



HAL
open science

Machine learning portfolios with equal risk contributions: Evidence from the Brazilian market

Alexandre Rubesam

► To cite this version:

Alexandre Rubesam. Machine learning portfolios with equal risk contributions: Evidence from the Brazilian market. *Emerging Markets Review*, 2022, 51 (Part B), pp.100891. 10.1016/j.ememar.2022.100891 . hal-03707365

HAL Id: hal-03707365

<https://hal.science/hal-03707365v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Machine learning portfolios with equal risk contributions: evidence from the Brazilian market

Alexandre Rubesam^{a,*}

^a*Department of Finance, IESEG School of Management (LEM-CNRS 9221), Paris, France*

Abstract

We investigate the use of machine learning (ML) to forecast stock returns in the Brazilian market using a rich proprietary dataset. While ML portfolios can easily outperform the local market, the performance of long-short strategies using ML is hampered by the high volatility of the short portfolios. We show that an Equal Risk Contribution (ERC) approach significantly improves risk-adjusted returns. We further develop an ERC approach that combines multiple long-short strategies obtained with ML models, equalizing risk contributions across ML models, which outperforms, on a risk-adjusted basis, all individual ML long-short strategies, as well as alternative combinations of ML strategies.

Keywords: emerging markets, machine learning, stock market prediction, portfolio optimization, equal risk contribution, risk parity

JEL: C53, G11, G15.

*We would like to thank the Subject Editor and two anonymous referees for helpful comments. We also thank participants of the 2021 Annual Meeting of the European Financial Management Association in Leeds, the 2021 Annual Meeting of the Financial Management Association in Denver, the 57th Meeting of the Eastern Finance Association, the XXI Workshop on Quantitative Finance in Naples, the 37th International Conference of the French Finance Association (AFFI) in Nantes, and seminar participants in the Finance Hub research seminar at INSPER in São Paulo, Brazil, and the Norsk Forening for Kvantitativ Finans in Oslo, Norway, for useful comments and suggestions.

*Corresponding author. IESEG School of Management, 1 Parvis de La Défense - F-92044 Paris La Défense, France. Tel: +33 (0)1435 882622.

Email address: a.rubesam@ieseg.fr (Alexandre Rubesam)

1. Introduction

The literature on stock return predictability and empirical asset pricing has identified hundreds of characteristics that appear to predict stock returns (Jacobs, 2015; Harvey et al., 2016; Green et al., 2017). Historically, most of these studies have relied on linear econometric techniques, sorts on firm characteristics, or rankings based on ad-hoc combinations of fundamental variables. As noted by Gu et al. (2020), these methods are not well-suited to deal with the high dimensionality of modern financial applications, or potential nonlinearities. While machine learning (ML) models have long been used in financial applications in the operations research and machine learning literature, the recent success of ML in many areas, including finance, has renewed interest in the topic within the finance community, including both academics and practitioners, stimulating new research at the intersection of finance and ML.

Recent studies have demonstrated the potential of ML models in various financial applications, including understanding the drivers of U.S. stock returns (Gu et al., 2020; Kelly et al., 2019; Kozak et al., 2020) and bond risk premia (Bianchi et al., 2020); forecasting cryptocurrencies (Akyildirim et al., 2020); and developing profitable statistical arbitrage strategies (Krauss et al., 2017; Fischer and Krauss, 2018; Huck, 2019). However, fewer studies have examined the use of ML to predict stock returns in emerging markets, and the existing studies often have limitations regarding the number and type of predictors used, as well as the ML models tested. Emerging markets tend to be extremely volatile (Bekaert and Harvey, 1997; Bekaert et al., 1998; Hwang and Satchell, 1999*a*), and therefore it is not reasonable to assume that conclusions from studies using U.S. data carry over to emerging markets. Moreover, the abundance of ML models creates a dilemma for portfolio managers seeking to leverage these technologies: which model(s) should be used? Should forecasts or portfolios obtained using different ML models be combined, and if so, how?

Given this context, the present paper provides three main contributions. The first one, of an empirical nature, is to investigate the use of ML to predict stock returns and create long-short investment strategies in an emerging market, namely the Brazilian equities market. We use a high-quality proprietary data set of 86 technical and fundamental predictors, provided by a local asset manager specialized in quantitative investing. This data set is comparable in scope to those used in studies with U.S. data, and includes many variables found to be relevant for return prediction by researchers, reflecting well-known empirical

regularities in the cross-section of returns. Using these data, we train a variety of linear and nonlinear ML models to predict individual stock returns. The models include linear regression with and without regularization (via LASSO and ridge), linear models with dimension reduction via principal components regression and partial least squares, linear models using Bayesian variable selection, and nonlinear models such as random forests, gradient boosting, and neural networks with different numbers of hidden layers. To our knowledge, this is the first systematic exploration of ML for investment management in an emerging market using a large number of technical and fundamental predictors. Our results suggest that most ML models are able to produce portfolios which outperform the market, even after accounting for exposures to common risk factors, suggesting that ML models are able to capture additional sources of average returns embedded in the predictors.

The second contribution of this paper is related to the questions of which ML model(s) to use, and whether and how to combine portfolios obtained with different ML models. In real life, the portfolio manager cannot know, *ex-ante*, which ML model will lead to portfolios with superior *ex-post* performance.¹ Two immediate possibilities emerge. The first one consists in combining the return forecasts from multiple ML models, and then constructing a portfolio based on the combined forecasts (the *predictive ensemble* approach). The second one entails combining portfolios obtained with different ML models (the *portfolio ensemble* approach). It is unclear whether either approach can dominate the other in practice. A predictive ensemble approach can reduce variance and reduce the risk of pre-selecting a specific ML model. On the other hand, the approach of combining multiple portfolios may provide diversification benefits, because different stocks will end up being selected in the long and short portfolios for each ML model. We investigate this empirically, by comparing portfolios obtained using these two approaches with different types of long-short portfolios. Our results suggest that there is no statistically significant difference between the two approaches when forecasts or portfolios are combined naively, i.e. using an equally-weighted approach.

¹In principle, one could expect that models with better out-of-sample accuracy would lead to portfolios with better performance. However, we cannot know *ex-ante* which model will produce lower out-of-sample errors, and trying to select models based on their past out-of-sample errors invites overfitting. Additionally, a lower out-of-sample error does not necessarily translate into better portfolio performance, because portfolio performance depends on transaction costs, and therefore, on portfolio turnover (see also Cenesizoglu and Timmermann, 2012). Therefore, this is an empirical question worth investigating.

The third contribution of this paper is motivated by our initial results with simple long-short portfolios constructed using forecasts from different ML models. Our results show that, while ML models are able to systematically identify stocks that outperform the local market, the portfolios of stocks with the lowest return forecasts obtained with most ML models still earn positive average returns and are extremely volatile, making traditional long-short strategies unattractive. A natural solution to deal with this issue is to employ an equal risk contribution (ERC) approach (e.g., Qian, 2005; Maillard et al., 2010) in the portfolio construction step, in order to balance the risk contributions of the long and short positions. Indeed, a simple application of the ERC approach to the portfolios in the highest and lowest quintiles of predicted returns obtained with each ML model leads to higher returns, while also significantly reducing volatility. This motivates the third contribution of this paper, which is of a methodological nature. We develop a multi-strategy Equal Risk Contribution (MS-ERC) approach that combines the long and short portfolios obtained with different ML forecasts. The objective is to construct a portfolio that invests in all ML long and short portfolios, such that (i) the risk contributions of all individual ML long-short strategies are equal, and (ii) the risk contribution of the overall long positions equals that of the short positions. This approach can be thought of as a forecast-free solution to the problem of deciding how much capital to allocate to individual long-short strategies, in which the portfolio manager assigns equal risk budgets to all strategies, with the additional requirement of maintaining balanced risk contributions between the long and short positions of the overall portfolio.

Our empirical results show that, in addition to attaining its objectives of equalizing risk contributions, the MS-ERC approach outperforms, on a risk-adjusted basis, all individual ML long-short strategies, as well as strategies based on forecasts from a simple equally-weighted ensemble of ML models (with or without the use of the ERC approach), and a multi-strategy approach that invests equally in all ML strategies. The MS-ERC approach also drastically reduces the volatility and maximum drawdown relative to these other approaches, while earning a higher average return and having lower exposure to factors such as size, value and momentum. The results remain economically and statistically significant after accounting for transaction costs. Our results suggest that the ERC approach is an attractive alternative to portfolio construction based on ML forecasts, which offers several advantages. First, it does not require choosing a specific class of models, providing a natural diversification against model risk. Second,

the formulation is intuitive, assigning equal risk budgets to different ML models. Finally, it allows the portfolio manager to achieve balanced risk contributions across long and short positions.

The remainder of this paper is organized as follows. Section 2 is a short literature review on stock return prediction and the recent use of ML in financial applications. Section 4.1 presents the data set used in the study. Section 3 explains the methodology used to train the various models used to forecast returns. Section 4 presents the empirical results. Our conclusions are presented in section 5.

2. Literature review

There have been many studies on stock return predictability in the finance and accounting literature, using a variety of methods and different sets of predictors. Historically, these studies have focused on linear models (e.g. Haugen and Baker, 1996; Campbell and Thompson, 2007; Lewellen, 2014; Green et al., 2017), procedures to sort stocks into portfolios based on firms' characteristics (e.g. studies following the approach introduced by Fama and French, 1993), or the creation of ad hoc measures combining different fundamental variables (e.g., Piotroski, 2000; Mohanram, 2005). The focus in this literature is often the identification of pervasive, priced risk factors, and the quantification of the corresponding risk premia. Although there is no consensus on the exact set of priced factors, researchers have identified hundreds of firm characteristics or "anomalies" that supposedly predict returns in a way that cannot be explained by risk exposures to widely accepted risk factors. There are now hundreds of such anomalies, see for example Jacobs (2015), Harvey et al. (2016), and Green et al. (2017).²

The approaches developed in the finance and accounting literature are not well-suited to deal with the large (and increasing) number of predictors, vast amounts of data, or the potential for nonlinear dynamics between returns and the predictors. ML methods, which have been developed precisely to deal with these issues, have long been used in financial applications. However, given financial economists' preference for linear econometric methods, much of this literature is published in operational research and machine learning journals, as highlighted by Huck (2019).³ Some examples of financial applications of ML in

²Despite the large number of predictors, recently published papers in top finance journals still rely on linear regression via ordinary least squares, see for example Green et al. (2017), although this is quickly changing.

³See also the surveys on the use of ML for prediction of financial time series by Atsalakis and Valavanis (2009); Henrique et al. (2019), and Sezer et al. (2020).

this literature include, among many others: prediction of changes in exchange rates (Sermpinis et al., 2013); return forecasting and pairs selection in pairs trading and statistical arbitrage strategies (Huck, 2009, 2010); development of trading systems using technical indicators and ML (Kaucic, 2010); and prediction of the direction of price of cryptocurrencies (Akyildirim et al., 2020), among many others.

Recently, there has been a surge of interest in ML within the finance and economics fields.⁴ Recent studies in empirical asset pricing, for example, focus on the application of ML techniques like regularization or dimension reduction via principal component analysis (PCA) to identify the most relevant drivers of asset returns (see for example Feng et al., 2020; Freyberger et al., 2020; Kelly et al., 2019; Kozak et al., 2020; Lettau and Pelger, 2020). Gu et al. (2020) examine a number of ML models to forecast monthly individual stock returns in the U.S. market using about 100 fundamental and technical features, and their interactions with macroeconomic variables. Their results suggest that nonlinear models such as neural networks can significantly improve predictions relative to simple linear regression approaches, and that long-short strategies based on several ML models appear to remain profitable in recent periods. Bianchi et al. (2020) find that ML models, particularly neural networks, provide strong evidence of bond return predictability.

Another fertile area of research is the application of ML models and concepts to portfolio formation. Ban et al. (2016) apply the concepts of regularization and cross-validation to portfolio optimization. DeMiguel et al. (2019) use a LASSO technique to select characteristics in a parametric portfolio problem. Heaton et al. (2017) explore deep learning to form portfolios, and provide an application of this approach to create portfolios to track or outperform an index. Kolm and Ritter (2019) provide an overview of applications of reinforcement learning in finance, including mean-reversion trading strategies, derivatives pricing and optimal hedging. De Spiegeleer et al. (2018) apply ML to various problems in quantitative finance including calculation of option prices and greeks (sensitivities of option prices to model inputs). Kyriakou et al. (2019) apply simple ML models to forecast annual stock returns to use as benchmarks for pension planning. Chen et al. (2019) develop a sparse-group LASSO methodology for portfolio

⁴See Varian (2014) and Mullainathan and Spiess (2017) for discussions on the role of big data and ML in the econometrics toolbox. Hsu et al. (2016) contrast studies using ML with those using regular (linear) econometric techniques, concluding that the best ML models are superior to the best econometric methods in terms of accuracy.

selection that allows investors to express preferences over equity sectors, and show its connection to robust portfolio selection (Kim et al., 2018). Li and Rossi (2020), DeMiguel et al. (2021), and Wu et al. (2020) use ML to select mutual funds and hedge funds.

Whereas, in asset pricing, the focus is in identifying a small set of priced risk factors that determine asset returns at lower frequencies (monthly or annual), practitioners are usually interested in exploiting ML to develop profitable trading strategies, which often operate at a higher frequency (measured from days to fractions of a second). Some recent studies suggest ML models can be used to build profitable long-short statistical arbitrage strategies, although the profitability seems to be decreasing or even negative in recent periods, consistent with arbitrageurs increasingly exploiting the market inefficiencies uncovered by these methods. Examples of these studies include Krauss et al. (2017), Fischer and Krauss (2018), and Huck (2019), who apply ML models including deep learning, gradient boosted trees, and random forests to forecast daily stock returns using different lags of individual stock returns.

The decline in the profitability of ML strategies in the U.S. market is in line with that reported by studies such as Green et al. (2017), who document a decrease in the profitability of a long-short strategy using linear regression forecasts since the early 2000s. The authors link this to changes in the regulatory and trading environments, which have made it cheaper and easier to implement quantitative long-short trading strategies exploiting a large number of signals. This is also consistent with the results of McLean and Pontiff (2016), who document significant out-of-sample and post-publication declines in the returns of predictors published in academic journals, suggesting arbitrageurs actively exploit new predictors as they become known.

The situation in emerging markets is less clear, due to the much smaller number of studies in these markets and the fact that most studies focus either on the prediction of stock market indices, or use a limited number of features. For example, out of the more than one hundred studies surveyed by Atsalakis and Valavanis (2009) and Hsu et al. (2016), only a handful investigate emerging markets. A similar situation is reported by other studies surveying the use of ML for financial market prediction (Sezer et al., 2020; Henrique et al., 2019). In an earlier work, Campbell (2000) investigated the use of neural networks to predict emerging market stock indices using lagged returns, concluding that an active strategy based on neural network forecasts beats a passive strategy or an active strategy based on linear regression

forecasts. A few studies have investigated the use of ML models to predict individual stock returns in emerging markets. Cao et al. (2005, 2011) use neural networks to forecast individual stock returns in China, however, the number of predictors is limited to the factors in the Fama and French (1993) models. Raposo and Cruz (2002) used fuzzy neural networks to predict individual stock returns in the Brazilian market using fundamental indicators. However, their data is limited to five fundamental indicators, and the analysis is focused on stocks belonging only to one sector. To our knowledge, ours is the first study to systematically compare multiple ML models to predict individual stock returns in an emerging market, using a large number of technical and fundamental predictors.

3. Methodology

3.1. Training, validation, and test windows

Our objective is to build regression models to forecast stock returns at time $t + 1$, based on the value of predictors at time t . We apply a sequential, expanding-window approach to divide the data into training, validation, and test sets, similar to the approaches used by Gu et al. (2020) and Bianchi et al. (2020). At each iteration, the training and validation samples are consecutive and not randomly selected, in order to preserve the time-series structure. We start with a training window of 24 months, a validation window of 12 months, and a test set of six months. We then expand the training window by six months, and move the validation and testing windows six months forward. This process is repeated until the end of the sample. The first two iterations of this sample-splitting strategy are shown in Figure 1. The training set is used to fit the models. The validation set is used for hyperparameter tuning within a class of models (for example, to choose the optimal penalty parameter in LASSO). We apply a grid-search approach to select hyperparameters. Specifically, at each iteration of the expanding-window approach, we calculate the mean squared error (MSE) of forecasts in the validation set, and choose the value of the hyperparameter(s) with the lowest MSE. Finally, out-of-sample predictions are made for observations in the test set.⁵

We use a pooled-data approach, stacking individual stock returns over all months of each training

⁵Table A.1 in the Appendix provides details of the ranges of hyperparameters for each method.

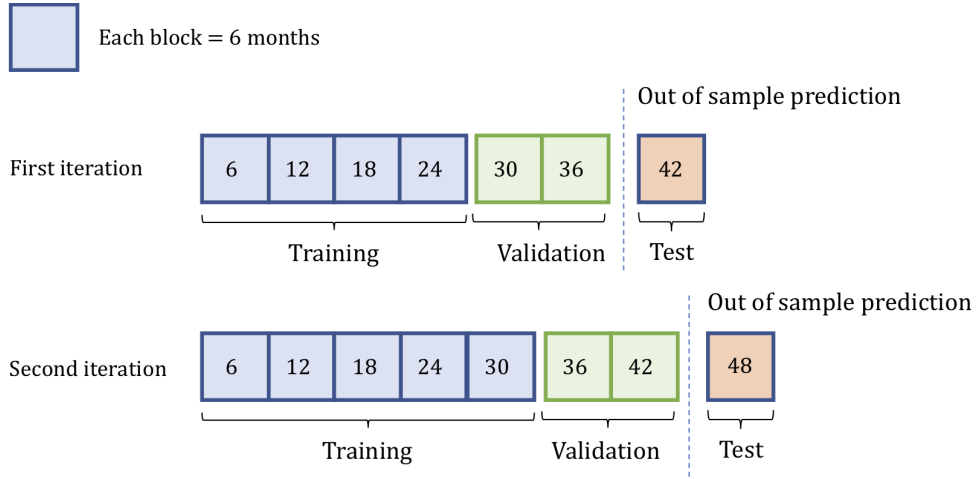


Figure 1: Schematic representation of the sample splitting strategy.

window into a single vector, and their predictors into a single matrix.⁶ To represent this more formally, let $y_{i,t+1}$ denote the return of stock i at month $t + 1$, n_t be the number of eligible stocks in month t , and let τ_k contain the indices corresponding to the months in the k -th training window. Pooling all returns yields a vector $\mathbf{y}_{Tr,k}$ of dimension $(\sum_{t \in \tau_k} n_t) \times 1$. Likewise, assume that there are $p_{Tr,k}$ predictors for which there are available data for the whole training window. Let x_{it} be the column vector of the $p_{Tr,k}$ predictors for stock i , at month t . Stacking all predictors produces a matrix $\mathbf{X}_{Tr,k}$ of size $(\sum_{t \in \tau_k} n_t) \times p_{Tr,k}$. This approach ensures that, even in the beginning of the sample period, thousands of observations are used to train the ML models. The regression models for the k -th training window thus have the following specification:

$$\mathbf{y}_{Tr,k} = f(\mathbf{X}_{Tr,k}) + \boldsymbol{\varepsilon}_{Tr,k}, \quad (1)$$

where $f(\cdot)$ represents a functional form and $\boldsymbol{\varepsilon}$ is an error term. We next describe specific choices for f .

3.2. Classes of models for f

We briefly review the types of models employed in our study. Most models are explained in standard textbooks such as Friedman et al. (2001). For simplicity, we drop the subscripts and use \mathbf{y} and \mathbf{X} for the

⁶We transform all variables to the $(-1, 1)$ interval at each month. This transformation removes outliers and mitigates the necessity of applying robust loss functions, simplifying the hyperparameter tuning. The details are explained in section 4.1.

response variable and matrix of predictors, respectively, assuming that $\mathbf{y} = (y_1, \dots, y_n)'$ has n elements, and \mathbf{X} is an $n \times p$ matrix.

3.2.1. Linear models via OLS

The usual regression model estimated via ordinary least squares (OLS) corresponds to $f(\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$. There are no hyperparameters.

3.2.2. Ridge regression and LASSO

Both ridge regression and LASSO shrink regression coefficients by imposing a penalty on their size, but ridge regression imposes an L_2 penalty, whereas LASSO uses an L_1 penalty. The coefficients are obtained as the solution to the following problem:

$$\hat{\boldsymbol{\beta}}^{\text{ridge/LASSO}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda J(\boldsymbol{\beta}) \right\}, \quad (2)$$

where the penalty term is $J(\boldsymbol{\beta}) = \sum_{j=1}^p \beta_j^2$ for ridge regression or $J(\boldsymbol{\beta}) = \sum_{j=1}^p |\beta_j|$ for LASSO. The model is estimated for a grid of values of the penalty parameter λ . The optimal parameter is the one that minimizes the MSE in the validation set.

3.2.3. Principal Components Regression (PCR)

Principal Components Regression (PCR) is a dimension reduction technique that uses linear transformations of the data (the principal components) as the predictors. Let Z_1, \dots, Z_M represent $M < p$ linear combinations of the original variables:

$$Z_m = \sum_{j=1}^p \phi_{jm} X_j, \quad m = 1, \dots, M. \quad (3)$$

A linear regression of \mathbf{y} on the transformed variables \mathbf{Z} can be represented as a linear combination of the original variables:

$$\begin{aligned}
y_i &= \theta_0 + \sum_{m=1}^M \theta_m z_{im}, i = 1, \dots, n \\
&= \theta_0 + \sum_{j=1}^p \beta_j x_{ij},
\end{aligned} \tag{4}$$

where $\beta_j = \sum_{m=1}^M \theta_m \phi_{mj}$. For PCR, the Z_m are the principal components of the data. The number of principal components (M) in the regression model (4) is chosen based on the validation set MSE.

3.2.4. Partial Least Squares (PLS)

Partial Least Squares (PLS), like PCR, considers linear combinations of the inputs for the regression, but it also makes use of the response variable \mathbf{y} . PLS starts by setting $\phi_{j1}, j = 1, \dots, p$ in equation (3) as the linear regression coefficient of \mathbf{y} onto each X_j . Once Z_1 , the first partial least squares direction, is obtained in this way, the first set of coefficients $\hat{\theta}_1$ is obtained as the regression coefficient of \mathbf{y} on Z_1 . Then, X_1, \dots, X_p are orthogonalized with respect to Z_1 . This process is continued until $M \leq p$ directions are obtained. The number of partial least squares dimensions (M) in the regression model (4) is chosen based on the validation set MSE.

3.2.5. Bayesian Variable Selection

There are many methods for Bayesian variable selection in regression models, see e.g. George and McCulloch (1997) and O'Hara and Sillanpää (2009). We focus on the method proposed by Smith and Kohn (1996) for the linear regression model, due to its simplicity and efficiency. Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$, with \mathbf{I} the identity matrix. The variable selection method works by introducing a vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)'$ of latent dummy variables, where $\gamma_j = 1$ if the j -th variable is included in the model, and zero otherwise. For a given value of $\boldsymbol{\gamma}$, let \mathbf{X}_γ represent the matrix of regressors corresponding to those elements of $\boldsymbol{\gamma}$ that are equal to one, and let $\boldsymbol{\beta}_\gamma$ contain the corresponding elements of $\boldsymbol{\beta}$. A hierarchical prior is assumed for $\boldsymbol{\beta}_\gamma | \boldsymbol{\gamma}, \sigma^2$, of the form $\boldsymbol{\beta}_\gamma | \boldsymbol{\gamma} \sim N(0, c\sigma^2(\mathbf{X}'_\gamma \mathbf{X}_\gamma)^{-1})$, where c is a

hyperparameter which Smith and Kohn (1996) suggest be set between 10 and 1000. The prior for σ^2 given γ is $p(\sigma^2|\gamma) \propto 1/\sigma^2$.

Given these choices, it is possible to simulate the relevant conditional posterior distributions using the Gibbs sampler. We consider three values for c : 100, 500, and 1000. For each value of c , we simulate the distribution 10,000 times, discarding the first 5,000 simulations as a “burn-in” period. The forecast from this method is obtained as the average of the forecasts across each simulated value of γ . We choose the value of c that minimizes the MSE in the validation set.

3.2.6. *Tree-based Ensemble Methods*

We apply two methods that make use of regression trees: random forests and boosting. A single regression tree works by partitioning the feature space into a set of distinct and non-overlapping rectangular regions, by sequentially creating splits in individual features. The predicted value of a regression tree within each of these rectangles is the average of the observations in that region. Trees are estimated using a recursive binary splitting algorithm which determines, at each step, a combination of a feature and a split point that minimize the forecast error at that stage. The complexity of a tree is thus a function of the number of splits and regions in the tree. Several strategies have been developed to decide when to stop growing a tree, or to prune a large tree in order to avoid overfitting, see for instance Breiman (1984) and Quinlan (1993).

Random Forests. Despite their conceptual simplicity and interpretability, trees are methods with high variance: small perturbations to the data usually lead to very different trees. Random Forests (Breiman, 2001) are an ensemble method that reduces variance and increases accuracy by combining many trees. In the regression setting, the predicted value is the average prediction across all trees in the ensemble. Each tree is trained on a bootstrapped sample of the training data using a randomly selected subset of predictors, in order to produce less correlated trees. Besides the parameters controlling the growth of individual trees, the hyperparameters of random forests are the number of bootstrap samples or trees and the number of variables to sample. Because random forests are quite robust to overfitting regarding the number of trees, we set the number of trees to a large value (1,000), and choose the number of variables to sample using the validation sample.

Boosting. The second tree-based method we use is boosting, which was originally developed for classification problems (Schapire, 1990), and relies on the idea of combining many weak classifiers (models whose error rates are just slightly better than random guessing), trained on sequentially modified versions of the data, to obtain a powerful ensemble with better performance. At each iteration of the algorithm, observations in the training set are multiplied by a weight, and a model (usually a tree) is fitted to the modified data. Observations that were misclassified in the previous iteration have their weights increased, while the opposite is true for correctly classified observations. Many boosting algorithms have been developed, see e.g. Friedman et al. (2000) and Freund and Schapire (1997). Friedman (2001) proposed a paradigm for function approximation based on additive expansions using steepest descent called gradient boosting, which can be used for classification and regression problems. We use the LS_Boost algorithm proposed in that paper for regression problems with a MSE criterion. The hyperparameters of the methods are the number of splits of the tree used in each iteration, the number of iterations, and a shrinkage or learning rate, which controls the contribution of each tree added to the ensemble. These parameters are chosen based on the MSE in the validation sample.

3.2.7. Neural Networks

Neural networks are a class of very flexible non-linear models that were developed in the artificial intelligence and statistics literature. We focus on feed-forward neural networks for regression problems.⁷ These models are typically represented using a network diagram, such as the one on Figure 2, which represents a multi-layer neural network with six inputs (x_1 to x_6), two hidden layers with five and three neurons, whose outputs are represented by $z_1^{(1)}$ to $z_5^{(1)}$ and $z_1^{(2)}$ to $z_3^{(2)}$, respectively, and a single output y . The output of a neuron of a given layer is obtained by applying an *activation function* g to a linear combination of the values reaching that neuron from the previous layer. A commonly used activation function for regression problems is the sigmoid function, $g(x) = 1/(1 + e^{-x})$. The complexity of a neural network is thus a function of the number of hidden layers, the number of neurons in each layer, and the activation functions used. It can be shown that neural networks are *universal approximators*: they can approximate arbitrarily well any continuous function if the number of neurons or layers is allowed to

⁷We provide only a brief overview of the type of neural network used in this study. More details on neural networks can be found, e.g. in Bishop (1995) and Ripley (1996).

increase.⁸

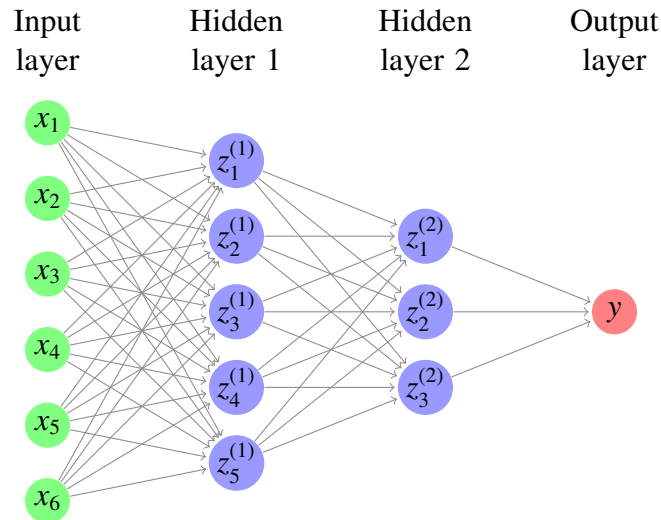


Figure 2: Diagram of a feedforward neural network with two hidden layers. The figure shows a neural network with two hidden layers, and a single output. The input layer has six inputs, labelled x_1 to x_6 . The first and second hidden layers have five and three neurons, respectively. The outputs from hidden layer neurons are represented by $z_j^{(l)}$, where $l \in 1, 2$ represents the hidden layer, and j identifies the neuron. They are calculated by applying an activation function to a linear combination of the preceding connections.

Consider a neural network with L hidden layers, where the number of neurons in layer l is $M^{(l)}$. Let $z_j^{(l)}$ be the output of neuron j of the hidden layer l . For the first hidden layer, these are simply the original inputs, i.e. $z_j^{(0)} = x_j, j = 1, \dots, p$. Assuming that $x_1 = 1$, a linear combination of the x s includes an intercept term, and so we can write $z_j^{(l)} = g(\boldsymbol{\beta}^{(l)T} \mathbf{z}^{(l-1)})$, where $\boldsymbol{\beta}^{(l-1)}$ is a vector of weights to be estimated, and $\mathbf{z}^{(l-1)} = (z_1^{(l-1)}, \dots, z_{M^{(l-1)}}^{(l-1)})$. Fitting a neural network consists in finding the weights $\boldsymbol{\beta}^{(l-1)}$ in each layer by minimizing a criterion such as MSE. This is done using a gradient descent method and what is known as the *backpropagation* algorithm. Because neural networks can be extremely flexible, regularization techniques such as adding a penalty term to the optimization criterion or early stopping are usually applied, to avoid overfitting. In this paper, we apply the latter, using the MSE in the validation sample to decide when to stop the training.

We consider neural networks similar to those in Gu et al. (2020), i.e. networks with one (NN1) to

⁸This has been proved by Cybenko (1989) for single-layer networks with sigmoid activation functions. See Lu et al. (2017) for deep networks.

five (NN5) hidden layers, with a pyramidal structure for the number of neurons in each hidden layer. The number of neurons in the first hidden layer is 32, and the number of neurons in each subsequent layer (if any) is half the number of neurons in the previous layer. Thus, NN1 has 32 neurons in its single hidden layer, NN2 has 32 neurons in the first layer and 16 in the second layer, and so on. We use a sigmoid activation function. In addition to early stopping, we also employ an ensemble approach. For each topology of neural network, 50 independent neural networks are trained, and their results are averaged to form a combined forecast. This approach reduces the variance of the individual models due to the fact that the weights are randomly initialized.

3.2.8. Ensembles of ML models

The use of ML models to forecast stock returns comes with the caveat that, in practice, it is not possible to know in advance which ML model will perform better in the future, as different modeling approaches may perform better under certain market conditions. One approach to attempt to improve the accuracy of the forecasts, or at least to reduce the risk of using a single model, is to create an ensemble that combines forecasts from multiple models. Ensembles are widely used in ML and are well justified for statistical and computational reasons (Dietterich, 2000; Polikar, 2012).⁹

Recent papers that apply ML to portfolio management have shown that combining forecasts from different types of ML model can be beneficial. For example, Huck (2009) combine return forecasts from Elman networks with a multi-decision criteria method to rank pairs in a pairs trading strategy. Krauss et al. (2017) aggregate probability forecasts from three ML models (deep neural networks, boosting, and random forests) to rank stocks in a statistical arbitrage strategy. Akyildirim et al. (2020) use an equally-weighted ensemble of four ML model to forecast cryptocurrencies.¹⁰ We adopt a similar approach and create a simple equally-weighted ensemble that averages the predictions from the 13 ML models. We refer to this ensemble method as ENS_{ML} .

⁹We note that some of the ML models we use are themselves based on ensembles. This is the case of random forests and boosting. We have also applied a simple averaging ensemble approach in the training of neural networks, as discussed previously.

¹⁰See Jiang et al. (2020) and references therein for more sophisticated ensemble approaches.

3.3. Portfolio construction and allocation

We consider the point of view of a portfolio manager seeking to apply ML to create a long-short investment strategy. We describe two approaches to create a long-short strategy using portfolios of stocks based on ML return forecasts. The first is a traditional long-short strategy, which buys stocks with the highest predicted returns and shorts stocks with the lowest predicted returns. The second is an Equal Risk Contribution (ERC) long-short strategy, which uses a variable multiplier to adjust the weights of the short positions in order to obtain equal risk contributions from long and short positions. We then explore approaches to create long-short strategies that combine portfolios obtained with all ML models, and propose an approach to do this in order to obtain equal risk contributions across all ML models, and across all long and short positions at the strategy level.

3.3.1. Baseline ML portfolios

For each ML model, we follow the common practice of grouping stocks into equally-weighted portfolios according to their predicted returns (e.g., Haugen and Baker, 1996; Lewellen, 2014; Green et al., 2017; Krauss et al., 2017; Gu et al., 2020).¹¹ Due to the significantly smaller number of stocks in the Brazilian market, compared with the U.S. market, we form portfolios by grouping stocks into quintiles based on their predicted returns.¹² The resulting portfolios, which we label P_1 (lowest quintile of predicted returns) to P_5 (highest quintile), form the basis of all the long-short strategies we consider.

3.3.2. Long-short strategies based on individual ML models

A traditional long-short strategy for a given ML model is obtained by going long the stocks in the top quintile of predicted returns (i.e., going long the P_5 portfolio) and short the stocks in the lower quintile (i.e., P_1). The gross excess return of this strategy is simply the difference between the return on the P_5 and P_1 portfolios. These long-short strategies have a net exposure of zero (i.e., they are cash neutral) and a gross leverage ratio of 2.

¹¹We note that, although we obtain different forecasts of individual stock returns, we do not attempt to solve a traditional mean-variance problem using these forecasts, due to the high level of noise in these estimates and the fact that mean-variance portfolio optimization is extremely sensitive to these inputs, i.e. Best and Grauer (1991); Michaud (1989).

¹²The results obtained using different percentiles to define the long and short portfolios are qualitatively similar.

If there are large differences between the volatilities of the P_1 and P_5 portfolios, a cash-neutral long-short strategy will likely have an undesirable risk-adjusted return profile. In this case, the portfolio manager might choose to reduce exposure to the portfolio with higher risk. We explore a formal way to achieve this through the ERC approach, also known as risk parity, in order to form a long-short strategy with balanced risk contributions from the long and short positions. ERC has emerged as a popular risk-based approach to build portfolios (see for example Qian, 2005; Maillard et al., 2010; Roncalli, 2016).¹³ Its underlying idea - creating portfolios with balanced risk contributions from different assets - is intuitive and appealing, and ERC portfolios have been shown to historically outperform (at least for long-only portfolios) mean-variance tangency portfolios, as well as other popular heuristic portfolio construction methods (Chaves et al., 2011).

In the long-short context, there has been little exploration of the ERC approach.¹⁴ In this subsection, we consider a simple setting in which there are two assets. In the next subsection, we will expand this to the case with multiple assets. The portfolio manager would like to form a long-short strategy which is long asset L and short asset S . Let $x_L > 0$ be the weight allocated to asset L , and let $x_S = -\kappa x_L$ be the weight in asset S , where κ is a positive number. The volatility of the resulting portfolio is $\sigma_P = \sqrt{x_L^2 \sigma_L^2 + x_S^2 \sigma_S^2 + 2x_L x_S \sigma_{LS}}$, where σ_L and σ_S are the volatilities of the two assets, and σ_{LS} is their covariance. The marginal risk contributions of assets L and S to the portfolio are defined as

$$MRC_L = \frac{\partial \sigma_P}{\partial x_L} = \frac{1}{\sigma_P} (x_L \sigma_L^2 + x_S \sigma_{LS}) \quad (6)$$

$$MRC_S = \frac{\partial \sigma_P}{\partial x_S} = \frac{1}{\sigma_P} (x_S \sigma_S^2 + x_L \sigma_{LS}) \quad (7)$$

The risk contributions of each asset are calculated by multiplying each weight by the corresponding

¹³See also Bertrand and Lapointe (2018) for an application of other risk-based strategies for portfolio construction using socially responsible investments.

¹⁴If short sales are allowed, but the assets to be shorted are not identified, multiple ERC solutions may exist, as shown by Bai et al. (2016). This is not the case in our applications, as the long and short portfolios are known in advance.

marginal risk contributions:

$$\mathcal{RC}_L = x_L \mathcal{MRC}_L = \frac{1}{\sigma_P} (x_L^2 \sigma_L^2 + x_L x_S \sigma_{LS}) \quad (8)$$

$$\mathcal{RC}_S = x_S \mathcal{MRC}_S = \frac{1}{\sigma_P} (x_S^2 \sigma_S^2 + x_L x_S \sigma_{LS}) \quad (9)$$

The portfolio volatility can be decomposed as the sum of the risk contributions: $\sigma_P = \mathcal{RC}_L + \mathcal{RC}_S$. An ERC portfolio requires that the two assets have the same risk contributions, i.e., $\mathcal{RC}_L = \mathcal{RC}_S$. Substituting $x_S = -\kappa x_L$ and equating the risk contributions leads to $\kappa = \sigma_L / \sigma_S$, i.e., we should short asset S in proportion to the ratio of the two volatilities. If asset L is more volatile than asset S , we will have $\kappa > 1$, resulting in a portfolio that is net short. Conversely, if asset S is more volatile, $\kappa < 1$ and the portfolio will be net long. We note that, similarly to the long-only risk parity problem with two assets (Maillard et al., 2010), the solution does not depend on the correlation between assets 1 and 2.

We refer to this as the “simple ERC” approach, and apply it to the P_1 and P_5 portfolios for each ML model. We assume that the portfolio manager will go long 100% of P_5 (i.e., $x_L = 1$) and short $-\kappa\%$ of P_1 , where κ is obtained as described above.

3.3.3. Long-short strategies using all ML models

As discussed previously, in general the portfolio manager cannot know in advance which ML model will lead to a superior portfolio. One way to mitigate this model uncertainty is to create an ensemble of different ML models, as explained in section 3.2.8. Another possibility is to create a strategy that combines the portfolios obtained with different ML models. Combining asset allocation strategies is a common approach to deal with the parameter uncertainty inherent in optimized portfolios (Kan and Zhou, 2007; Tu and Zhou, 2011; Peter et al., 2015; Bonaccolto and Paterlini, 2019). Even naïve diversification allocation strategies, such as the equally-weighted ($1/n$) portfolio, are surprisingly difficult to beat (DeMiguel et al., 2009). The two approaches are schematically represented in Figure 3 in the context of using ML models to create portfolios.

It is unclear whether either approach can dominate the other in practice. While a predictive ensemble approach can lead to better forecasts and reduce the risk of selecting a specific ML model, an approach of combining multiple portfolios may be advantageous due to potential diversification gains, since each

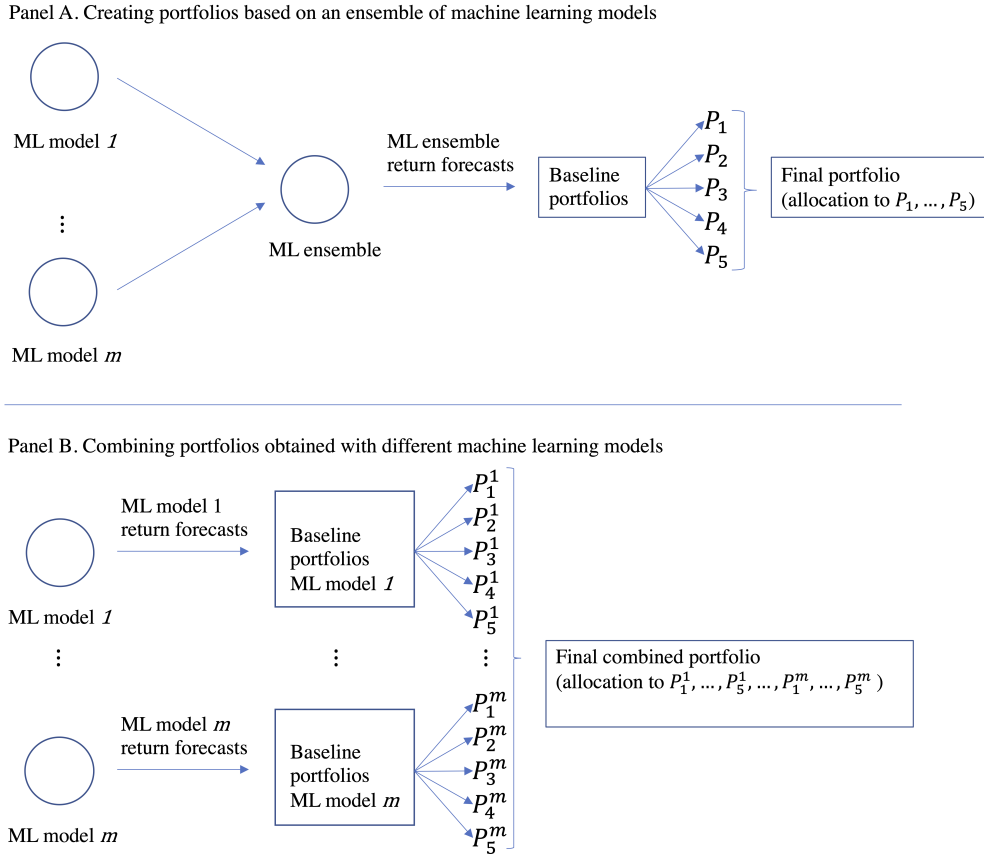


Figure 3: Schematic representation of two possible approaches to portfolio creation using machine learning models. In Panel A, an ensemble of machine learning models is initially created. Assets are assigned to different portfolios based on their return forecasts. A final portfolio is created by allocating to these portfolios. In Panel B, different portfolios are created for each machine learning model, and a final combined portfolio is obtained by allocating to these portfolios.

ML model may select different stocks in the P_5 and P_1 portfolios. To investigate this issue, we compare portfolios based on the two approaches.

Strategies based on an ensemble of ML models. First, we consider the long-short and simple ERC long-short strategies using the ML ensemble described in subsection 3.2.8. We refer to the resulting strategies as ENS_{ML} -EW and ENS_{ML} -ERC.

Strategies that combine portfolios obtained with different ML models. We consider two approaches to combine the P_5 and P_1 portfolios obtained with all the ML models. The first approach is an equally-weighted average of the quintile long-short strategies. We refer to this as the Multi-Strategy Equally-

Weighted (MS-EW) approach. This approach is conceptually identical to the metastrategy used by Gu et al. (2020). The second approach, which is the main methodological contribution of this paper, is a novel application of the ERC principle to the set of long-short strategies obtained with the different ML models. Specifically, we propose an approach to combine the individual long-short strategies such that (i) the risk contributions of all individual long-short strategies are equal, and (ii) the overall risk contributions of the long positions equals that of the short positions. We refer to this as the Multi-Strategy Equal Risk Contribution (MS-ERC) approach. We can think of this approach by considering a portfolio manager who is allocating capital to m long-short strategies. The allocations are decided by giving each long-short strategy the same risk budget, while simultaneously requiring that the risk contributions of the long and short positions be balanced.

In order to explain the MS-ERC approach, we introduce some notation. Let \mathbf{r}_t be a $2m \times 1$ vector of returns, such that the first (last) m columns contain the returns of the long (short) portfolios obtained with each ML model at time t , and let Σ_t be its covariance matrix at time t . We drop time subscripts to simplify the notation in what follows. A long-short combination of the $2m$ portfolios can be represented by the $2m \times 1$ vector of weights $\mathbf{x} = (x_1^+, \dots, x_m^+, x_1^-, \dots, x_m^-)'$. The first m elements of \mathbf{x} are positive, while the remaining m elements are negative. We can think of the combination as a portfolio of pair trades, where each pair trade is long the P_5 portfolio and short the corresponding P_1 portfolio for a given ML model. The return on the combination portfolio is $r_C = \mathbf{x}'\mathbf{r}$, and its volatility is given by $\sigma_C = \sqrt{\mathbf{x}'\Sigma\mathbf{x}}$.

We consider forming a long-short strategy with the following characteristics. First, the total long allocation is 100%, i.e. $\sum_{j=1}^m x_j^+ = 1$. Second, the allocation to the short leg (i.e. the P_1 portfolio) obtained with each ML model is a fraction κ of the long allocation: $x_j^- = -\kappa x_j^+$. As a result, the overall short allocation is equal to κ .¹⁵ Third, let RC_j^+ and RC_j^- be the risk contributions of the long and short

¹⁵In principle, it could be possible to find a solution that does not impose this restriction. We attempted such a solution in this study and found that, in general, it is not feasible, due to the high volatilities of the P_1 portfolios.

legs of the ML long-short strategy j to the combined portfolio. These are calculated as

$$RC_j^L = x_j^+ \frac{\partial \sigma_C}{\partial x_j^+} = x_j^+ \frac{(\Sigma \mathbf{x})_j}{\sqrt{\mathbf{x}' \Sigma \mathbf{x}}}, \quad (10)$$

$$RC_j^S = x_j^- \frac{\partial \sigma_C}{\partial x_j^-} = x_j^- \frac{(\Sigma \mathbf{x})_j}{\sqrt{\mathbf{x}' \Sigma \mathbf{x}}}. \quad (11)$$

The total risk contribution of the long-short strategy associated with ML model j is $RC_j^{LS} = RC_j^+ + RC_j^-$, and the total risk contribution of all long and short positions are $RC^+ = \sum_{j=1}^m RC_j^+$ and $RC^- = \sum_{j=1}^m RC_j^-$, respectively. We then impose the following conditions:

$$RC_j^{LS} = RC_{j'}^{LS}, \quad j, j' \in \{1, \dots, m\} \quad (12)$$

$$RC^+ = RC^-. \quad (13)$$

Condition (12) states that the contribution of each long-short strategy is equal for all ML long-short strategies. Condition (13) states that the overall risk contributions of all long and short positions are the same. In order to solve this non-standard ERC problem, we define the following quantities. Let $\boldsymbol{\eta} = (x_1^+, \dots, x_{m,t}^+, \kappa)'$ denote a vector containing the weights of the long portfolios and the short multiplier κ . Since the weights on the short portfolios are defined as $x_j^- = -\kappa x_j^+$, the vector $\boldsymbol{\eta}$ determines the full allocation to the long and short portfolios. We collect the risk contributions in a $(m+2) \times 1$ vector $\mathbf{RC} = (RC_1^{LS}, \dots, RC_m^{LS}, RC^+, RC^-)'$. Next, we define the risk budgets as follows. The risk budget for each long-short strategy is $b_j^{LS} = 1/m$, $j = 1, \dots, m$. The risk budget for all long positions is equal to the risk budget of all short positions: $b^+ = b^- = 0.5$. The risk budgets are collected in a $(m+2) \times 1$ vector $\mathbf{b} = (b_1^{LS}, b_2^{LS}, \dots, b_m^{LS}, b^+, b^-)'$. We then solve the following optimization problem:

$$\begin{aligned} & \underset{\boldsymbol{\eta}}{\text{minimize}} && f(\boldsymbol{\eta}, \mathbf{b}) \\ & \text{subject to} && \sum_{s=1}^m x_s^+ = 1 \\ & && 0 \leq x_j^+ \leq 1, \quad s = 1, \dots, m \end{aligned} \quad (14)$$

where

$$f(\boldsymbol{\eta}, \mathbf{b}) = \sum_{i=1}^m \sum_{j=1}^m \left(\frac{RC_i^{LS}}{b_i^{LS}} - \frac{RC_j^{LS}}{b_j^{LS}} \right)^2 + \left(\frac{RC^+}{b^+} - \frac{RC^-}{b^-} \right)^2. \quad (15)$$

The function $f(\boldsymbol{\eta}, \mathbf{b})$ is minimized when conditions (12) and (13) are met. In the empirical part of the paper, we allow the short multiplier κ to vary between 0 and 2, a range which allows for higher or lower weights in the short portfolios, relative to the long portfolios. In general, if the volatilities of the short portfolios are substantially higher than those of the long portfolios, we should expect $\kappa < 1$, which is what we observe empirically.

3.3.4. Portfolio calculations

Let N be the total number of stocks in the universe, T be the total number of months, and w_{it}^P denote the weights of a generic portfolio P . We calculate the average monthly turnover of a portfolio P over T months as

$$\text{Turnover}^P = \frac{1}{T} \sum_{t=2}^T \sum_{i=1}^N |w_{it}^P - w_{i,t-1}^P|. \quad (16)$$

The turnover in each month is used to estimate the transaction costs. Higher turnovers imply higher transaction costs and therefore negatively impact the net results of a given portfolio or strategy. To account for transaction costs, we consider a fixed cost of 15 basis points (bps) for all trades, implying a bid-ask spread of 0.30%.¹⁶ We also include a fixed annual borrowing cost of 4.5%, applied on the total amount of short positions for long-short strategies.

Because the ERC portfolios have variable leverage, we calculate the average leverage ratio of a long-short strategy as

$$\text{Leverage}^P = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N |w_{it}^P|. \quad (17)$$

Note that for traditional long-short strategies, the leverage ratio is constant and equal to 2.

¹⁶This bid-ask spread is about twice the post-decimalization average quoted bid-ask spread in the U.S. reported by Bessembinder (2003) for large stocks. Author's discussions with asset managers and data providers operating in the Brazilian market have suggest this is a conservative estimate for the universe of liquid stocks considered in this paper. We discuss results with more conservative transaction costs in section 4.5.

4. Empirical Results

4.1. Data

Our data set includes 572 Brazilian stocks over the period from January 2003 to December 2018. In order to have a set of reasonably tradable stocks at each period, we impose a few minimum requirements. First, a minimum of one year of trading data and two years of accounting data is required. Second, a minimum liquidity requirement is enforced, based on the average daily traded volume of each stock in the local currency, to eliminate stocks which are too small or illiquid to use in a realistic trading strategy.¹⁷ These restrictions are applied at each month, and thus result in a variable set of eligible stocks.

Features. For each stock and each month, we have data on 62 features, which include fundamental indicators (i.e. derived from financial statement data) as well as technical indicators (i.e. constructed from market data). We also add 24 dummy variables to represent the firms' sectors, bringing the total number of variables to 86. The features are listed in Table 1. Due to the proprietary nature of the data, the names of some of the features are redacted (e.g. "Quality indicator 1"), and the exact definitions of all the features are not disclosed. We recognize that the lack of precise definitions of some of the features may raise questions regarding the replicability of the study. However, we believe that this is compensated by a few factors. First, we note that, even though we do not provide precise definitions of all the features, they are all constructed from public sources (i.e., market or financial statement data). Thus, it is reasonable to expect that other researchers working with the same publicly available data would obtain similar results in terms of the relative performances of the ML portfolios.¹⁸ The second factor is the exceptional quality of this proprietary data set, which has undergone rigorous quality checks and is free of survivorship biases. Finally, the fact that the data are actually used by the asset manager for the purposes of developing and running quantitative investment strategies suggests that our work offers a unique insight into the potential benefits of using ML for portfolio management in this emerging market. Despite the opaqueness in the description of the data, generally speaking, the features are similar to the set of variables used in large-scale stock predictability studies. Indeed, many are identical to those used by Green et al. (2017), and

¹⁷At each month, stocks whose 10-day average daily volume are inferior to R\$500,000 are excluded from the set of eligible stocks.

¹⁸We thank the Subject Editor for this comment.

reflect characteristics that have been identified in the empirical asset pricing and accounting literature as important drivers of returns.¹⁹ One important difference is that our data set does not include data based on analysts' forecasts, as these are not widely available for most companies in the Brazilian market. We also do not use interactions with macroeconomic variables, as in Gu et al. (2020), because of the large increase in the dimensionality of the feature space and the much smaller sample in terms of the number of stocks and the length of the sample period.

It is important to note that some of the predictors rely on fundamental data, obtained through the companies' financial statements. These data are updated on quarterly basis, whereas market data is available on a daily basis. Consequently, predictors that rely only on market data (those listed as technical indicators in Table 1) are updated every month using the most recent data, while predictors that rely exclusively on fundamental data (for example, EBIT growth) are updated on a quarterly frequency. Predictors that rely on both types of data (e.g. multiples such as the price-to-book and price-to-earnings ratios) are updated every month based on the available market data, and every quarter based on the available fundamental data. We note that other studies, such as Gu et al. (2020), have used a similar approach, and have also investigated the sensitivity of their results to the frequency of the variables by using annual (instead of monthly) returns in the regressions. Their results suggest that the same variables tend to be significant, which provides support for using the fundamental-based predictors, despite the lower update frequency.

Data cleaning/pre-processing. In order to mitigate the effect of outliers and deal with missing data, we apply a transformation to all features when creating the stacked matrix described in section 3.1. Each month, for each feature, we cross-sectionally rank all eligible stocks according to the value of the feature, and calculate a normalized value using the cross-sectional rank of that feature minus the cross-sectional average rank of that feature, as in Asness et al. (2013). Formally, let $x_{t,i,j}$ denote the value of feature j for stock i in month t . Then the value of the transformed feature is given by the expression below, where $c_{i,t}$ is chosen to scale the feature such that the sum of all positive values equals one, and the sum of all

¹⁹Please refer to the Appendix in Green et al. (2017) for references.

Table 1 Summary of features used in this study

Fundamental indicators	Technical indicators
Accruals	Downside volatility 1
CAPEX to sales	Downside volatility 2
CAPEX to total assets	Downside volatility 3
Capitalization rate	Momentum indicator 1
Change in inventory to total assets	Momentum indicator 2
Change in leverage	Momentum indicator 3
Change in liquidity	Momentum indicator 4
Coverage ratio	Money flow
Dividend payout ratio	Technical indicator 1
Dividend yield	Technical indicator 2
Dividend yield stability	Technical indicator 3
EBIT growth	Technical indicator 4
EBIT to total assets	Technical indicator 5
EBIT volatility	Volatility indicator 1
EPS growth	Volatility indicator 2
Earnings volatility	Volatility indicator 3
Growth indicator 1	
Growth indicator 2	
Growth indicator 3	
Growth indicator 4	
Leverage	
Liquidity	
Net debt to EBITDA	
Price to book	
Price to earnings	
Price to earnings growth	
Quality indicator 1	
Quality indicator 2	
Quality indicator 3	
Quality indicator 4	
Quality indicator 5	
Quality indicator 6	
Quality indicator 7	
Quality indicator 8	
ROE	
ROIC	
Sales growth	
Sales volatility	
Sector	

EBIT = Earnings Before Interest and Taxes; EBITDA = Earnings Before Taxes, Interest, Taxes, Depreciation and Amortization;
 ROE = Return On Equity; ROIC = Return On Invested Capital;
 EPS = Earnings Per Share; CAPEX = Capital Expenditure.

negative values equals minus 1.

$$x_{t,i,j}^* = c_{i,t} \left(\text{rank}(x_{t,i,j}) - \sum_i \text{rank}(x_{t,i,j})/n_t \right). \quad (18)$$

This normalization mitigates the effect of outliers and also scales all features to have zero mean. We replace missing observations with the mean value of zero.²⁰

Table 2 reports summary statistics for the stocks which fulfill the eligibility criteria. The average number of eligible stocks per month varies over time from 54 in 2003 to 189 in 2018. Over the period from January 2003 to December 2018, an equally-weighted portfolio of the eligible stocks in the universe earned on average 1.63% per month, with a standard deviation of 5.56%. For comparison purposes, we also report these statistics for the IBOVESPA index, a widely used gauge of the Brazilian equity market, as well as the Brazilian interbank certificate of deposit rate, or CDI, which is the risk-free benchmark rate in Brazil.²¹ The IBOVESPA earned on average 1.29% per month, with a standard deviation of 6.09%. The average cross-sectional volatility of the eligible stocks is 11.71%, suggesting that individual stocks in the Brazilian market are much more volatile than those in developed markets. For example, Hou et al. (2020) report a cross-sectional standard deviation of 10.52% for the entire U.S. market, excluding micro-cap stocks, over the period from January 1967 to December 2016. This feature of emerging markets has been reported in several studies (Bekaert and Harvey, 1997; Bekaert et al., 1998; Hwang and Satchell, 1999b).

It is worth noting that, although the average returns of the equally weighted and IBOVESPA benchmarks seem high compared to developed markets, the standard deviations are also high. In addition, the Brazilian risk-free rate has been historically high, particularly during the earlier part of the sample, peaking at over 26% per year in early 2003. The average monthly CDI rate over the 2003-2018 period was 0.99% per month, which is over 12.54% on an annual basis). As a result, the annualized Sharpe ratios of the equally weighted and IBOVESPA benchmarks over this period are relatively low: only 0.31 and 0.20, respectively, over the 2003-2018 period, and even lower in the 2006-2018 period.

²⁰We have also trained models using winsorized versions of the features; the results were very similar.

²¹The IBOVESPA is a total return, value-weighted index that tracks the performance of the more actively traded stocks in the Brazilian equity market.

Table 2 Average monthly returns, average monthly standard deviations, and cross-sectional standard deviations of eligible stocks in the investment universe. The table reports statistics for the universe of eligible stocks used in the paper, as well as the IBOVESPA index and the risk-free rate of return (the Brazilian interbank certificate of deposit rate, or CDI). All numbers are on a monthly basis, except Sharpe ratios, which are annualized. The table reports average returns (Ave), standard deviations (Std), and Sharpe ratios (SR) for each period.

Year	Eligible stocks used in this study						
	Number of firms	Equal weighted returns		Cross-sectional	IBOVESPA		Risk-free
		Ave	Std	Std	Ave	Std	Ave
2003	54	6.88%	6.03%	12.85%	6.03%	6.65%	1.79%
2004	67	3.26%	4.79%	10.16%	1.51%	5.37%	1.25%
2005	75	1.67%	6.16%	11.14%	2.31%	7.46%	1.46%
2006	87	3.73%	5.36%	10.31%	2.57%	6.19%	1.22%
2007	124	2.93%	4.78%	11.29%	3.15%	4.33%	0.95%
2008	163	-4.30%	9.47%	12.49%	-3.85%	9.89%	0.95%
2009	178	7.30%	7.75%	12.74%	5.29%	5.79%	0.82%
2010	198	1.60%	3.86%	10.14%	0.22%	5.46%	0.76%
2011	198	-1.27%	3.68%	10.45%	-1.54%	4.89%	0.92%
2012	192	1.53%	4.08%	10.51%	0.75%	5.84%	0.71%
2013	186	-0.92%	3.25%	10.54%	-1.30%	4.42%	0.63%
2014	187	-1.46%	4.38%	10.75%	-0.05%	6.47%	0.84%
2015	179	-1.69%	4.88%	13.84%	-1.03%	5.90%	1.02%
2016	172	2.71%	8.32%	15.99%	3.08%	8.20%	1.10%
2017	179	3.22%	5.09%	11.72%	2.08%	4.11%	0.84%
2018	189	0.84%	7.12%	12.45%	1.36%	6.48%	0.52%
Ave (2003-2018)	152	1.63%	5.56%	11.71%	1.29%	6.09%	0.99%
Ave (2006-2018)	172	1.09%	5.54%	11.79%	0.83%	6.00%	0.87%
SR (2003-2018)			0.31			0.20	
SR (2006-2018)			0.07			-0.03	

4.2. Performance of baseline ML portfolios

Table 3 reports return statistics for the P_1 , P_5 , and $P_5 - P_1$ baseline ML portfolios.²² Since the first training window comprises the first two years, and we use a one-year validation window, the results comprise the period from January 2006 to December 2018. All numbers are on a monthly basis, except for Sharpe ratios, which are annualized. For each method, we report average gross and net (i.e. after costs) returns, the standard deviation of returns, gross and net Sharpe ratios, maximum drawdowns, and the average portfolio turnover.²³ The table also reports the out-of-sample (OOS) R^2 for each ML model, calculated as in Gu et al. (2020). We report the number of stocks in each quintile portfolio at each month in Figure A.1 in the Appendix. We test the hypothesis of equality between the Sharpe ratios of each portfolio and that of the IBOVESPA index using the Ledoit and Wolf (2008) test. The results are shown in Table A.3 and in subsequent tables using the usual one, two and three stars convention for p-values below 0.10, 0.05, and 0.01, respectively.

As shown on Table A.3 in the Appendix, returns increase monotonically across quintiles for all ML models, indicating that, at the quintile portfolio level, the ML return forecasts align with future realized returns. The portfolios formed on the highest quintile of return forecasts (P_5) deliver gross average monthly returns that range from 1.61% (for the NN1 model) to 2.13% (for the NN4 model). On a net basis, i.e. after transaction costs, these returns range from 1.47% (NN1) to 2.02% (NN5). The models that deliver the highest P_5 returns are the neural networks with four and three hidden layers (gross monthly average returns of 2.13% and 1.94%, respectively).²⁴ As shown on the last row of Table 2, over the same period an equally-weighted portfolio of all the eligible stocks achieved a monthly average return of 1.09%, while the IBOVESPA index earned 0.83% per month. Therefore, at first glance, all ML models seem to be successful at producing portfolios that outperform the market, out of sample and net

²²Statistics for all the quintile ML portfolios are reported in Table A.3 in the Appendix.

²³To calculate Sharpe ratios, we use as the risk-free rate the Brazilian interbank certificate of deposit rate, or CDI, which represents the average rate of all interbank overnight transactions in Brazil. When calculating Sharpe ratios, we assume the investor deploys any excess cash in an investment that yields the risk-free rate.

²⁴We note that there is only a weak relationship between the statistical accuracy of the models, measure by the OOS R^2 , and the performance of portfolios. For example, while LASSO and Ridge both have much higher OOS R^2 compared to OLS, the net SR of the $P_5 - P_1$ portfolios are similar. As shown for example by Cenesizoglu and Timmermann (2012), statistical performance may be a misleading indicator of economic performance.

Table 3 Performance of Baseline Machine Learning Portfolios. The table reports out-of-sample performance metrics for equally-weighted quintile portfolios formed on predicted returns using machine learning methods. P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of predicted returns. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports average monthly return before (Ave) and after costs (Ave (net)), the monthly standard deviation (Std), the annualized Sharpe Ratio before (SR) and after costs (SR (net)), the maximum drawdown (Max. DD), the monthly turnover. The table also reports, for each machine learning method, the out-of-sample (OOS) R^2 . One, two and three asterisks denote a statistically significant difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period at the 10%, 5% and 1% significance level, respectively, using the Ledoit and Wolf (2008) test.

	OLS (OOS $R^2 = -0.0338$)			LASSO (OOS $R^2 = 0.0079$)			Ridge (OOS $R^2 = 0.0079$)			PLS (OOS $R^2 = 0.0017$)		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.30	1.91	1.61	0.22	1.94	1.72	0.07	1.92	1.86	0.03	1.92	1.90
Ave (net)	0.17	1.77	0.97	0.13	1.84	1.16	-0.02	1.83	1.32	-0.05	1.84	1.38
Std	8.65	6.33	5.17	9.32	5.99	6.23	8.85	6.07	5.80	9.77	5.13	7.18
SR	-0.22	0.57**	1.08*	-0.24	0.63**	0.96	-0.31	0.61**	1.11	-0.30	0.72**	0.91
SR (net)	-0.28	0.50**	0.65	-0.27	0.57*	0.65	-0.34	0.55*	0.79	-0.32	0.66**	0.66
Max.DD	82.87	50.53	53.83	85.09	51.97	65.67	84.94	52.13	54.65	88.08	42.31	62.91
Turnover	86.24	91.57	177.81	59.37	66.35	125.72	57.04	63.30	120.35	48.54	53.78	102.32
	PCR (OOS $R^2 = 0.0049$)			Bayes (OOS $R^2 = -0.0010$)			Boost (OOS $R^2 = 0.0075$)			RF (OOS $R^2 = -0.0030$)		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.26	1.69	1.43	0.16	1.84	1.68	0.25	1.85	1.60	0.15	1.86	1.71
Ave (net)	0.19	1.61	0.90	0.04	1.71	1.06	0.16	1.73	1.02	0.02	1.72	1.07
Std	9.75	4.70	7.21	8.96	5.87	5.92	9.30	5.51	6.58	9.29	5.84	6.10
SR	-0.21	0.61**	0.69	-0.27	0.58**	0.98	-0.23	0.62**	0.84	-0.27	0.59**	0.97
SR (net)	-0.24	0.55**	0.43	-0.32	0.50*	0.62	-0.26	0.55**	0.54	-0.31	0.51**	0.61
Max.DD	88.51	39.76	66.59	84.68	51.14	62.80	90.64	40.64	67.78	88.85	46.65	58.29
Turnover	51.50	54.34	105.84	79.14	86.91	166.06	62.38	74.78	137.16	85.21	94.92	180.13
	NN1 (OOS $R^2 = -0.0802$)			NN2 (OOS $R^2 = -0.0180$)			NN3 (OOS $R^2 = 0.0033$)			NN4 (OOS $R^2 = 0.0092$)		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.28	1.61	1.33	0.17	1.67	1.49	-0.01	1.94	1.95	0.07	2.13	2.06
Ave (net)	0.15	1.47	0.69	0.05	1.52	0.86	-0.12	1.81	1.33	-0.03	2.02	1.48
Std	7.92	6.42	4.97	8.45	6.35	5.29	8.59	6.08	5.51	9.21	5.79	5.96
SR	-0.25	0.40*	0.93*	-0.28	0.44**	0.98	-0.35	0.61**	1.22*	-0.30	0.76**	1.20*
SR (net)	-0.31	0.33	0.48	-0.33	0.36*	0.56	-0.40	0.54**	0.84	-0.34	0.69**	0.86
Max.DD	83.45	55.14	39.77	85.13	50.66	52.43	87.73	49.04	55.28	87.78	48.07	55.61
Turnover	89.28	93.85	183.13	84.41	94.14	178.55	77.01	86.12	163.13	66.94	76.47	143.41
	NN5 (OOS $R^2 = 0.0093$)			ENS_{ML} (OOS $R^2 = 0.0096$)								
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$						
Ave	0.17	1.89	1.72	0.02	2.02	2.00						
Ave (net)	0.07	1.78	1.15	-0.08	1.90	1.41						
Std	9.50	5.46	6.38	9.21	5.76	6.11						
SR	-0.25	0.65**	0.93	-0.32	0.69**	1.13*						
SR (net)	-0.29	0.58**	0.63	-0.36	0.62**	0.80						
Max.DD	87.82	45.94	62.42	89.79	47.54	59.00						
Turnover	60.64	70.70	131.34	69.39	79.29	148.68						

of transaction costs.²⁵ We also note that most P_5 portfolios produce significantly higher Sharpe ratios relative to the IBOVESPA, on both a gross and net basis. In order to build a long-short strategy, it would be desirable to be able to predict which stocks will have negative returns in the future. However, with the exception of the NN3 model, all P_1 portfolios earn low, but positive returns, and have significantly higher volatility than the corresponding P_5 portfolios (the average volatility of the P_1 portfolios is 57% higher than that of P_5 portfolios). This limits the potential profitability of traditional long-short strategies, especially on a risk-adjusted basis.

The average turnovers are generally high, but vary significantly across ML models. For example, for the $P_5 - P_1$ portfolios, the average turnover varies from about 102% for the PLS method to 183% for the NN1 method. These numbers are higher than those reported by Gu et al. (2020) for the U.S. market. Notwithstanding the high turnovers, most of the transaction costs of the $P_5 - P_1$ strategies come from the borrowing costs of the short positions in the P_1 portfolio. While each 10% of turnover adds 1.5 basis points of transaction cost per month, the borrowing cost is fixed and amounts to 37.5 basis points each month. So, in addition to being unprofitable on average, the short positions are very costly.

All portfolios have very high maximum drawdowns, typically higher than 50% and much higher for the P_1 portfolios. The maximum drawdown of the $P_5 - P_1$ strategies range from approximately 40% for the NN1 model to approximately 68% for the boosting model. In terms of risk-adjusted returns, the best performers are the neural network models with four and three hidden layers. The net Sharpe ratio (SR) of the $P_5 - P_1$ strategies using these models are 0.86 and 0.84, respectively. The next best method is ridge regression, with a net SR of 0.79. The other nonlinear methods tested in this study, boosting and random forests, do not seem to be very competitive, slightly underperforming most linear methods after transaction costs are taken into account. Most of the $P_5 - P_1$ portfolios do not produce significantly higher Sharpe ratios compared to the IBOVESPA, and none do when the net Sharpe ratio is considered.

Table 3 also reports results for the equally-weighted ensemble that averages the forecasts from all ML models (ENS_{ML}). The ensemble achieves results which are close to those obtained with the best meth-

²⁵In Table A.2 in the Appendix, we report results for equally-weighted portfolios formed using the individual predictors. The highest Sharpe ratio obtained for individual long-short portfolios is close to 0.64, or about half that of the more competitive ML portfolios that use all the predictors, which confirms that ML models are able to improve by combining the individual signals.

ods. The $P_5 - P_1$ long-short strategy using the ensemble achieves a gross (net) average return of 2.00% (1.41%), which is inferior only to the neural network with four layers. Its net SR of 0.80 is comparable to those of the long-short strategies using the NN4 and NN3 models (0.86 and 0.84, respectively), but it is also not statistically different from that of the IBOVESPA index. Therefore, at a first glance, the predictive ensemble approach appears to be beneficial.

A relevant question is which variables are useful to predict stock returns. In order to investigate this, we calculate, for each ML model and each feature, a measure of variable importance, calculated as the reduction in the predictive (out-of-sample) R^2 from setting all values of that feature to zero, while holding the remaining model estimates fixed, as in Gu et al. (2020) and Kelly et al. (2019). We aggregate these measures across the whole test period. Figure 4 summarizes this variable importance measure across all models.²⁶ The features are shown in order from top (most influential) to bottom (least influential) according to their overall contribution, measured by the sum of the ranks of the corresponding variable importance measures across all models. Columns show individual ML models, with color gradients within each column indicating the most influential (dark blue) to the least influential (white) features. Of the top ten most relevant predictors across all models, four are fundamental (Quality Indicator 1, dividend yield, Growth indicator 1 and Growth indicator 2) and six are technical (related to momentum (Jegadeesh and Titman, 1993) and volatility (Ang et al., 2006)). In contrast, Gu et al. (2020) show only one fundamental variable in the top ten predictors (ratio of sales to market capitalization), with the most important variables being of a technical nature. Although direct comparisons are challenging due to differences in sample periods and the set of variables used, these results suggest that fundamental indicators play a more prominent role in the prediction of stock returns in the Brazilian equity market, in comparison with a much more mature market, the United States. This is sensible; the higher volatility observed in emerging markets would suggest that investors need to pay closer attention to fundamentals. Interestingly, the sector dummy variables seem to be extremely relevant for neural networks, in comparison with other ML models. This suggests that neural networks are better able to incorporate interactions of the sector dummies with other variables. The figure also shows that neural networks tend to make use of

²⁶Figure A.2 in the Appendix provides the normalized variable importance measures for the 15 most relevant variables within each model.

more variables, similarly to the result found by Gu et al. (2020) for the U.S. market.

Our results bear some similarities with those of Gu et al. (2020), which is perhaps the closest paper in terms of the ML models explored and the type and frequency of the data used. Like them, we also find that neural networks are the best performers among ML models, with performance peaking at three to four hidden layers. However, in our application, other nonlinear ML models, such as boosting and random forests, are not very competitive. It is possible that other choices for hyperparameter optimization for these models might have produced better results. However, the fact that only neural networks were able to uncover patterns that produce long-short strategies that beat linear models suggests that non-linearities do play a strong role in this market. In comparison with other studies that use daily stock return data, our results show important differences. For example, Krauss et al. (2017) reports highest Sharpe ratios with random forests, followed by gradient boosting and deep networks. Huck (2019) also finds that random forests outperform other methods such as deep belief networks and the elastic net. Comparison with these studies is challenging, however, due to differences in data and methodology. Particularly, these studies use a classification approach to predict the direction of daily price changes, while we use a regression approach to predict monthly returns.

Overall, our results show that ML is not a panacea for portfolio management, and highlight the importance of the portfolio construction and risk management processes, as well as the context of the market in which it is applied. Additionally, since many ML models produce portfolios with high turnover, taking into account transaction costs when comparing ML portfolios is essential.

4.3. Performance of long-short strategies using individual ML models

The results in Table 3 show that traditional long-short strategies based on ML forecasts in the Brazilian equity market have some drawbacks. First, the short positions do not produce additional returns, since the P_1 portfolios typically earn positive returns. Second, the high volatility of the P_1 portfolios increases the volatility of the long-short strategies. Third, the long-short strategies have high maximum drawdowns. In this subsection, we explore the use of the simple ERC approach to combine portfolios based on the P_5 and P_1 portfolios for each ML model. At each month, for each ML model, we first identify the stocks in the corresponding P_1 and P_5 portfolios. We use an exponentially weighted moving average estimator with a decay factor of $\lambda = 0.96$ to estimate the covariance matrix of these stocks,

using one year of daily returns.²⁷ We then estimate the volatilities of the P_1 and P_5 portfolios, which are required to obtain κ , the short weight multiplier, as described in Section 3.3.2.

Table 4 reports the results. Panel A shows results for the traditional long-short strategies, i.e. the $P_5 - P_1$ portfolios in Table 3. Panel B reports results for the simple ERC long-short strategies for each ML model. For most methods, the average returns are higher for ERC long-short strategies compared with traditional long-short strategies, since the short (P_1) portfolios are underweighted to balance the risk contributions. The volatilities and maximum drawdowns of the ERC long-short strategies are greatly reduced, compared to traditional long-short strategies. The average reduction in volatility is 33%, while maximum drawdowns are reduced by over 47% on average. The maximum drawdowns of the best ERC long-short strategies are less than 30%. The ERC long-short strategies have variable leverage, depending on the evolution of the volatilities of the baseline portfolios. The average leverage ratios of these strategies varies from 1.61 for the PCR model to 1.86 for the RF model. On average, the simple ERC strategy reduces short exposures by 25% across all ML models.

In terms of net SR, all ERC portfolios outperform the traditional long-short strategies. Similarly to the traditional long-short strategies, the highest net SR using individual models are achieved by the NN3 and NN4 models (1.15 and 1.11, respectively), while the net SR of the ENS_{ML} ensemble is 1.14. All of these are significantly higher compared to the IBOVESPA at the 10% significance level (and at the 5% level for NN3). Figure 5 plots the average (net) monthly return and monthly standard deviation of the different long-short strategies for each ML model. It is clear from the figure that, for almost all ML models, the ERC approach reduces risk, while simultaneously increasing returns.

An important question is whether the returns of ML strategies can be explained by exposure to well-known risk factors. To investigate this issue, we regress the excess returns of the different long-short and simple ERC long-short strategies on the Fama and French (1996) three factors and a momentum factor.²⁸ Table 5 reports the estimates of intercepts and factor loadings, as well as the associated p-values, calcu-

²⁷Our results are not dependent on this choice. Results with other values of λ , rolling-window sample covariance estimates, or the shrinkage estimator of Ledoit and Wolf (2004) are qualitatively similar.

²⁸The factors are constructed for the Brazilian equity universe following the methodology in Fama and French (2017). The average monthly returns of the SMB , HML , and WML factors during the sample period are -0.03% , 0.47% , and 1.02% , respectively (not tabulated).

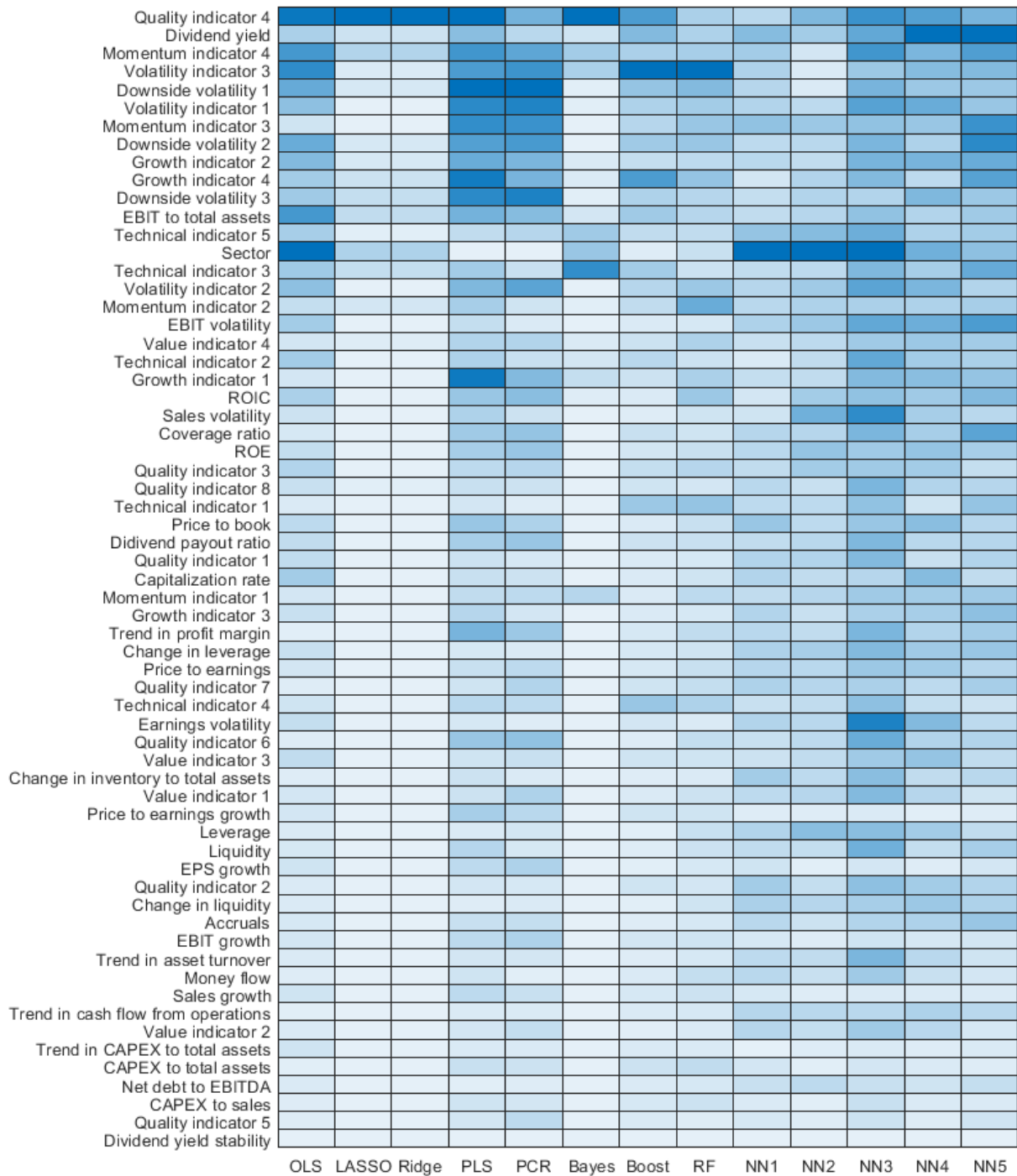


Figure 4: Summary of variable importance across different ML models. EBIT = Earnings Before Interest and Taxes; EBITDA = Earnings Before Taxes, Interest, Taxes, Depreciation and Amortization; ROE = Return On Equity; ROIC = Return On Invested Capital; EPS = Earnings Per Share; CAPEX = Capital Expenditure.

Table 4 Risk and return statistics of long-short strategies obtained with different machine learning methods. The table reports out-of-sample performance metrics for different long-short strategies formed on predicted returns using machine learning methods. Panel A shows results for traditional long-short strategies which are long (short) the highest (lowest) decile of predicted returns with each machine learning method. Panel B shows results for Equal Risk Contribution (ERC) long-short strategies, which equalize the risk contributions of the long and short legs. The table reports the average monthly return before (Ave) and after costs (Ave (net)), the monthly standard deviation (Std), the annualized Sharpe Ratio before (SR) and after costs (SR (net)), the maximum drawdown (Max. DD), the average monthly turnover, and the average leverage. One, two and three asterisks denote a statistically significant difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period at the 10%, 5% and 1% significance level, respectively, using the Ledoit and Wolf (2008) test.

Panel A: Traditional long-short strategies														
	OLS	LASSO	Ridge	PLS	PCR	Bayes	Boost	RF	NN1	NN2	NN3	NN4	NN5	ENS_{ML}
Ave	1.61	1.72	1.86	1.90	1.43	1.68	1.60	1.71	1.33	1.49	1.95	2.06	1.72	2.00
Ave (net)	0.97	1.16	1.32	1.38	0.90	1.06	1.02	1.07	0.69	0.86	1.33	1.48	1.15	1.41
Std	5.17	6.23	5.80	7.18	7.21	5.92	6.58	6.10	4.97	5.29	5.51	5.96	6.38	6.11
SR	1.08*	0.96	1.11	0.91	0.69	0.98	0.84	0.97	0.93*	0.98	1.22*	1.20*	0.93	1.13*
SR (net)	0.65	0.65	0.79	0.66	0.43	0.62	0.54	0.61	0.48	0.56	0.84	0.86	0.63	0.80
Max.DD	53.83	65.67	54.65	62.91	66.59	62.80	67.78	58.29	39.77	52.43	55.28	55.61	62.42	59.00
Ave. turnover	177.81	125.72	120.35	102.32	105.84	166.06	137.16	180.13	183.13	178.55	163.13	143.41	131.34	148.68
Ave. leverage	2.00	2.00	1.96	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00

Panel B: Equal risk contribution long-short strategies														
	OLS	LASSO	Ridge	PLS	PCR	Bayes	Boost	RF	NN1	NN2	NN3	NN4	NN5	ENS_{ML}
Ave	1.76	1.72	1.76	1.72	1.51	1.87	1.56	1.74	1.41	1.50	1.98	1.98	1.71	1.95
Ave (net)	1.19	1.25	1.31	1.33	1.14	1.31	1.12	1.23	0.83	0.94	1.45	1.51	1.26	1.48
Std	3.95	4.18	4.11	4.35	4.02	4.20	3.84	3.96	3.67	3.88	3.86	4.05	4.08	3.81
SR	1.42**	1.29*	1.32**	1.16*	1.01	1.43**	1.17*	1.32*	1.21**	1.21**	1.62**	1.52**	1.27**	1.56**
SR (net)	0.92*	0.90	0.94	0.85	0.70	0.98	0.78	0.87	0.66	0.70	1.15**	1.11*	0.88	1.14*
Max.DD	28.38	36.29	27.85	30.35	39.00	34.82	35.46	40.08	16.76	30.70	26.15	23.79	29.70	30.10
Ave. turnover	172.59	119.84	114.54	93.87	92.81	161.66	125.63	164.15	178.79	172.26	155.44	133.29	121.75	136.83
Ave. leverage	1.83	1.79	1.76	1.68	1.61	1.84	1.68	1.72	1.86	1.82	1.80	1.75	1.73	1.71

lated using heteroscedasticity and autocorrelation-consistent standard errors. The first important result from this table is that the estimates of alphas are all positive and statistically significant. The best ML long-short strategy, NN3, has an alpha of 1.79% per month, with a p-value of less than 0.01. The results on Panel A show that almost none of the ML long-short strategies have a significant market exposure, as expected. On the other hand, all ML long-short strategies have negative - and typically high significant - loadings on the *SMB* and *HML* factors, indicating that the strategies tend to tilt towards large-cap stocks which are classified as growth stocks in terms of book-to-market ratios. Finally, the positive and typically significant loadings on the momentum factor reveal that most ML strategies are tilted towards stocks which have been on positive trends. This is not surprising, given the well-documented existence of momentum in general, and in particular the importance of momentum features in return prediction seem in Figure 4. The results for the simple ERC ML strategies on Panel B are largely similar, with usually higher market beta, which is expected as these strategies are on average net long, and lower loadings on the other factors.

These results show that the ML strategies are delivering returns that exceed what would be expected given their exposures to the market, size, value and momentum factors. Indeed, since most strategies have a negative loading on the value factor, which has a positive average monthly return of 0.47%, this tilt towards growth stocks represents a drag on the risk-adjusted performance. On the other hand, the consistently positive loading on the momentum factor, which has an even higher return of 1.02%, does not explain the returns of the ML strategies, suggesting that they are successfully capturing other sources of expected return.

4.4. Performance of long-short strategies using multiple ML models

In this section, we explore long-short strategies that combine multiple ML models, either by using the equally-weighted average ensemble, or by combining the long and short portfolios obtained with different models. We compare the four long-short strategies discussed in section 3.3.3. The first one is a traditional long-short strategy using the predictive ensemble model that averages forecasts from all ML models. We label the corresponding long-short portfolio as ENS_{ML-EW} . The second is the simple ERC strategy, applied to the P_5 and P_1 portfolios obtained with this ensemble, which we denoted by

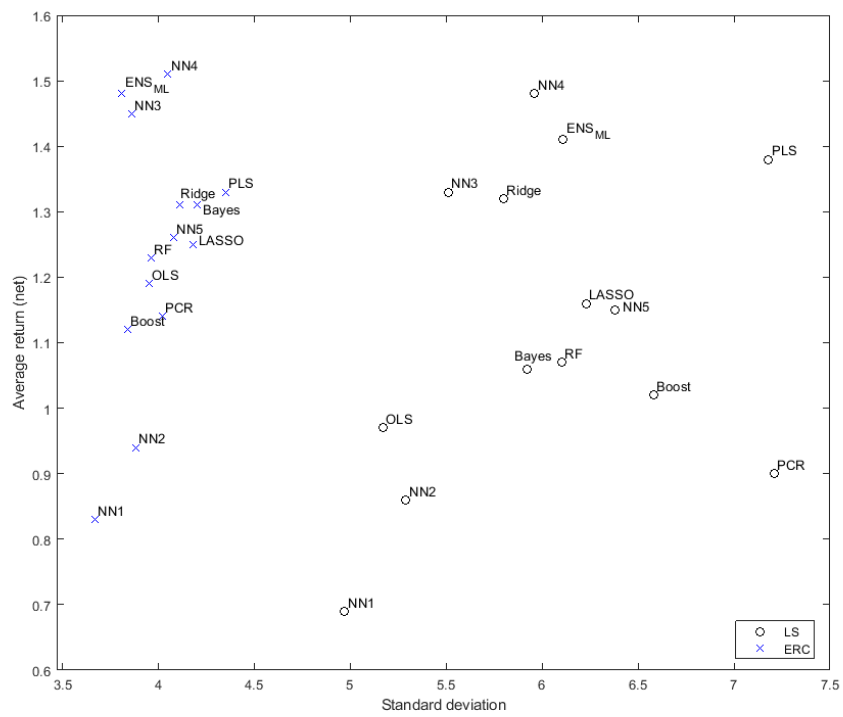


Figure 5: Monthly standard deviation and net average return of different long-short strategies

Table 5 Factor loadings and risk-adjusted returns of long-short strategies obtained with different machine learning methods. The excess returns of different strategies formed on predicted returns using machine learning methods are regressed on the Fama and French (1996) three factors plus a momentum factor:

$$r_i = \alpha_i + \beta_i r_{mt} + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \beta_i^{WML} WML_t + \epsilon_t,$$

where r_{mt} is the excess return on a value-weighted market portfolio, SMB_t is the small-minus-big or size factor, HML_t is the high-minus-low book-to-market factor, and WML_t is the winner-minus-loser or momentum factor. Standard errors are adjusted for heteroscedasticity and autocorrelation. Panel A shows results for traditional long-short strategies which are long (short) the highest (lowest) decile of predicted returns with each machine learning method. Panel B shows results for Equal Risk Contribution (ERC) long-short strategies, which equalize the risk contributions of the long and short legs.

Panel A: Traditional long-short strategies														
	OLS	LASSO	Ridge	PLS	PCR	Bayes	Boost	RF	NN1	NN2	NN3	NN4	NN5	Ens1
Intercept (%)	1.4***	1.41***	1.35***	1.36***	0.95**	1.35***	1.24***	1.4***	1.53***	1.5***	1.79***	1.77***	1.37***	1.73***
p-value	< 0.01	< 0.01	< 0.01	< 0.01	0.02	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
r_{mt}	0.01	-0.03	0	-0.11*	-0.22***	0.01	-0.06	-0.07	-0.08	-0.05	0.01	-0.04	-0.11*	-0.05
p-value	0.46	0.37	0.48	0.06	< 0.01	0.47	0.19	0.20	0.21	0.26	0.46	0.32	0.09	0.27
SMB_t	-0.23**	-0.38***	-0.2**	-0.5***	-0.55***	-0.31***	-0.44***	-0.39***	-0.05	-0.15	-0.29***	-0.39***	-0.42***	-0.37***
p-value	0.02	< 0.01	0.05	< 0.01	< 0.01	0.01	< 0.01	< 0.01	0.32	0.11	< 0.01	< 0.01	< 0.01	< 0.01
HML_t	-0.38***	-0.51***	-0.24*	-0.46***	-0.39***	-0.41***	-0.55***	-0.44***	-0.54***	-0.5***	-0.45***	-0.41***	-0.42***	-0.51***
p-value	< 0.01	< 0.01	0.07	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
WML_t	0.22**	0.31***	0.46***	0.47***	0.41***	0.32***	0.35***	0.3***	-0.04	0.1	0.17**	0.27***	0.32***	0.28***
p-value	0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.39	0.23	0.04	< 0.01	< 0.01	< 0.01
Adj. R^2	0.32	0.49	0.38	0.61	0.62	0.39	0.57	0.47	0.24	0.28	0.34	0.44	0.51	0.48
Panel B: Equal risk contribution long-short strategies														
	OLS	LASSO	Ridge	PLS	PCR	Bayes	Boost	RF	NN1	NN2	NN3	NN4	NN5	Ens1
Intercept (%)	1.47***	1.31***	1.19***	1.06***	0.93***	1.48***	1.04***	1.35***	1.48***	1.38***	1.67***	1.53***	1.23***	1.53***
p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
r_{mt}	0.09*	0.1*	0.11*	0.11**	0.07	0.04	0.12**	0.13**	-0.01	0.05	0.09	0.1*	0.03	0.09*
p-value	0.08	0.06	0.05	0.03	0.13	0.29	0.01	0.01	0.47	0.27	0.11	0.09	0.34	0.08
SMB_t	-0.16**	-0.23***	-0.11	-0.21**	-0.24***	-0.23***	-0.17**	-0.17**	-0.03	-0.09	-0.2***	-0.21**	-0.23***	-0.2**
p-value	0.03	0.01	0.13	0.01	< 0.01	0.01	0.02	0.04	0.34	0.18	0.01	0.01	< 0.01	0.01
HML_t	-0.17**	-0.19**	0	-0.15	-0.21**	-0.07	-0.17**	-0.23***	-0.31***	-0.26***	-0.14*	-0.11	-0.09	-0.16**
p-value	0.05	0.03	0.49	0.12	0.02	0.27	0.02	0.01	< 0.01	0.01	0.06	0.10	0.15	0.05
WML_t	0.12**	0.19***	0.3***	0.31***	0.2***	0.18**	0.22***	0.14**	-0.1	0.01	0.08	0.17**	0.19***	0.14**
p-value	0.04	< 0.01	< 0.01	< 0.01	< 0.01	0.01	< 0.01	0.02	0.16	0.45	0.13	0.01	< 0.01	0.02
Adj. R^2	0.12	0.24	0.17	0.30	0.29	0.16	0.24	0.17	0.12	0.11	0.13	0.17	0.24	0.19

ENS_{ML} -ERC.²⁹ The third strategy is the equally-weighted average of the individual long-short strategies obtained with each model (MS-EW). Finally, the fourth strategy results from applying the MS-ERC approach explained in section 3.3.3 to the set of individual long and short portfolios obtained with each ML model. To solve the ERC problem (14), we require an estimate of the covariance matrix of the P_1 and P_5 portfolios for each of the ML models. At each month, we do this using an exponentially-weighted moving average estimator with a decay factor of $\lambda = 0.96$, applied to the daily returns of the P_1 and P_5 portfolios for all models.³⁰

Panel A of Table 6 shows the results for the two strategies using the equally-weighted ensemble of ML models, while Panel B shows the results of the strategies that combine the individual ML portfolios. Since we initially require one year of daily returns for each strategy in order to estimate covariance matrices, these results comprise the period from January 2007 to December 2018. The first question we address is whether it is better to first create an ensemble of ML models, and then create a long-short strategy based on the ensemble forecasts (as done, for example, by Krauss et al., 2017), or whether it is better to aggregate the long-short strategies obtained with different ML models (the “metastrategy” approach of Gu et al., 2020). The long-short strategy using the forecasting ensemble method (ENS_{ML} -EW, first column of Panel A) earns on average 2.04% per month (1.45% on a net basis) with a monthly standard deviation of 6.18%, resulting in a Sharpe ratio of 1.14 (0.81 on a net basis). The strategy that invests equal weights in the individual ML long-short strategies (MS-EW, first column of Panel B) earns 1.81% per month (1.24% on net basis), with a standard deviation of 5.93%, which translates into a SR of 1.06 (0.72 on a net basis). Despite the slightly superior performance of the ensemble approach, the difference in Sharpe ratios is not statistically significant (not reported). The test of equality of SRs of each strategy relative to the IBOVESPA also fails to detect any significant difference. Both strategies have maximum drawdowns of approximately 60%, which is similar to those of the individual long-short strategies.

As with the individual ML models, the simple ERC long-short strategy using the equally-weighted ensemble (ENS_{ML} -ERC, second column of Panel A) is able to drastically reduce the volatility and the

²⁹These two strategies are the ones on the last columns of Panels A and B of Table 4.

³⁰Our results are not sensitive to this choice. Results with other values of λ , rolling-window sample covariance estimates, or the shrinkage estimator of Ledoit and Wolf (2004) are qualitatively similar.

maximum drawdown in comparison with the traditional long-short strategy, while earning a slightly lower gross return (1.99% per month, compared to 2.04% for the ENS_{ML} -EW strategy). In addition, because of the lower borrowing costs of this strategy, which on average is short 70% of the total amount in the long portfolio, its average net return is higher than that of the ENS_{ML} -EW strategy (1.53% compared with 1.45%). The maximum drawdown of the ENS_{ML} -ERC strategy is roughly half that of the ENS_{ML} -EW strategy. The SR of the ENS_{ML} -ERC strategy is 1.59 before costs (significantly higher compared to the IBOVESPA at the 5% level) and 1.18 after costs (significantly higher compared to the IBOVESPA at the 10% level). The net SR is also significantly higher compared to the ENS_{ML} -EW strategy at the 1% significance level (not reported).

The last column of Panel B reports results for the MS-ERC approach, which balances the risk contributions of all individual ML long-short strategies, as well as the total risk contributions of the total long and short positions. This strategy earns on average 2.06% per month (1.64% on a net basis), with a monthly standard deviation of 2.94%, resulting in a SR, before costs, of 2.10, which is significantly higher compared to the IBOVESPA at the 1% level. The net SR of 1.61 is significantly higher compared to the IBOVESPA at the 5% level, and is also significantly higher than the SRs of the other strategies in Table 6 at the 1% significance level. Additionally, the MS-ERC approach results in a maximum drawdown of approximately 16%, which is almost half that of the ENS_{ML} -ERC strategy, and about a quarter that of the two other strategies. The leverage ratio of the MS-ERC strategy is 1.66, which indicates that on average, the short exposure is 66% of the long exposure, i.e. the strategy has, on average, a net long exposure of approximately 34%. As shown in Figure 6, there is substantial variation in the value of κ_t over time. Across the entire sample period, κ_t is always lower than 1, reaching a minimum of less than 30% in mid-2011. These results show that the MS-ERC strategy has at all times a net long exposure, and indicate that the P_1 portfolios are consistently more volatile than the P_5 portfolios over this period, although this varies over time. The MS-ERC approach is able to dynamically account for this feature, as well as for the correlations between the individual ML long and short portfolios, to obtain balanced risk contributions at the level of each ML strategy, as well as over all long and short positions.

Table 7 reports estimates of alphas and factor loadings for each of the combined ML strategies and their corresponding p-values, calculated using heteroscedasticity and autocorrelation-consistent standard

Table 6 Risk and return statistics of long-short strategies obtained using an ensemble of machine learning methods, or by combining long-short strategies obtained with various machine learning methods. The table reports average monthly return before (Ave) and after costs (Ave (net)), the monthly standard deviation (Std), the annualized Sharpe Ratio before (SR) and after costs (SR (net)), the maximum drawdown (Max. DD), the average monthly turnover, and the average leverage. One, two and three asterisks denote a statistically significant difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period at the 10%, 5% and 1% significance level, respectively, using the Ledoit and Wolf (2008) test.

	Panel A. Long-short strategies using an ensemble of machine learning models		Panel B. Long-short strategies combining portfolios obtained with different machine learning models	
	ENS_{ML-EW}	ENS_{ML-ERC}	MS-EW	MS-ERC
Ave	2.04	1.99	1.81	2.06
Ave (net)	1.45	1.53	1.24	1.64
Std	6.18	3.78	5.93	2.94
SR	1.14	1.59**	1.06	2.10***
SR (net)	0.81	1.18*	0.72	1.61**
Max.DD	59.00	30.10	60.54	15.67
Ave. turnover	147.65	134.86	133.46	117.12
Ave. leverage	2.00	1.70	2.00	1.66

errors. As with the individual ML strategies, we find that alphas are positive and statistically significant. The market exposure of the ENS_{ML-ERC} (respectively, MS-ERC) strategy is positive and significant at the 10% (respectively, 5%) level, which is not surprising given the net long exposure implied by the ERC approach in this market. On the other hand, the market exposure of the ENS_{ML-EW} and MS-EW strategies are not significant. As with the individual ML strategies, the combined strategies have negative loadings on the size (SMB) and value (HML) factors, and positive loadings on the momentum (WML) factor. Interestingly, the MS-ERC has a similar alpha compared to the other combined strategies, while having lower exposure to the size, value and momentum factors, and also a much lower maximum drawdown, as shown on Table 6.

These results strongly support the idea of combining strategies obtained using various ML models into a single strategy with balanced risk contributions. The MS-ERC approach outperforms, on a risk-adjusted basis, all individual ML portfolios, the ensemble of ML portfolios, and the equally-weighted combination of the ML portfolios, while keeping the maximum drawdown at a more acceptable level. Other studies that apply ensembles of forecasts obtained with different ML models, such as Krauss et al. (2017), also find benefits to aggregating different forecasts, however the underlying characteristics of the resulting portfolio, such as the maximum drawdown, do not change as much.

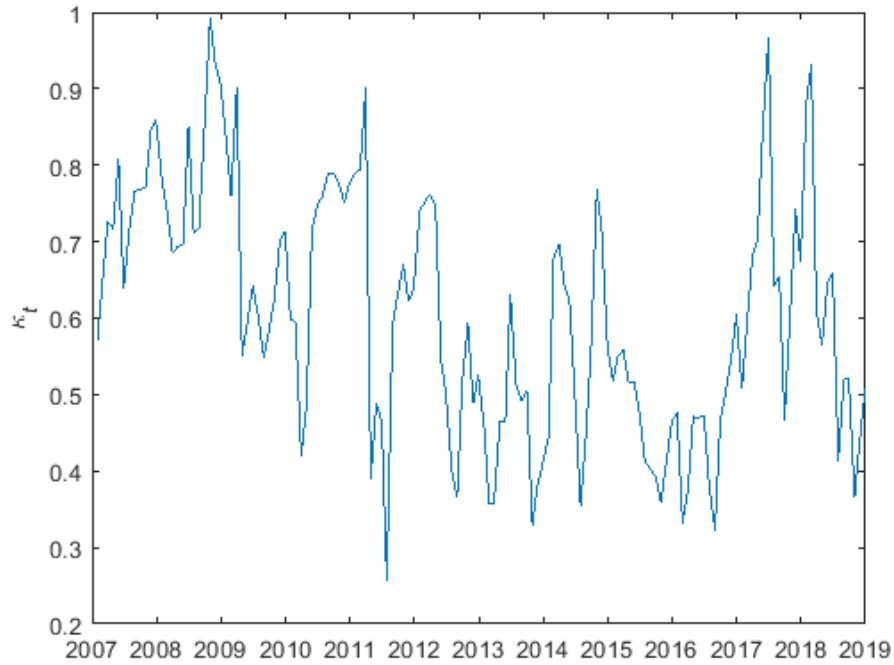


Figure 6: Evolution of the short weight multiplier ($-\kappa_t$) for a long-short equal risk contribution strategy combining machine learning portfolios.

The risk reduction benefits obtained through the MS-ERC approach translate into higher average and cumulative returns. Figure 7 plots the cumulative gross returns of the four strategies in Table 6, as well as the excess market return (EMR) and the momentum (WML) factor. All series are scaled to the realized volatility of the MS-ERC strategy. The total cumulative gross return earned by the MS-ERC strategy is several times that of the other strategies.

4.5. Robustness with respect to transaction costs

In this section, we address potential concerns related to the robustness of our results with respect to transaction costs, which throughout the paper, were assumed to consist of 0.15% for all trades, i.e. a bid-ask spread of 0.30%. While the universe of stocks used in this study consists of the most liquid stocks in the Brazilian market, this level of transaction costs may be too low, particularly for the earlier part of the sample. Therefore, in Tables A.4, A.5 and A.6 in the Appendix, we replicate the main results of the paper, i.e. Tables 3, 4 and 6, using a much more conservative estimate of 0.50%, which translates

Table 7 Factor loadings and risk-adjusted returns of long-short strategies obtained using an ensemble of machine learning methods, or by combining long-short strategies obtained with various machine learning methods. Excess returns are regressed on the Fama and French (1996) three factors plus a momentum factor:

$$r_i = \alpha_i + \beta_i r_{mt} + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \beta_i^{WML} WML_t + \epsilon_t,$$

where r_{mt} is the excess return on a value-weighted market portfolio, SMB_t is the small-minus-big or size factor, HML_t is the high-minus-low book-to-market factor, and WML_t is the winner-minus-loser or momentum factor. Standard errors are adjusted for heteroscedasticity and autocorrelation.

	Panel A. Long-short strategies using an ensemble of machine learning models		Panel B. Long-short strategies combining portfolios obtained with different machine learning models	
	ENS_{ML-EW}	ENS_{ML-ERC}	MS-EW	MS-ERC
Intercept (%)	1.73***	1.53***	1.56***	1.52***
p-value	< 0.01	< 0.01	< 0.01	< 0.01
r_{mt}	-0.05	0.09 *	-0.06	0.1**
p-value	0.27	0.08	0.20	0.02
SMB_t	-0.37***	-0.2***	-0.37***	-0.13**
p-value	< 0.01	0.01	< 0.01	0.02
HML_t	-0.51***	-0.16**	-0.48***	-0.16***
p-value	< 0.01	0.05	< 0.01	< 0.01
WML_t	0.28***	0.14**	0.28***	0.12**
p-value	< 0.01	0.02	< 0.01	0.02
Adj. R^2	0.48	0.19	0.52	0.25

into a bid-ask spread of 1% for all trades. While, as expected, the net returns and Sharpe ratios of the strategies decrease accordingly, the main results of the paper, namely the relative performance of the ML portfolios, and the benefit from incorporating multiple ML portfolios using the ERC approach, remain. In particular, Table A.6 shows that the MS-ERC continues to outperform the other portfolios, with a net SR above 1, which is still statistically superior to the Sharpe ratio of the IBOVESPA index, albeit only at the 10% significance level. Since these calculations assume no costs whatsoever for investing in the IBOVESPA, these results are quite conservative.

We also consider that readers may have different opinions about the appropriate level of transaction costs, and note that net returns can be easily calculated for any level of transaction costs as follows:

$$\text{net return} = \text{gross return} - \frac{\text{avg. turnover}}{100} \times TC - (\text{avg. leverage} - 1) \times BC, \quad (19)$$

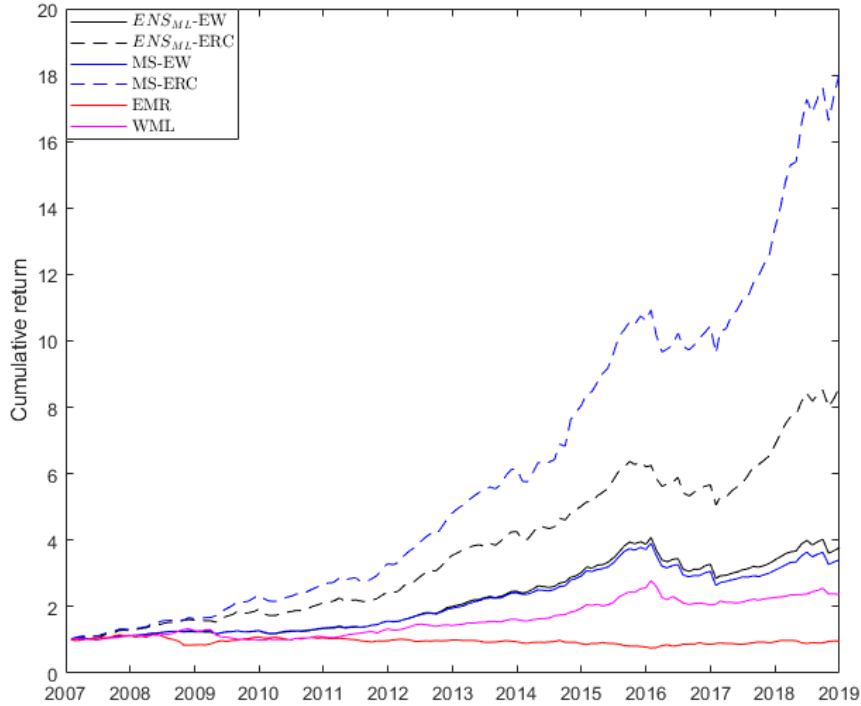


Figure 7: Cumulative returns of various long-short strategies obtained using machine learning methods. The Ensemble Long-Short strategy (ENS_{ML-EW}) goes long (short) the stocks in the top (bottom) quintile of return forecasts using an equally-weighted ensemble of all machine learning forecasts. The Ensemble Equal Risk Contribution (ENS_{ML-ERC}) strategy equalizes the risk contributions of the long and short portfolios obtained with the ensemble. The Multi-Strategy Equally Weighted (MS-EW) approach invests equal amounts in long-short strategies obtained with individual machine learning methods. The Multi-Strategy Equal Risk Contribution approach (MS-ERC) combines the long and short portfolios to achieve equal risk contributions across machine learning methods and across the overall long and short exposures.

where TC is the trading transaction cost and BC is the borrowing cost.³¹ As an example of the utilization of this formula, we apply it to recover the net returns of the ENS_{ML-ERC} portfolio in Table 6. The average return before costs is 1.99%, the average turnover is 134.86, and the average leverage is 1.70. Using $TC = 0.15\%$ and $BC = 0.375\%$, we have net return = $1.99 - 1.3486 \times 0.15 - 0.70 \times 0.375 = 1.525\%$, which is rounded to the 1.53% shown in the table. Small differences may occur due to rounding.

³¹We thank an anonymous referee for the suggestion to include results using more conservative transaction costs, and the Subject Editor for the suggestion to include the formula for the calculation of net returns.

5. Concluding remarks

In this paper, we explore the use of machine learning (ML) models to forecast stock returns in the Brazilian equity market, using a rich dataset of technical and fundamental indicators. While all ML models we test are able to produce portfolios that outperform the local market, the performance of traditional long-short strategies based on ML models are hampered by the high volatility of the short portfolios. We find that a simple Equal Risk Contribution (ERC) approach, which balances the risk contributions of the long and short components of individual ML long-short strategies, significantly improves their risk-adjusted returns. We further develop a multi-strategy ERC approach that combines multiple long-short strategies obtained with various ML models, such that the risk contributions of all individual long-short strategies are equal, and the aggregate risk contribution of all long positions equals that of all short positions. We find that the multi-strategy ERC approach outperforms all individual ML long-short strategies, as well as strategies based on forecasts from an ensemble of ML models (with or without the use of ERC), and a multi-strategy approach that invests equally in all ML strategies.

Our work has important managerial implications related to investment management, especially for portfolio managers seeking to incorporate ML into their investment processes. Specifically, the multi-strategy ERC approach delivers a solution to two problems faced in practice by a portfolio manager applying ML models to create a long-short investment strategy. First, it provides a formal way to define allocations to each ML long-short strategy, which does not require forecasts of which ML model will outperform in the future. In this sense, each ML long-short strategy can be thought of as a trading desk or unit, to which the portfolio manager allocates an equal risk budget. Second, the multi-strategy ERC approach automatically balances the risk contributions of the overall long and short positions, by allowing a variable degree of leverage in the short positions. We have shown this to be considerably relevant and effective in the Brazilian market, due to the much higher risk of the short portfolios. Although we apply the approach to combinations of long and short portfolios obtained using ML models, this approach is general, and can be applied to any combination of long and short portfolios.

References

- Akyildirim, E., Goncu, A. and Sensoy, A. (2020), 'Prediction of cryptocurrency returns using machine learning', *Annals of Operations Research* pp. 1–34.
- Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X. (2006), 'The cross-section of volatility and expected returns', *Journal of Finance* **61**, 259–299.
- Asness, C. S., Moskowitz, T. J. and Pedersen, L. H. (2013), 'Value and momentum everywhere', *The Journal of Finance* **68**(3), 929–985.
- Atsalakis, G. S. and Valavanis, K. P. (2009), 'Surveying stock market forecasting techniques—part ii: Soft computing methods', *Expert Systems with Applications* **36**(3), 5932–5941.
- Bai, X., Scheinberg, K. and Tutuncu, R. (2016), 'Least-squares approach to risk parity in portfolio selection', *Quantitative Finance* **16**(3), 357–376.
- Ban, G.-Y., El Karoui, N. and Lim, A. E. (2016), 'Machine learning and portfolio optimization', *Management Science* **64**(3), 1136–1154.
- Bekaert, G., Erb, C. B., Harvey, C. R. and Viskanta, T. E. (1998), 'Distributional characteristics of emerging market returns and asset allocation', *Journal of Portfolio Management* **24**(2), 102–+.
- Bekaert, G. and Harvey, C. R. (1997), 'Emerging equity market volatility', *Journal of Financial Economics* **43**(1), 29–77.
- Bertrand, P. and Lapointe, V. (2018), 'Risk-based strategies: the social responsibility of investment universes does matter', *Annals of Operations Research* **262**(2), 413–429.
- Bessembinder, H. (2003), 'Trade execution costs and market quality after decimalization', *Journal of Financial and Quantitative Analysis* pp. 747–777.
- Best, M. J. and Grauer, R. R. (1991), 'On the sensitivity of mean-variance efficient portfolios to changes in asset means: Some analytical and computational results', *Review of Financial Studies* **4**, 315–342.

- Bianchi, D., Büchner, M. and Tamoni, A. (2020), ‘Bond risk premiums with machine learning’, *The Review of Financial Studies* .
- Bishop, C. M. (1995), *Neural networks for pattern recognition*, Oxford university press.
- Bonaccolto, G. and Paterlini, S. (2019), ‘Developing new portfolio strategies by aggregation’, *Annals of Operations Research* pp. 1–39.
- Breiman, L. (1984), *Classification and regression trees*, Wadsworth.
- Breiman, L. (2001), ‘Random forests’, *Machine learning* **45**(1), 5–32.
- Campbell, J. Y. (2000), ‘Asset Pricing at the Millennium’, *The Journal of Finance* .
- Campbell, J. Y. and Thompson, S. B. (2007), ‘Predicting excess stock returns out of sample: Can anything beat the historical average?’, *The Review of Financial Studies* **21**(4), 1509–1531.
- Cao, Q., Leggio, K. B. and Schniederjans, M. J. (2005), ‘A comparison between fama and french’s model and artificial neural networks in predicting the chinese stock market’, *Computers & Operations Research* **32**(10), 2499–2512.
- Cao, Q., Parry, M. E. and Leggio, K. B. (2011), ‘The three-factor model and artificial neural networks: predicting stock price movement in china’, *Annals of Operations Research* **185**(1), 25–44.
- Cenesizoglu, T. and Timmermann, A. (2012), ‘Do return prediction models add economic value?’, *Journal of Banking & Finance* **36**(11), 2974–2987.
- Chaves, D., Hsu, J., Li, F. and Shakernia, O. (2011), ‘Risk parity portfolio vs. other asset allocation heuristic portfolios’, *The Journal of Investing* **20**(1), 108–118.
- Chen, J., Dai, G. and Zhang, N. (2019), ‘An application of sparse-group lasso regularization to equity portfolio optimization and sector selection’, *Annals of Operations Research* pp. 1–20.
- Cybenko, G. (1989), ‘Approximation by superpositions of a sigmoidal function’, *Mathematics of control, signals and systems* **2**(4), 303–314.

- De Spiegeleer, J., Madan, D. B., Reyners, S. and Schoutens, W. (2018), ‘Machine learning for quantitative finance: fast derivative pricing, hedging and fitting’, *Quantitative Finance* **18**(10), 1635–1643.
- DeMiguel, V., Garlappi, L., Nogales, F. J. and Uppal, R. (2009), ‘A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms’, *Management Science* **55**, 798–812.
- DeMiguel, V., Gil-Bazo, J., Nogales, F. J. and AP Santos, A. (2021), Can machine learning help to select portfolios of mutual funds? Mimeo, Available at SSRN: 3768753.
- DeMiguel, V., Martin-Utrera, A., Nogales, F. J. and Uppal, R. (2019), A portfolio perspective on the multitude of firm characteristics.
- Dietterich, T. G. (2000), Ensemble methods in machine learning, *in* ‘International workshop on multiple classifier systems’, Springer, pp. 1–15.
- Fama, E. F. and French, K. R. (1993), ‘Common risk factors in the returns on stocks and bonds’, *Journal of financial economics* **33**(1), 3–56.
- Fama, E. F. and French, K. R. (1996), ‘Multifactor Explanations of Asset Pricing Anomalies’, *The Journal of Finance* **51**(1), 55–84.
- Fama, E. F. and French, K. R. (2017), ‘International tests of a five-factor asset pricing model’, *Journal of Financial Economics* **123**(3), 441–463.
- Feng, G., Giglio, S. and Xiu, D. (2020), ‘Taming the factor zoo: A test of new factors’, *The Journal of Finance* **75**(3), 1327–1370.
- Fischer, T. and Krauss, C. (2018), ‘Deep learning with long short-term memory networks for financial market predictions’, *European Journal of Operational Research* **270**(2), 654–669.
- Freund, Y. and Schapire, R. E. (1997), ‘A decision-theoretic generalization of on-line learning and an application to boosting’, *Journal of computer and system sciences* **55**(1), 119–139.

- Freyberger, J., Neuhierl, A. and Weber, M. (2020), ‘Dissecting characteristics nonparametrically’, *The Review of Financial Studies* **33**(5), 2326–2377.
- Friedman, J. H. (2001), ‘Greedy function approximation: a gradient boosting machine’, *Annals of statistics* pp. 1189–1232.
- Friedman, J., Hastie, T. and Tibshirani, R. (2001), *The elements of statistical learning*, Vol. 1, Springer series in statistics New York.
- Friedman, J., Hastie, T., Tibshirani, R. et al. (2000), ‘Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)’, *The annals of statistics* **28**(2), 337–407.
- George, E. I. and McCulloch, R. E. (1997), ‘Approaches for Bayesian Variable Selection’, *Statistica Sinica* **7**, 339–373.
- Green, J., Hand, J. R. and Zhang, X. F. (2017), ‘The characteristics that provide independent information about average us monthly stock returns’, *The Review of Financial Studies* p. hhx019.
- Gu, S., Kelly, B. and Xiu, D. (2020), ‘Empirical asset pricing via machine learning’, *The Review of Financial Studies* **33**(5), 2223–2273.
- Harvey, C. R., Liu, Y. and Zhu, H. (2016), ‘... and the cross-section of expected returns’, *The Review of Financial Studies* **29**(1), 5–68.
- Haugen, R. A. and Baker, N. L. (1996), ‘Commonality in the determinants of expected stock returns’, *Journal of Financial Economics* **41**(3), 401–439.
- Heaton, J., Polson, N. and Witte, J. H. (2017), ‘Deep learning for finance: deep portfolios’, *Applied Stochastic Models in Business and Industry* **33**(1), 3–12.
- Henrique, B. M., Sobreiro, V. A. and Kimura, H. (2019), ‘Literature review: Machine learning techniques applied to financial market prediction’, *Expert Systems with Applications* **124**, 226–251.
- Hou, K., Xue, C. and Zhang, L. (2020), ‘Replicating anomalies’, *The Review of Financial Studies* **33**(5), 2019–2133.

- Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T. and Johnson, J. E. (2016), 'Bridging the divide in financial market forecasting: machine learners vs. financial economists', *Expert Systems with Applications* **61**, 215–234.
- Huck, N. (2009), 'Pairs selection and outranking: An application to the s&p 100 index', *European Journal of Operational Research* **196**(2), 819–825.
- Huck, N. (2010), 'Pairs trading and outranking: The multi-step-ahead forecasting case', *European Journal of Operational Research* **207**(3), 1702–1716.
- Huck, N. (2019), 'Large data sets and machine learning: Applications to statistical arbitrage', *European Journal of Operational Research* **278**(1), 330–342.
- Hwang, S. and Satchell, S. E. (1999a), 'Modelling Emerging Market Risk Premia Using Higher Moments', *Politics* **296**, 271–296.
- Hwang, S. and Satchell, S. E. (1999b), 'Modelling emerging market risk premia using higher moments', *International Journal of Finance & Economics* **4**(4), 271–296.
- Jacobs, H. (2015), 'What explains the dynamics of 100 anomalies?', *Journal of Banking & Finance* **57**, 65–85.
- Jegadeesh, N. and Titman, S. (1993), 'Returns to buying winners and selling losers: Implications for stock market efficiency', *The Journal of Finance* **48**(1), 65–91.
- Jiang, M., Jia, L., Chen, Z. and Chen, W. (2020), 'The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm', *Annals of Operations Research* pp. 1–33.
- Kan, R. and Zhou, G. (2007), 'Optimal portfolio choice with parameter uncertainty', *Journal of Financial and Quantitative Analysis* **42**(3), 621–656.
- Kaucic, M. (2010), 'Investment using evolutionary learning methods and technical rules', *European Journal of Operational Research* **207**(3), 1717–1727.

- Kelly, B. T., Pruitt, S. and Su, Y. (2019), ‘Characteristics are covariances: A unified model of risk and return’, *Journal of Financial Economics* **134**(3), 501–524.
- Kim, J. H., Kim, W. C. and Fabozzi, F. J. (2018), ‘Recent advancements in robust optimization for investment management’, *Annals of Operations Research* **266**(1-2), 183–198.
- Kolm, P. N. and Ritter, G. (2019), ‘Modern perspectives on reinforcement learning in finance’, *The Journal of Machine Learning in Finance* **1**(1).
- Kozak, S., Nagel, S. and Santosh, S. (2020), ‘Shrinking the cross-section’, *Journal of Financial Economics* **135**(2), 271–292.
- Krauss, C., Do, X. A. and Huck, N. (2017), ‘Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500’, *European Journal of Operational Research* **259**(2), 689–702.
- Kyriakou, I., Mousavi, P., Nielsen, J. P. and Scholz, M. (2019), ‘Forecasting benchmarks of long-term stock returns via machine learning’, *Annals of Operations Research* .
- Ledoit, O. and Wolf, M. (2004), ‘Honey, I Shrunk the Sample Covariance Matrix’, *Journal of Portfolio Management* **30**, 110–119.
- Ledoit, O. and Wolf, M. (2008), ‘Robust performance hypothesis testing with the sharpe ratio’, *Journal of Empirical Finance* **15**(5), 850–859.
- Lettau, M. and Pelger, M. (2020), ‘Factors that fit the time series and cross-section of stock returns’, *The Review of Financial Studies* **33**(5), 2274–2325.
- Lewellen, J. (2014), ‘The cross section of expected stock returns’, *Forthcoming in Critical Finance Review* .
- Li, B. and Rossi, A. G. (2020), Selecting mutual funds from the stocks they hold: A machine learning approach. Available at SSRN: 3737667.

- Lu, Z., Pu, H., Wang, F., Hu, Z. and Wang, L. (2017), The expressive power of neural networks: A view from the width, *in* ‘Advances in Neural Information Processing Systems’, pp. 6231–6239.
- Maillard, S., Roncalli, T. and Teiletche, J. (2010), ‘The properties of equally weighted risk contribution portfolios’, *The Journal of Portfolio Management* **36**(4), 60–70.
- McLean, R. D. and Pontiff, J. (2016), ‘Does academic research destroy stock return predictability?’, *The Journal of Finance* **71**(1), 5–32.
- Michaud, R. O. . (1989), ‘The Markowitz optimization enigma: Is optimized optimal’, *Financial Analysts Journal* **45**, 31–42.
- Mohanram, P. S. (2005), ‘Separating winners from losers among lowbook-to-market stocks using financial statement analysis’, *Review of accounting studies* **10**(2), 133–170.
- Mullainathan, S. and Spiess, J. (2017), ‘Machine learning: an applied econometric approach’, *Journal of Economic Perspectives* **31**(2), 87–106.
- O’Hara, R. B. and Sillanpää, M. J. (2009), ‘A Review of Bayesian Variable Selection Methods: What, How and Which’, *Bayesian Analysis* **4**(1), 85–118.
- Peter, S. et al. (2015), ‘Averaging across asset allocation models’, *Journal of Economics and Statistics* **235**(1), 61–81.
- Piotroski, J. D. (2000), ‘Value investing: The use of historical financial statement information to separate winners from losers’, *Journal of Accounting Research* pp. 1–41.
- Polikar, R. (2012), Ensemble learning, *in* ‘Ensemble machine learning’, Springer, pp. 1–34.
- Qian, E. (2005), ‘Risk parity portfolios: Efficient frontiers through true diversification’, *PanAgora Asset Management White Paper* .
- Quinlan, J. R. (1993), *C4. 5: Programs for Machine Learning*, Morgan Kaufmann.

- Raposo, R. and Cruz, A. D. O. (2002), Stock market prediction based on fundamentalist analysis with fuzzy-neural networks, *in* 'Proceedings of 3rd WSES International Conference on Fuzzy Sets'.
- Ripley, B. D. (1996), *Pattern recognition and neural networks*, Cambridge university press.
- Roncalli, T. (2016), *Introduction to risk parity and budgeting*, Chapman and Hall/CRC.
- Schapire, R. E. (1990), 'The strength of weak learnability', *Machine learning* **5**(2), 197–227.
- Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E. F. and Dunis, C. (2013), 'Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and particle swarm optimization', *European Journal of Operational Research* **225**(3), 528–540.
- Sezer, O. B., Gudelek, M. U. and Ozbayoglu, A. M. (2020), 'Financial time series forecasting with deep learning: A systematic literature review: 2005–2019', *Applied Soft Computing* **90**, 106181.
- Smith, M. and Kohn, R. (1996), 'Nonparametric regression using bayesian variable selection', *Journal of Econometrics* **75**(2), 317–343.
- Tu, J. and Zhou, G. (2011), 'Markowitz meets talmud: A combination of sophisticated and naive diversification strategies', *Journal of Financial Economics* **99**(1), 204–215.
- Varian, H. R. (2014), 'Big data: New tricks for econometrics', *Journal of Economic Perspectives* **28**(2), 3–28.
- Wu, W., Chen, J., Yang, Z. and Tindall, M. L. (2020), 'A cross-sectional machine learning approach for hedge fund return prediction and selection', *Management Science* .

Appendix A.

Table A.1 Range of hyperparameters for different ML models

Model	Hyperparameters/range	Grid/Range
LASSO/Ridge	Regularization parameter (λ)	Default grid with 100 values
PLS/PCR	Number of principal components	Grid from 1 to p (number of features)
Bayesian Variable Selection	Dispersion of prior distribution (c)	100, 500, 1000
Boosting	Maximum number of splits for individual trees	1 to 50
	Learning rate	Fixed at 0.0025
	Number of training iterations	1 to 500
Random Forests	Maximum number of splits for individual trees	1 to 50
	Number of variables to sample	Equally-spaced grid with 10 values
Neural networks	Number of epochs for each neural network	Early stopping according to MSE in validation sample

Table A.2 Performance of Portfolios Formed on Individual Firm Characteristics. The table reports metrics for equally-weighted, monthly rebalanced portfolios formed on individual firm characteristics. P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of each firm characteristic. The characteristics are ordered such that P_5 will produce a higher return than P_1 . $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports the average monthly return (Ave), the monthly standard deviation (Std), the annualized Sharpe Ratio (SR), the monthly alpha from four-factor model with the three Fama and French (1993) factors plus momentum, the maximum drawdown (Max. DD), and the monthly turnover. One, two and three asterisks denote statistical significance at the 10%, 5% and 1% significance level. For Sharpe ratios, the test is based on the difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period, using the Ledoit and Wolf (2008) test.

	Accruals			CAPEX to sales			CAPEX to total assets			Capitalization rate		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.73	1.07	0.34	0.80	0.81	0.01	0.85	0.99	0.14	0.51	0.85	0.33
Std	7.72	6.98	5.09	6.44	7.20	4.12	7.18	6.05	4.12	8.29	6.62	4.57
SR	-0.06	0.1	0.15	-0.03	-0.03	0.01	0	0.07	0.07	-0.15	-0.01	0.14
Alpha	-0.12	0.41	0.53	0.09	-0.12	-0.21	-0.06	0.3	0.37	-0.23	0.09	0.31
Max.DD	74.06	58.13	45.78	55.30	69.39	70.39	64.12	48.47	44.45	81.89	61.61	60.96
Turnover	29.18	24.46	53.64	19.67	22.84	42.51	23.81	21.46	45.27	30.41	35.43	65.83
	Change in inventory to total assets			Change in leverage			Change in liquidity			Coverage ratio		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.74	1.07	0.34	0.71	1.11	0.40	1.08	1.12	0.04	0.42	1.30	0.88
Std	7.67	7.78	3.96	6.68	7.51	3.77	6.37	7.77	3.82	8.74	5.77	5.34
SR	-0.06	0.1	0.15	-0.08	0.11	0.21	0.12	0.12	0.02	-0.18	0.27	0.35
Alpha	0.02	0.46	0.43	-0.01	0.43	0.44	0.26	0.54	0.28	-0.27	0.47**	0.75*
Max.DD	68.55	60.68	51.53	59.20	63.46	47.63	50.43	60.05	52.30	83.38	51.39	53.44
Turnover	45.42	46.93	92.35	34.74	33.63	68.37	37.19	35.40	72.60	23.26	17.13	40.38
	Dividend payout ratio			Dividend yield			Dividend yield stability			Downside volatility 1		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.84	1.22	0.38	0.70	1.46	0.76	1.16	1.19	0.03	0.51	1.39	0.88
Std	9.21	4.94	6.63	7.48	5.44	4.11	5.03	7.63	5.60	9.86	4.38	7.47
SR	-0.01	0.25	0.14	-0.07	0.38**	0.35	0.21	0.15	0.02	-0.12	0.42*	0.31
Alpha	0.28	0.35*	0.07	-0.03	0.66***	0.69**	0.24	0.46	0.22	-0.05	0.42**	0.47
Max.DD	75.72	39.42	55.17	69.04	35.73	42.21	39.04	63.31	79.13	88.60	37.92	65.21
Turnover	47.35	28.46	75.81	21.07	25.49	46.55	41.48	33.75	75.23	26.72	24.34	51.06

Table A.2 Continued: Performance of Portfolios Formed on Individual Firm Characteristics. The table reports metrics for equally-weighted, monthly rebalanced portfolios formed on individual firm characteristics. The characteristics are ordered such that P_5 will produce a higher return than P_1 . P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of each firm characteristic. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports the average monthly return (Ave), the monthly standard deviation (Std), the annualized Sharpe Ratio (SR), the monthly alpha from four-factor model with the three Fama and French (1993) factors plus momentum, the maximum drawdown (Max. DD), and the monthly turnover. One, two and three asterisks denote statistical significance at the 10%, 5% and 1% significance level. For Sharpe ratios, the test is based on the difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period, using the Ledoit and Wolf (2008) test.

	Downside volatility 2			Downside volatility 3			EBIT growth			EBIT to total assets		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.57	1.61	1.04	0.80	1.54	0.74	0.34	1.66	1.32	0.29	1.19	0.90
Std	9.58	4.41	7.08	9.95	4.49	7.47	8.00	7.11	4.30	8.86	5.43	5.69
SR	-0.1	0.59***	0.38	-0.02	0.53**	0.26	-0.22	0.39**	0.57*	-0.22	0.21	0.35
Alpha	-0.02	0.71***	0.73*	0.27	0.58***	0.31	-0.43	0.92**	1.34***	-0.39	0.35*	0.74*
Max.DD	86.62	33.44	60.94	87.94	39.95	65.46	80.25	55.46	41.79	86.10	46.48	57.86
Turnover	49.72	54.15	103.87	35.15	35.48	70.63	28.40	30.53	58.93	22.91	15.35	38.26
	EBIT volatility			EPS growth			Earnings volatility			Growth indicator 1		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	1.03	1.32	0.29	0.69	1.40	0.71	0.88	1.26	0.38	0.31	1.65	1.34
Std	6.20	7.02	4.61	8.00	7.11	4.70	6.26	6.86	4.50	8.28	7.03	4.68
SR	0.09	0.23	0.16	-0.07	0.26	0.31	0.01	0.2	0.21	-0.23	0.39**	0.56
Alpha	0.09	0.55	0.46	-0.03	0.63*	0.65*	0.03	0.47	0.44	-0.37	0.84**	1.22***
Max.DD	46.37	55.54	67.26	75.79	56.15	41.81	48.29	54.54	62.05	78.10	55.22	50.70
Turnover	11.71	21.74	33.45	30.03	30.78	60.81	11.79	20.15	31.93	48.31	50.34	98.65
	Growth indicator 2			Growth indicator 3			Growth indicator 4			Leverage		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.33	1.48	1.15	0.66	1.16	0.51	0.16	1.67	1.51	1.03	1.42	0.39
Std	7.57	6.42	4.38	7.12	6.94	3.62	8.21	6.91	4.49	6.92	6.72	3.65
SR	-0.24	0.33**	0.53	-0.1	0.15	0.25	-0.3	0.41**	0.64*	0.08	0.29	0.19
Alpha	-0.24	0.54*	0.78**	-0.07	0.4	0.47*	-0.5	0.91**	1.41***	0.33	0.71***	0.39
Max.DD	79.89	48.68	43.89	60.99	56.20	34.47	80.88	54.96	33.48	66.39	52.51	43.31
Turnover	50.81	54.49	105.30	61.64	60.15	121.79	44.47	46.18	90.65	20.99	16.72	37.72

Table A.2 Continued: Performance of Portfolios Formed on Individual Firm Characteristics. The table reports metrics for equally-weighted, monthly rebalanced portfolios formed on individual firm characteristics. The characteristics are ordered such that P_5 will produce a higher return than P_1 . P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of each firm characteristic. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports the average monthly return (Ave), the monthly standard deviation (Std), the annualized Sharpe Ratio (SR), the monthly alpha from four-factor model with the three Fama and French (1993) factors plus momentum, the maximum drawdown (Max. DD), and the monthly turnover. One, two and three asterisks denote statistical significance at the 10%, 5% and 1% significance level. For Sharpe ratios, the test is based on the difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period, using the Ledoit and Wolf (2008) test.

	Liquidity			Momentum indicator 1			Momentum indicator 2			Momentum indicator 3		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.46	1.13	0.67	0.78	1.50	0.72	1.01	1.18	0.17	0.51	1.75	1.24
Std	7.06	7.25	4.35	8.02	6.58	5.15	6.84	8.25	4.91	9.67	5.96	7.23
SR	-0.2	0.13	0.33	-0.03	0.34*	0.31	0.07	0.13	0.09	-0.13	0.52**	0.44
Alpha	-0.31	0.49*	0.81**	0.09	0.67*	0.58	0.24	0.48	0.24	0.22	0.52*	0.29
Max.DD	72.11	60.12	38.13	72.49	57.02	58.32	62.58	68.43	76.92	86.08	39.70	62.41
Turnover	23.85	20.59	44.44	147.17	153.59	300.77	155.16	149.39	304.55	43.82	46.55	90.37
	Momentum indicator 4			Money flow			Net debt to EBITDA			Price to book		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.41	1.89	1.48	1.05	1.05	0.00	0.75	1.01	0.25	0.84	1.18	0.33
Std	9.80	6.07	7.37	7.17	6.23	3.87	7.40	6.95	3.71	5.80	9.32	6.52
SR	-0.16	0.59**	0.52	0.09	0.11	0	-0.05	0.07	0.12	-0.01	0.12	0.2
Alpha	0.01	0.71**	0.7	0.24	0.34	0.09	0	0.31	0.31	0.02	0.55	0.52
Max.DD	86.38	36.70	64.62	62.86	54.22	80.03	66.05	65.38	44.20	52.15	76.68	83.84
Turnover	58.13	61.05	119.18	156.28	150.28	306.56	25.79	26.54	52.33	20.09	27.05	47.13
	Price to earnings			Price to earnings growth			Quality indicator 1			Quality indicator 2		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.72	0.81	0.09	0.81	0.98	0.17	0.82	1.10	0.28	1.02	1.08	0.06
Std	8.05	5.92	4.45	5.74	5.39	2.60	6.21	6.57	2.82	6.70	5.82	2.87
SR	-0.06	-0.03	0.04	-0.03	0.08	0.11	-0.02	0.13	0.16	0.08	0.13	0.03
Alpha	0.03	0.02	-0.01	0.07	0.14	0.07	0.06	0.33	0.26	0.25	0.27	0.02
Max.DD	77.26	54.86	63.72	48.22	51.15	39.97	54.53	55.08	53.85	55.98	46.58	54.85
Turnover	28.01	32.47	60.48	40.71	38.21	78.91	47.91	49.11	97.03	37.41	48.38	85.79

Table A.2 Continued: Performance of Portfolios Formed on Individual Firm Characteristics. The table reports metrics for equally-weighted, monthly rebalanced portfolios formed on individual firm characteristics. The characteristics are ordered such that P_5 will produce a higher return than P_1 . P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of each firm characteristic. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports the average monthly return (Ave), the monthly standard deviation (Std), the annualized Sharpe Ratio (SR), the monthly alpha from four-factor model with the three Fama and French (1993) factors plus momentum, the maximum drawdown (Max. DD), and the monthly turnover. One, two and three asterisks denote statistical significance at the 10%, 5% and 1% significance level. For Sharpe ratios, the test is based on the difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period, using the Ledoit and Wolf (2008) test.

	Quality indicator 3			Quality indicator 4			Quality indicator 5			Quality indicator 6		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.56	1.22	0.66	0.38	1.48	1.10	0.62	1.43	0.80	0.75	1.42	0.67
Std	8.11	5.64	4.58	8.30	6.97	4.82	6.99	6.38	3.38	7.05	6.20	3.40
SR	-0.13	0.22	0.28	-0.2	0.31*	0.46	-0.12	0.31*	0.4*	-0.05	0.31**	0.33
Alpha	-0.12	0.33	0.45	-0.27	0.68*	0.95*	-0.19	0.73***	0.92***	0.02	0.68***	0.66**
Max.DD	74.28	38.93	47.79	76.70	53.65	51.28	57.75	48.91	25.19	58.43	46.63	24.96
Turnover	26.97	22.24	49.21	47.40	50.08	97.48	30.84	32.10	62.94	48.42	51.28	99.71
	Quality indicator 7			Quality indicator 8			ROE			ROIC		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.79	1.41	0.62	1.11	1.15	0.04	0.46	0.98	0.52	0.34	1.13	0.80
Std	6.80	7.63	4.23	6.81	6.74	2.74	8.39	5.90	4.82	9.59	5.44	6.41
SR	-0.04	0.25	0.32	0.13	0.15	0.02	-0.16	0.07	0.21	-0.19	0.17	0.29
Alpha	0.11	0.6*	0.5	0.37	0.4	0.03	-0.23	0.22	0.44	-0.28	0.3	0.58
Max.DD	61.99	66.40	58.10	53.62	55.43	62.09	84.64	49.08	52.44	86.88	44.73	56.26
Turnover	25.87	27.19	53.06	60.72	61.48	122.20	24.23	18.08	42.31	27.43	16.48	43.91
	Sales growth			Sales volatility			Technical indicator 1			Technical indicator 2		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.55	1.01	0.47	1.01	1.14	0.14	0.50	0.53	0.03	0.55	1.92	1.37
Std	7.60	7.43	5.79	6.93	6.32	4.78	7.04	6.50	4.13	7.96	6.21	4.88
SR	-0.14	0.07	0.21	0.07	0.16	0.07	-0.18	-0.18	0.01	-0.14	0.59***	0.6
Alpha	-0.34	0.47	0.81	0.25	0.27	0.01	-0.25	-0.22	0.03	-0.09	1.09***	1.17***
Max.DD	71.69	62.79	56.05	60.79	46.89	72.35	78.26	64.43	79.37	77.70	41.03	46.50
Turnover	25.59	25.56	51.15	22.03	13.01	35.04	162.08	162.74	324.82	137.55	139.11	276.66

Table A.2 Continued: Performance of Portfolios Formed on Individual Firm Characteristics. The table reports metrics for equally-weighted, monthly rebalanced portfolios formed on individual firm characteristics. The characteristics are ordered such that P_5 will produce a higher return than P_1 . P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of each firm characteristic. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports the average monthly return (Ave), the monthly standard deviation (Std), the annualized Sharpe Ratio (SR), the monthly alpha from four-factor model with the three Fama and French (1993) factors plus momentum, the maximum drawdown (Max. DD), and the monthly turnover. One, two and three asterisks denote statistical significance at the 10%, 5% and 1% significance level. For Sharpe ratios, the test is based on the difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period, using the Ledoit and Wolf (2008) test.

	Technical indicator 3			Technical indicator 4			Technical indicator 5			Trend in CAPEX to total assets		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.54	2.04	1.50	0.21	0.92	0.71	0.36	1.05	0.69	0.72	1.14	0.41
Std	8.21	6.07	5.39	7.37	6.17	4.47	8.14	5.54	5.32	6.71	6.86	3.39
SR	-0.13	0.67***	0.63	-0.31	0.03	0.33	-0.21	0.12	0.29	-0.07	0.14	0.21
Alpha	-0.08	1.14***	1.22***	-0.42	0.03	0.46	-0.23	0.08	0.31	-0.01	0.46*	0.47
Max.DD	78.40	37.55	47.12	78.24	64.41	54.42	83.18	52.41	52.21	56.49	50.55	51.11
Turnover	123.49	124.64	248.13	135.08	140.24	275.32	96.90	100.27	197.17	51.56	55.63	107.19
	Trend in asset turnover			Trend in cash flow from operations			Trend in profit margin			Value indicator 1		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	1.09	1.13	0.04	0.84	1.04	0.20	0.33	1.35	1.02	0.99	1.75	0.76
Std	6.65	7.08	3.73	6.75	7.54	3.71	8.41	6.98	4.59	6.27	7.93	4.63
SR	0.12	0.13	0.02	-0.01	0.08	0.1	-0.22	0.24	0.42	0.07	0.39**	0.42
Alpha	0.4	0.34	-0.07	0.11	0.39	0.28	-0.24	0.52	0.76*	0.22	0.97***	0.75**
Max.DD	60.45	52.69	58.63	58.07	61.71	58.09	77.77	57.95	44.45	58.99	61.08	69.70
Turnover	54.23	54.76	108.99	58.59	57.60	116.19	49.61	50.87	100.48	17.86	26.64	44.49
	Value indicator 2			Value indicator 3			Value indicator 4			Volatility indicator 1		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.75	0.93	0.17	0.65	0.96	0.31	0.44	1.45	1.01	0.51	1.31	0.81
Std	6.42	7.62	3.72	8.16	6.58	4.14	6.92	6.93	4.27	9.89	4.59	7.37
SR	-0.06	0.03	0.09	-0.09	0.05	0.13	-0.21	0.3	0.51*	-0.12	0.34*	0.28
Alpha	0	0.28	0.28	-0.1	0.3	0.4	-0.39	0.66**	1.05***	-0.09	0.42**	0.51
Max.DD	61.83	66.26	63.71	78.36	61.70	48.41	64.45	55.73	34.36	89.31	39.38	65.52
Turnover	31.07	25.83	56.90	28.64	35.73	64.37	18.06	20.79	38.85	23.94	20.73	44.67

Table A.2 Continued: Performance of Portfolios Formed on Individual Firm Characteristics. The table reports metrics for equally-weighted, monthly rebalanced portfolios formed on individual firm characteristics. The characteristics are ordered such that P_5 will produce a higher return than P_1 . P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of each firm characteristic. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports the average monthly return (Ave), the monthly standard deviation (Std), the annualized Sharpe Ratio (SR), the monthly alpha from four-factor model with the three Fama and French (1993) factors plus momentum, the maximum drawdown (Max. DD), and the monthly turnover. One, two and three asterisks denote statistical significance at the 10%, 5% and 1% significance level. For Sharpe ratios, the test is based on the difference between the Sharpe ratio of the portfolio and that of the IBOVESPA index over the same period, using the Ledoit and Wolf (2008) test.

	Volatility indicator 2			Volatility indicator 3		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.93	1.29	0.37	0.61	1.32	0.70
Std	9.78	4.48	7.27	10.12	4.54	7.50
SR	0.02	0.33*	0.13	-0.08	0.35*	0.24
Alpha	0.33	0.38*	0.05	0.05	0.43**	0.38
Max.DD	86.03	39.60	71.04	90.53	38.84	69.37
Turnover	46.84	49.97	96.82	31.59	32.39	63.98

Table A.3 Performance of Baseline Machine Learning Portfolios. The table reports out-of-sample performance metrics for equally-weighted quintile portfolios formed on predicted returns using machine learning methods. P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of predicted returns. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports average monthly return before (Ave) and after costs (Ave (net)), the monthly standard deviation (Std), the annualized Sharpe Ratio before (SR) and after costs (SR (net)), the maximum drawdown (Max. DD), the monthly turnover.

	OLS (out-of-sample $R^2 = -0.0338$)						LASSO (out-of-sample $R^2 = 0.0079$)						Ridge (out-of-sample $R^2 = 0.0079$)						PLS (out-of-sample $R^2 = 0.0017$)					
	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$
Ave	0.30	0.78	1.05	1.41	1.91	1.61	0.22	0.61	0.97	1.64	1.94	1.72	0.07	0.55	0.87	1.55	1.92	1.86	0.03	0.87	1.10	1.52	1.92	1.90
Ave (net)	0.17	0.58	0.84	1.21	1.77	0.97	0.13	0.46	0.81	1.49	1.84	1.16	-0.02	0.41	0.72	1.41	1.83	1.32	-0.05	0.74	0.95	1.38	1.84	1.38
Std	8.65	6.53	6.12	5.91	6.33	5.17	9.32	6.37	6.04	6.16	5.99	6.23	8.85	6.27	6.11	6.07	6.07	5.80	9.77	7.19	6.40	5.74	5.13	7.18
SR	-0.22	-0.04	0.1	0.32*	0.57**	1.08*	-0.24	-0.14	0.06	0.44**	0.63**	0.96	-0.31	-0.17	0.01	0.4**	0.61**	1.11	-0.3	0.01	0.13	0.4*	0.72**	0.91
SR (net)	-0.28	-0.15	-0.01	0.2	0.5**	0.65	-0.27	-0.22	-0.03	0.35*	0.57*	0.65	-0.34	-0.25	-0.08	0.31*	0.55*	0.79	-0.32	-0.06	0.05	0.31	0.66**	0.66
Max.DD	82.87	64.46	51.17	49.26	50.53	53.83	85.09	66.99	50.66	46.62	51.97	65.67	84.94	64.01	52.03	46.24	52.13	54.65	88.08	72.31	54.99	46.15	42.31	62.91
Turnover	86.24	131.89	140.14	136.71	91.57	177.81	59.37	99.31	107.42	102.54	66.35	125.72	57.04	93.55	100.10	96.80	63.30	120.35	48.54	86.83	96.85	92.99	53.78	102.32
	PCR (out-of-sample $R^2 = 0.0049$)						Bayes (out-of-sample $R^2 = -0.0010$)						Boost (out-of-sample $R^2 = 0.0075$)						RF (out-of-sample $R^2 = -0.0030$)					
	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$
Ave	0.26	1.22	0.94	1.32	1.69	1.43	0.16	0.80	1.22	1.41	1.84	1.68	0.25	0.69	1.17	1.23	1.85	1.60	0.15	0.96	1.14	1.33	1.86	1.71
Ave (net)	0.19	1.08	0.80	1.18	1.61	0.90	0.04	0.61	1.02	1.22	1.71	1.06	0.16	0.54	1.00	1.07	1.73	1.02	0.02	0.76	0.93	1.13	1.72	1.07
Std	9.75	7.45	6.09	5.98	4.70	7.21	8.96	6.74	6.10	6.23	5.87	5.92	9.30	7.22	6.25	5.83	5.51	6.58	9.29	6.25	6.30	6.21	5.84	6.10
SR	-0.21	0.16	0.05	0.27	0.61**	0.69	-0.27	-0.03	0.2	0.3	0.58**	0.98	-0.23	-0.08	0.17	0.22	0.62**	0.84	-0.27	0.05	0.15	0.26	0.59**	0.97
SR (net)	-0.24	0.1	-0.04	0.19	0.55**	0.43	-0.32	-0.13	0.09	0.2	0.5*	0.62	-0.26	-0.15	0.08	0.12	0.55**	0.54	-0.31	-0.06	0.04	0.15	0.51**	0.61
Max.DD	88.51	64.44	51.17	51.00	39.76	66.59	84.68	65.99	48.84	48.30	51.14	62.80	90.64	66.72	56.69	56.38	40.64	67.78	88.85	56.80	54.87	50.19	46.65	58.29
Turnover	51.50	88.16	97.04	91.42	54.34	105.84	79.14	123.98	131.50	124.97	86.91	166.06	62.38	100.81	110.16	104.19	74.78	137.16	85.21	132.78	140.03	135.79	94.92	180.13
	NN1 (out-of-sample $R^2 = -0.0802$)						NN2 (out-of-sample $R^2 = -0.0180$)						NN3 (out-of-sample $R^2 = 0.0033$)						NN4 (out-of-sample $R^2 = 0.0092$)					
	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$
Ave	0.28	1.02	1.24	1.29	1.61	1.33	0.17	1.18	1.14	1.28	1.67	1.49	-0.01	1.06	1.14	1.31	1.94	1.95	0.07	0.93	0.93	1.37	2.13	2.06
Ave (net)	0.15	0.82	1.03	1.09	1.47	0.69	0.05	0.98	0.93	1.08	1.52	0.86	-0.12	0.87	0.94	1.12	1.81	1.33	-0.03	0.76	0.74	1.18	2.02	1.48
Std	7.92	6.50	6.70	6.26	6.42	4.97	8.45	6.58	6.22	6.23	6.35	5.29	8.59	6.75	6.69	5.75	6.08	5.51	9.21	7.01	5.95	5.90	5.79	5.96
SR	-0.25	0.08	0.2	0.24	0.4*	0.93*	-0.28	0.17	0.15	0.23	0.44**	0.98	-0.35	0.1	0.14	0.27	0.61**	1.22*	-0.3	0.04	0.04	0.3	0.76**	1.2*
SR (net)	-0.31	-0.02	0.09	0.13	0.33	0.48	-0.33	0.06	0.04	0.12	0.36*	0.56	-0.4	0	0.04	0.15	0.54**	0.84	-0.34	-0.05	-0.07	0.19	0.69**	0.86
Max.DD	83.45	51.68	55.44	50.61	55.14	39.77	85.13	53.11	52.62	57.61	50.66	52.43	87.73	55.46	54.58	49.93	49.04	55.28	87.78	66.39	51.50	50.36	48.07	55.61
Turnover	89.28	135.20	143.34	136.44	93.85	183.13	84.41	133.03	140.29	137.10	94.14	178.55	77.01	126.49	134.69	130.56	86.12	163.13	66.94	115.07	126.05	124.85	76.47	143.41
	NN5 (out-of-sample $R^2 = 0.0093$)						ENS_ML (out-of-sample $R^2 = 0.0096$)																	
	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$	P_1	P_2	P_3	P_4	P_5	$P_5 - P_1$												
Ave	0.17	0.89	1.13	1.38	1.89	1.72	0.02	0.73	1.26	1.42	2.02	2.00												
Ave (net)	0.07	0.72	0.94	1.20	1.78	1.15	-0.08	0.55	1.06	1.24	1.90	1.41												
Std	9.50	7.00	6.18	5.76	5.46	6.38	9.21	6.72	6.52	5.76	5.76	6.11												
SR	-0.25	0.01	0.15	0.31	0.65**	0.93	-0.32	-0.07	0.21	0.34	0.69**	1.13*												
SR (net)	-0.29	-0.07	0.05	0.21	0.58**	0.63	-0.36	-0.16	0.11	0.23	0.62**	0.8												
Max.DD	87.82	65.87	54.12	52.95	45.94	62.42	89.79	63.68	49.43	52.64	47.54	59.00												
Turnover	60.64	109.11	122.50	117.74	70.70	131.34	69.39	119.62	131.18	123.65	79.29	148.68												

Table A.4 Performance of Baseline Machine Learning Portfolios (50 basis points transaction costs). The table reports out-of-sample performance metrics for equally-weighted quintile portfolios formed on predicted returns using machine learning methods, considering a conservative transaction cost assumption of 50 basis points per round trip transaction. P_1 (P_5) is the portfolio formed on the lowest (highest) quintile of predicted returns. $P_5 - P_1$ is a hedge portfolio long the stocks in P_5 and short the stocks in P_1 . The table reports average monthly return before (Ave) and after costs (Ave (net)), the monthly standard deviation (Std), the annualized Sharpe Ratio before (SR) and after costs (SR (net)), the maximum drawdown (Max. DD), the monthly turnover. The table also reports, for each machine learning method, the out-of-sample (OOS) R^2 .

	OLS (OOS $R^2 = -0.0338$)			LASSO (OOS $R^2 = 0.0079$)			Ridge (OOS $R^2 = 0.0079$)			PLS (OOS $R^2 = 0.0017$)		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.30	1.91	1.61	0.22	1.94	1.72	0.07	1.92	1.86	0.03	1.92	1.90
Ave (net)	-0.13	1.45	0.35	-0.07	1.61	0.72	-0.22	1.61	0.90	-0.22	1.65	1.02
Std	8.65	6.33	5.17	9.32	5.99	6.23	8.85	6.07	5.80	9.77	5.13	7.18
SR	-0.22	0.57**	1.08*	-0.24	0.63**	0.96	-0.31	0.61**	1.11	-0.3	0.72**	0.91
SR (net)	-0.4	0.32	0.23	-0.35	0.43*	0.4	-0.42	0.43*	0.54	-0.38	0.53**	0.49
Max.DD	86.27	52.06	57.21	88.19	52.70	67.42	87.64	52.83	56.08	89.35	43.08	64.37
Turnover	86.24	91.57	177.81	59.37	66.35	125.72	57.04	63.30	120.35	48.54	53.78	102.32
	PCR (OOS $R^2 = 0.0049$)			Bayes (OOS $R^2 = -0.0010$)			Boost (OOS $R^2 = 0.0075$)			RF (OOS $R^2 = -0.0030$)		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.26	1.69	1.43	0.16	1.84	1.68	0.25	1.85	1.60	0.15	1.86	1.71
Ave (net)	0.01	1.42	0.53	-0.23	1.40	0.48	-0.06	1.47	0.54	-0.28	1.39	0.44
Std	9.75	4.70	7.21	8.96	5.87	5.92	9.30	5.51	6.58	9.29	5.84	6.10
SR	-0.21	0.61**	0.69	-0.27	0.58**	0.98	-0.23	0.62**	0.84	-0.27	0.59**	0.97
SR (net)	-0.3	0.41*	0.26	-0.42	0.32	0.28	-0.35	0.38	0.28	-0.43	0.31	0.25
Max.DD	89.86	40.57	68.31	88.93	52.28	65.90	91.93	41.75	70.30	91.15	47.62	64.63
Turnover	51.50	54.34	105.84	79.14	86.91	166.06	62.38	74.78	137.16	85.21	94.92	180.13
	NN1 (OOS $R^2 = -0.0802$)			NN2 (OOS $R^2 = -0.0180$)			NN3 (OOS $R^2 = 0.0033$)			NN4 (OOS $R^2 = 0.0092$)		
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$
Ave	0.28	1.61	1.33	0.17	1.67	1.49	-0.01	1.94	1.95	0.07	2.13	2.06
Ave (net)	-0.17	1.14	0.05	-0.25	1.19	0.23	-0.39	1.51	0.76	-0.26	1.75	0.98
Std	7.92	6.42	4.97	8.45	6.35	5.29	8.59	6.08	5.51	9.21	5.79	5.96
SR	-0.25	0.4*	0.93*	-0.28	0.44**	0.98	-0.35	0.61**	1.22*	-0.3	0.76***	1.2*
SR (net)	-0.45	0.15	0.03	-0.45	0.18	0.15	-0.51	0.37*	0.48	-0.42	0.53**	0.57
Max.DD	86.68	57.15	44.45	88.19	52.24	55.96	90.89	50.84	58.43	90.04	49.06	58.05
Turnover	89.28	93.85	183.13	84.41	94.14	178.55	77.01	86.12	163.13	66.94	76.47	143.41
	NN5 (OOS $R^2 = 0.0093$)			ENS_{ML} (OOS $R^2 = 0.0096$)								
	P_1	P_5	$P_5 - P_1$	P_1	P_5	$P_5 - P_1$						
Ave	0.17	1.89	1.72	0.02	2.02	2.00						
Ave (net)	-0.14	1.53	0.69	-0.33	1.62	0.89						
Std	9.50	5.46	6.38	9.21	5.76	6.11						
SR	-0.25	0.65**	0.93	-0.32	0.69**	1.13*						
SR (net)	-0.36	0.42*	0.38	-0.45	0.46*	0.5						
Max.DD	89.90	47.06	64.43	91.43	48.69	61.18						
Turnover	60.64	70.70	131.34	69.39	79.29	148.68						

Table A.5 Risk and return statistics of long-short strategies obtained with different machine learning methods (50 basis points transaction costs). The table reports out-of-sample performance metrics for different long-short strategies formed on predicted returns using machine learning methods, considering a conservative transaction cost assumption of 50 basis points per round trip transaction. Panel A shows results for traditional long-short strategies which are long (short) the highest (lowest) decile of predicted returns with each machine learning method. Panel B shows results for Equal Risk Contribution (ERC) long-short strategies, which equalize the risk contributions of the long and short legs. The table reports the average monthly return before (Ave) and after costs (Ave (net)), the monthly standard deviation (Std), the annualized Sharpe Ratio before (SR) and after costs (SR (net)), the maximum drawdown (Max. DD), the average monthly turnover, and the average leverage.

Panel A: Traditional long-short strategies														
	OLS	LASSO	Ridge	PLS	PCR	Bayes	Boost	RF	NN1	NN2	NN3	NN4	NN5	ENS_{ML}
Ave	1.61	1.72	1.86	1.90	1.43	1.68	1.60	1.71	1.33	1.49	1.95	2.06	1.72	2.00
Ave (net)	0.35	0.72	0.90	1.02	0.53	0.48	0.54	0.44	0.05	0.23	0.76	0.98	0.69	0.89
Std	5.17	6.23	5.80	7.18	7.21	5.92	6.58	6.10	4.97	5.29	5.51	5.96	6.38	6.11
SR	1.08*	0.96	1.11	0.91	0.69	0.98	0.84	0.97	0.93*	0.98	1.22*	1.2*	0.93	1.13*
SR (net)	0.23	0.4	0.54	0.49	0.26	0.28	0.28	0.25	0.03	0.15	0.48	0.57	0.38	0.5
Max.DD	57.21	67.42	56.08	64.37	68.31	65.90	70.30	64.63	44.45	55.96	58.43	58.05	64.43	61.18
Turnover	177.81	125.72	120.35	102.32	105.84	166.06	137.16	180.13	183.13	178.55	163.13	143.41	131.34	148.68
Leverage	2.00	2.00	1.96	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00

Panel B: Equal risk contribution long-short strategies														
	OLS	LASSO	Ridge	PLS	PCR	Bayes	Boost	RF	NN1	NN2	NN3	NN4	NN5	ENS_{ML}
Ave	1.76	1.72	1.76	1.72	1.51	1.87	1.56	1.74	1.41	1.50	1.98	1.98	1.71	1.95
Ave (net)	0.59	0.83	0.91	1.00	0.82	0.75	0.68	0.65	0.20	0.34	0.91	1.04	0.83	1.00
Std	3.95	4.18	4.11	4.35	4.02	4.20	3.84	3.96	3.67	3.88	3.86	4.05	4.08	3.81
SR	1.42**	1.29**	1.32**	1.16*	1.01	1.43**	1.17*	1.32*	1.21**	1.21**	1.62**	1.52**	1.27**	1.56**
SR (net)	0.39	0.55	0.6	0.59	0.42	0.51	0.38	0.37	0.07	0.17	0.66	0.71	0.52	0.7
Max.DD	33.60	39.74	32.28	33.04	41.34	39.76	39.43	44.75	29.06	35.71	30.41	27.79	33.52	34.02
Turnover	172.59	119.84	114.54	93.87	92.81	161.66	125.63	164.15	178.79	172.26	155.44	133.29	121.75	136.83
Leverage	1.83	1.79	1.76	1.68	1.61	1.84	1.68	1.72	1.86	1.82	1.80	1.75	1.73	1.71

Table A.6 Risk and return statistics of long-short strategies obtained using an ensemble of machine learning methods, or by combining long-short strategies obtained with various machine learning methods (50 basis points transaction costs), considering a conservative transaction cost assumption of 50 basis points per round trip transaction. The table reports average monthly return before (Ave) and after costs (Ave (net)), the monthly standard deviation (Std), the annualized Sharpe Ratio before (SR) and after costs (SR (net)), the maximum drawdown (Max. DD), the average monthly turnover, and the average leverage.

	Panel A. Long-short strategies using an ensemble of machine learning models		Panel B. Long-short strategies combining portfolios obtained with different machine learning models	
	ENS_{ML-EW}	ENS_{ML-ERC}	MS-EW	MS-ERC
Ave	2.04	1.99	1.81	2.06
Ave (net)	0.93	1.05	0.77	1.23
Std	6.18	3.78	5.93	2.94
SR	1.14	1.59**	1.06	2.10**
SR (net)	0.52	0.74	0.45	1.13*
Max.DD	61.18	34.02	62.68	19.27
Turnover	147.65	136.83	133.46	117.12
Leverage	2.00	1.71	1.99	1.66

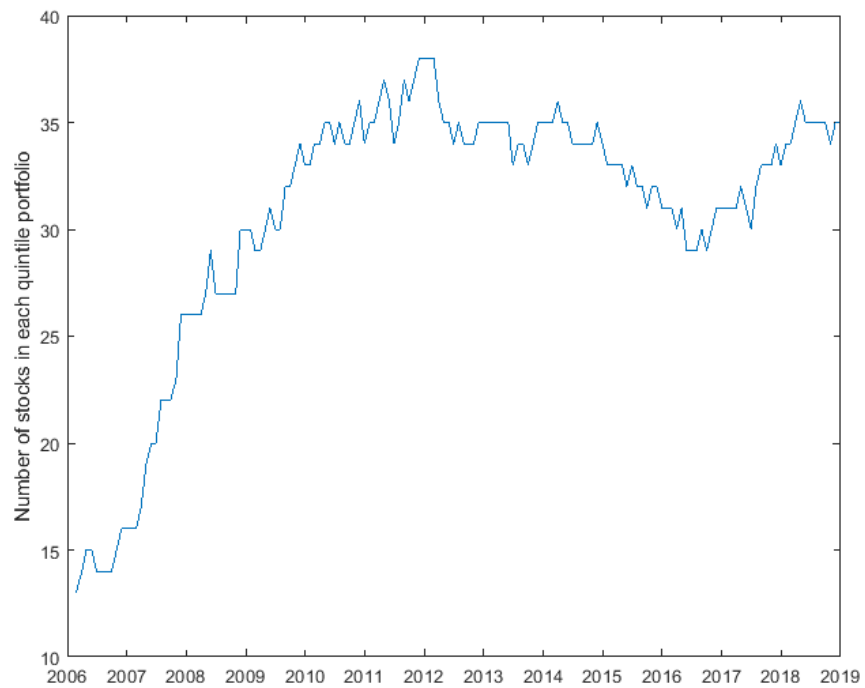


Figure A.1: Number of stocks in each quintile portfolio.

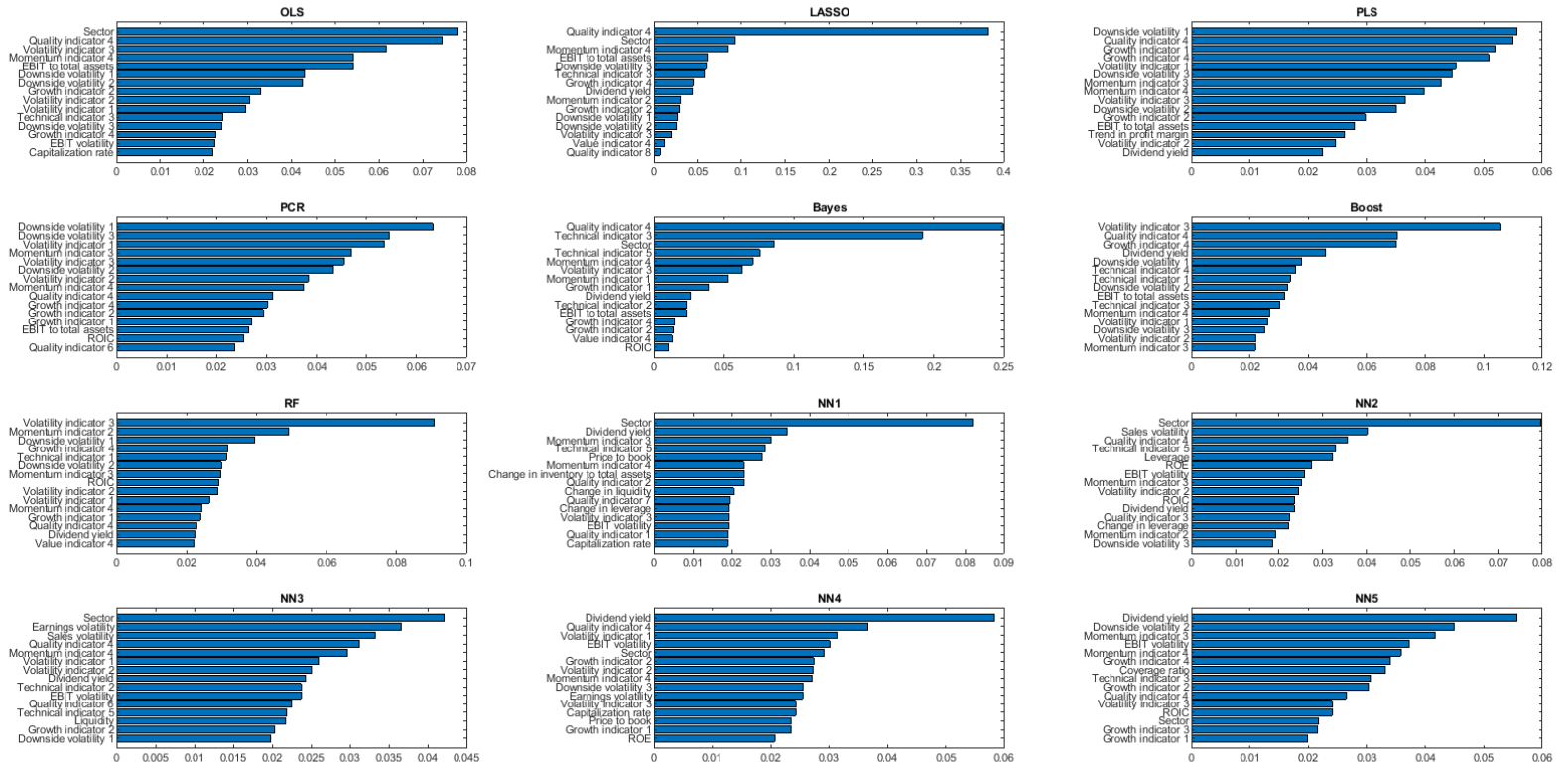


Figure A.2: Variable importance measure by model

Obs.: We omit the graph for Ridge as it is essentially identical to that of LASSO.