a global perspective.

Keywords Private equity · Computational methods · Fund return drivers · Feature selection · Global financial market

✉ Sergio Manuel Fernández-Miguélez
sergiofernandezmiguelez@uma.es

1 Autonomous University of Madrid, Madrid, Spain

2 University of Málaga, Málaga, Spain

3 Abdelmalek Essaâdi University, Avenue Khenifra, 93000 Tetouan, Morocco

Published online: 03 April 2024

1 Introduction

In the last twenty years, private equity (PE) has acquired particular importance in the global financial system and already exceeds 1.1 trillion dollars in assets under management. PE investment involves the purchase of shares in companies that are not listed on the stock market. It is distinguished by its often long-term approach to enhance the performance and value of the company, seeking significant benefits when carrying out the divestment (Cumming et al., 2023). The main investment strategies in PE include the acquisition of companies, investment in growing companies, and restructuring companies with financial problems (Gredil et al., 2023; Gupta & Nieuwerburgh, 2021). Over the years since the founding of the venture capital industry, PE transactions have emerged as an increasingly crucial mechanism for carrying out rapid and significant restructurings in organizations globally (Cumming et al., 2007; Wright & Bruining, 2008). On the other hand, knowing the PE return drivers has greatly interested researchers and academics (Brown et al., 2021a, 2021b). Understanding the factors that drive profitability enables PE investors and managers to detect opportunities that maximize the returns on their investments and improve risk assessment. This ability is essential given the often illiquid and long-term nature inherent to these investments. Furthermore, deep knowledge of the profitability drivers enables PE funds to optimize their investment strategies, focusing on sectors or companies with a higher potential for growth and profitability (Cumming et al., 2023; Harris et al., 2023). In this regard, some studies have developed statistical models to determine PE return drivers. Still, the explanatory capacity of these models has certain limitations related to their precision levels and exclusive focus on groups of countries located in Europe and the EE.UU. (Dai, 2022). For example, Diller and Kaserer (2009), Aigner et al. (2008), and Achleitner et al. (2010) studied the European PE, and their accuracy levels were at most 64.10%. On the other hand, Manigart et al. (2002) and Jegadeesh et al. (2015) used PE samples from Europe and EE.UU. Their accuracy was between 42.60 and 60.10%. Finally, Franzoni et al. (2012) addressed samples corresponding to EE.UU., Europe and Asia, but their precision regression did not exceed 12.50%. Consequently, the current literature demands new analysis models for PE return drivers that provide a better fit in scenarios worldwide (Caporale et al., 2024; Dai, 2022; Easton et al., 2020).

To cover the aforementioned research gap, this study aims to develop models capable of identifying the PE returns with high accuracy. We use computational methods that provide good precision and a sample with PE funds worldwide. The computational methodology has previously been used successfully in studies related to PE. Calafiore et al. (2020) apply Multilayer Perceptron for PE that carries out public share offerings. Using computational techniques, Sugathan and Baid (2013) also analyze the various factors affecting PE investment decisions. The models built here can be used in all regions, achieving accuracy above 86%. These models are built from a sample of 1,606 PE funds with a geographical focus on the world´ five regions. Our study includes a much larger sample in number and regions than those usually used in previous studies. For example, Manigart et al. (2002) used a sample of 200 PE funds in four European countries and the EE.UU. Diller and Kaserer

(2009) selected 791 European PE funds. Jegadeesh et al. (2015) used a sample of 26 PE funds from Europe, EE.UU. and Australia. Therefore, we make at least two additional contributions to the literature. First, we improve the precision of identifying PE return drivers concerning that obtained in previous studies using innovative methodologies. Second, our study has addressed the return of PE globally and is therefore not restricted to Europe or EE.UU.

This study is structured as follows. Section 2 reviews the empirical research literature on PE return drivers. Section 3 establishes the methodology used. Section 4 details the data and variables used in the study. Finally, Sect. 5 analyzes the results obtained. The article concludes by presenting the conclusions of the study, its implications, and future research.

2 Background

Previous studies on PE have addressed various issues, including the return, investment criteria, valuation of PE funds, and the dynamics of collaboration with partners (Caporale et al., 2024). For its part, the research focused on PE return drivers has been carried out from theoretical and empirical perspectives. From a theoretical perspective, different hypotheses about the functioning of private capital markets impact PE return. The first refers to the specialization effect. By specializing, PE managers are more likely to work on a significant number of similar transactions. In addition, thanks to this, the manager creates a more extensive network of contacts also specialized in a single niche, improving profitability (Manigart et al., 2002). This implies that greater specialization by industry implies higher returns (Ewens et al., 2013). The second hypothesis is linked to the speed effect on investment. According to Phalippou and Gottschalg (2009), funds with a slower investment rate obtain worse returns. Thirdly, the hypothesis on economies of scale and perimeter postulates that when PE management teams´ ability to create value and expertise have to be distributed among several investments, their returns begin to decrease. Therefore, a lesser specialization of the PE manager among several industries can have the same negative effect (Lopez-de-Silanes et al., 2015). The persistence of return hypothesis maintains that PE managers who perform better than the industry are more likely to continue with that positive spread in their subsequent managed funds (Kaplan & Schoar, 2005). The hypothesis on macroeconomic conditions also presupposes an impact on PE return (Kwabi et al., 2022). PE return is characterized by being procyclical and improves with GDP growth, and deteriorates with increases in interest rates. Likewise, it´s directly correlated with the evolution of listed markets (Phalippou & Gottschalg, 2009). Finally, according to the Resource-Based Theory, the increase in value in PE is explained by the accumulation of different internal resources, synergies and degrees of specialization. Thus, the greater the number of managed companies and the smaller they are, the less profitability they usually require. A possible reason that reinforces this hypothesis is that, in these cases, the numerous investments represent a hedge for the manager regarding the variation in returns (Manigart et al., 2002).

From an empirical perspective, the study of PE return drivers has gained importance from the seminal work of Gompers (2000), who coined money-chasing transactions, a concept that establishes the flow of funds in the PE industry as the most crucial factor driving the valuation. Subsequently, various studies have used statistical methods such as regression to identify PE return drivers. For example, Manigart et al. (2002) studied PE returns in Europe and the EE.UU from 1994 to 1997. They concluded that fund location, degree of participation, and shorter holding periods would produce higher returns. With a sample of 200 funds, their regression results achieved an accuracy of 42.6%. For their part, Diller and Kaserer (2009) confirmed the findings of Gompers (2000). Using a regression approach on a sample of 791 PE funds from 1980 to 2003, they obtained a precision of 47% and attributed a significant portion of the variation in PE returns to general fund inflows, partner skills, and independent income. Furthermore, they found that PE returns are unrelated to stock markets and negatively related to economic growth rates.

Kaplan and Schoar (2005) find evidence of fund performance persistence in their PE study. Fund managers who have been successful in the industry are more likely to repeat the success in the next fund. Aigner et al. (2008) also found persistence in PE performance and showed that past performance determines future return. In addition, other factors that drive returns include fund experience, stock market movement, economic trends, year of seniority, stage of funding, and the number of portfolio companies. They obtained an accuracy of 48.2% in their regression models with a sample of 358 funds. Jegadeesh et al. (2015) used regression and samples of 26 PE funds from Europe, EE.UU., and Australia from 1994 to 2008, achieving a precision of 65.65%. Franzoni et al. (2012), with global data from 1975 to 2006, obtained an accuracy of 18.9%. Both studies concluded that abnormal PE returns are due to risk factors stemming from a lack of liquidity. Scarpati and Ng (2013) pointed out that this abnormal profitability is due to traditional financial theory, rational factors that are not risk factors, irrational factors, behavioural factors, and internal cost factors.

Achleitner et al. (2010) investigated the drivers of PE value creation in European leveraged buyouts. They attributed one-third of returns to leverage and two-thirds to trading and market effects. Their accuracy levels did not exceed 64.10%. For Gohil and Vyas (2016), PE returns are affected by skill and market factors, and petite are affected by structural elements. These factors include the size of the investment, industry, sponsor, type of exit and stage of investment. The study by Steger (2017) indicates that macroeconomic conditions influence PE yields. It finds that weak economic growth, low bond yields, and low stock market valuations during the period in which investments are made favour returns.

Korteweg (2019) recently reviewed empirical methods to assess the risk and PE return. His findings indicate that profitability estimates vary substantially by method, period time, and data source. For their part, Roggi et al. (2019) investigated the relationship between performance and characteristics in EE.UU. Using linear and polynomial regressions, they detected a concave relationship between fund size and performance. Morri et al. (2021) investigated the performance of unlisted European real estate PE between 2001 and 2014. Their results show the importance of size and duration in the performance of funds, emphasizing that the effects of the independent variables on performance do not change significantly in different business cycles. Finally, Brown et al. (2021a, 2021b) have provided evidence of continued PE outperformance of public companies.

None of the studies discussed above examine the PE return using computational methods and for a wide range of geographic areas. Furthermore, the results presented by these studies are far from reaching a high level of precision in estimating of PE return drivers. On the contrary, the present article adopts such computational methods to obtain a global view of PE returns and to establish with high precision which drivers are important.

3 Methods

3.1 Research Design

This section discusses selecting the most critical drivers in the PE return. For this purpose, we use two PE return proxies as dependent variables: *Internal Rate of Return* and *Cash Multiple* (Gompers et al., 2016). An illustration of this procedure appears in Fig. 1. First, we prepare the database eliminating outliers and cases with missing values. Second, to check the robustness of the PE return drivers model, we use a set of computational classifiers that have provided excellent accuracy results in previous empirical studies on financial investments (Patel et al., 2015; Subasi & Cankurt, 2019; Ullah et al., 2020). These classifiers are Multilayer Perceptron, Random Forest, Näive Bayes, Vector Support Machine, and C4.5 Algorithm, which have been finally combined by the Voting Algorithm using a majority approach. Third, we use Filter and Wrapper methods for feature selection, namely Info Gain Attribute Eval, Cfs Subset Eval, Correlation Attribute Eval, Gain Ratio Attribute Eval, ReliefF Attribute Eval, Symmetrical Uncert Attribute Eval, and Classifier Attribute Eval.

3.2 Classifiers Theory

Multilayer Perceptron (MLP) is one of the widespread examples of feedforward neural networks. In MLP, the initial processing elements are unidirectionally biased. In these networks, information evolution occurs as a function of communications between three types of overlapping layers: input, hidden, and output layers. The networks between these layers are associated with weighting values, performing two functions in each MLP node, which are called summation and activation functions (Ojha et al., 2017). For this, the weights W are adjusted with the information from the sample set, considering that both the architecture and the network´s connections are known, and the objective is to obtain those weights that minimize the learning error. Given, then, a set of pairs of learning patterns $\{(x_1, y_1), (x_2, y_2)\dots (x_p, y_p)\}$ and an error function ε (W, X, Y), the training process implies the search for the set of weights that minimizes the learning error $E(W)$ (Shang & Benjamin, 1996), according to the expression (1).

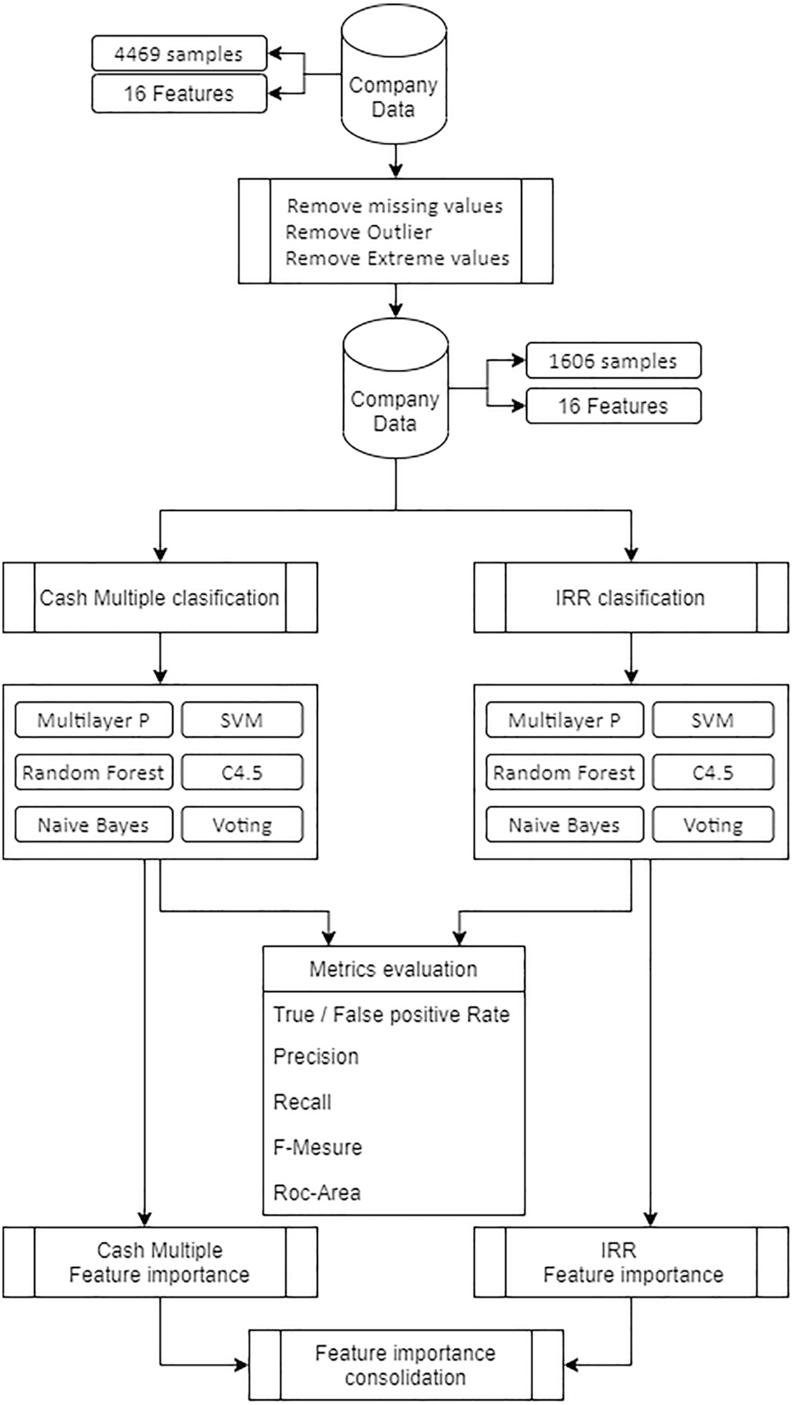

Fig. 1 Research design

$$min_w E(W) = min_w \sum_{i=1}^{p} E(W) \tag{1}$$

For its part, Random Forest (RF) is a data mining tool to solve problems related to classification and regression. Growing a set of trees and deciding the class type by voting has significantly improved classification accuracy (Breiman, 2001). To that end, random vectors are built to grow these sets. Each tree is generated from one of the random vectors. Classification problems are solved by analyzing the output of trees. A majority of the class votes determine the RF prediction. The training algorithm can be summarized as follows: A sample of the training data is drawn for each tree in the set. By growing the tree T_b over Z, the available features as candidates for splitting at the respective node are randomly selected. Finally, the grown tree $\mathrm{T_b}$ is added to the set. During inference, each tree makes a prediction $\hat{c}_b(x)$ for the class label of the new observation x. The final prediction of the random forest $\hat{c}_{RF}(x)$ is then the majority vote of the trees, as expressed in (2).

$$\hat{c}_{RF}(x) = majority\,vote\left\{\hat{c}_b(x)\right\}. \tag{2}$$

On the other hand, the Naive Bayes (NB) classifier is considered one of the probabilistic classifiers. This classifier assumes that probability distributions govern the attributes to be classified and that the optimal decision can be made considering counts of these probabilities together with the observed data (Joachims, 1998; Lewis & Ringuette, 1994). To predict the probability, it uses the concept of Bayes' theorem, which is helpful because it provides a way to compute the posterior probability, $P(C|X)$, from $P(C)$, $P(X|C)$, and $P(X)$. Bayes' theorem states Eq. (3).

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \tag{3}$$

where $P(C|X)$ indicates the posterior probability that hypothesis C is true given that event X has occurred.

Also, Support Vector Machines (SVM) have been shown to achieve good generalization performance over several classification problems. In geometric terms, SVM is seen as the attempt to find a hyperplane that separates the positive examples from the negative ones by the widest possible margin (Xu et al., 2009). The final decision is to find the maximum margin hyperplane. Assume that $x_i \in R^d$, $i=1, 2,\ldots,$ N forms a set of input vectors with class labels $y_i \in \{+1; -1\}$, $i=1, 2,\ldots, N$. SVM can map the input vectors $x_i \in R^d$ in a high-dimensional feature space $\boldsymbol{\theta}(x_i) \in H$. A kernel function $K\ (x_i, x_j)$ performs the mapping θ (.). The resulting decision boundary is defined as (4).

$$f(x) = \mathrm{sgn}\left(\sum_{1}^{N} y_i\alpha_i.K\left(x, x_i\right) + b\right) \tag{4}$$

Another of the computational classifiers used in this study is C4.5 Algorithm (C4.5), designed as an extension of ID3 Algorithm described by Quinlan (1986).

The latter is part of the classifiers known as decision trees, which are represented by trees where their internal nodes are labelled as attributes, and the outgoing branches of each node represent tests for the attribute values. The leaves of the tree identify the categories. C4.5 is formed in the following order: (a) select the attributes as roots, (b) create a branch for each value, and (c) repeat the process for each branch until all instances of the branches have the same class. The highest gain is used to select attributes, such as the root, according to the Eq. (5).

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} = Entropy(S) \tag{5}$$

where S is the set of cases, A is the attributes, n is the partition number of attribute A, and Si is the number of patients in the i-th partition. For its part, the value of *Entropy* is expressed according to Eq. (6).

$$Entropy(S) = \sum_{i=1}^{n} -p_i * log_2 * p_i \tag{6}$$

where n is the number of partitions of S, and p_i is the proportion of S.

Finally, the Voting Algorithm matched each basic-level classifier using a voting approach. The most straightforward voting approach is majority voting, in which the base-level classifier votes for its predictions. The instance is ranked in the class that gets the most votes. The plurality voting method is modified for the situation where basic classifiers estimate class probability distributions (Dietterich, 1997). This algorithm calculates the weights based on the distance between the module outputs (7).

$$W_i = \frac{1}{1 + \prod_{\substack{i=1, j=1, \\ i \neq j}}^{N} \frac{d^2(x_i, x_j)}{a^2}} \tag{7}$$

where d (xi,xj) is the distance between the output values of modules i and j, and a is a scaling factor. After assigning the weights, the output of the voter is calculated according to (8).

$$x_0 = \sum_{i=1}^{N} \left(\frac{w_i}{S}\right) \cdot x_i \tag{8}$$

where S is the sum of all the weights.

Additionally, this study uses the statistical Logit classifier to compare its precision to computational classifiers. The Logit model specification may be represented according to Eq. (9).

$$y_i^* = x_i + \varepsilon_i \tag{9}$$

with

$$y_i = \begin{cases} 1\, if\ y_i^* > 0.5 \\ 0\, if\ y_i^* \leq 0.5 \end{cases}$$

Then

$$P[y_i = 1] = P[x_i\beta + \varepsilon_i > 0.5] = F(x_i\beta) \tag{10}$$

$$P[y_i = 0] = 1 - F(x_i\beta) \tag{11}$$

Frequently, models involving discrete dependent variables are presented as index function models. In this context, discrete choice is interpreted as a reflection of an underlying regression. (Alaminos et al., 2016). The model was calculated using the backward stepwise method, in which variables are eliminated based on the probability of the plausibility statistic. As such, from Eq. (9), we obtain (12).

$$P(y_i = 1) = \frac{e^{\beta' x}}{1 + e^{\beta' x}} = \frac{1}{1 + e^{-(\beta' x)}} \tag{12}$$

Thus, the ratio between the two probabilities (known as the Odds ratio) is established as follows in Eq. (13).

$$Odds = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \frac{1/[1 + e^{-(\beta' x)}]}{1/[1 + e^{(\beta' x)}]} = \frac{1 + e^{(\beta' x)}}{1 + e^{-(\beta' x)}} = e^{(\beta' x)} \tag{13}$$

The estimated coefficients (β) represent measurements of the changes in the odds ratio. The Odds ratio may be interpreted as the number of times the phenomenon is more likely to occur than it is not (Hair et al., 1999). By applying the logarithms in (13), we obtain (14), a linear expression of the model under consideration.

$$y_i^* = \ln \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \ln e^{\beta' x} = \beta' x \tag{14}$$

3.3 Feature Selection Methods

Feature selection is the process of selecting the most important and relevant features from a data set to improve the prediction performance of predictors, provide faster and more cost-effective predictors, and provide a better understanding of the process. The feature selection methods used in this study were of two classes: Filter and Wrappers. Wrapper methods perform better than filter methods because the feature selection process is optimized for a given classifier. However, they are usually too expensive to use if the number of features is significant because they must evaluate each set of elements considered with the given classifier. On the other hand, Filter methods are much faster than wrapper methods. They are independent of any learning method since they focus on the data characteristics, making them very useful for large data sets. In this study, a set of Filter methods has been applied (Info Gain Attribute Eval, Cfs Subset Eval, Correlation Attribute Eval, Gain Ratio Attribute

Eval, ReliefF Attribute Eval, Symmetrical Uncert Attribute Eval) and Wrapper (Classifier Attribute Eval), that cater to most feature selection techniques (Omuya et al., 2021).

Info Gain Attribute Eval (IGA) evaluates the worth of an attribute by measuring the information gained to the class. It provides a way to use entropy to calculate how a change in the dataset impacts the distribution of types. Information Gain ($GAIN(X|Y)$ is expressed through the Eq. (15).

$$GAIN(X|Y) = H(X) - H(X|Y) \tag{15}$$

where $H(X)$ describe the entropy of a discrete random variable X, and $H(X|Y)$ is the conditional entropy of the class-given attribute. The higher the value of mutual information between types and attributes, the higher the relevance between types and attributes (Ashraf et al., 2010).

Cfs Subset Eval (CFS) identifies a subset of the attributes highly correlated with the class without being strongly correlated. By default, it searches through the space of possible attribute subsets for "the best" using a specified search method. It does this by evaluating a subset of attributes by calculating each attribute´s correlations (Pearson) against the class and the correlations between attributes (Chandrashekar, 2014; Megha, 2013). The equation used to filter out the irrelevant, redundant feature that leads to the class´s poor prediction is defined as (16).

$$F_s = \frac{N * r_a}{N + N(N-1)r_n} \tag{16}$$

For its part, Correlation Attribute Eval (COA) evaluates the attributes of the target class. Consider nominal attributes based on value, each value acts as an indicator. In addition, using the Pearson correlation coefficient measures the correlation between each of the attributes and the attribute of the target class (Gnanambal et al., 2018).

Gain Ratio Attribute Eval (GAA) evaluates the worth of an attribute by measuring the gain ratio to the class. It was proposed to reduce a bias towards multi-valued attributes by considering the number and size of branches when choosing an attribute (Quinlan, 1986). The Gain Ratio ($GainR$) is given through the Eq. (17).

$$GainR(Class, Attribute) = \frac{H(Class) - H(Class|Attribute)}{H(Attribute)} \tag{17}$$

ReliefF Attribute Eval (RFA) evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. It consists of assigning a weight to each attribute and selecting the attributes whose weight exceeds a predetermined threshold (Chandrashekar, 2014). If X and Y are taken as two instances, then the differences in feature values between these two instances are defined by Eq. (18).

$$diff\left(x_k, y_k\right) = \left(x_k - y_k\right)/nu_k \tag{18}$$

where nu_k is a normalization unit to normalize the values of *diff* into the interval [0,1].

Symmetrical Uncert Attribute Eval (SUA) overcomes the IGA bias towards features with more values by normalizing its value to the range [0,1] (Hall & Smith, 1998). Symmetric uncertainty is expressed through the Eq. (19).

$$SU(X, Y) = \frac{2 * GAIN(X|Y)}{H(X) + H(Y)} \tag{19}$$

where $GAIN(X|Y)$ describe the mutual information, and $H(X)$ represents the entropy of a discrete random variable X. A value 1 of $SU(X,Y)$ indicates stronger dependence/correlation between X and Y. In contrast, value 0 indicates the independence/no correlation between X and Y.

Finally, Classifier Attribute Eval (CLA) is a wrapper algorithm that uses an individual classifier as a function to evaluate the subsets and cross-validation to estimate the accuracy of the classifier (Diao et al., 2014; Kohavi & John, 1997).

4 Sample and Variables

This study aims to determine which variables explain PE return at a global level. For this purpose, we have chosen a sample composed of PE funds corresponding to five world regions between 1980 and 2020. This sample comes from the Preqin Dataset and is made up of 1606 PE funds. Of the total sample, 1.72% are from Asia, 0.16% are from Australia, 24.94% are from Europe, 72.41% are from America, and 0.77% are from Africa. In addition, among the strategies, Buyout (34.19%), Fund of funds (18.94%), Venture (17.66%) and Early Stage (12.17%) stand out. Likewise, the PE funds in the sample cover 11 industries, including Diversified (58.95%), Information Technology (19.86%), and Healthcare (7.94%) (Table 1).

On the other hand, the present study uses a set of variables selected from the previous literature (Aigner et al., 2008; Dai, 2022). As dependent variables, two main measures are used to assess the returns generated by a PE in a portfolio company exit: *Internal Rate of Return* (*IRR*) and *Cash Multiple*. The *IRR* is the most common means of assessing the performance of PE and represents the discount rate that makes the cash return multiple equal to the initial investment for a given period, according to (20).

$$IRR = \left(\frac{CM}{[1 + IRR]^t}\right) - 1 = 0 \tag{20}$$

where CM is the cash multiple of the transaction and t is the holding period.

Cash Multiple is the ratio of cash received from income divided by cash invested by PE investors and calculated according to (21).

$$Cash\,Multiple = \frac{Cash\,received\,from\,income}{Cash\,invested\,by\,PE\,investors} \tag{21}$$

Table 1 Sample characteristics

Characteristic		n	%
Geographic focus	America	4585	72.41
	Asia	109	1.72
	Australia	10	0.16
	Europe	1580	24.94
	Africa	49	0.77
Strategy	Buyout	2163	34.19
	Co-Investment	169	2.67
	Early Stage	771	12.17
	Expansion	184	2.90
	Fund of funds	1200	18.94
	Growth	422	6.66
	Secondaries	305	4.81
	Venture	1119	17.66
Core industry	Business Services	62	0.98
	Consumer Discretionary	193	3.04
	Diversified	3732	58.95
	Energy & Utilities	123	1.94
	Financial & Insurance Services	62	0.98
	Healthcare	503	7.94
	Information Technology	1258	19.86
	Industrials	183	2.89
	Raw Materials & Natural Resources	24	0.38
	Real Estate	37	0.58
	Telecoms & Media	156	2.46

Regarding the independent variables, those related to the ability factors of the fund managers to execute a certain agreement have been considered. Specifically, they refer to the experience of the managers (*Assets manager*) (Gohil & Vyas, 2016), the size of the fund (*Fund size* and *Log fund size*) (Aigner et al., 2008; Farooq et al., 2021; Gohil & Vyas, 2016), and the location of the administrators (*Region fund manager*) (Caporale et al., 2024). Also, variables refer to market factors and related to the country´s financial market where the PE is located. Seven factors have been included in the study as market drivers: three-month average interest rates in Germany and the EE.UU. over the life of a fund (*Germany interest and EE.UU. interest*) (Steger, 2017), nominal GDP growth in Germany and EE.UU. (*Germany GDP and EE.UU. GPD*) (Aigner et al., 2008), the performance of the MSCI World Performance Index as a proxy for developing the stock market (*MSCI return*) (Aigner et al., 2008; Steger, 2017) and the amount of money committed with PE worldwide (*Log world PE* and *World PE*) (Aigner et al., 2008). Finally, in third place, the variables that include the variables related to the PE structure: diversification between regions, industrial sectors and stages of investment, measured by the

Table 2 Econometric variables

Variable	Description
Internal rate of return	Discount rate, which makes the cash return multiple equal to the initial investment
Cash multiple	The ratio of cash received from income divided by cash invested
Assets manager	Assets under management (USD MN)
Fund size	Fund capital (USD MN)
Log fund size	The logarithm of Fund size
Region fund manager	Location of managers
Germany interest	The three-month interest rate in Germany
EEUU interest	The three-month interest rate in EE UU
Germany GDP	GDP annual growth rate in Germany
EEUU GPD	GDP annual growth rate in EEUU
MSCI return	The gross annual return of MSCI World Performance Index
World PE	Amount of money committed to PE worldwide
Log world PE	The logarithm of World PE
Diversification	Herfindahl–Hirschman Index
Geography focus	Geographical area in which the fund has made its investment
Industry focus	Specific industry in which the PE has made the investment
Type of fund	The fund was classified as venture capital or buyout fund
Lifetime	Duration of the fund in years

Herfindahl–Hirschman index (*Diversification*) (Ick, 2005), the geographical area in which the fund has made its investment (*Geography focus*) (Caporale et al., 2024), the specific PE industry (*Industry focus*), the fund categories according to whether they are venture capital and private capital (*Type of fund*), and the duration of the fund in years (*Lifetime*) (Aigner et al., 2008). Table 2 shows the description of the variables used in the research.

5 Empirical Results

This section presents PE return drivers analysis results for the two proposed independent variables (*IRR* and *Cash Multiple*).

5.1 Descriptive Analysis

The objective of the descriptive analysis is to examine the variables to be used and to know their main statistical parameters. A summary of the quantitative variables used in the research is shown in Table 3. All the variables present a moderate dispersion to the mean values, except for *Assets manager* and *Fund size*, due to the broad spectrum of PE in the sample.

Table 3 Descriptive statistics

Variable	Mean	S.D	Max	Min	K–S
Internal Rate of Return	15.3664	12.7678	106.9000	− 15.4000	0.371
Cash Multiple	1.7720	1.0185	19.9300	0.0600	0.488
Assets manager	710.5792	1957.8558	26,864.2000	0.1769	0.206
Fund size	912.8143	2058.3590	26,200.0000	0.7000	0.178
Log fund size	2.4921	0.6402	4.4183	− 0.1549	0.025
Germany interest	− 0.3703	0.3529	0.4644	− 0.6871	0.001
EEUU interest	2.0587	0.2760	2.6319	1.1390	0.002
Germany GDP	0.6613	0.5287	1.3455	− 2.0000	0.013
EEUU GPD	1.3599	0.2834	1.7364	− 0.5500	0.004
MSCI return	11.4560	1.8683	22.4500	8.8300	0.277
World PE	500.5368	59.1543	587.6667	408.1538	0.134
Log world PE	2.6963	0.0517	2.7691	2.6108	0.016
Diversification	0.6282	0.4950	1.0000	0.0000	–
Lifetime	9.1687	4.0587	16.0000	2.0000	0.319

K–S: Kolmogorov–Smirnov test significance

Table 4 Classification results for *IRR*

Classifier	Global accuracy (%)	Class	TP	FP	Precision (%)	Recall (%)	F-measure	MCC	ROC Area
MLP	85.01	0	0.872	0.182	0.873	0.872	0.873	0.690	0.903
		1	0.818	0.128	0.817	0.818	0.818	0.690	0.903
RF	85.81	0	0.717	0.109	0.904	0.717	0.800	0.599	0.853
		1	0.891	0.283	0.687	0.891	0.776	0.599	0.853
NB	78.84	0	0.717	0.109	0.904	0.717	0.800	0.599	0.853
		1	0.891	0.283	0.687	0.891	0.776	0.599	0.853
SVM	84.85	0	0.846	0.148	0.891	0.846	0.868	0.692	0.849
		1	0.852	0.154	0.794	0.852	0.822	0.692	0.849
C4.5	86.13	0	0.886	0.174	0.88	0.886	0.883	0.713	0.882
		1	0.826	0.114	0.834	0.826	0.830	0.713	0.882
Voting	86.05	0	0.851	0.125	0.907	0.851	0.880	0.718	0.918
		1	0.875	0.149	0.803	0.875	0.840	0.718	0.918
Logit	71.43	0	0.702	0.148	0.860	0.702	0.789	0.583	0.782
		1	0.825	0.316	0.612	0.825	0.735	0.583	0.782

TP: rate of true positives; FP: rate of false positives; Precision: proportion of instances that are true of a class divided by the total instances classified as that class; Recall: proportion of instances classified as a given class divided by the actual total in that class; F-Measure: combined measure for precision and recall calculated as 2 * Precision * Recall/(Precision + Recall); MCC: a standard of the quality of binary (two-class) classifications. ROC area: Receiver Operating Characteristics area

Table 5 Classification results for *Cash Multiple*

Classifier	Global accuracy	Class	TP	FP	Precision	Recall	F-measure	MCC	ROC Area
MLP	80.36	0	0.910	0.355	0.792	0.910	0.847	0.587	0.848
		1	0.645	0.090	0.829	0.645	0.726	0.587	0.848
RF	81.81	0	0.903	0.309	0.813	0.903	0.856	0.618	0.871
		1	0.691	0.097	0.828	0.691	0.754	0.618	0.871
NB	74.43	0	0.757	0.275	0.804	0.757	0.780	0.477	0.828
		1	0.725	0.243	0.668	0.725	0.695	0.477	0.828
SVM	81.57	0	0.930	0.355	0.796	0.93	0.858	0.615	0.788
		1	0.645	0.07	0.862	0.645	0.738	0.615	0.788
C4.5	81.41	0	0.940	0.373	0.789	0.94	0.858	0.614	0.821
		1	0.627	0.060	0.875	0.627	0.731	0.614	0.821
Voting	81.80	0	0.945	0.369	0.792	0.945	0.862	0.625	0.869
		1	0.631	0.055	0.885	0.631	0.737	0.625	0.869
Logit	70.08	0	0.737	0.206	0.800	0.737	0.715	0.462	0.736
		1	0.684	0.362	0.604	0.684	0.709	0.462	0.736

TP: rate of true positives; FP: rate of false positives; Precision: proportion of instances that are true of a class divided by the total instances classified as that class; Recall: proportion of instances classified as a given class divided by the actual total in that class; F-Measure: combined measure for precision and recall calculated as 2 * Precision * Recall/(Precision + Recall); MCC: a standard of the quality of binary (two-class) classifications. ROC area: Receiver Operating Characteristics area

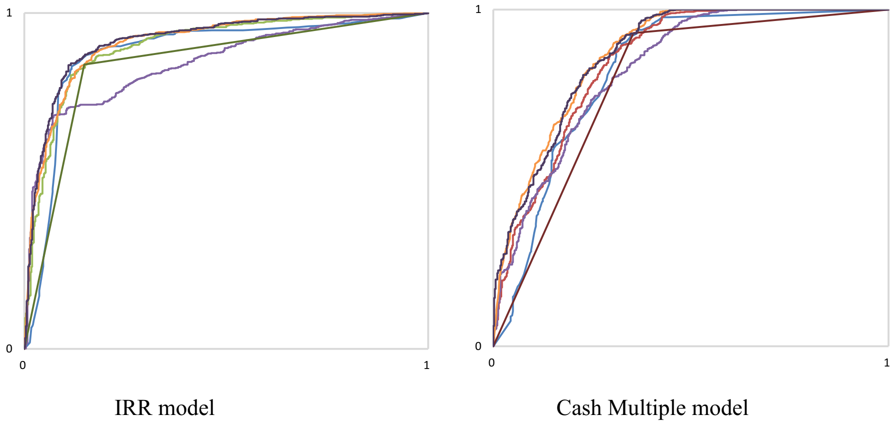

Fig. 2 ROC curve analysis

5.2 Classification Fit

Tables 4 and 5 report the classification results obtained in analyzing PE return drivers for IRR and Cash Multiple. Considering all goodness-of-fit criteria (Global accuracy, Precision, Recall, F-Measure, and MCC), the classifier that has provided the highest accuracy for IRR is C4.5, with 86.13% (Table 4). For Cash Multiple, the highest accuracy has been obtained with RF (81.81%) (Table 5). For its part, Fig. 2 illustrates the ROC curve of all the classifiers, indicating that, in all cases, the classification fit has been acceptable.

5.3 Feature Selection

Tables 6 and 7 show the variables selected by each feature selection method for *IRR* and *Cash Multiple* variables, respectively. For a better understanding of the ranked variables obtained, three groups have been established based on the importance of the variables. Thus, group G1 (green colour) is identified with the highest-ranked variables. Group G2 (pink colour) with those of a moderate ranking. And finally, the group G3 (grey colour), with the lowest rated. Next, the frequencies of each variable were counted according to the feature selection method used, obtaining a summary of the importance of the independent variables (Table 8). The results indicate that

Table 6 Results feature selection for *IRR*.

Info gain attribute		Cfs subset		Correlation attribute		Gain ratio attribute		ReliefF attribute		Symmetrical uncert		Classifier attribute	
Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable
0.3795	*Lifetime*	–	*MSCI return*	0.5475	*Germany interest*	0.1699	*EEUU interest*	0.3159	*Lifetime*	0.2347	*EEUU interest*	0.2267	*Germany interest*
0.3785	*Germany GDP*	–	*EEUU GPD*	0.4288	*MSCI return*	0.1560	*Germany interest*	0.1240	*Industry focus*	0.2220	*Germany interest*	0.2267	*MSCI return*
0.3785	*Log world PE*	–	*Germany interest*	0.4221	*EEUU interest*	0.1424	*MSCI return*	0.1084	*Germany interest*	0.2073	*MSCI return*	0.2267	*EEUU interest*
0.3785	*World PE*	–	*EEUU interest*	0.4075	*Log world PE*	0.1350	*World PE*	0.0678	*EEUU interest*	0.2003	*World PE*	0.2267	*Germany GDP*
0.3758	*Germany interest*	–	*Industry focus*	0.3982	*World PE*	0.1350	*Log world PE*	0.0655	*MSCI return*	0.2003	*Log world PE*	0.2267	*Log world PE*
0.3717	*MSCI return*	–	*Assets manager*	0.1852	*Lifetime*	0.1308	*EEUU GPD*	0.0499	*EEUU GPD*	0.1945	*EEUU GPD*	0.2267	*EEUU GPD*
0.3704	*EEUU interest*	–	–	0.1696	*Assets manager*	0.1294	*Germany GDP*	0.0492	*Diversification*	0.1940	*Germany GDP*	0.2267	*World PE*
0.3700	*EEUU GPD*	–	–	0.1438	*Industry focus*	0.1022	*Lifetime*	0.0418	*Log world PE*	0.1618	*Lifetime*	0.2267	*Lifetime*
0.0802	*Industry focus*	–	–	0.1314	*Diversification*	0.0580	*Assets manager*	0.0388	*World PE*	0.0640	*Assets manager*	0.0508	*Industry focus*
0.0698	*Assets manager*	–	–	0.1147	*Germany GDP*	0.0471	*Industry focus*	0.0347	*Geography focus*	0.0599	*Industry focus*	0.0363	*Assets manager*
0.0133	*Diversification*	–	–	0.0994	*EEUU GPD*	0.0138	*Diversification*	0.0337	*Type of fund*	0.0137	*Diversification*	0.0040	*Region fund manager*
0.0104	*Region fund manager*	–	–	0.0845	*Type of fund*	0.0109	*Region fund manager*	0.0315	*Region fund manager*	0.0108	*Region fund manager*	0.0008	*Log fund size*
0.0061	*Geography focus*	–	–	0.0662	*Log fund size*	0.0063	*Type of fund*	0.027	*Germany GDP*	0.0062	*Geography focus*	0.0008	*Fund size*
0.0051	*Type of fund*	–	–	0.0179	*Region fund manager*	0.0062	*Geography focus*	0.0232	*Log fund size*	0.0057	*Type of fund*	0.0004	*Geography focus*
0.0000	*Log fund size*	–	–	0.0176	*Geography focus*	0.0000	*Fund size*	0.0122	*Fund size*	0.0000	*Fund size*	0.0000	*Diversification*
0.0000	*Fund size*	–	–	0.0033	*Fund size*	0.0000	*Log fund size*	0.0018	*Assets manager*	0.0000	*Log fund size*	0.0000	*Type of fund*

Classifier Attribute results using C4.5

Table 7 Results feature selection for *Cash multiple*.

Info gain attribute		Cfs subset		Correlation attribute		Gain ratio attribute		ReliefF attribute		Symmetrical uncert		Classifier attribute	
Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable	Ranked	Variable
0.2626	*Lifetime*	–	*EEUU GPD*	0.4106	*EEUU GPD*	0.1699	*EEUU interest*	0.2077	*Lifetime*	0.1751	*EEUU GPD*	0.1794	*Germany interest*
0.2607	*Germany GDP*	–	*Germany interest*	0.2735	*Germany GDP*	0.1560	*Germany interest*	0.1234	*Industry focus*	0.1528	*Germany interest*	0.1794	*World PE*
0.2596	*MSCI return*	–	*Fund size*	0.1335	*Log fund size*	0.1424	*MSCI return*	0.0570	*EEUU GPD*	0.1528	*EEUU interest*	0.1794	*EEUU GDP*
0.2559	*World PE*	–	*Industry focus*	0.1295	*Lifetime*	0.1350	*World PE*	0.0510	*Diversification*	0.1403	*World PE*	0.1794	*Germany GDP*
0.2559	*Log world PE*	–	–	0.1286	*Industry focus*	0.1350	*Log world PE*	0.0450	*MSCI return*	0.1403	*Log world PE*	0.1794	*MSCI return*
0.2548	*Germany interest*	–	–	0.1225	*World PE*	0.1308	*EEUU GDP*	0.0432	*Germany GDP*	0.1325	*MSCI return*	0.1794	*EEUU interest*
0.2548	*EEUU interest*	–	–	0.1210	*Diversification*	0.1294	*Germany GDP*	0.0387	*World PE*	0.1230	*Germany GDP*	0.1794	*Log world PE*
0.2474	*EEUU GPD*	–	–	0.1156	*Germany interest*	0.1022	*Lifetime*	0.0381	*Log world PE*	0.1120	*Lifetime*	0.1758	*Lifetime*
0.0619	*Industry focus*	–	–	0.1105	*Log world PE*	0.0580	*Assets manager*	0.0328	*EEUU interest*	0.0462	*Industry focus*	0.0552	*Industry focus*
0.0155	*Log fund size*	–	–	0.0979	*Type of fund*	0.0471	*Industry focus*	0.0274	*Type of fund*	0.0158	*Fund size*	0.0128	*Geography focus*
0.0155	*Fund size*	–	–	0.0431	*MSCI return*	0.0138	*Diversification*	0.0227	*Geography focus*	0.0158	*Log fund size*	0.0048	*Region manager*
0.0129	*Geography focus*	–	–	0.0421	*Fund size*	0.0109	*Region fund manager*	0.0217	*Germany interest*	0.0132	*Geography focus*	0.0036	*Log fund size*
0.0120	*Region fund manager*			0.0338	*Assets manager*	0.0063	*Type of fund*	0.0195	*Region fund manager*	0.0125	*Region fund manager*	0.0036	*Fund size*
0.0115	*Diversification*			0.0332	*Geography focus*	0.0062	*Geography focus*	0.0170	*Log fund size*	0.0118	*Diversification*	0.0000	*Assets manager*
0.0068	*Type of fund*			0.0138	*Region fund manager*	0.0000	*Fund size*	0.0042	*Fund size*	0.0077	*Type of fund*	0.0000	*Diversification*
0.0000	*Assets manager*			0.0114	*EEUU interest*	0.0000	*Log fund size*	0.0019	*Assets under the manager*	0.0000	*Assets under the manager*	0.0000	*Type of fund*

Classifier Attribute results using RF

Table 8 Variables frequency in the feature selection

Variable	IRR			Cash multiple		
	Frequency in G1	Frequency in G2	Frequency in G3	Frequency in G1	Frequency in G2	Frequency in G3
Type of fund	0	2	5	0	1	8
Lifetime	7	0	0	9	0	0
World PE	6	1	0	8	0	0
Log world PE	6	1	0	8	0	0
MSCI return	8	0	0	7	1	0
Germany GDP	0	6	1	8	0	1
EEUU GPD	6	2	0	9	0	0
Germany interest	8	0	0	8	1	0
EEUU interest	8	1	0	7	1	0
Fund size	0	0	7	1	3	5
Log fund size	0	1	6	1	3	4
Geography focus	0	2	5	0	6	2
Region fund manager	0	6	1	1	3	5
Diversification	1	4	2	2	2	4
Industry focus	3	5	0	3	6	0
Assets manager	2	4	2	0	0	8

the most important variables for *IRR* are, in the first place (G1), *Lifetime*, World PE, *Log world PE*, *MSCI return*, *EE.UU. GPD*, *Germany interest*, *EE.UU. interest*, and *Industry focus*. In second place (G2), *Germany GDP*, *Diversification*, and *Assets manager*. The least important ones (G3) are *Fund size*, *Log fund size*, and *Geography focus*. On the other hand, and for *Cash Multiple*, it is confirmed that the most important variables (G1) are *Lifetime*, *World PE*, *Log world PE*, *MSCI return*, *Germany GDP*, *EE.UU. GPD*, *Germany interest*, and *EE.UU. interest*. In second place (G2), *Geography focus*, *Region fund manager*, and *Industry focus*. And in third place (G3), Type of fund, *Fund size*, *Log fund size*, and *Assets manager*.

Globally, considering the results obtained for *IRR* and *Cash Multiple*, the variables of the most significant importance for the analysis of PE return have turned out to be the set made up of *Lifetime*, *World PE*, *Log world PE*, *MSCI return*, *EE.UU. GDP*, *Germany interest*, and *EE.UU. interest* (Fig. 3).

5.4 Discussion

The results of the present study show that the computational methods exceed the accuracy of previous studies on PE return using statistical methods. For example, Diller and Kaserer (2009), Aigner et al. (2008), and Achleitner et al. (2010) did not exceed 64.10% accuracy with European PE samples, and Manigart et al. (2002) and Jegadeesh et al. (2015), with samples from Europe and EE.UU., were only

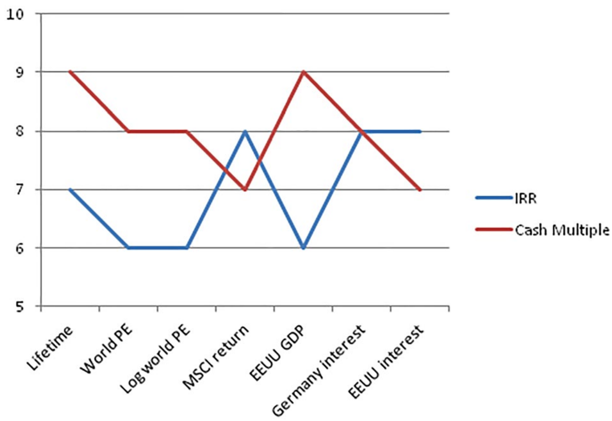

Fig. 3 Frequency of the most important variables

between 42.60 and 60.10%. Likewise, Franzoni et al. (2012) study, with samples from EE.UU., Europe, and Asia, only achieved a regression fit of 12.50%. However, our models on PE return have provided accuracy of 86.13% and 81.81% for IRR and Cash Multiple models, respectively. Specifically, C4.5 and RF have been the classifiers with the best classification fit. This greater precision of computational classifiers that traditional statistical classifiers have also been verified with data from our sample. The LOGIT classifier applied in the present study has provided the lowest level of precision. Therefore, computational classifiers offer greater accuracy in modelling the PE return factors with global samples from various world regions. The nonlinear relationship of the variables that explain PE return is possibly better measured with computational methods because they do not require previous assumptions about the relationship between variables (Núñez de Castro & Von Zuben, 1998).

Our results provide very robust global models to identify PE return drivers. This represents a significant advance in the generalization of PE return models. However, significant differences in economic environment, regulation, business culture, and other factors between the five regions studied could provide different results when applying such global models with data from a single region (Morri et al., 2021). Perhaps the use of a worldwide sample is more directed towards a broad and complete vision of the trends and patterns that may be present worldwide, that is, when the objective is to obtain a more holistic and diverse understanding of the problem studied.

Regarding the most critical variables in the analysis, our results suggest that market factors are highly significant, specifically those related to interest rates (*Germany interest* and *EE.UU. interest*), with GDP growth (*EE.UU. GPD*), and market development (*MSCI return*, *World PE*, and *Log world PE*). These results are similar to those obtained by Phalippou and Gottschalg (2009) and Steger (2017) since they indicated that macroeconomic conditions affect PE return and that it is characterized by being pro-cyclical. However, our results differ from those obtained by Morri et al. (2021) and Diller and Kaserer (2009), who found that PE returns are unrelated to macroeconomic factors. Perhaps using more extensive databases in the present study has provided new information on the variables that affect PE return.

On the other hand, in our study, the variable that refers to the years of PE duration (*Lifetime*) has turned out to be highly significant, and this is in line with what was obtained by Aigner et al. (2008), as they found that duration was a variable directly related to PE return. In general, our results partially confirm the previous findings of Gohil and Vyas (2016) by verifying that PE returns are mainly affected by market factors. However, although Gohil and Vyas (2016), Roggi et al. (2019) and Morri et al. (2021) also pointed out the importance of variables related to the ability of PE managers, in our study, this point has not been relevant. Such is the case of variables such as *Assets manager*, *Fund size*, *Log fund size*, and *Region fund manager*.

Finally, our results confirm what Gohil and Vyas (2016) proposed by indicating that some variables related to the PE structure have not shown significance in explaining PE return either. These structural variables that have shown little importance in the analysis refer to the distribution of PE between regions and industrial sectors (*Diversification*, *Geography focus*, and *Industry focus*).

6 Conclusions and Implications

In the present study, a comparison of computational methodologies has been carried out. Consequently, new models have been developed that determine the PE return drivers at a global level with high precision. Therefore, the models built here can be used in all countries, achieving accuracy above 86%. These models have been built from a sample of 1,606 PE funds with a geographical focus on the world´s five regions. For this, different computational classifiers and feature selection methods have been applied, which have provided robustness in determining the PE return drivers. Specifically, the objective has been to improve the accuracy of the models developed in previous studies using different methodologies and increasing the sample size to all world regions. The results obtained in this study are significantly higher than those obtained in previous literature, with a precision range of 81.81–86.13% for selecting PE return drivers. These results suggest that market factors are of great importance in the selection of PE return drivers, especially the interest rate in Germany and EE.UU., GDP growth in the EE.UU., the performance of the MSCI World Performance Index, and the amount of money committed with PE around the world. In addition, the variable related to the PE duration has also been significant, including one of the drivers of PE structural characteristics. In short, the variables selected in this study constitute a unique set of PE return drivers to estimate the PE performance globally with high precision.

The conclusions obtained present important theoretical and practical implications. Unlike previous research, this study has expanded knowledge about PE return drivers beyond the experience with funds focused on Europe and EE.UU, offering a global analysis model. Our study suggests a unique set of significant explanatory variables to forecast PE return. This research also demonstrates that the C4.5 and RF classifiers are the most accurate for PE return analysis, thus contributing to existing knowledge in financial forecasting. These results can be used as a reference to establish better PE management decisión-making. In addition, it has also been possible to verify that the PE returns are linked to macroeconomic and market factors. PE benefits from a growing economy over the life of the fund. Therefore, investors should commit when the markets are weak and the economy begins to recover. Finally, contrary to what might be expected, we have yet to find any significant advantage in specialization for a PE manager. Industrial or regional specialization effects are only somewhat relevant in the constructed models. One explanation for this phenomenon could be that PE managers are often specialized within their organization. The managing entity of the fund is not a single person but a management team that, together, creates a well-established organization and good performance.

This study has limitations that suggest future research lines. Our models on PE return drivers have considered both specific variables of the PE funds and others referring to the market context. However, variables related to the countries' environmental policies have yet to be taken into account. Future research in this field could examine whether these new variables impact PE performance, as it seems particularly important considering that governments' capital investments related to sustainable technologies are a vital issue. Secondly, the present study has used a set of robust computational classifiers to measure the models´ precision. Still, other computational methods have not been considered. This is the case, for example, of quantum computing methods, which recently also show an interesting trend in finance. Therefore, future studies could evaluate PE return models with these new computational techniques. Finally, only three variables on the characteristics of the fund manager have been considered in the present study. Further research could shed more light on the skill factors that promote the return of PE.

Funding Funding for open access publishing: Universidad Málaga/CBUA. The authors have not disclosed any funding.

Declarations

Conflict of interest The authors have not disclosed any conflict of interest.

References

Achleitner, A. K., Braun, R., Engel, N., Figge, C., & Tappeiner, F. (2010). Value creation drivers in private equity buyouts: empirical evidence from Europe. *The Journal of Private Equity, 13*(2), 17–27.

Aigner, P., Albrecht, S., Beyschlag, G., Friederich, T., Kalepky, M., & Zagst, R. (2008). What drives pe? Analyses of success factors for private equity funds. *The Journal of Private Equity, 11*(4), 63–85.
Alaminos, D., del Castillo, A., & Fernández, M. A. (2016). A global model for bankruptcy prediction. *PLoS ONE, 11*(11), e0166693.
Ashraf, M., Le, K., & Huang, X. (2010). Information gain and adaptive neuro-fuzzy inference system for breast cancer diagnoses. In *Computer Sciences and Convergence Information Technology* (ICCIT), 5th International Conference on IEEE.
Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
Brown, G., Harris, R. S., Hu, W., Jenkinson, T., Kaplan, S. N., & Robinson, D. T. (2021b). Can investors time their exposure to private equity? *Journal of Financial Economics, 139*(2), 561–577.
Brown, G., Harris, R., & Munday, S. (2021a). Capital structure and leverage in private equity buyouts. *Journal of Applied Corporate Finance, 33*(3), 42–58.
Calafiore, G. C., Morales, M. H., Tiozzo, V., Fracastoro, G., & Marquie, S. (2020). Survival and neural models for private equity exit prediction. *IFAC-PapersOnLine, 53*(2), 16983–16988.
Caporale, G. M., Gil-Alana, L. A., & Puertolas, F. (2024). Modelling profitability of private equity: A fractional integration approach. *Research in International Business and Finance, 67*, 102087.
Chandrashekar, G. (2014). A survey on feature selection methods. *Computers and Electrical Engineering, 40*, 16–28.
Cumming, D., Kumar, S., Lim, W. M., & Pandey, N. (2023). Mapping the venture capital and private equity research: A bibliometric review and future research agenda. *Small Business Economics, 61*, 173–221.
Cumming, D., Siegel, D. S., & Wright, M. (2007). Private equity, leveraged buyouts and governance. *Journal of Corporate Finance, 13*, 439–460.
Dai, N. (2022). Empirical research on private equity funds: A review of the past decade and future research opportunities. *Review of Corporate Finance*, *Forthcoming*, Available at SSRN: https://ssrn.com/abstract=4088907.
Diao, R., Chao, F., Peng, T., Snooke, N., & Shen, Q. (2014). Feature selection inspired classifier ensemble reduction. *IEEE Transactions on Cybernetics, 44*(8), 1259–1268.
Dietterich, T. G. (1997). Machine learning research: Four current directions. *AI Magazine, 18*(4), 97–101.
Diller, C., & Kaserer, C. (2009). What drives private equity returns? Fund inflows, skilled GPs, and/or risk? *European Financial Management Journal., 15*(3), 643–675.
Easton, P., Larocque, S., & Stevens, J. (2020). Private equity valuation before and after ASC 820. *Journal of Investment Management, 19*(4), 1–14.
Ewens, M., Jones, D. M., & Rhodes-Kropf, M. (2013). The price of diversifiable risk in venture capital and private equity. *The Review of Financial Studies, 26*(8), 1854–1889.
Farooq, U., Nasir, A., Bilal, & Quddoos, M. U. (2021). The impact of COVID-19 pandemic on abnormal returns of insurance firms: a cross-country evidence. *Applied Economics, 53*(31), 3658–3678.
Franzoni, F., Nowak, E., & Phalippou, L. (2012). Private equity performance in liquidity risk. *Journal of Finance, 67*, 2341–2373.
Gnanambal, S., Thangaraj, M., Meenatchi, V. T., & Gayathri, V. (2018). Classification algorithms with attribute selection: An evaluation study using WEKA. *International Journal of Advanced Networking and Applications, 9*(6), 3640–3644.
Gohil, R. K., & Vyas, V. (2016). Factors driving abnormal returns in private equity industry: A new perspective. *The Journal of Private Equity, 19*(3), 30–36.
Gompers, P. A. (2000). Money chasing deals? The impact of fund inflows on private equity valuation. *Journal of Financial Economics, 55*, 281–325.
Gompers, P., Kaplan, S. N., & Mukharlyamov, V. (2016). What do private equity firms say they do? *Journal of Financial Economics, 121*(3), 449–476.
Gredil, O. R., Griffiths, B., & Stucke, R. (2023). Benchmarking private equity: The direct alpha method. *Journal of Corporate Finance, 81*, 102360.
Gupta, A., & Nieuwerburgh, S. V. (2021). Valuing private equity investments strip by strip. *The Journal of Finance, 76*(6), 3255–3307.
Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1999). *Análisis multivariante, 5a edición*. Editorial Prentice Hall.
Hall, M. A., & Smith, L. A. (1998). Practical Feature subset selection for machine learning. In *Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98*. Springer, pp. 181–191.

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, 81*, 102361.
Jegadeesh, N., Kräussl, R., & Pollet, J. (2015). Risk and expected returns of private equity investments: Evidence based on market prices. *The Review of Financial Studies, 28*(12), 3269–3302.
Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the Tenth European Conference on Machine Learning (ECML'98)*. Lecture Notes in Computer Science, vol. 1398, pp. 137–142.
Kaplan, S. N., & Schoar, A. (2005). Private equity performance: Returns, persistence, and capital flows. *The Journal of Finance, 60*(4), 1791–1823.
Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273–324.
Korteweg, A. (2019). Risk adjustment in private equity returns. *Annual Review of Financial Economics, 11*, 131–152.
Kwabi, F. O., Owusu-Manu, S., Boateng, A., Ezeani, E. B., & Du, M. (2022). Economic policy uncertainty and cost of capital: The mediating effects of foreign equity portfolio flow. *Review of Quantitative Finance and Accounting, 59*, 457–481.
Lewis, D. D., & Ringuette, M. (1994). A comparison of two learning algorithms for text categorization. *Third Annual Symposium on Document Analysis and Information Retrieval, 33*, 81–93.
Lopez-de-Silanes, F., Phalippou, L., & Gottschalg, O. (2015). Giants at the gate: Investment returns and diseconomies of scale in private equity. *Journal of Financial and Quantitative Analysis, 50*(3), 377–411.
Manigart, S., De Waele, K., Wright, M., Robbie, K., Desbrièresc, P., Sapienzad, H. J., & Beekmane, A. (2002). Determinants of required return in venture capital investment: A five-country study. *Journal of Business Venturing, 17*(4), 291–312.
Megha, A. (2013). Performance analysis of different feature selection methods in intrusion detection. *International Journal of Scientific & Technology Research, 2*, 225–231.
Morri, G., Perini, U., & Anconetani, R. (2021). Performance determinants of European private equity real estate funds. *Journal of European Real Estate Research, 14*(2), 209–228.
Núñez de Castro, L., & von Zuben, F.J. (1998). *Optimised training techniques for feedforward neural networks technical report DCA RT 03/98*. Department of Computer Engineering and Industrial Automation. FEE/UNICAMP, Brasil.
Ojha, V. K., Abraham, A., & Snásel, V. (2017). Metaheuristic design of feedforward neural networks: A review of two decades of research. *Engineering Applications of Artificial Intelligence, 60*, 97–116.
Omuya, E. O., Okeyo, G. O., & Kimwele, M. W. (2021). Feature selection for classification using principal component analysis and information gain. *Expert Systems with Applications, 174*, 114765.
Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications, 42*(1), 259–268.
Phalippou, L., & Gottschalg, O. (2009). The performance of private equity funds. *The Review of Financial Studies, 22*(4), 1747–1776.
Quinlan, J. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106.
Roggi, O., Giannozzi, A., Baglioni, T., & Pagliai, F. (2019). Private equity characteristics and performance: An analysis of North American venture capital and buyout funds. *Economic Notes, 48*(2), 12128.
Scarpati, F., & Ng, W. (2013). What really drives risk premium and abnormal returns in private equity funds? A new perspective. *The Journal of Private Equity, 16*(4), 8–20.
Shang, Y., & Benjamin, W. W. (1996). Global optimization for neural network training. *IEEE Computer, 29*(3), 45–54.
Steger, D. (2017). Macroeconomic conditions and private equity fund returns a Swiss perspective. *The Journal of Private Equity, 21*(1), 20–30.
Subasi, A., & Cankurt, S. (2019). Prediction of default payment of credit card clients using data mining techniques. *International Engineering Conference (IEC), 2019*, 115–120.
Sugathan, A., & Baid, A. (2013). Private equity decision making using a risk-weighted optimal decision-making paradigm and neural networks. *The Journal of Private Equity, 16*(3), 15–18.
Ullah, I., Xuefeng, Q., Shah, M. H., Rehman, A., Ali, S., & Ahmed, Z. (2020). Forecasting wages inequality in response of trade openness in Pakistan: An artificial neural network approach. *The Singapore Economic Review*. https://doi.org/10.1142/s0217590820500058
Wright, M., & Bruining, H. (2008). *Private equity and management buyouts*. Edward Elgar.

Xu, X., Zhou, C., & Wang, Z. (2009). Credit scoring algorithm based on link analysis ranking with support vector machine. *Expert Systems with Applications, 36*, 2625–2632.