Insper
Programa de Mestrado Profissional em Economia

Nelly Pires Colnaghi

A Parametric Portfolio Optimization Model using Signals Generated via Machine Learning

São Paulo
2021

Nelly Pires Colnaghi

A Parametric Portfolio Optimization Model using Signals Generated via Machine Learning

Dissertação apresentada ao Programa de Mestrado Profissional em Economia do Insper como parte dos requisitos para obtenção do título de Mestre em Economia.

Área de concentração: Economia dos Negócios

Linha de pesquisa: Finanças

Insper
Programa de Mestrado Profissional em Economia

Supervisor: Adriana Bruscato

São Paulo
2021

Colnaghi, Nelly Pires.

A Parametric Portfolio Optimization Model using Signals Generated via Machine Learning. / Nelly Pires Colnaghi. – São Paulo, 2021.

86f.

Dissertação (Mestrado) – Insper

Programa de Mestrado Profissional em Economia, 2021.

Orientador: Adriana Bruscato

1. Machine Learning. 2. Otimização de Portfólio. 3. Análise Fundamentalista. I. Adriana Bruscato. II. Insper. III. Departamento de Economia. IV. A Parametric Portfolio Optimization Model using Signals Generated via Machine Learning

Nelly Pires Colnaghi

A Parametric Portfolio Optimization Model using Signals Generated via Machine Learning

Dissertação apresentada ao Programa de Mestrado Profissional em Economia do Insper como parte dos requisitos para obtenção do título de Mestre em Economia.

Área de concentração: Economia dos Negócios

Linha de pesquisa: Finanças

DATA DE APROVAÇÃO: São Paulo, 02 de fevereiro de 2021.

BANCA EXAMINADORA

Adriana Bruscato
Orientador

Michael Viriato Araújo
Convidado Interno

Maurício Mesquita
Convidado Externo

À minha família e amigos.

ACKNOWLEDGEMENTS

Agradeço a Prof. Dra. Adriana Bruscato, orientadora desta dissertação, pelo direcionamento e suporte no desenvolvimento deste trabalho.

A minha família, Jorge e Lecy Colnaghi, e aos meus irmãos, Ana Elisa e Jorge, por todo o apoio e incentivo durante o mestrado. Contar com vocês foi fundamental.

Aos meus colegas de turma, pelo companheirismo durantes esses dois anos de mestrado. Alexandre, Gabriel e Débora, meus queridos amigos e parceiros de estudos. Obrigada por terem feito esses dois anos muito especiais.

ABSTRACT

This work consists on a portfolio optimization model based on trading signals generated from financial ratios for the identification of mispriced stocks. The signals stem from machine learning models, fitted to predict future stock returns using the aforementioned indicators. These models comprehend classifiers such as Logit, Random Forest, Support Vector Classifier (SVC), k-NN, and Naïve Bayes. The trading signals produced specify which stocks should outperform the market median, and which ones should underperform. To incorporate these signals into the optimization model, the parameterization technique proposed by Brandt, Santa-Clara and Valkanov (2009) was chosen, allowing the weights of the portfolio stocks to be expressed as a function of each stock characteristic, i.e. the corresponding trading signals. An empirical analysis using data from the Brazilian stock market was designed to verify the performance of the proposed optimization model. The results show that the selected machine learning models attain an average out-of-sample accuracy of 64% in predicting stock price direction. In addition, the optimal portfolio outperforms all the benchmarks in terms of risk-adjusted returns. Finally, the results also demonstrate that the combination of machine learning and the approach developed by Brandt, Santa-Clara and Valkanov (2009) is superior to a portfolio built solely using fundamental ratios without converting them into trading signals through machine learning models.

Keywords: Machine Learning. Portfolio Optimization. Fundamental Analysis.

RESUMO

Este trabalho consiste em um modelo de otimização de portfólio baseado em sinais de *trading*, gerados por indicadores financeiros para a identificação de ações mal precificadas. Os sinais partem de modelos de *machine learning*, ajustados para prever os retornos futuros das ações a partir dos indicadores mencionados. Esses modelos compreendem classificadores como Logit, Random Forest, Support Vector Classifier (SVC), k-NN e Naïve Bayes. Os sinais de *trading* produzidos especificam quais ações devem superar o desempenho mediano do mercado e quais devem apresentar um desempenho inferior. Para incorporar esses sinais ao modelo de otimização, foi escolhida a técnica de parametrização proposta por Brandt, Santa-Clara and Valkanov (2009), permitindo que os pesos das ações do portfólio sejam expressos em função das características de cada ação, ou seja, os respectivos sinais de *trading*. Uma análise empírica utilizando dados do mercado de ações brasileiro foi elaborada para verificar o desempenho do modelo de otimização proposto. Os resultados mostram que os modelos de *machine learning* selecionados atingem uma precisão média fora da amostra de 64% na previsão da direção do preço das ações. Além disso, o portfólio otimizado supera todos os benchmarks em termos de retorno ajustado ao risco. Por fim, os resultados também demonstram que a combinação de *machine learning* e a abordagem desenvolvida por Brandt, Santa-Clara and Valkanov (2009) é superior a um portfólio construído apenas usando indicadores fundamentalistas, sem convertê-los em sinais de *trading* através de modelos de *machine learning*.

Palavras-chave: Machine Learning. Otimização de Portfólio. Análise Fundamentalista.

EXECUTIVE SUMMARY

Despite the solid theoretical foundations behind the semi-strong form of EMH, there is empirical evidence that supports the use of fundamental analysis to actively manage investment portfolios, as discussed in Kosowski, Naik and Teo (2007). In this sense, fundamental analysis is one of the most commonly used methods for estimating the price behavior of stocks, as emphasized by Baresa, Bogdan and Ivanovic (2013). In addition, machine learning techniques have become increasingly more popular in the last few years, providing encouraging results, according to Jordan and Mitchell (2015). Therefore, the combination of fundamental analysis with machine learning in the context of portfolio optimization seems appropriate and likely to be successful.

Therefore, the purpose of the present work involves developing a portfolio optimization model based on signals inferred from financial ratios, given their predictive power to explain stock returns. The approach advocated relies on machine learning methods, for they have been successfully used in previous studies; see Huang, Nakamori and Wang (2005), Kara, Boyacioglu and Baykan (2011), Ballings et al. (2015), and Tsai et al. (2011). These machine learning models comprehend classifiers such as Logit, Random Forest, Support Vector Classifier (SVC), k-NN, and Naïve Bayes. The trading signals produced specify which stocks should outperform the market median, and which ones should underperform.

Thus, to build a portfolio optimization incorporating these signals, the work is based on the method introduced by Brandt, Santa-Clara and Valkanov (2009), where the authors present an equation allowing the parameterization of the weights of the portfolio stocks as a function of the characteristics of each stock. In the case of this present work, the characteristics consists on their trading signals. This way, the equation is incorporated into the optimization model, with the objective of maximizing the portfolio expected utility.

With respect to the data employed in the present work, the database comprises a total of 46 stocks of Brazilian companies listed on B3, which are compiled in Appendix A. These stocks were selected from the composition of the Bovespa index portfolio, which provided an initial universe of stocks that was later filtered in terms of availability of historical financial ratios. To validate the effectiveness of the model, the portfolio was backtested against Bovespa index and also against other benchmarks designed to assess the individual value added by the parameterization and machine learning models considered herein.

Financial ratios have great informational content on the financial health and the perspectives of a given company. It is reasonable to assume that applying them as inputs for portfolio optimization should generate excess returns instead of simply using the market value of each company as a weighting criterion. Besides, given that optimization models using financial ratios are scarce in the literature, this dissertation provides a positive contribution.

The main contribution of this work is to deliver an application of machine learning techniques to estimate trading signals extracted from financial ratios, using this strategy to generate alpha in asset allocation. Ultimately, it should be noted that no similar work joining machine learning and optimization models has been implemented for the Brazilian stock market, further justifying the decisions made here.

Briefly, the results of the work show that the selected machine learning models attain an average out-of-sample accuracy of 64% in predicting stock price direction. In addition, the optimal portfolio outperforms all the benchmarks in terms of risk-adjusted returns. Also, the results also demonstrate that the combination of machine learning and the approach developed by Brandt, Santa-Clara and Valkanov (2009) is superior to a portfolio built solely using fundamentals without converting them into trading signals.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Fundamental analysis is often seen with some skepticism by investors, largely due the conflicts with the semi-strong efficient market hypothesis (EMH), as suggested in Bernstein (1975). The hypothesis posits that because all public information is reflected on current stock prices, investors cannot generate excess returns by applying technical or fundamental analysis (BODIE; KANE; MARCUS, 2014). Therefore, only information that is not readily available to the public can help investors boost their returns and outperform the market.

Despite the solid theoretical foundations behind the semi-strong form of EMH, there is empirical evidence that supports the use of fundamental analysis to actively manage investment portfolios. For instance, several hedge funds and prominent investors are known to persistently beat the market and, thus, generate alpha in excess of the return required by their exposure to systematic risk factors; see Kosowski, Naik and Teo (2007). Equivalently, the authors demonstrate that active portfolio management based on fundamental analysis is capable of generating alpha.

In this sense, it should be noted that EMH in semi-strong form does not claim that an investor cannot outperform a given benchmark – in fact, it states that the investor cannot generate alpha in relation to the risk factors to which it is exposed (BODIE; KANE; MARCUS, 2014). In the case of the Brazilian market, there are even doubts as to whether EMH is valid in its weak form in all equity markets segments, as discussed by Ely (2011).

Baresa, Bogdan and Ivanovic (2013) suggest that fundamental analysis is widely used to verify the impact of micro and macroeconomic factors on the firm business with the purpose of predicting future economic and financial impacts. For that reason, the authors claim that fundamental analysis is one of the most commonly used methods for estimating the price behavior of stocks. Moreover, as emphasized by Spooner (1984), fundamental analysis is an assessment of financial and non-financial factors that involve the company's environment to measure its intrinsic value. Wadhwa (2019) also reinforces the argument that financial ratios have the capacity to summarize the fundamentals of any firm and, thus, possess compelling analytical abilities.

In this context, the objective of the present work involves developing a portfolio optimization model based on signals emerging from financial ratios, given their predictive power to explain stock returns. The approach advocated relies on machine learning methods, for they have been successfully used in previous studies; see Huang, Nakamori and Wang (2005), Kara, Boyacioglu and Baykan (2011), Ballings et al. (2015), and Tsai et al. (2011). These machine learning models comprehend classifiers such as Logit, Random Forest, Support Vector Classifier (SVC), k-NN, and Naïve Bayes. The trading signals produced specify which stocks should outperform the market median, and which ones should underperform.

According to Jordan and Mitchell (2015), machine learning can be described as a collection of models that automatically improve through experience, joining techniques from Computer Science and Statistics. Due to its success and recent technological advance, it has experienced significant growth in the last few years.

Thus, to build a portfolio optimization incorporating these signals, the work is based on the method introduced by Brandt, Santa-Clara and Valkanov (2009), where the authors present an equation allowing the parameterization of the weights of the portfolio stocks as a function of the characteristics of each stock. In the case of this present work, the characteristics consists on their trading signals. This way, the equation is incorporated into the optimization model, with the objective of maximizing the portfolio expected utility.

Indisputably, the approach proposed by Brandt, Santa-Clara and Valkanov (2009) is computationally simple, can be easily modified, and offers robust performance in- and out-of-sample. In addition, it efficiently permits the optimization of portfolios even when the investment universe is large and diverse. Furthermore, in their experiments using a CRRA utility function and, as firm's characteristics, the market capitalization, book-to-market ratio, and one-year lagged returns, the results obtained are sound. Indeed, the certainty-equivalent gains from incorporating the firm characteristics, relative to holding the market portfolio, is an annualized 11% return. For these reasons, Brandt, Santa-Clara and Valkanov (2009) is the main reference for the present work.

Medeiros, Passos and Vasconcelos (2014), Lyle and Yohn (2020) and Hand and Green (2011) are examples of papers that used this parameterization method as a reference for the construction of their models. For instance, Lyle and Yohn (2020) implement fundamental analysis integrated with mean-variance portfolio optimization by using a fundamentals-based technique to estimate expected returns, coupled with a state-of-the-art nonlinear shrinkage estimator to estimate the covariance matrix of individual stock returns and build mean-variance optimized portfolios. Medeiros, Passos and Vasconcelos (2014) and Hand and Green (2011) follow similar approaches in their study.

With respect to the data employed in the present work, the database comprises a total of 46 stocks of Brazilian companies listed on B3, which are compiled in Appendix A. These stocks were selected from the composition of the Bovespa index portfolio, which provided an initial universe of stocks that was later filtered in terms of availability of historical financial ratios. To validate the effectiveness of the model, the portfolio was backtested against Bovespa index and also against other benchmarks designed to assess the individual value added by the parameterization and machine learning models considered herein.

In that sense, financial ratios have great informational content on the financial health and the perspectives of a given company. It is reasonable to assume that applying them as inputs for portfolio optimization should generate excess returns instead of simply using the market value of each company as a weighting criterion. Besides, given that optimization models using financial ratios are scarce in the literature, this dissertation provides a positive contribution.

The main contribution of this work is to deliver an application of machine learning techniques to estimate trading signals extracted from financial ratios, using this strategy to generate gains in asset allocation. Ultimately, it should be noted that no similar work joining machine learning and optimization models has been implemented for the Brazilian stock market, further justifying the decisions made here.

Briefly, the results of the work show that the selected machine learning models attain an average out-of-sample accuracy of 64% in predicting stock price direction. In addition, the optimal portfolio outperforms all the benchmarks in terms of risk-adjusted returns. Also, the results also demonstrate that the combination of machine learning and the approach developed by Brandt, Santa-Clara and Valkanov (2009) is superior to a portfolio built solely using fundamentals without converting them into trading signals.

Finally, the work is structured as follows. In chapter 2, it is presented the literature review on the theoretical framework that underlies the use of fundamental analysis, the machine learning models and portfolio optimization models. Next, data and methodology are discussed in chapter 3, while the presentation and discussion of the results obtained are carried out in chapter 4. The main conclusions are summarized in chapter 5.

2 LITERATURE REVIEW

In this section, the literature review to build the theoretical foundations for this work is detailed. The themes are divided into fundamental analysis, machine learning models, and portfolio optimization model.

2.1 Fundamental Analysis

In this section, it is explored the relevance of the fundamental analysis in forecasting future stock returns. Also, the cross-sectional analysis of these returns in terms of financial ratios is presented and debated.

As claimed by Spooner (1984), fundamental analysis is an assessment of financial and nonfinancial factors that influence the company's environment to measure its intrinsic value. Financial ratios are part of this context, since they provide a snapshot of the company performance, cash flow generation, and balance sheet, transforming accounting data into ratios (HORRIGAN, 1965; MURESAN; WOLITZER, 2004). Also, as put by Wadhwa (2019), financial ratios have the capacity to explain the fundamentals of an entity for any sector or nature and thus possess pervasive analytical abilities. Therefore, these indicators convey information regarding current and future performance of the company, making them a relevant technique to be employed for financial analysis, in which Whittington (1980) identified two principal uses: the normative use of the measurement of a firm's ratio compared with a benchmark and the application for predictive purposes.

Baresa, Bogdan and Ivanovic (2013) suggest that the fundamental analysis is widely used to verify the impact of micro and macroeconomic factors on the business of the corporation for the purpose of predicting future economic and financial effects. For that reason, the authors affirm that fundamental analysis is one of the most commonly used methods for estimating price movements of securities.

The effectiveness of the fundamental analysis has also been explored in several econometric studies and published in specialized journals. Table 1 presents a compilation of recent works on the topic, carried out based on the stock market of different countries, and using distinct techniques. Next, it is discussed some of the studies already carried out, addressing their similarities and differences.

Abrokwa and Nkansah (2015), one of the most recent of these studies, explores the fundamental analysis effectiveness through the study of the predictive ability of five firm's variables (dividends per share, dividend price ratio, price earnings ratio, book to market ratio, and size) on common stock returns using the emerging market country of Ghana as a dataset. The authors implemented a Fama and Macbeth (1973) model with cross-sectional regression setup for the analysis. Their results bring evidence that, of the five variables of the study, only dividends per share had a significant predictive effect on stock returns.

Table 1 – Studies involving fundamental analysis.

Methods	**Papers**
Regression Model	Abrokwa and Nkansah (2015) - Ghana Avramov, Kaplanski and Subrahmanyam (2020) - United States Kheradyar, Ibrahim and Nor (2011) - Malasya LaPorta (1996) - United States Lewellen (2004) - United States Rapach and Wohar (2005) - United States Sehgal (2010) - BRICKS
Other Discussions and Techniques	Abarbanell and Bushee (1998) - United States | Portfolio Optimization] Shen and Tzeng (2015) - Taiwan | Portfolio Optimization Grossman and Stiglitz (1980) - United States | EMH Analysis Lim and Brooks (2011) - Worldwide | EMH Analysis Lo (2004) - United States | EMH Analysis Ely (2011) - Brazil | EMH Analysis Kosowski, Naik and Teo (2007) - Global| Active Portfolio Management Hsu, Kalesnik and Kose (2019) - Worldwide* | Multi-factor equity strategy Kumar (2013) - India | Questionnaire Survey

*United States, Global Developed, Japan, Europe, and Asia Pacific excluding Japan.

Similar to this study, LaPorta (1996) explored this topic also by performing a cross-sectional regression. The author discusses the idea that stock returns have a predictable component, seeking to answer the question of why these returns can be predicted by using survey data on the expectations of stock market analysts, using for that the Fama and Macbeth (1973) procedure, just like Avramov, Kaplanski and Subrahmanyam (2020). The author's results suggest that investment strategies seeking to exploit errors in analysts' forecasts earn superior returns because expectations about future growth in earnings are too extreme – that is, less sophisticated investors may become excessively pessimistic about future earnings growth after a series of lackluster ones. In subsequent periods, however, the prices of undervalued stocks rise as these investors are positively surprised and consequently revise the expectations for their future growth upwards.

Still using regression models for the US stock market, Lewellen (2004) and Rapach and Wohar (2005) carried out this study with a focus on forecasting returns with financial ratios, both using dividend yield and price-earnings multiples in their analysis. In their respective windows of analysis (1946-2000 and 1872–1997), both works found significant evidence of return predictability, suggesting the effectiveness of fundamental analysis.

The other studies listed in the first block in the table focus on regression models based on fundamental analysis applied to the stock exchange of different countries. In a comparative way, all of these attempt to verify the efficiency of forecasting the return on stocks using fundamental analysis in their respective markets and use econometric models for this purpose.

Kheradyar, Ibrahim and Nor (2011) use generalized least squares (GLS) techniques to estimate the predictive regressions in form of simple and multiple models of panel data, verifying that financial ratios are able to enhance stock return predictability when the ratios are combined in the multiple regression model in Malaysia stock exchange. Finally, Sehgal (2010), who studies the BRICKS countries (BRICS plus South Korea), uses two performance evaluation criteria to evaluate the efficacy of popular value indicators: i) root mean squared error, and ii) Thail inequality coefficient. The author then verifies that price-to-book value is the best standalone price multiple for the Asian economies (India, China and South Korea) while price-to-earnings does a better job for equity valuation in case of Brazil and South Africa.

The second block in the table focuses on different approaches to discussing the relevance of fundamental analysis. The first two works listed, Abarbanell and Bushee (1998) and Shen and Tzeng (2015), explore portfolio optimization based on fundamental analysis. While Abarbanell and Bushee (1998) verifies evidence that the fundamental-based signals provide information about stock future returns, Shen and Tzeng (2015) proposes a dominance-based rough set approach (DRSA) methodology – extended data mining technique for ordinal classification problems in a data set – using ratios as financial criteria for the purpose of stock selection for the Taiwan stock market.

Next we have the works of Grossman and Stiglitz (1980), Lim and Brooks (2011) and Lo (2004), which examine the Efficient Market Hypothesis (EMH). Lim and Brooks (2011) conducted a literature review on the topic, defining the concepts of weak, semi-strong and strong form, respectively:

- Weak form - says that asset prices today should reflect historical prices;
- Semi-strong form - says that, besides reflecting historical prices, prices today also reflect current public information;
- Strong form - says that, in addition to current prices reflecting price history and contemporary public information, contemporary private information would also be reflected in market prices.

Thus, according to the strong form of EMH, there would be no possibility of insider trading, since the price would already reflect all possible information, which is something that does not happen in practice.

The paper by Grossman and Stiglitz (1980) presents a paradox, which is nowadays known as Grossman-Stiglitz Paradox, reinforcing the hypothesis that markets cannot be efficient in the semi-strong form. According to the paradox, for a market to be efficient, prices need to reflect their fundamentals. For this to occur, it is necessary for investors to carry out research and trade based on that. However, if the market were efficient, there would be no incentive to conduct financial research, as the theory says that, in these situations, fundamental analysis is not capable of generating alpha in relation to the risk factors and, thus, reward analysts. Thus, the paradox, also represented in Figure 1, supports the hypothesis that markets are not efficient.

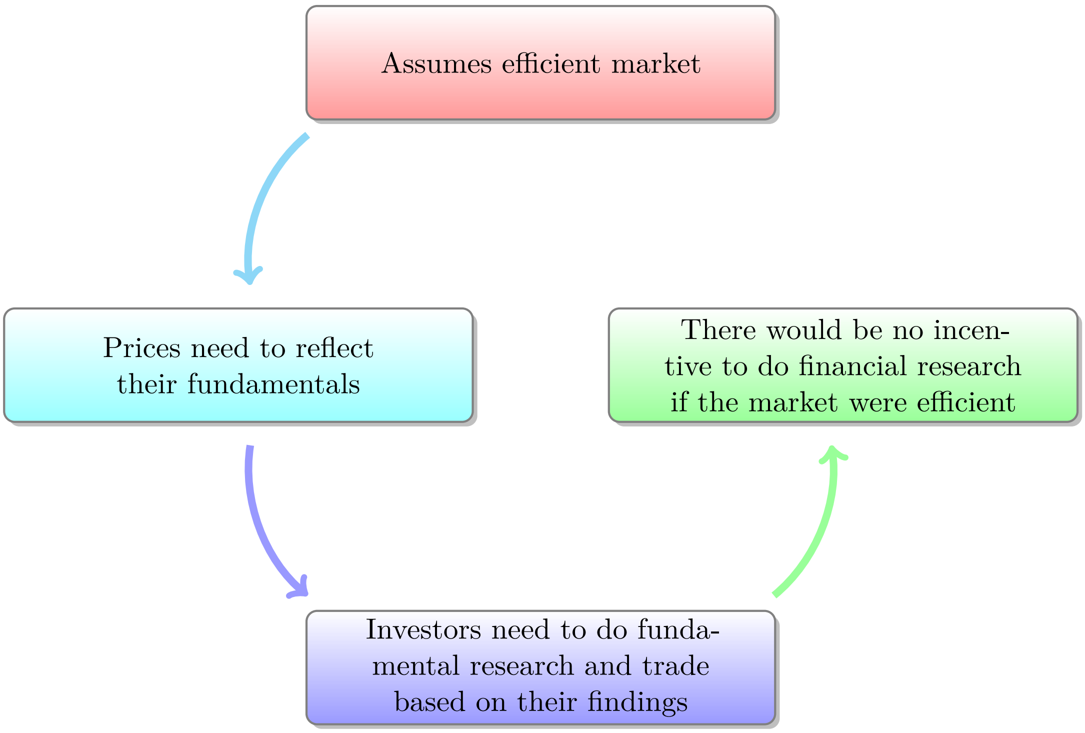

Figure 1 – Grossman and Stiglitz paradox.

The adaptation covered in Lim and Brooks (2011) and Lo (2004) is an evolution of EMH, known as the Adaptive Markets Hypothesis (AMH), explored in both works. In this evolutionary hypothesis, it is considered that market efficiency is not a binary variable, that is, there are more states other than total efficiency or total inefficiency. The hypothesis presents as an argument that the markets have degrees of efficiency and that these vary over time. This new theory is consistent with fundamental analysis, since, in times of crisis and high market volatility, asset prices detach from their fundamentals, compromising the market efficiency. This creates an opportunity for a fundamental-based analyst to reap gains. In more stable moments, with a more efficient market, the analyst would have fewer opportunities to generate value, since the fundamentals would be more in line with asset prices. In this hypothesis, it is also stated that no market ever reaches its maximum efficiency level.

The work by Ely (2011) is another example of a study that analyzes the validity of EMH in its weak form, with application to the Brazilian market. In his results, the author questions its validity in all equity markets segments. Reinforcing this idea, there is also the empirical evidence of several hedge funds and prominent investors that are known to persistently beat the market and, thus, generate alpha in excess of the return required by their exposure to systematic risk factors, as shown in Kosowski, Naik and Teo (2007). As an implication of these findings, active portfolio management based on fundamental analysis seems to be able to generate alpha for investors.

The remaining studies approach different methodologies applied in the context of fundamental analysis. With a distinct proposal from the works already mentioned, Hsu, Kalesnik and Kose (2019) explore the question of whether there is a commonly accepted definition for quality, unlike factors such as value and momentum. The authors thus carry out a multi-factor equity strategy analysis to verify which characteristics of the company tend to capture more factors related to quality. The conclusion is that profitability and investment-related factors, which are popular in fundamental analysis, are precisely the ones that capture the idea of quality more closely.

Finally, the work presented by Kumar (2013) refers to an online questionnaire survey about the perception by brokers of the importance of technical and fundamental analysis and its performance in forecasting stock prices on the Bombay Stock Exchange, India. His results suggest that brokers tend to use technical analysis more in the short term and more fundamental analysis in the long term, since they believe that prices tend to reflect their fundamentals over time.

In conclusion, this literature review about the existing works regarding fundamental analysis reinforces its predictive ability and its relevance for stock picking when actively managing equity portfolios.

2.2 Machine Learning Models

According to Jordan and Mitchell (2015), machine learning can be described as a collection of models that automatically improve through experience, joining techniques from Computer Science and Statistics. Due to its success and recent technological advance, it has experienced significant growth in the last few years.

In this section, it is covered a selection of machine learning models to be incorporated into the portfolio optimization process. The choice of these methods was based on an investigation of the literature, which suggested that the Logistic Ridge, Logistic LASSO, Random Forest, k-NN, SVM, and Naïve Bayes would be effective for the purpose of generating trading signals indicating, in each period, which stocks should outperform the market median, and which ones should underperform. In the sequence, some illustrative applications are detailed.

The use of Ridge regression for forecasting stock prices has already been explored in the paper by Yang, Liu and Wu (2018). In the work, the authors, after selecting stock valuation ratios with strong explanatory power, employ 5 frequently used statistical methods - including Ridge regression, Random Forests, among others - to determine, in each period, the one that produces the lowest mean squared error. Subsequently, the selected model is used to rank the stocks according to their expected returns. Afterwards, they adopt a dynamic buy-and-hold strategy for the top 20% stocks. The results found

Table 2 – Examples of studies using machine learning models to predict stock price direction. In the table, the models are identified as follows: Ridge Regression (RR), LASSO Regression (LAR), Random Forest (RF), K-nearest neighbor (k-NN), Logistic Regression (LR), Logistic Ridge Regression (LRR), Logistic LASSO Regression (LLAR), Support Vector Machine (SVM), and Naïve Bayes (NB).

Paper	**RR**	**LAR**	**RF**	**k-NN**	**LR**	**LRR**	**LLAR**	**SVM**	**NB**
Ballings et al. (2015)				x	x		x	x	
Chen and Hao (2017)				x					
Dutta, Bandopadhyay and Sengupta (2012)					x				
Ghazanfar et al. (2017)				x				x	x
Panagiotidis, Stengos and Vravosinos (2018)		x							
Patel et al. (2015)			x					x	x
Pereira, Basto and Silva (2016)						x	x		
Shihavuddin et al. (2010)									x
Wolff and Echterling (2020)			x			x	x		
Wu, Yang and Liu (2014)		x							
Yang, Liu and Wu (2018)	x								
Zhang, Lin and Shang (2017)				x					
This Study	x	x	x	x	x	x	x	x	x

show that the strategy outperforms the S&P500 in terms of Sharpe ratio and cumulative return, proving the effectiveness of these methods for forecasting stock returns using financial ratios as input.

In the case of LASSO regression, examples of applications of the methodology are provided in the works of Wu, Yang and Liu (2014) and Panagiotidis, Stengos and Vravosinos (2018). In the first case, the authors use an implementation with nonnegative constraints on the coefficients to track the CSI 300 index. The results confirm that LASSO is capable of delivering a smaller tracking error than OLS and it is effective in asset selection. In the case of Panagiotidis, Stengos and Vravosinos (2018), the authors scrutinize the significance of twenty-one possible explanatory drivers of bitcoin returns for the period 2010–2017. The results corroborate that search intensity and gold returns are the most relevant variables for bitcoin returns.

Meanwhile, in the work of Ballings et al. (2015), the authors carry out a comparative study on Random Forest against other conventional methods in machine learning in the context of forecasting the direction of the stock prices in European markets. The results show the superiority of the method in relation to the others based on the measurement of performance through the area under the receiver operating characteristic curve (AUC).

For the k-nearest neighbors (k-NN) method, an example of application is the paper by Chen and Hao (2017), where the authors combine this method with support vector machine (SVM) to predict stock prices. The experiment conducted in the Chinese stock market shows that the model yields a positive excess return in different time horizons. Furthermore, Zhang, Lin and Shang (2017) bring a similar application to the North American and Singapore markets, concluding that k-NN coupled with other methods is also capable of yielding a positive out-of-sample performance.

It is worth mentioning that, while the authors cited are concerned with forecasting asset returns for different horizons, the work in question will focus on the estimation of which stocks will have, relative to the others, a superior performance. This way, it is possible to indicate to the portfolio optimization model which stocks tend to outperform according to their fundamentals.

2.2.1 Shrinkage Methods

In this section, the application of shrinkage methods for linear regressions is discussed. Under the Gauss-Markov assumptions, the OLS regression estimators are BLUE – Best Linear Unbiased Estimates. Therefore, these are the estimators of minimum variance found for non-biased models. Despite these theoretical results, in practice the estimates produced suffer from high variance due to the existence of highly correlated variables.

This problem can be addressed by inserting a bias to improve the out-of-sample performance of the model, even if there is a worsening in the in-the-sample performance. The purpose is to mitigate the effects of the correlation between the independent variables of the model, which is the essence of shrinkage models.

According to James et al. (2013), the two best-known techniques for shrinking the regression coefficients towards zero are the Ridge and the LASSO regression. Both techniques are briefly presented in the sequence.

2.2.1.1 Ridge Regression

According to Hastie, Tibshirani and Friedman (2008), the Ridge regression adds to the objective function of OLS (minimization of the sum of the squared errors) a penalty associated with the magnitude of the β_j coefficients, thus managing to reduce the variance of these estimators. Mathematically, the problem is formulated as follows:

$$\hat{\beta}^{\text{Ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{2.1}$$

where N represents the number of observations, p is the number of independent variables, and λ denotes a parameter that determines the intensity of the shrinkage. The higher its value, the stronger the shrinkage effect, inducing the betas to zero.

As previously mentioned, the high correlation of the independent variables in the equation can generate multicollinearity, worsening the quality of the regression by increasing the variance of the betas. By imposing a penalty on the size of the coefficients, the issue can be mitigated without affecting the consistency of the estimators (HASTIE; TIBSHIRANI; FRIEDMAN, 2008).

In the case of Ridge regression, it is necessary that the inputs be normalized before its implementation. Additionally, it is worth mentioning that the intercept β_0 is not considered in the penalty.

The matrix form of the Ridge regression loss (RSS – Residual Sum of Squares) is given as follows:

$$\text{RSS}(\lambda) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^T\beta \tag{2.2}$$

whose minimization is achieved when:

$$\hat{\beta}^{\text{Ridge}} = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{2.3}$$

where I denotes the $p \times p$ identity matrix. In Equation 2.3, it is possible to see that, in the case of multicollinearity, which implies that $\mathbf{X}^T\mathbf{X}$ be singular, the term λI ensures the invertibility of the first factor. Thus, the problem will always admit a solution.

2.2.1.2 LASSO Regression

Following Hastie, Tibshirani and Friedman (2008), the LASSO regression is a shrinkage method similar to Ridge regression, with minor, but important changes. The problem is formulated as follows:

$$\hat{\beta}^{\text{LASSO}} = \underset{\beta}{\operatorname{argmin}}\left\{\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p}x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{2.4}$$

Unlike Ridge, the penalty introduced is comprised by the ℓ_1 norm, which makes the solution nonlinear, with no closed-form expression. Still, efficient algorithms are available for numerical approximation. Due to the nature of the constraint, as λ grows, some coefficients will be exactly zero, meaning that LASSO can be used for variable selection.

In the LASSO, analagously to the penalty parameter in Ridge regression, the λ should be adaptively chosen to minimize an estimate of expected prediction error. More specifically, λ may be determined via cross-validation or through information criteria, conciliating the goodness-of-fit with parsimony.

2.2.2 Logistic Regression

According to James et al. (2013), the logistic regression model can be understood as an analog of linear regression for classification problems, where it is desired to categorize some variable by class. In this model, the variable to be predicted is always discrete. The forecast provided by logistic regression will always lie between 0 and 1, so that its results can be interpreted as a probability of a certain event occurring.

An illustration of this classification is provided by James et al. (2013). In their example, the authors consider a credit card customers dataset, where the probability of default by credit card is verified considering a monthly credit card balance as a variable. The default response falls into the Yes (1) or No (0) category. Thus, the logistic regression model indicates the probability that the dependent variable Y belong to a category, and not directly the value of the dependent variable.

In Figure 2, the orange tick indicates the values of No and Yes (0 and 1) in case of default. In the example, these default cases are concentrated in higher credit card balance scenarios. In the left-hand plot, the probability is estimated using linear regression. Due to the binary behavior of the dependent variable Y, the quality of the regression is compromised, where some of the probabilities estimated in the illustration are even negative for certain credit card balance scenarios. In the case of forecasting probabilities using a logistic regression, the method produces default probabilities that converge asymptotically to one when the credit card balance goes to infinity, and to zero when the credit card balance goes to minus infinity. Thus, for any given value of credit card balance, a prediction of the probability of default can be made.

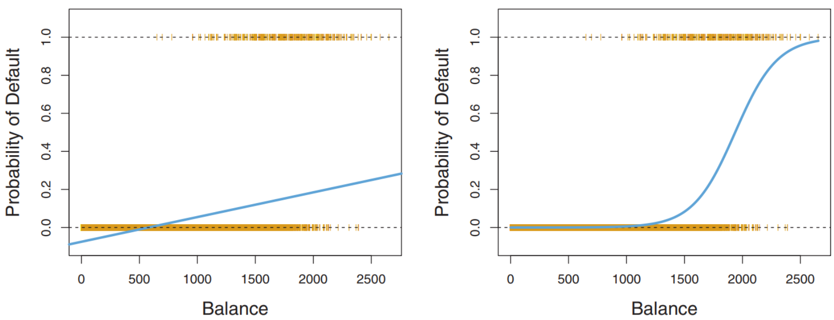

Figure 2 – Examples of linear and logistics regression models. Source: James et al. (2013).

An important feature of this type of regression is that it is not influenced by outliers, since the focus of the algorithm is concentrated on the frontier region of the classification. According to James et al. (2013), the linear regression model has the following form:

$$p(X) = \beta_0 + \beta_1 X \tag{2.5}$$

In the case of logistic regression, the output of the dependent variable must lie between 0 and 1 for all values of the independent variable X. Therefore, the linear regression is adjusted and can be expressed by:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \tag{2.6}$$

Through this new equation, it is possible to observe that the probability forecast can never lie below zero nor above one. Also, the logistic function will always produce a S-shaped curve, making it possible to obtain a sensible prediction regardless of the value of X. After manipulation of Equation 2.6, it can be verified that:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X} \tag{2.7}$$

where the quantity $p(X)/[1-p(X)]$ is called *the odds*, and can take on any value between zero and infinite. Also, the values of the odds that are close to zero indicate very low probability of default and, similarly, the values of the odds that are close to infinite indicate very high probabilities of default. By taking the logarithm of both sides of Equation 2.7, it can be verified that:

$$\ln\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X \tag{2.8}$$

where the left-hand side is called the *log-odds* or *logit*. That way, it is possible to observe that the logistic regression model (Equation 2.8) has a logit that is linear in X. To fit the logistic regression model, the maximum likelihood is used. According to Hastie, Tibshirani and Friedman (2008), the log-likelihood ($\ell(\beta)$) for N observations can be written as:

$$\begin{aligned}\ell(\beta) &= \sum_{i=1}^{N}\left\{y_i \ln p\left(x_i;\beta\right) + \left(1-y_i\right)\ln\left(1-p\left(x_i;\beta\right)\right)\right\} \\ &= \sum_{i=1}^{N}\left\{y_i\beta^T x_i - \ln\left(1+e^{\beta^T x_i}\right)\right\}\end{aligned} \tag{2.9}$$

Finally, to maximize the log-likelihood, we set its derivatives to zero.

2.2.2.1 Logistic LASSO

The LASSO regression model presents an ℓ_1 penalty, implemented with the aim of limiting the variance of the model, as previously mentioned. In the case of the logistic regression model, the Logistic LASSO is obtained by maximizing the likelihood function with a penalized parameter applied to all coefficients except the intercept. Thus, a penalized version of the likelihood must be maximized (HASTIE; TIBSHIRANI; FRIEDMAN, 2008):

$$\max_{\beta_0,\beta}\left\{\sum_{i=1}^{N}\left[y_i\left(\beta_0+\beta^T x_i\right) - \ln\left(1+e^{\beta_0+\beta^T x_i}\right)\right] - \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{2.10}$$

The objective function in Equation 2.10 is concave and, that way, it can be found a solution using nonlinear programming methods, as shown by Koh, Kim and Boyd (2007), for instance.

2.2.2.2 Logistic Ridge Regression

Analogous to the LASSO model, the Logistic Ridge regression uses a ℓ_2 penalty instead of the ℓ_1 used in LASSO regression. According to Cessie and Houwelingen (2007), the Ridge estimators in logistic regression are the values that maximize the following log-likelihood function, resulting in the following constrained maximization equation:

$$\ell_{\lambda}^{R}(\beta) = \sum_{i=1}^{n} \left[y_i x_i \beta - \ln\left(1 + e^{x_i \beta}\right) \right] - \lambda \sum_{j=1}^{p} \beta_j^2 \tag{2.11}$$

It is worth mentioning that the Ridge coefficient estimates approach zero as shrinkage penalty λ increases. Also, the penalty introduced in the log-likelihood function shrinks all of the coefficients towards zero, without setting any of them exactly to zero (PEREIRA; BASTO; SILVA, 2016). That way, the Ridge regression includes all the predictors in the final model, which can difficult the model interpretation.

2.2.3 Random Forest

Prior to discussing Random Forest, it is necessary an overview of *regression trees* (also known as *decision trees*). In line with Vasconcelos (2018), they consist in series of decisions based on observed values of the features selected to explain the dependent variable. At each level, the tree divides into new branches based on which feature is contributing most to the explanation of the variable. This flow goes through the leaves of the tree, where the dependent variable is forecast.

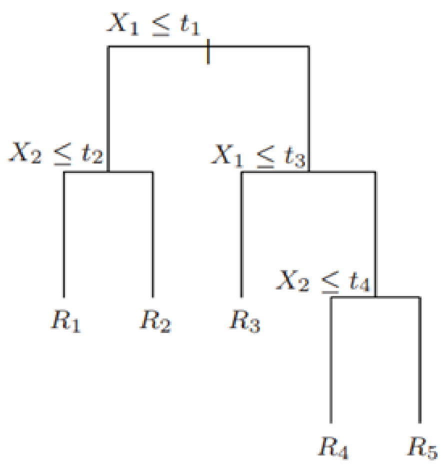

Figure 3 – Illustration of regression tree. The variables X_i denote the features of the sample, and the values t_i determine how the features space is partitioned. Source: Hastie, Tibshirani and Friedman (2008).

A negative aspect of the decision trees, however, is the tendency of overfitting, where the tree uses data for training and adjusts itself excessively to them, losing generalization power for data out-of-sample. According to Hastie, Tibshirani and Friedman (2008), one way to solve this problem involves the *bagging technique*, through which a series of trees is trained, each of them with unique training and test samples. This strategy of combining trees is one of the pillars of the Random Forest model.

Another important pillar involves the so-called *Random Subspaces*, whose objective is to mitigate the problem of overimportance that can be given to certain features for subsamples, considering only a subset of features chosen randomly.

As stated by Hastie, Tibshirani and Friedman (2008), the use of Random Forest for regression can be defined as follows:

$$\hat{f}_{\mathrm{rf}}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{2.12}$$

where B is the number of trees produced using the bagging strategy, and $T_b(x)$ is the forecast generated by tree b for the new feature x observed. Therefore, the forecast provided by Random Forest is the arithmetic mean of the individual forecasts for each tree.

2.2.4 k-Nearest Neighbors

In line with Hastie, Tibshirani and Friedman (2008), k-nearest neighbors (k-NN) is a technique used for classification or regression problems. In the first case, the technique involves classifying points out the training sample based on their neighborhood. In this sense, for each point, the classifications provided by the nearest k points are analyzed and the most frequent result is selected as the estimated classification (majority vote).

At the same time, for regression problems, k-NN makes predictions considering for each point the k closest neighbors and employing summary statistics (for example, its average) as a forecast for that point. Song et al. (2017) provide an illustrative example of empirical analysis based on regressions via k-NN. The authors describe the k-NN given a distance metric d, used to determine the k points closest to a point x in the test sample. The prediction of this point is then given by the following expression:

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i(x) \tag{2.13}$$

where $y_i(x)$ is the observed value of the dependent variable at each of the k neighbor points.

2.2.5 Support Vector Machine

According to James et al. (2013), the *Support Vector Machine* (SVM) can be described as a generalization of a classifier called the *maximum margin classifier*. To better describe the classifier, it is necessary to introduce the concepts of *hyperplane* and *optimal separating hyperplane*.

James et al. (2013) describe the hyperplane as a flat affine subspace of dimension $p-1$, for a p-dimensional space. For instance: in two dimensions, a hyperplane can be described as a line (a flat one-dimensional subspace) and be defined by:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0 \tag{2.14}$$

for parameters β_0, β_1, and β_2. In a p-dimensional space, that equation is extended to:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p = 0 \tag{2.15}$$

It is important to highlight that what defines the hyperplane is that if a point $X = (X_1, X_2, ..., X_p)^T$ in the p-dimensional space satisfies Equation 2.15, then X is in the hyperplane. When this equation is greater than zero, X is on one side of the hyperplane and, when the equation is less than zero, X is on the other side. In this way, the hyperplane divides the p-dimensional space into two halves.

For illustration, the following figure (Figure 4) by James et al. (2013) shows a hyperplane in two-dimensional space, whose equation is $1 + 2X_1 + 3X_2 = 0$. In this plot, it is possible to observe that the blue region is the set of points whose left side of the equation is greater than zero and the purple side and region is the set of points whose left side of the equation is less than zero.

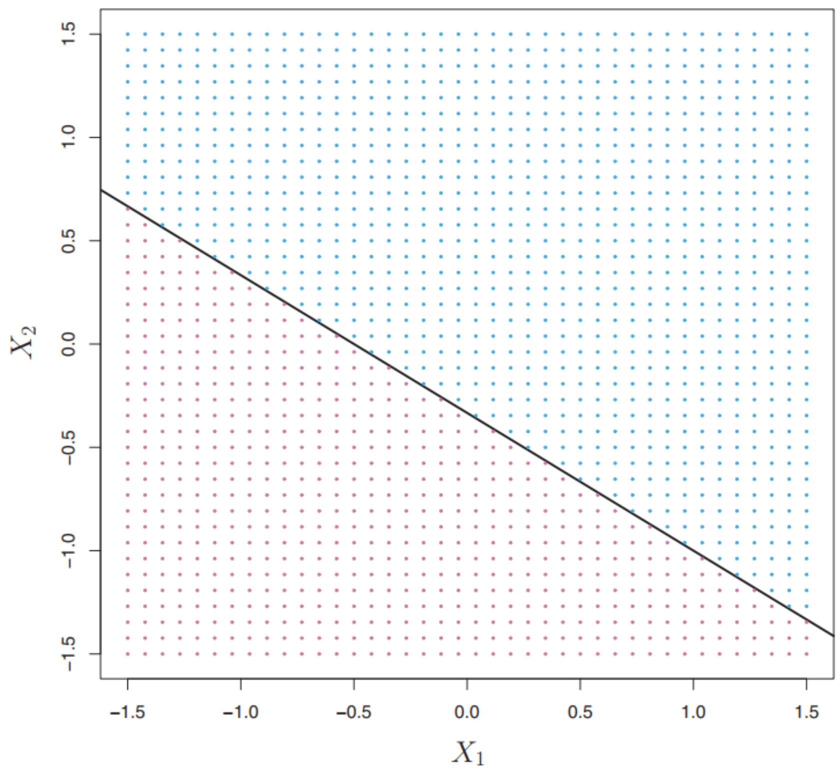

Figure 4 – Hyperplane in two-dimensional space. Source: James et al. (2013).

The next step is to develop a classifier based on the training data that will correctly classify the test observation using its feature measurements. For that, it is assumed an $n \times p$ data matrix X that consists of n training observations in p-dimensional space:

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \ldots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix} \tag{2.16}$$

Thus, these observations belong to either one of two classes, class -1 and class 1, so that $y_1, ..., y_n \in \{-1, 1\}$. In addition to n training observations, there is also a test observation, which consists on a p-vector of observed features $x^* = (x_1^* \dots x_p^*)^T$. This approach is based upon the concept of a separating hyperplane, which attempts to build a hyperplane that seggregates training observations according to their class labels.

Given the two classes mentioned above (-1 and 1), a separating hyperplane obeys:

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} > 0 \text{ if } y_i = 1 \tag{2.17}$$

and also:

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} < 0 \text{ if } y_i = -1 \tag{2.18}$$

which is equivalent to the following equation:

$$y_i \left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}\right) > 0 \tag{2.19}$$

for all $i = 1, ..., n$. That way, the test observation x^* is classified based on the sign of $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + ... + \beta_p x_p^*$. If positive, the test is assigned to class 1. If negative, then it is assigned to class -1.

Thus, in order to construct a classifier based upon a separating hyperplane, it is necessary to define which separating hyperplanes to use. For this, the *maximal margin hyperplane* method is used, in which the separating hyperplane that is farthest from the training observations is employed. In this method, it is possible to calculate the smallest distance from each training observation to a given separating hyperplane, which is known as *margin*. Thus, the maximal margin hyperplane is the separating hyperplane that has the farthest minimum distance to the training observations, being able to classify a test observation based on which side of the maximal margin hyperplane it lies. Therefore, this method is known as *maximal margin classifier*. Finally, if β_0, β_1,. . . , β_p are the coefficients of the maximal margin hyperplane, then the *maximal margin classifier* classifies the test observation x^* based on the sign of $f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \dots + \beta_p x_p^*$ (JAMES et al., 2013).

The training observations that are equidistant from the maximal margin hyperplane are known as *support vectors*. The idea that these vectors support the maximum hyperplane margin comes from the observation that, if these points are slightly moved, the maximal margin hyperplane would also move.

Finally, the solution to the maximum margin hyperplane optimization problem, based on a set of n training observations $x_1, ..., x_n \in \mathbb{R}^p$ and associated class labels $y_1, ..., y_n \in \{-1, 1\}$, is given by (JAMES et al., 2013):

$$\begin{aligned} &\underset{\beta_0,\beta_1,\dots,\beta_p}{\text{maximize}} M \\ &\text{subject to } \textstyle\sum_{j=1}^{p} \beta_j^2 = 1 \\ &y_i \left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}\right) \geq M \quad \forall i = 1, \dots, n \end{aligned} \tag{2.20}$$

The constraints ensure that each observation is at a minimum distance M from the hyperplane, in addition to ensuring that they are on its correct side.

It is important to highlight that there are many cases where no separating hyperplane exists, which implies the inexistence of the maximal margin classifier. In other cases, there are instances in which a classifier based on a separating hyperplane might not be desirable (JAMES et al., 2013).

In this sense, it is possible to extend the concept of a separating hyperplane, developing a hyperplane that almost separates the classes, but not totally. This concept is called a *soft margin*. The generalization of the maximal margin classifier to the non-separable case is known as the *Support Vector Classifier* (SVC) (JAMES et al., 2013).

2.2.6 Support Vector Classifiers

The idea of the *Support Vector Classifier* (SVC) is that it could be worthwhile to misclassify some training observations incorrectly in order to do a better job in classifying the remaining observations. Therefore, the method allows some observations to be closer to the hyperplane than required by the margin, or even on the incorrect side of the hyperplane (JAMES et al., 2013).

Thereby, the hyperplane correctly separates most of the training observations into the two classes, but may misclassify a few of them. Then, the SVC classifies a test observation depending on which side of the hyperplane it lies.

The optimization problem is given by (JAMES et al., 2013):

$$\begin{aligned} &\underset{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n}{\text{maximize}} \; M \\ &\text{subject to } \textstyle\sum_{j=1}^{p} \beta_j^2 = 1 \\ &y_i\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}\right) \geq M\left(1 - \epsilon_i\right), \\ &\epsilon_i \geq 0, \textstyle\sum_{i=1}^{n} \epsilon_i \leq C \end{aligned} \tag{2.21}$$

where C is a nonnegative tuning parameter related to the fraction of training observations that are misclassified by the SVC.

2.2.7 Naïve Bayes Classifier

Consider $x = [x_1, ..., x_n]^T$ a feature vector. Then, to label it in one of the c classes of the problem, denoted $w_1, ..., w_p$, a posterior probability $p\left(\omega_k \mid x\right)$ is defined, which is called the Bayes classifier. The corresponding error of it is then called the Bayes error. The posterior probabilities can be calculated as $p\left(\omega_k \mid x\right) = \frac{p(x|\omega_k)p(\omega_k)}{p(x)}$, with $p\left(x \mid \omega_k\right)$ being the class-conditional probability density function conditioned on w_k, $p\left(\omega_k\right)$ being the prior probability for w_k and $p(x)$ the unconditional probability density function (KUNCHEVA, 2006).

The Naïve Bayes classifier then assumes conditional independence between the features and calculates the class-conditional probability density function as a product of n individual probability density functions (KUNCHEVA, 2006):

$$p(x \mid \omega_k) = \prod_{i=1}^{n} p(x_i \mid \omega_k) \tag{2.22}$$

The Naïve Bayes often delivers a satisfactory result, which is attributed mainly to the fact that the conditional independence is a sufficient, but not a necessary condition for optimality of the method. Applications of Naïve Bayes can be seen in Hand and Yu (2001) and Domingos and Pazzani (1997).

2.3 Portfolio Optimization Model

Portfolio optimization is an approach that helps designing optimal portfolios according to certain objectives and constraints. This framework has been fundamental in the development of financial markets and financial decision making (KOLM; TUTUNCU; FABOZZI, 2014). The chapter begins by presenting the Modern Portfolio Theory (MPT) developed by Markowitz (1952), and demonstrating the mean-variance optimization (MVO) problem formulated by the author. In the sequence, it posits the limitations of the method and ways of overcoming them, which is the focus of the present work.

2.3.1 Markowitz Model

The Modern Portfolio Theory is defined in Markowitz (1952). According to the author, this theory refers to how to build portfolios to maximize the expected return for risk-averse investors, based on a certain level of market risk. The paper proposes that risk and return should not be seen in isolation, but rather evaluated together. Markowitz (1952) also suggests using mean and variance as measures of return and risk for the purpose of portfolio optimization (KOLM; TUTUNCU; FABOZZI, 2014).

The Markowitz (1952) model, called mean–variance optimization (MVO), proposes building a portfolio with the objective of maximizing its return for a predefined level of risk. The model also assumes that investors are risk-averse, that is, for a given level of return, the investor will always prefer a less risky portfolio (MARKOWITZ, 1952).

The formulation considers an investment universe of n assets $S_1, S_2, ..., S_n$ and a vector of uncertain future returns denoted by $r = [r_1, ..., r_n]^T$. The proportion of the total funds invested in security i in the portfolio is denoted by w_i, forming a vector of portfolio weights $w = [w_1, ..., w_n]^T$. Therefore, the return of the portfolio, r_p, can be expressed by (KOLM; TUTUNCU; FABOZZI, 2014):

$$r_p(w) = w_1 r_1 + ... + w_n r_n = w^T r \tag{2.23}$$

The standard deviation of r_i is denoted by σ_i and the correlation coefficient of the returns of assets S_i and S_j (for $i \neq j$), is represented by ρ_i. The symmetric $n \times n$ covariance matrix of the returns of all the assets can be expressed by Σ, which is given by (KOLM; TUTUNCU; FABOZZI, 2014):

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} \tag{2.24}$$

where $\sigma_{ii} = \sigma_i^2$ and $\sigma_{ij} = \sigma_{ji} = \rho_{ij}\sigma_i\sigma_j$ (for $i \neq j$). Also, for a covariance matrix to be valid, it must be positive semidefinite. Then, for a given portfolio with weight vector w, the standard deviation of the portfolio return can be written as (KOLM; TUTUNCU; FABOZZI, 2014):

$$\sigma(w) = \sqrt{w^\top \Sigma w} \tag{2.25}$$

The standard deviation of the portfolio return $\sigma(w)$ is also referred to as portfolio volatility. The assets expected returns are given by:

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \tag{2.26}$$

where $\mu_i = E(r_i)$ for all $i = 1, ..., n$. Finally, using this notation, the MVO can be written as:

$$\max_{w \in \Omega} \mu^\top w - \lambda \cdot w^\top \Sigma w \tag{2.27}$$

where λ denotes the investor specific risk aversion parameter (KOLM; TUTUNCU; FABOZZI, 2014).

Despite the intuitive appeal of the method, there are many concerns associated with the application of MVO in real world. Some of these are pointed out by Kolm, Tutuncu and Fabozzi (2014), who indicate the non trivial relationship between MVO inputs and outputs. Other concerns are related to the sensitivity of the target portfolio allocation to changes in inputs, which are difficult to estimate. Some practitioners also consider that the Modern Portfolio Theory to be ill-defined due to the lack of real world considerations in its formulation. For example, the original MVO does not consider transaction costs and their effects on portfolio returns and allocation.

As previously mentioned, the Markowitz (1952) model is based on the hypothesis that investors are risk-averse and use variance and mean, respectively, to assess the risk and return on their investments. This hypothesis is far from reality, as mentioned by Luenberger (1998), creating problems in the optimization process. Some of them are linked to the fact that the returns on financial assets do not always follow an elliptical distribution, which implicitly supports the Markowitz model. In addition, Fabozzi et al. (2007) emphasize that return distributions often exhibit heavy tails and asymmetries around the mean, characteristics that cannot be captured and explained solely using expected values and variance, for they are associated to moments of higher order, i.e. skewness and kurtosis.

Therefore, with the aim of incorporating these effects in investment portfolio selection models, one possibility is to incorporate moments of order higher than two of the distribution of portfolio returns. Also, so as to mitigate the aforementioned problems in portfolio management, some extensions are added to the formulation proposed by Markowitz (1952), in line with illustrative examples brought forth by Kolm, Tutuncu and Fabozzi (2014). As an alternative solution that also allows the inclusion of financial ratios to identify mispriced stocks when optimizing equity portfolios, there is the parameterization approach advocated by Brandt, Santa-Clara and Valkanov (2009), which is discussed subsequently.

2.3.2 Parametric Portfolio Framework

As discussed before, the present work consists on a portfolio optimization model which uses signals extracted from financial ratios as input. In order for these signals to be useful and aid in the optimization, it is necessary to use a model that allows their incorporation. Based on this requirement, the approach presented in Brandt, Santa-Clara and Valkanov (2009) was chosen to help the construction of the portfolio optimization of the present work.

The selected technique allows the financial ratios of each stock to be directly included in the model as parameters. In their work, the authors begin with the investor's problem of choosing the portfolio weights $w_{i,t}$ to maximize the conditional expected utility of the portfolio return, $r_{p,t+1}$:

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} \mathrm{E}_t\left[u\left(r_{p,t+1}\right)\right] = \mathrm{E}_t\left[u\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right)\right] \tag{2.28}$$

where N_t represents the number of stocks for each date t, i represents each stock whose return is $r_{i,t+1}$, and u is a predefined utility function. The authors then present a parameterization of the portfolio weights $w_{i,t}$ as a function of the stocks characteristics, which are represented by a vector of firm characteristics $x_{i,t}$, as displayed in sequence:

$$w_{i,t} = f(x_{i,t};\theta) \tag{2.29}$$

The paper then focus on the following simple linear specification for the portfolio weight function:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\theta^\top \hat{x}_{i,t} \tag{2.30}$$

where $\bar{w}_{i,t}$ is the weight of stock i at date t in a benchmark portfolio, such as the value-weighted market portfolio, θ is the vector of coefficients to be estimated, and $\hat{x}_{i,t}$ is the characteristics of stock i, standardized cross-sectionally to have zero mean and unit standard deviation across all stocks at date t. In the formulation presented, it can be observed that the intercept is the weight of the stock in the benchmark portfolio, and the term $\theta^\top \hat{x}_{i,t}$ represents the deviations of the optimal portfolio weight from this benchmark. Therefore, the deviation can be interpretated as a directly characteristic of the stock.

Brandt, Santa-Clara and Valkanov (2009) present an empirical implementation of their novel approach of optimizing portfolios for the universe of all stocks in the CRSP – Compustat data set. The investor in their exercise is assumed to have constant relative risk aversion (CRRA) preferences. The portfolio weight in each stock is modeled as a function of the firm's market capitalization, book-to-market ratio, and lagged return. The coefficients of the function are calculated by optimizing the investor's average utility of the portfolio's return over the given sample period.

Scrutinizing the results obtained with the application of the model, the authors suggests that their empirical results document the importance of the firm's analyzed characteristics for explaining deviations of the optimal portfolio for a CRRA investor from the market. The results also suggests that, for a relative risk aversion of five ($\gamma = 5$), the certainty-equivalent gain from incorporating the firm characteristics, relative to holding the market portfolio, is an annualized 11% return, surpassing the benchmarks considered. Also, the results are robust out-of-sample. Besides, their idea can easily be applied to other asset classes, making the model ideal for the context of this dissertation.

An application of the Brandt, Santa-Clara and Valkanov (2009) can be found in Medeiros, Passos and Vasconcelos (2014), which explores a parametric portfolio optimization in the Brazilian market and presents results which show that the technique is efficient out-of-sample. As firm characteristics, the authors employed the book-to-market ratio, the market cap, and the one-year momentum. The optimal portfolio delivered a performance considerably superior than the Markowitz portfolio out-of-sample, even when transaction costs are included. In particular, its certainty equivalent return was 11.4%, against 8% of the Markowitz case.

Another application of Brandt, Santa-Clara and Valkanov (2009) can be seen in Hand and Green (2011). In the study, the authors examine an extension of Brandt, Santa-Clara and Valkanov (2009), incorporating accounting-based variables in addition to price-based characteristics already used in the study. The results show that, with this addition, the parametric portfolio policy offers improved performance. Lyle and Yohn (2020) also perform an implementation of the Brandt, Santa-Clara and Valkanov (2009) parameterization, with the aim of checking the usefulness of financial ratios, providing a model that directly links fundamental-based ratios to expected returns. According to their findings, a fully optimized fundamental-based portfolio yields significant gains in terms of returns and Sharpe ratios over naive equal-weighted portfolios.

As discussed earlier, the literature lacks portfolio optimization studies that are essentially based in fundamental analysis. The work by Silva (2014), for instance, uses both fundamental and technical analysis to optimize portfolios using evolutionary algorithms. According to his findings, the optimizer produces portfolios concentrated on stocks with high return on investment that are identified by choosing companies exhibiting high net profit growth rates and above-average profit margins.

In a similar way, the paper by Palazzo et al. (2018) tested a value investing strategy for the Brazilian market, selecting stocks based on the criteria suggested by Graham (2007) so that lower quality companies with potential risks not captured by the traditional risk models were eliminated. Next, a Markowitz portfolio optimization is implemented. The portfolios obtained were able to offer higher risk-adjusted returns than the Bovespa Index in the period, which confirmed the validity of the value investing strategy in the domestic market.

Despite using fundamental analysis to select the stocks that entered its portfolio, the optimization was performed following a traditional Markowitz model, not incorporating fundamental analysis as an allocation tool. From this perspective, it is possible to conclude the importance of an optimization model that also encompasses fundamental analysis when determining portfolio weights, not focusing exclusively on screening stocks.

2.3.3 Extensions of the Parametric Portfolio Framework

A more modern application of the equation by Brandt, Santa-Clara and Valkanov (2009) was presented in the work of DeMiguel et al. (2020), where the author made an extension of their parametric portfolio framework.

The work by DeMiguel et al. (2020) analyzes how many characteristics matter jointly for an investor who are interested about portfolio risk and transaction costs. To study this theme from a portfolio perspective, the author then extends the parametric portfolio methodology presented in Brandt, Santa-Clara and Valkanov (2009). In their findings, the authors expose that transaction costs significantly increase the dimension of the cross section of stock returns.

For this extension, it is necessary to recall Equation 2.30. From there, it is possible to verify that the return of the parametric portfolio at time $t+1$, which we denote as $r_{p,t+1}(\theta)$, is equal to (DEMIGUEL et al., 2020):

$$r_{p,t+1}(\theta) = w_{b,t}^{\top} r_{t+1} + \theta^{\top} X_t^{\top} r_{t+1}/N_t = r_{b,t+1} + \theta^{\top} r_{c,t+1} \tag{2.31}$$

where $r_{t+1} \in \mathbb{R}^{N_t}$ is the return vector at time $t+1$, $r_{b,t+1} = w_{b,t}^{\top} r_{t+1}$ is the benchmark portfolio return at time $t+1$, and $r_{c,t+1} = X_t^{\top} r_{t+1}/N_t$ is the characteristic return vector at time $t+1$. The authors, then, assume that the investor optimizes mean-variance utility, solving the following problem as follows:

$$\min_{\theta} \frac{\gamma}{2} \operatorname{var}_t \left[r_{p,t+1}(\theta) \right] - E_t \left[r_{p,t+1}(\theta) \right] \tag{2.32}$$

where γ is the risk-aversion parameter and $\text{var}_t\left[r_{p,t+1}(\theta)\right]$ and $E_t\left[r_{p,t+1}(\theta)\right]$ are the variance and mean of the parametric portfolio return, respectively. Then, considering T historical observations of returns and characteristics, the parametric portfolio problem is formulated as a quadratic optimization problem:

$$\min_{\theta}(\gamma/2)\theta^{\top}\widehat{\Sigma}_c\theta + \gamma\theta^{\top}\widehat{\sigma}_{bc} - \theta^{\top}\widehat{\mu}_c \tag{2.33}$$

where $\widehat{\Sigma}_c$ is the sample covariance matrix of the characteristic-return vector r_c; $\widehat{\mu}_c$ is the mean of the characteristic-return vector r_c; and $\widehat{\sigma}_{bc}$ is the sample vector of covariances between the benchmark portfolio return r_b and the characteristic-return vector r_c.

Then, combining Equation 2.33 with a transaction cost function $TC(\theta)$, the mean-variance parametric portfolio problem can be written as:

$$\min_{\theta}(\gamma/2)\theta^{\top}\widehat{\Sigma}_c\theta + \gamma\theta^{\top}\widehat{\sigma}_{bc} - \theta^{\top}\widehat{\mu}_c + \text{TC}(\theta) \tag{2.34}$$

Also, the mean-variance parametric portfolio problem with transaction costs can be rewritten by decomposing the variance of the characteristic portfolio return into a term associated with the characteristic own-variances and a term associated with the characteristic covariances - which means, the term $\theta^{\top}\widehat{\Sigma}_c\theta$ can be decomposed in $\theta^{\top}\,\text{diag}\left(\widehat{\Sigma}_c\right)\theta$ (own-variances characteristics) and in $\theta^{\top}\left(\widehat{\Sigma}_c - \text{diag}\left(\widehat{\Sigma}_c\right)\right)\theta$ (covariances characteristics). Therefore, the parametric portfolio problem can be expressed by:

$$\min_{\theta} \underbrace{(\gamma/2)\theta^{\top}\,\text{diag}\left(\widehat{\Sigma}_c\right)\theta}_{\text{own-var(char)}} + \underbrace{(\gamma/2)\theta^{\top}\left(\widehat{\Sigma}_c - \text{diag}\left(\widehat{\Sigma}_c\right)\right)\theta}_{\text{cov(char)}} + \underbrace{\theta^{\top}\gamma\widehat{\sigma}_{bc}}_{\text{cov(bench)}} - \\ - \underbrace{\theta^{\top}\widehat{\mu}_c}_{\text{mean}} + \underbrace{\text{TC}(\theta)}_{\text{tran. costs}} \tag{2.35}$$

To deal with the large number of characteristics considered, the authors propose regularized parametric portfolios, which are obtained by imposing a LASSO constraint on the parametric portfolio. The problem then can be written as:

$$\begin{array}{ll} \min_{\theta} & \frac{\gamma}{2}\theta^{\top}\widehat{\Sigma}_c\theta + \theta^{\top}\gamma\widehat{\sigma}_{bc} - \theta^{\top}\widehat{\mu}_c + \text{TC}(\theta) \\ \text{s.t.} & \|\theta\|_1 \leq \delta \end{array} \tag{2.36}$$

where $\|\theta\|_1 = \sum_{k=1}^{K}|\theta_k|$ is the 1-norm of the parameter vector, and δ is the threshold parameter. It is important to notice that, for $\delta = \infty$, it is recovered the standard parametric portfolios, and, for $\delta = 0$, it is recovered the benchmark portfolio

Compared to Brandt, Santa-Clara and Valkanov (2009), the extension proposed by DeMiguel et al. (2020) applies the LASSO constraint to the optimization problem, which helps to avoid overfitting, reducing the impact of estimation error. Besides, the LASSO constraint is a variable-selection method where only the relevant characteristics receive a nonzero parameter, which improves the quality of the model.

3 DATA AND METHODOLOGY

In this section, the dataset used to conduct the experiment to test the proposed optimization model is detailed. Subsequently, the methodology adopted to implement this model and build the benchmark portfolios is discussed.

3.1 Data

The database comprises a total of 46 stocks of Brazilian companies listed on B3, which are compiled in Appendix A. These stocks were selected from the composition of the Bovespa index portfolio, which provided an initial universe of stocks that was later filtered in terms of availability of historical financial ratios. According to this criterion, stocks of high capitalization companies with satisfactory liquidity for trading and adequate data availability were selected.

The historical data of these stocks were collected through the Bloomberg terminal. The base comprises historical series of prices, from which daily returns and covariance matrices were calculated from September 10, 2013 to December 30, 2020. Besides, financial ratios listed in Appendix B were compiled with data ranging from August 28, 2014 to December 30, 2020. The choice of this investment horizon reflects the avalility of financial ratios for all the stocks of the universe considered.

Additionally, for the implementation of the models, a risk-free rate is required, which was approximated by the DI rate. The series was downloaded from Cetip's website, responsible for its calculation and official publication. The DI rate was also used to compute excess returns whenever required.

The indicators collected were normalized so that they presented a zero mean and unit variance. The objective of this procedure was to ensure that all series had the same scale, a characteristic that influences the adjustment of machine learning models. These were used to generate the signals input to the optimization model.

3.2 Methodology

This chapter presents the model developed in this work, which is a portfolio optimization model that uses as input trading signals extracted from financial ratios. The signals stem from machine learning models, fitted to predict future stock returns using the aforementioned indicators. These models comprehend classifiers such as Logit, Random Forest, Support Vector Classifier (SVC), k-NN, and Naïve Bayes. The trading signals produced specify which stocks should outperform the market median, and which ones should underperform.

In the optimization model, to generate the necessary inputs, a combination of machine learning methods is used to produce more accurate predictions about future stock price direction, a methodology already implemented in other works focused on time series forecasting. An example of this approach is provided by Vasconcelos (2018), who verify that enhanced accuracy can be attained in the context of inflation modeling by merging individual forecasting methods, such as LASSO, Random Forest, Factor Models, among others.

It should be emphasized here that the combination is carried out entirely by the parametrization proposed by Brandt, Santa-Clara and Valkanov (2009). Although the extension demonstrated by DeMiguel et al. (2020) brings a more modern approach, the present work will use the equation by Brandt, Santa-Clara and Valkanov (2009) in its original form, since this was the one used by Medeiros, Passos and Vasconcelos (2014), who also carried out an application of the methodology for the Brazilian case, with promising results. Thus, this equation will also be chosen for comparative purposes.

Another example is the work of Hansen, Lunde and Nason (2011), whose methodology of model confidence set (MCS) allows the combination and optimal selection of competitor models for time series forecasting. In their study, the authors also present applications for inflation forecasting and Taylor rule construction, in which the MCS produced a set of models that correctly captured the dynamics of the analyzed time series.

Therefore, in the following sections, the computational implementation of the portfolio optimization is explored, introducing the procedures employed for building the model. In addition, the benchmarks elected for performance appraisal are also presented.

3.2.1 Computational Implementation

In this section, the computational implementation of the model developed will be discussed. It will be detailed aspects such as: i) the objective function of the model; ii) the use of the framework proposed by Brandt, Santa-Clara and Valkanov (2009) to incorporate fundamental analysis in the optimization process; and iii) changes made to include restrictions and transaction costs.

3.2.1.1 Objective Function

The proposed portfolio optimization model assumes utility-maximizing investors. For this purpose, a CRRA utility function was adopted, which can be expressed as follows (BRANDT; SANTA-CLARA; VALKANOV, 2009):

$$u(c) = \begin{cases} \frac{c^{1-\gamma}-1}{1-\gamma} & \gamma \geq 0, \gamma \neq 1 \\ \ln(c) & \gamma = 1 \end{cases} \tag{3.1}$$

where γ represents the investor's level of relative risk aversion, and c represents his wealth, here defined as the final value of a portfolio that begins with \$ 1 and which grows according to the performance of the stocks selected by the optimization model.

The main feature of the CRRA function is a constant relative risk aversion. In other applications of the model by Brandt, Santa-Clara and Valkanov (2009), exposed for example by Medeiros, Passos and Vasconcelos (2014) and Hand and Green (2011), this utility function was also chosen.

In the case of Medeiros, Passos and Vasconcelos (2014), the choice of the objective function was based on three desirable characteristics, namely: i) assimilation of high-order preferences without the need to include additional parameters beyond the level of relative risk aversion; ii) the twice continuous differentiable function, helping in the optimization process; and iii) the CRRA function is optimal for the partially myopic investor.

The optimization will follow a multiperiod approach in such a way that the maximization of the utility function will occur following the rebalancing windows that form the investment horizon considered.

Inspired by Medeiros, Passos and Vasconcelos (2014), in the experiments conducted herein, in the base scenario, it was considered $\gamma = 5$. The impact of this choice is further investigated in chapter 4.

3.2.1.2 Model Constraints

The optimization model implemented in the work considers two constraints applied to its objective function. The first one imposes that the sum of the weights of the assets must be equal to 1. This constraint is justified for two reasons: i) the fact that it is not beneficial to keep resources allocated in cash, since the objective of the model is to produce an optimal allocation in stocks based on a fundamental analysis; and ii) the limitations that exist in practice for investors to leverage their portfolios.

The other constraint involves individual limits on the minimum and maximum weights of each stock in the portfolio. In the context of the Brazilian stock market, there are regulatory requirements that prevent very concentrated positions in portfolios held by open-ended funds, which are the reference for this work. Thus, an allocation constraint of a minimum of -20% and a maximum of 20% was inserted for each stock in the portfolio. It should be emphasized that the possibility of short-selling is assumed.

The model developed considers an adaptation of the equation defined by Brandt, Santa-Clara and Valkanov (2009) for the parameterization of their weights. In their work, an equation was developed that allows to link the weights w to external variables that are related to the future performance of each stock and, consequently, of the portfolio. The equation is described as follows:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t} \theta^\top \hat{x}_{i,t} \tag{3.2}$$

where $\bar{w}$ is the weight of stock i at date t in a benchmark portfolio, such as the value-weighted market portfolio – in the case of Brazil, that being the Bovespa index; N_t is the number of stocks available in the investment universe; θ is the coefficient vector to be estimated; and x_t is the vector of characteristics of each stock. In the present work, x_t consists on a vector of signals derived from machine learning models that relate financial ratios to the likelihood that a particular stock will outperform the market in the future. More specifically, each binary signal indicates whether the stock will outperform or underperform.

Another factor that influences the objective function is the transaction cost. In real implementations, the transaction cost is a factor that negatively impacts the return of any investment portfolio. In fact, it entails a cost associated with changes in the asset allocation on each rebalancing date, and the more frequent and volumous the reallocations, the more impacting this cost will be. For this reason, it must be included in the present context.

To incorporate the transaction costs in the model, it is used the approach that Medeiros, Passos and Vasconcelos (2014) explored in their study, also based on Brandt, Santa-Clara and Valkanov (2009). They consider, for a given portfolio policy, the turnover at each period t as the sum of all absolute changes in weights from $t-1$ to t, which is expressed by:

$$T_t = \sum_{i=1}^{N_t} |w_{i,t} - w_{i,t-1}| \tag{3.3}$$

It is important to highlight that, for a turnover of x at date t, a change of $x * 100\%$ of the wealth of the portfolio is made in the positions from $t-1$ to t. Therefore, the return of the portfolio net of trading costs is given by:

$$r_{p,t+1} = \sum_{i=1}^{N_t} \left(w_{i,t} r_{i,t+1} - c_{i,t} \left|w_{i,t} - w_{i,t-1}\right|\right) \tag{3.4}$$

where $c_{i,t}$ represents the proportional transaction cost for stock i at time t. That way, the transaction cost can be incorporated in the model as a function of the changes in the assets weights.

In this work, in the base scenario, it was assumed transaction costs equal to 100 bps, which is conservative even for less wealthy investors. The sensitivity of the results with respect to transaction costs is explored subsequently in chapter 4.

3.2.1.3 Configuration of Machine Learning Models

The configuration of the machine learning models, with the exception of Naïve Bayes, was carried out using grid search and cross-validation. Essentially, for each model considered, possible configurations of the hyperparameters were designed and combined. Later, these configurations were tested using cross-validation to check their out-of-sample performance. The most accurate configurations were selected for generating tradings signals for the portfolio optimization process.

In the case of Naïve Bayes, since there are no hyperparameters to be tuned, the preceding process was unnecessary. The only required decision was regarding the prior distribution, which was set to be Gaussian.

Another aspect to be considered is the conversion of probabilities into binary signals. Indeed, classifiers provide the probability that a given point belong to any of the possible classes available in the model. A threshold probability must be defined to transform these probabilities into signals. In the numerical exercise, a threshold of 65% was determined. Different values and the respective impacts on the optimal portfolio are covered in chapter 4.

3.2.2 Flowchart

The following flowchart (Figure 5) illustrates the portfolio optimization process, from the definition of its inputs to the performance analysis. It is a multi-period optimization model with rebalancing. Thus, there is an iterative optimization process on each rebalancing date, which constitutes the investment horizon.

In the next subsections, two specific procedures will be detailed: the rebalancing process and the performance analysis employing benchmarks.

3.2.3 Rebalancing Procedures

The rebalancing process consists of reallocating the portfolio's stocks on a given date, using the optimization model developed and the data available up to that date. Rebalancing is necessary because a static allocation would not allow updating the assets weights according to the evolution of returns and indicators.

In Figure 6, the rebalancing process for the first two periods of the investment horizon is illustrated. On the rebalancing date, the optimization model uses two years (504 business days) of past returns of the portfolio's stocks and a one-year window (252 business days) of financial ratios of these stocks. Once this optimization is executed, the portfolio is held for a window of 6 months (126 business days), until the next rebalancing date, when a new optimization is performed and the portfolio is reallocated. The window size was set following recommendations found in the literature; see Fong, Wong and Lean (2005), for instance. This window is called *holding period*.

It is important to highlight that the optimization is always done considering the returns and indicators available until the immediately preceding business day. This one-day lag is due to the fact that closing prices for stocks can only be traded the next day, when the market reopens. This process is then repeated every six months (126 business days), until the end of the investment horizon.

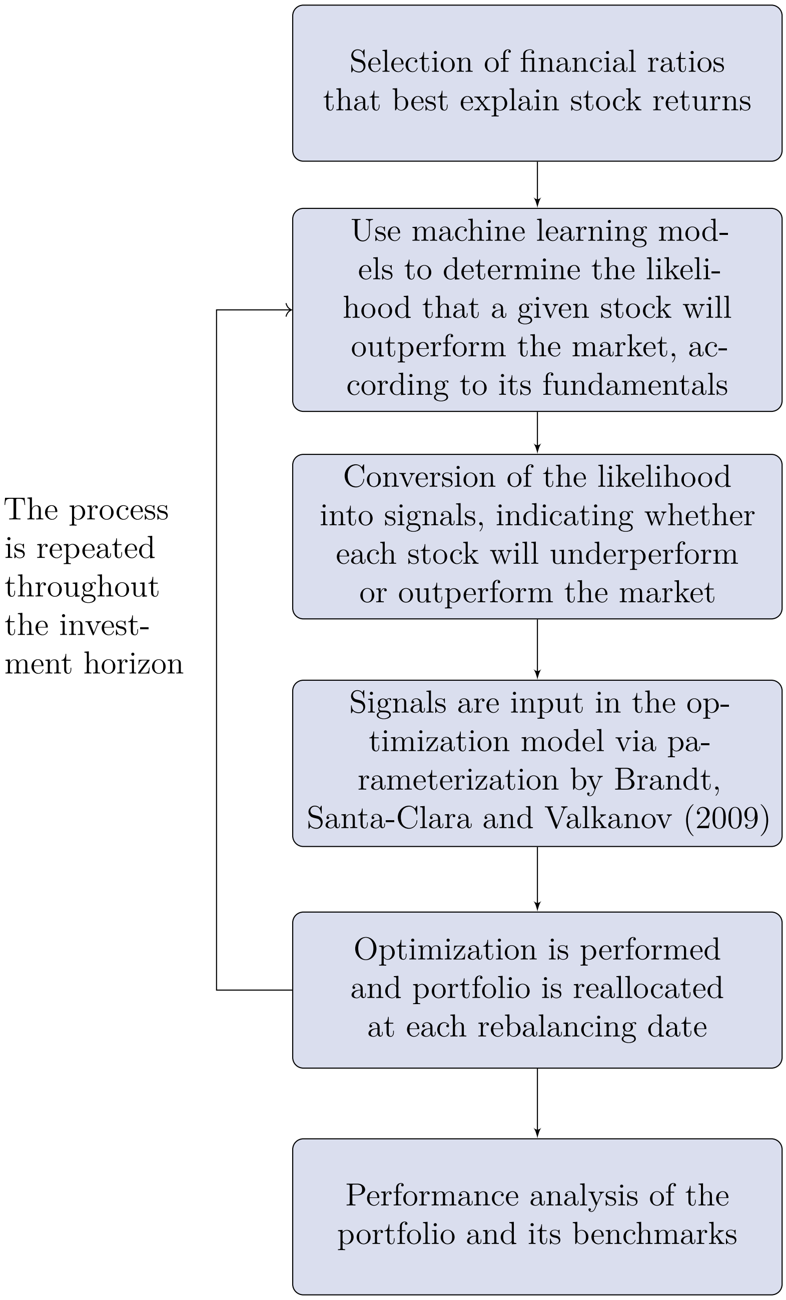

Figure 5 – Flowchart exhibiting the proposed optimization process.

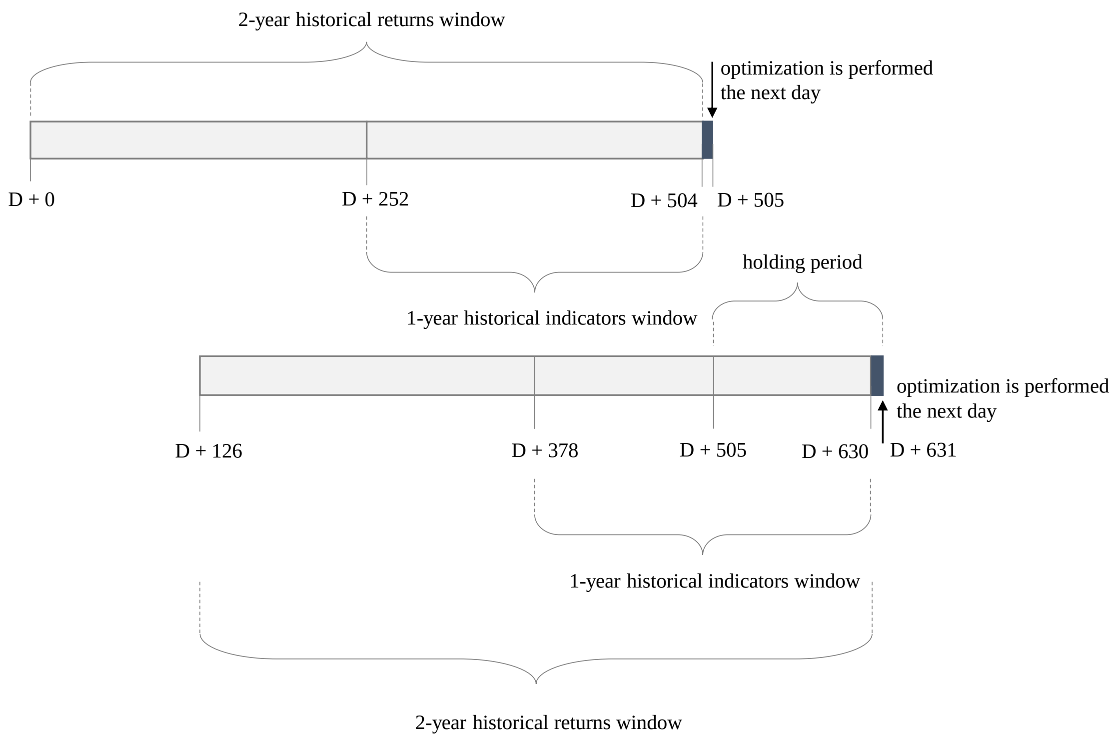

Figure 6 – Rebalancing procedure.

3.2.4 Benchmarks for Performance Appraisal

To evaluate the performance of the optimized portfolio, some benchmarks were selected. The choice prioritized models commonly used in the literature and in practical applications, such as the Markowitz, Contrarian, Momentum, and Equal-weighted models. The Bovespa index, the main reference for the Brazilian stock market, was also used, as it adequately portrays the behavior of the universe of stocks considered.

Additionally, an analysis of a simplified version of the proposed optimization model was also executed, which excludes statistical methods for generating returns signals and replaces them directly with financial ratios. Such analysis was implemented with the objective of verifying the incremental value added by the proposed model. All of these models will be detailed in the next subsections.

3.2.4.1 Markowitz Model

As described in subsection 2.3.1, the Markowitz model considers in its optimization only the mean and variance of the returns to produce a portfolio with the greatest risk-adjusted return. It is the simplest and most widespread model in the literature, which is why it is commonly used as a benchmark in academic work on portfolio optimization. An example of its implementation can be seen in Palazzo et al. (2018), where the authors screen stocks using value investing techniques and later insert the selected stocks into a Markowitz model.

In line with Figure 6, in the numerical exercise, the model was implemented considering rebalancing windows. In addition, the universe of stocks is the same as described in section 3.1.

3.2.4.2 Contrarian, Momentum, and Equal-Weighted Strategies

The contrarian strategy relies on the allocation on assets that, in the recent past, have not performed well, under the assumption that these assets will show an above-average return in the future, according to a mean-reversion process. In other words, the strategy overweight assets that performed poorly in the last period, and underweights the top performing ones. An implementation of this strategy can be found in Bondt and Thaler (1985).

In the present context, the strategy was implemented in each window (252 working days), ranking the stocks according to the return exhibited by each one. The portfolio comprises long and short positions, buying the 10 worst stocks in the last window and selling the 10 best stocks. Each stock was included in the portfolio with a 10% weight.

In the case of the momentum strategy, the same ranking process is executed, but implementing opposite positions, buying the 10 best stocks in the last window and selling the 10 worst ones. Each stock is added in the portfolio with a 10% weight. An implementation of this strategy can be found in Jegadeesh and Titman (1993).

Finally, the equal-weighted strategy involves a portfolio in which all assets have the same weight, with daily rebalancing. The intuition behind this strategy is that it conceives the most naive diversified portfolio available, equally distributing wealth across assets. Transactions costs are also included to make the rebalancing procedure realistic. An implementation of this strategy can be found in Plyakha, Uppal and Vilkov (2012).

3.2.4.3 Bovespa Index

The Bovespa index (Ibovespa) is a market-weighted total return index, designed to gauge the stock market's average performance. The index tracks changes in the prices of the more actively traded and better representative stocks of the Brazilian stock market. It is a popular benchmark for Brazilian stock portfolios as well. Ibovespa's eligibility criteria cover liquidity and negotiability metrics, being a proxy for the performance of large capitalization stocks listed in the Brazilian market.

The use of Ibovespa as a benchmark to assess the portfolio performance is a common approach in empirical studies with Brazilian market data; see Palazzo et al. (2018).

3.2.4.4 Simplified Optimization Model Version

This model consists on a simplified version of the proposed model in the present work, discarding machine learning methods for the generation of price direction signals. Thus, the characteristic vector x_t of the model by Brandt, Santa-Clara and Valkanov (2009) is formed here directly by the financial ratios. These are first normalized then inserted into the model, that is, with zero mean and unit standard deviation. Thus, the model determines that stocks with similar characteristics must have similar weights, even if their historical returns are different, concentrating their allocation on characteristics that maximize the expected portfolio utility. An example of this application can be seen in Medeiros, Passos and Vasconcelos (2014), Lyle and Yohn (2020) and Hand and Green (2011).

3.2.5 Programming Languages

The proposed model was implemented using the Python language (version 3.7.6) and executed on a PC with an Intel Core i5-4210U processor, running Microsoft Windows with 8 GB of RAM.

For optimization algorithms and machine learning models, Scipy (version 1.4.1) and SKLearn (version 0.21.1) were respectively used. SKLearn provides several regression models selected to generate returns signals, such as Ridge regression, LASSO, and random forest.

Table 3 – Libraries and functions used in Python.

Model	**Library**	**Function**
Ridge	SKLearn	LogisticRegressionCV
LASSO	SKLearn	LogisticRegressionCV
Random Forest	SKLearn	RandomForestClassifier
k-NN Regression	SKLearn	KNeighborsClassifier
Optimize	Scipy	Minimize
SVC	SKLearn	SVC
Naïve Bayes	SKLearn	GaussianNB

The github page with the models is https://github.com/nellycolnaghi/FinancialAnalysis. In the page, it can be found: the parametric model proposed in the present work ("Brandt.ipynb"); the simplified strategy of this proposed model ("Simplified_Brandt.ipynb"); the Markowitz model ("Markowitz.ipynb"); the Contrarian strategy ("Contrarian.ipynb"); the Momentum strategy ("Momentum.ipynb"), the Equal-weighted strategy ("EW.ipynb") and the performance analysis ("Performance_Analysis"). The sensitivity analysis was implemented using "Brandt.ipynb" code, changing only its parameters.

4 RESULTS

In this section, we present the main results of the numerical exercise conducted to demonstrate the superiority of the portfolio optimization model proposed in this works. Initially, in order to introduce the process adopted to evaluate the risk and the returns of each portfolio built in the simulations, a discussion on performance measurement is carried out. Subsequently, the out-of-sample accuracy and ROC curves of the machine learning models used to generate trading signals are analyzed. Finally, the performance of the simulated portfolios and the benchmarks are investigated.

4.1 Performance Measurement

Here, an overview of performance measurement of investment portfolios is briefly examined. Prior to the presentation of the main metrics available in the literature, an exploration of the empirical properties of the return distributions of financial assets is necessary, for the moments of these distributions play a determinant role when ranking multiple investment alternatives.

4.1.1 Skewness and Kurtosis

According to Cont (2001), who presents a set of stylized empirical facts arising from a careful econometric analysis of asset returns in various financial markets, two important properties are commonly observed in the probability density function of these returns. First, unconditional distributions seem to display a power-law or Pareto-like tail, implying tails are heavier than expected were the returns normally distributed. This behavior persists even in conditional returns corrected for time-varying, clustering volatilities. Second, it is usually observed that stock prices and stock indices quotes suffer large drawdowns sporadically, but not equally large upward movements, creating an asymmetry in the return distribution.

Statistically, as detailed in Tsay (2010), these properties are related to the moments of third and fourth order of the return distribution. These moments, when normalized, are denominated *skewness* and *kurtosis*, respectively. The skewness has as fundamental property the ability to measure the asymmetry of the probability density in relation to its expected value. With respect to the normal distribution, which is perfectly symmetric (i.e. skewness equals to 0), a positive or negative skewness shifts the peak of the curve to the left or to the right of the mean, respectively. For investors, the first case is more advantageous, since it implies that gains are typically larger than losses. On the other hand, negative skewness translates into losses of greater magnitude than gains.

Mathematically, the skewness is defined as:

$$S = E\left[\frac{(X-\mu_x)^3}{\sigma_x^3}\right] \tag{4.1}$$

where X is a random variable of expected value μ_x and standard deviation σ_x. In his empirical study, Cont (2001) discovers that assets returns usually display negative skewness, for losses are historically more intense than gains.

Turning the attention to the kurtosis, this moment captures the likelihood of events falling in the tails of a probability distribution, i.e. the chance of extreme events. Due to the fact that the normal distribution has a kurtosis of 3, the excess kurtosis over this level is commonly referred to in practice. Hence, a distribution with positive excess kurtosis has heavy tails, meaning that the distribution allocates more mass on its tails than the normal does. As a consequence, as concluded by Tsay (2010), kurtosis also measures the flatness of a distribution, since heavier tails reduce the height of the peak of the curve.

Formally, the kurtosis of a random variable X with mean μ_x and standard deviation σ_x is given by:

$$K = E\left[\frac{(X-\mu_x)^4}{\sigma_x^4}\right] \tag{4.2}$$

In financial markets, Cont (2001) identifies that assets are characterized by extreme events happening somewhat frequently, emerging from the existence of positive excess kurtosis. Analogously to the case of volatility, dispersion is not necessarily detrimental if it is more prominent for positive rather than negative returns. However, in general, volatility and extreme events are far more common during turbulent scenarios, amplifying losses and, thus, explaining why most investors tend to avoid them and demand higher risk premiums for assets displaying high volatility / kurtosis.

4.1.2 Sharpe Ratio

Following Bodie, Kane and Marcus (2014), under certain conditions, investors assess risky assets according to their expected excess return and volatility (in this case, risk), requiring higher returns as the volatility grows. This principle is the basis of the Sharpe ratio, which measures how much excess return an investment offers for a fixed increment in the dispersion. Mathematically, it is expressed as:

$$\text{Sharpe ratio} = \frac{\text{Risk premium}}{\text{SD of excess return}} \tag{4.3}$$

The Sharpe ratio is widely used to evaluate and rank investment alternatives and portfolio managers. Applications are provided by Tola et al. (2008), Konno and Yamazaki (1991), and DeMiguel et al. (2009), to name a few examples.

4.1.3 Maximum Drawdown

The maximum drawdown (MDD) of any asset is the largest accumulated loss observed from the peak to a trough of the historical price series of this asset. Therefore, the maximum drawdown is a measure of downside risk and, by definition, the higher the MDD, the riskier the asset. It is usually accompanied by other measures because it does not offer insights regarding the frequency of large losses nor details how long it took for the losses to be recovered. An example of application of MDD in the context of portfolio optimization is provided by Chekhlov, Uryasev and Zabarankin (2005).

4.1.4 VaR e CVaR

Another risk measure employed in the financial industry in several contexts is VaR (Value at Risk). According to Hull (2008), the VaR of a portfolio can be modeled as a function of two parameters: a time horizon and a confidence level $1 - \alpha$. Thereby, for this fixed confidence level, VaR represents the minimum loss the portfolio might suffer in the time horizon considered with probability α. In other words, with probability α, the portfolio will report a loss greater than the computed VaR. Mathematically, let L be a random variable denoting the cumulative loss exhibited by the portfolio in a period of N days and $(1 - \alpha)$ the confidence level. Then the VaR may be expressed as:

$$\mathrm{VaR}_\alpha = \inf\left\{\mathrm{L}_0 \in \mathbb{R} : \mathrm{p}\left(\mathrm{L} > \mathrm{L}_0\right) \leq \alpha\right\} \tag{4.4}$$

Applications of VaR in the context of portfolio management are provided by Fabozzi et al. (2007), Ferreira and Ribeiro (2005), and Ghaoui, Oks and Oustry (2003).

Despite its popularity, VaR presents some computational problems and theoretical limitations about which practitioners should be warned, as highlighted by Krokhmal, Palmquist and Uryasev (2002). The authors illustrate mentioning the unavailability of optimization algorithms that are able to efficiently deal simultaneously with risk measures based on percentiles of a portfolio value distribution, which is the case of VaR, and a high number of decision variables (i.e. asset weights). VaR also does not fulfill the requirements to be classified as a *coherent* risk measure, as defined in the seminal paper by Artzner et al. (1999).

Due to these disadvantages, Rockafellar and Uryasev (2000) propose an extension of the VaR called CVaR (Conditional Value at Risk) that overcomes the limitations of the former. By definition, the CVaR of a portfolio corresponds to the expected loss conditioned on it being greater than the VaR. The CVaR formula is given by:

$$\mathrm{CVaR}_\alpha = \mathrm{E}\left[\mathrm{L} | \mathrm{L} \geq \mathrm{VaR}_\alpha\right] \tag{4.5}$$

Applications of CVaR in portfolio optimization models can be found in the works by Bertsimas, Lauprete and Samarov (2004), Krokhmal, Palmquist and Uryasev (2002), and Rockafellar and Uryasev (2000). In these examples, the formulations fall into convex linear programming problems, for which numerical solutions can be efficiently obtained.

For the purpose of the numerical exercise, both VaR and CVaR were computed using a confidence level of 95% and two different methods: historical and nonparametric. The latter involved kernel density estimators to estimate the density from observed values.

4.1.5 Omega

As established in the previous sections, rational, utility-maximizing investors select investments by assessing all the moments of the return distribution, without constraining the analysis solely to the expected value and the volatility. Therefore, any performance measure that relies on these two moments is incomplete, producing misleading ranks. As one may observe, this is the case of the Sharpe ratio.

With the aim of addressing this issue, Keating and Shadwick (2002) introduced a novel performance measure called Omega (Ω). It is defined by the following equation:

$$\Omega(\mathrm{r}) = \frac{\int_{\mathrm{r}}^{\mathrm{b}} [1 - \mathrm{F}(\mathrm{x})] \mathrm{dx}}{\int_{\mathrm{a}}^{\mathrm{r}} \mathrm{F}(\mathrm{x}) \mathrm{dx}} \tag{4.6}$$

where $F(x)$ is the cumulative probability distribution of portfolio returns, $[a, b]$ is the interval corresponding to the domain of F, and r is a predefined threshold above which returns are interpreted by investors as gains and below which they are classified as losses. Hence, Omega takes into account the investor's perception of gains and losses when assessing investments (FAVRE-BULLE; PACHE, 2003). In this sense, the greater the Omega, the more attractive a portfolio is for an investor (KEATING; SHADWICK, 2002).

Although not explicit in Equation 4.6, all the moments of the probability distribution are reflected on the computation of Omega, for they influence the shape of F and, consequently, the value of the integrals appearing in the numerator and the denominator. Therefore, this metric overcomes the limitations of the Sharpe ratio and others that focus exclusively on few low-order moments of the return distribution.

The use of Omega in practical contexts can be found in Keating and Shadwick (2002), Bertrand and Prigent (2010), and Favre-Bulle and Pache (2003). In general, these authors conclude that Omega enhances and deepens the relative performance analysis of multiple investment alternatives. Also, this measure may be coupled with portfolio optimization models to produce more robust allocations.

In the experiments, Omega has been computed using a nonparametric estimation of the density using KDE, which was later numerically integrated to find the value of the numerator and the denominator of the formula.

4.2 Analysis of the Trading Signals Extracted from Financial Ratios

As discussed in the preceding sections, the trading signals derived from financial ratios, which indicate whether a given stock will outperform or underperform the market in the next period, are the cornerstone of the portfolio optimization model proposed. Hence, it is crucial to evaluate the performance of the machine learning models employed to produce these signals and verify if they are able to yield informative predictions out-of-sample.

For binary classifier methods, it is recurrent to appraise the performance using the Receiver Operating Characteristic (ROC) curve, which portrays the diagnostic ability of the classifier as its probability discrimination threshold is changed. More specifically, independently of the method considered, its output is essentially the likelihood that a given point belongs to any of two existing categories. The point is ascribed to one of these categories according to the predefined threshold. Thus, by varying this parameter, different classifications are conceived.

The behavior of a classifier in terms of the discrimination threshold can be visualized in the ROC curve, which is structured by plotting the true positive rate (also known as *sensitivity* or *recall*) against the false positive rate while varying the threshold. In machine learning, the false positive rate is equivalent to 1 minus the specificity (share of negatives correctly identified). The curve not only unveils the trade-off between these rates, but allows to compare models in terms of this trade-off.

For this purpose, it is worth noting that a noninformative classifier model that randomly attributes points to one of the two possible categories produces a straight 45° line in the plot. In this framework, the quality of the classifier can be checked by comparing its ROC curve against that line and verifying if the area under the curve (AUC) is significantly greater than 0.5, number representing the area under the line. Furthermore, a steep ROC curve is desirable, for it means that high true positive rates can be achieved without increasing false positives (type I error).

After this brief introduction, it is possible to proceed to the analysis of the ROC curves of each machine learning model. The curves plotted are shown in Figure 7, which contains the outcomes for a selected rebalancing date. The profile verified in this date is similar to those observed in other periods, explaining why not all the plot have been reproduced here. Still, on Table 4, the areas under the curve for all rebalancing dates are compiled.

Immediately, it is possible to see that the curves possess two important properties: the area under the curve is significantly above 0.5, indicating that these models have predictive power, and the steepness is strong, yielding high true positive rates while maintaining a low type I error. Another feature of the plot is that the Random Forest ROC curve is clearly superior to the remaining curves, outperforming the competing classifiers. The

Table 4 – AUC for each model and holding period.

Holding Period	RF	k-NN	LASSO	Ridge	SVC	Bayes
1	0.8682	0.7714	0.7106	0.6948	0.7210	0.7064
2	0.9833	0.8747	0.8773	0.8560	0.8672	0.8637
3	0.9620	0.8341	0.8352	0.7996	0.8233	0.8470
4	0.9326	0.7303	0.7034	0.6788	0.6872	0.6914
5	0.9522	0.7800	0.7578	0.7816	0.7893	0.7842
6	0.9972	0.8790	0.8996	0.8800	0.8894	0.9022
7	0.9714	0.8072	0.8132	0.8229	0.8375	0.8067
8	0.9294	0.7592	0.7704	0.7788	0.7634	0.7550
9	0.8051	0.6830	0.6832	0.6726	0.6602	0.6174
10	0.9833	0.9009	0.8884	0.8852	0.8951	0.8859
11	0.9955	0.8163	0.8017	0.8214	0.7793	0.8277
12	0.9181	0.6986	0.6590	0.6585	0.6665	0.6385

superiority is also verified on other rebalancing dates, as Table 5 reveals. This conclusion is consistent with other studies, such as Ballings et al. (2015).

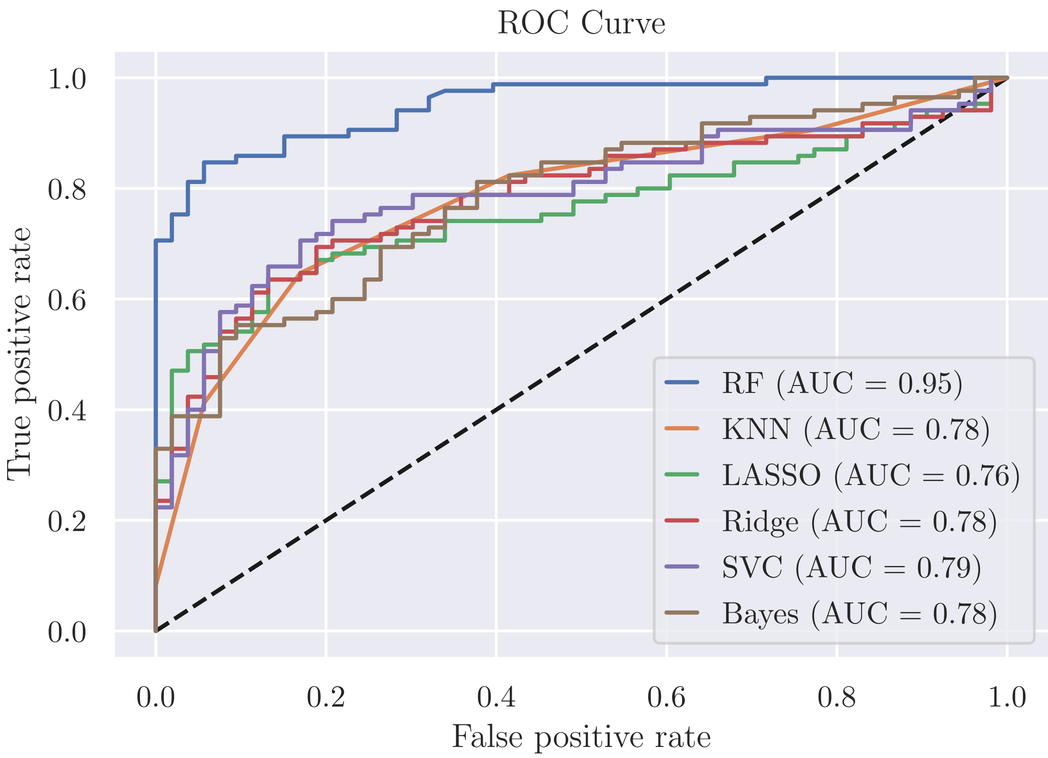

Figure 7 – ROC curve for each model.

The accuracy of each model is reported on Table 5. Overall, the results comply with those stemming from the analysis of the ROC curve, confirming that the machine learning models chosen produce informative out-of-sample forecasts with an average accuracy across time and models of 64%. However, it should be noted that accuracy oscillates through time,

Table 5 – Accuracy for each model and holding period.

Holding Period	RF	k-NN	LASSO	Ridge	SVC	Bayes
1	61%	57%	61%	52%	54%	54%
2	83%	63%	80%	83%	78%	78%
3	85%	83%	78%	76%	80%	85%
4	57%	50%	63%	61%	48%	54%
5	52%	43%	43%	43%	39%	50%
6	87%	85%	83%	72%	83%	87%
7	61%	59%	67%	65%	67%	67%
8	57%	57%	67%	65%	65%	63%
9	33%	35%	35%	37%	35%	37%
10	93%	91%	85%	87%	89%	93%
11	89%	85%	78%	80%	61%	91%
12	54%	35%	33%	33%	41%	33%

sometimes dropping below 50%. This behavior is attributed to the challenges imposed by a multitude of events that occurred during the time horizon analyzed, which led to periods of excessive volatility, harming the forecasting accuracy of every model considered.

4.3 Performance of the Simulated Portfolios

In this subsection, the performance of the simulated portfolios is presented. With this intent, in Figure 8, it is displayed the cumulative returns of each portfolio, while Figure 4.3 contains the summary statistics for each of them. Beginning with returns, the proposed portfolio, M1, managed to beat all the benchmarks in the investment horizon, reaching a higher total return net of costs. Both the average and the median return of this portfolio are superior in comparison to the benchmarks elected.

Moreover, the fraction of positive and negative returns unveils that gains are slightly more frequent than losses. The same is valid for excess returns. However, in line with Cont (2001), the skewness of the proposed portfolio is negative, meaning that the magnitude of these losses must be more intense than the magnitude of gains on average. All the portfolios exhibit negative skewness, except for the contrarian strategy, which could be explained by the fact that this strategy attempts to capture a mean-reversal in stock prices.

Moving towards kurtosis, the conclusion is that all the strategies generated distributions with heavy tails, although to different degrees. In the case of the proposed portfolio, the excess kurtosis is considerable and in line with the market portfolio. Markowitz, contrarian, and momentum strategies yielded lower kurtosis, while equal-weighted strategy displayed the highest value. In any situation, the outcomes corroborate that moments of order above two are essential when measuring the performance of stock portfolios.

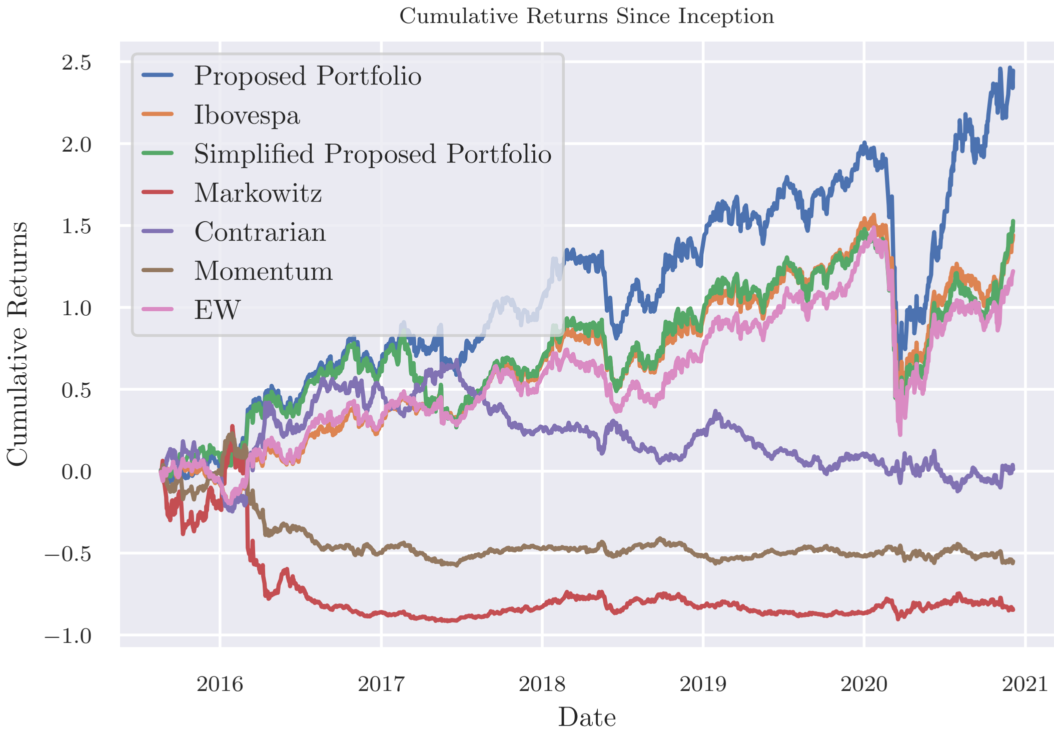

Figure 8 – Cumulative returns of each portfolio.

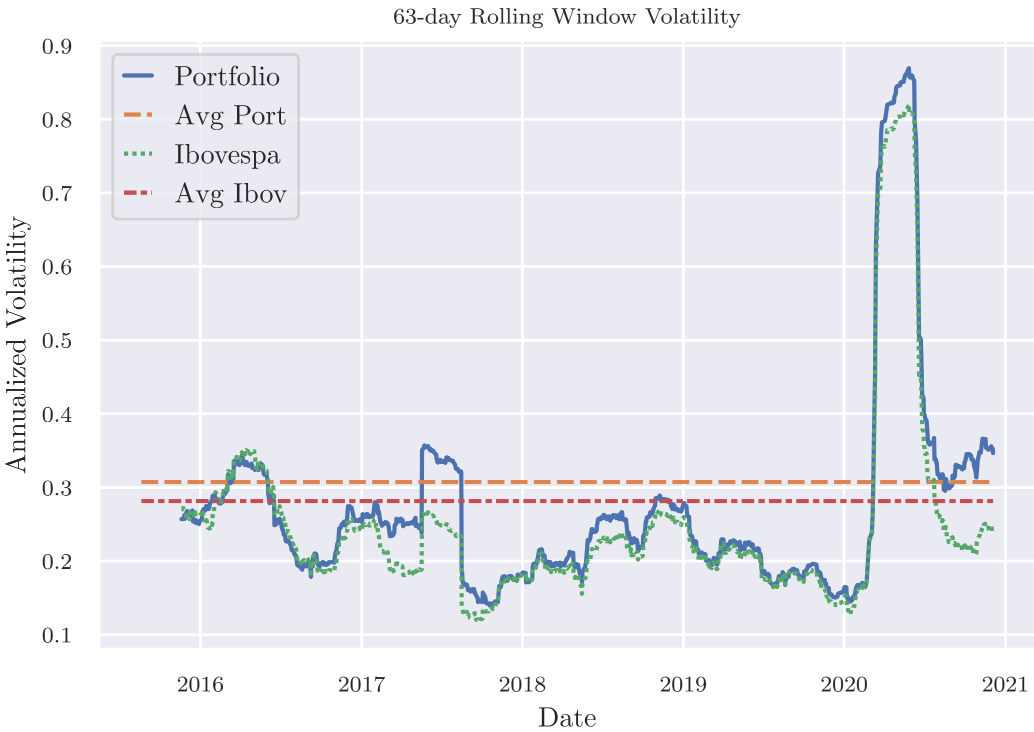

Figure 9 – Annualized volatility of the parametric portfolio proposed in the present work and Ibovespa.

Table 6 – Summary of the descriptive statistics of the return distributions of each portfolio. In the table, the models are identified as follows: M1 - the parametric model proposed in the present work; M2 - the simplified strategy of this proposed model, M3 - Markowitz model; M4 - Contrarian strategy; M5 - Bovespa index; M6 - Momentum strategy; and M7 - Equal-weighted strategy. Finally, “LTM” means Last Twelve Months.

Statistics	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**
Mean	0.11%	0.09%	-0.05%	0.01%	0.08%	-0.05%	0.08%
Daily Volatility	1.92%	1.94%	4.32%	1.57%	1.76%	1.57%	1.85%
Minimum	-15.08%	-15.08%	-28.54%	-6.39%	-14.78%	-12.92%	-16.01%
First Quartile	-0.82%	-0.83%	-2.53%	-0.85%	-0.74%	-0.82%	-0.76%
Median	0.13%	0.10%	0.08%	-0.03%	0.11%	0.02%	0.13%
Third Quartile	1.12%	1.08%	2.47%	0.80%	0.98%	0.82%	0.95%
Maximum	15.69%	15.69%	18.94%	12.92%	13.91%	6.39%	14.63%
Range	30.77%	30.77%	47.49%	19.31%	28.69%	19.31%	30.64%
Skewness	-0.72	-0.67	-0.39	0.81	-0.83	-0.84	-0.90
Excess Kurtosis	12.92	12.54	3.22	6.02	13.83	5.89	14.69
Hist. Volatility	30.48%	30.73%	68.60%	24.92%	27.91%	24.99%	29.37%
Volatility LTM	49.51%	47.70%	84.23%	27.80%	44.26%	27.94%	48.23%
Sharpe LTM	0.77	0.45	0.78	-0.17	0.30	-0.32	0.27
Hist. Sharpe	0.81	0.56	-0.29	-0.20	0.54	-0.79	0.46
MDD	-48.57%	-48.57%	-93.39%	-47.79%	-46.82%	-65.57%	-50.92%
Hist. VaR	-2.60%	-2.66%	-6.83%	-2.42%	-2.46%	-2.56%	-2.49%
Nonpar. VaR	-2.58%	-2.64%	-6.81%	-2.40%	-2.41%	-2.52%	-2.47%
Hist. CVaR	-4.34%	-4.38%	-10.03%	-3.29%	-4.02%	-4.04%	-4.30%
Nonpar. CVaR	-4.60%	-4.67%	-11.01%	-3.95%	-4.25%	-4.05%	-4.44%
Omega	1.28	1.23	0.99	0.98	1.22	0.91	1.23
Returns > 0	53.66%	52.75%	50.99%	48.78%	53.74%	50.69%	54.12%
Returns < 0	46.34%	47.25%	49.01%	51.22%	46.26%	49.31%	45.88%
Exc. Ret. > 0	52.90%	52.06%	50.84%	47.63%	52.98%	49.47%	53.21%
Exc. Ret. < 0	47.10%	47.94%	49.16%	52.37%	47.02%	50.53%	46.79%
Beta	0.97	1.03	0.54	0.10	1.00	-0.09	1.02

Relatively to the correlations between the strategies, Figure 10 presents the correlation heatmap for illustration. It is trivial to see that the proposed portfolio, its simplified version, the Ibovespa, and the equal-weighted strategy are highly correlated. This can be explained by the fact that the universe of stocks considered by these strategies is similar and they generate diversified portfolios. Momentum and contrarian produce more concentrated portfolios, and both implement a trading dynamics that naturally creates negative or weak correlation with the remaining portfolios.

In what concerns risk measures, the proposed portfolio yielded satisfactory results. The reported maximum drawdown is aligned with the Ibovespa and the other benchmarks, except for the Markowitz portfolio, which registered intense losses during certain periods of the investment horizon considered here. The VaR and CVaR demonstrate that the proposed portfolio conveys a controlled risk level, despite investing exclusive in equity,

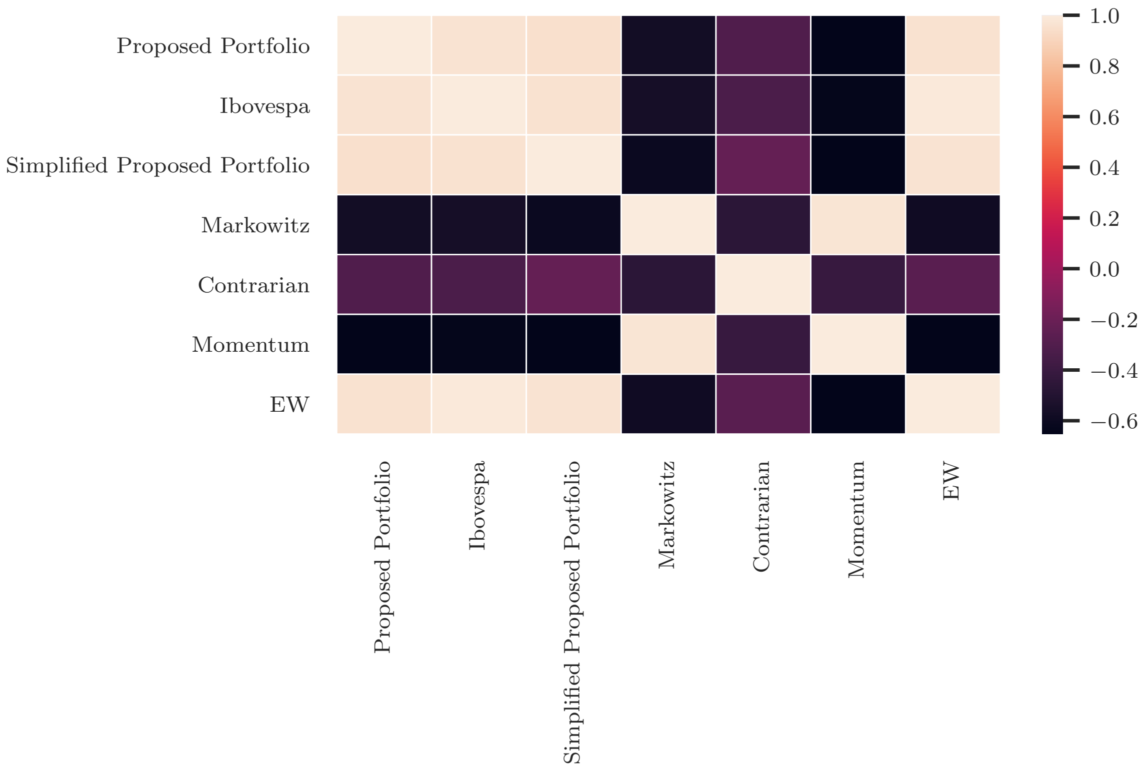

Figure 10 – Correlation heatmap of the portfolios.

which is naturally more volatile.

Together, these results suggest that the proposed portfolio must produce higher risk-adjusted returns, for it exhibits superior cumulative returns while consuming the same risk budget of its benchmarks. Indeed, when analyzing the performance metrics reported in the table, namely the Sharpe Ratio and Omega, it is possible to deduce that, in the investment horizon examined, the proposed portfolio outperformed its benchmarks.

Hence, the results categorically demonstrate that the approach developed for optimizing portfolios using signals derived from financial ratios accomplishes compelling risk-adjusted returns and surpasses its benchmarks. More specifically, it should be noted that both machine learning models for signal conception and the parameterization by Brandt, Santa-Clara and Valkanov (2009) contribute for the superior returns. Indeed, the simplified portfolio also generates convincing returns, which are enhanced when financial ratios are converted into signals prior to applying the parameterization.

4.4 Sensitivity Analysis

With the purpose of determining how sensitive are the results to the parameters of the proposed optimization model, additional simulations were run using distinct configurations, varying four parameters of the model: (1) the relative risk aversion (γ); (2) transaction costs;

(3) rebalancing window; and (4) the fixed threshold for stock price direction classification.

The impact of γ can be inferred from Table 7. Initially, it must be emphasized that increasing γ makes the investor more risk-averse. As expected, the simulations show that a higher gamma is associated with smaller returns and only slightly lower volatility. Consequently, the risk-adjusted returns suffer, as it may be seen from performance metrics such as Sharpe ratio and Omega.

Transaction costs also impose significant effects on portfolio optimization. Table 8 shows that these costs have the expected influence on returns, reducing them whenever the portfolio is rebalanced. Since volatility is virtually unchanged, the risk-adjusted returns are also harmed. However, the strategy developed here is robust to large variations in transaction costs, since even for a conservative value of 175 bps the strategy remains not only profitable, but with an attractive Omega.

The impact of the size of the rebalancing window is more subtle. The smaller the window, the more frequent rebalancing is, increasing transaction costs and making stock price direction forecasts more vulnerable to short-term noise / volatility. On the other hand, a longer time window may diminish these costs, but produce somewhat static portfolios that underperform due to changes in market conditions that are not readily reflected on allocations. The simulations unveil that the latter predominates, that is, longer rebalancing windows are associated with a lower return and slightly higher risk. Thus, risk-adjusted returns decrease when the rebalancing window increases.

Finally, in relation to the fixed threshold employed to transform forecast probabilities into classifications, the expected effects are also more complex to analyze. A higher threshold is more conservative, for it only classifies a stock into the category of outperformers if the classifiers are extremely certain that this stock belongs to that class. Thus, stocks that could also outperform, but with lower likelihood, are excluded, increasing the type I error. Nevertheless, a low threshold may be too flexible, classifying stocks as outperforms too leniently. Therefore, it is natural to believe that an optimal threshold exists, and Table 10 confirms this suspicion, showing that a threshold of 65% outperforms thresholds of 50% and 80%.

Table 7 – Sensitivity analysis of the optimal portfolio with respect to γ.

Statistics	γ = **1.1**	γ = **5**	γ = **100**
Mean	0.12%	0.11%	0.08%
Daily Volatility	1.93%	1.92%	1.86%
Minimum	-15.08%	-15.08%	-15.08%
First Quartile	-0.82%	-0.82%	-0.78%
Median	0.13%	0.13%	0.07%
Third Quartile	1.15%	1.12%	0.98%
Maximum	15.69%	15.69%	15.69%
Range	30.77%	30.77%	30.77%
Skewness	-0.71	-0.72	-0.70
Excess Kurtosis	12.68	12.92	14.03
Historical Volatility	30.61%	30.48%	29.48%
Volatility LTM	49.51%	49.51%	48.72%
Sharpe LTM	0.77	0.77	0.63
Historical Sharpe	0.83	0.81	0.43
MDD	-48.57%	-48.57%	-48.57%
Historical VaR	-2.61%	-2.60%	-2.58%
Nonparametric VaR	-2.61%	-2.58%	-2.53%
Historical CVaR	-4.35%	-4.34%	-4.26%
Nonparametric CVaR	-4.63%	-4.60%	-4.48%
Omega	1.28	1.28	1.19
Returns > 0	53.82%	53.66%	52.75%
Returns < 0	46.18%	46.34%	47.25%
Exc. Returns > 0	52.98%	52.90%	51.76%
Exc. Returns < 0	47.02%	47.10%	48.24%
Beta	0.96	0.97	0.98

Table 8 – Sensitivity analysis of the optimal portfolio with respect to transaction costs.

Statistics	**Cost = 0 bps**	**Cost = 100 bps**	**Cost = 175 bps**
Mean	0.12%	0.11%	0.11%
Daily Volatility	1.91%	1.92%	1.92%
Minimum	-15.08%	-15.08%	-15.08%
First Quartile	-0.82%	-0.82%	-0.82%
Median	0.15%	0.13%	0.13%
Third Quartile	1.12%	1.12%	1.11%
Maximum	15.69%	15.69%	15.69%
Range	30.77%	30.77%	30.77%
Skewness	-0.73	-0.72	-0.74
Excess Kurtosis	13.15	12.92	12.82
Historical Volatility	30.35%	30.48%	30.56%
Volatility LTM	49.51%	49.51%	49.67%
Sharpe LTM	0.87	0.77	0.70
Historical Sharpe	0.90	0.81	0.73
MDD	-48.57%	-48.57%	-48.57%
Historical VaR	-2.59%	-2.60%	-2.61%
Nonparametric VaR	-2.56%	-2.58%	-2.61%
Historical CVaR	-4.31%	-4.34%	-4.42%
Nonparametric CVaR	-4.57%	-4.60%	-4.63%
Omega	1.29	1.28	1.26
Returns > 0	53.74%	53.66%	53.44%
Returns < 0	46.26%	46.34%	46.56%
Exc. Returns > 0	53.05%	52.90%	52.67%
Exc. Returns < 0	46.95%	47.10%	47.33%
Beta	0.97	0.97	0.97

Table 9 – Sensitivity analysis of the optimal portfolio with respect to rebalancing window (in business days).

Statistics	**Window = 63**	**Window = 126**	**Window = 252**
Mean	0.09%	0.11%	-0.03%
Daily Volatility	2.02%	1.92%	2.15%
Minimum	-15.08%	-15.08%	-15.01%
First Quartile	-0.80%	-0.82%	-0.94%
Median	0.08%	0.13%	0.01%
Third Quartile	1.12%	1.12%	0.95%
Maximum	15.69%	15.69%	15.86%
Range	30.77%	30.77%	30.88%
Skewness	-0.56	-0.72	-0.44
Excess Kurtosis	10.23	12.92	9.13
Historical Volatility	32.11%	30.48%	34.07%
Volatility LTM	45.13%	49.51%	59.79%
Sharpe LTM	0.27	0.77	-1.17
Historical Sharpe	0.55	0.81	-0.48
MDD	-48.94%	-48.57%	-79.59%
Historical VaR	-2.81%	-2.60%	-3.24%
Nonparametric VaR	-2.79%	-2.58%	-3.15%
Historical CVaR	-4.75%	-4.34%	-5.42%
Nonparametric CVaR	-4.90%	-4.60%	-5.36%
Omega	1.20	1.28	1.01
Returns > 0	52.29%	53.66%	50.38%
Returns < 0	47.71%	46.34%	49.62%
Exc. Returns > 0	51.37%	52.90%	49.39%
Exc. Returns < 0	48.63%	47.10%	50.61%
Beta	0.91	0.97	1.02

Table 10 – Sensitivity analysis of the optimal portfolio with respect to thresholds (t).

Statistics	**t = 50%**	**t = 65%**	**t = 80%**
Mean	0.09%	0.11%	0.09%
Daily Volatility	1.91%	1.92%	1.90%
Minimum	-15.08%	-15.08%	-15.08%
First Quartile	-0.80%	-0.82%	-0.82%
Median	0.11%	0.13%	0.07%
Third Quartile	1.09%	1.12%	1.05%
Maximum	15.69%	15.69%	15.69%
Range	30.77%	30.77%	30.77%
Skewness	-0.65	-0.72	-0.60
Excess Kurtosis	12.71	12.92	12.56
Historical Volatility	30.31%	30.48%	30.23%
Volatility LTM	49.10%	49.51%	49.02%
Sharpe LTM	0.47	0.77	0.56
Historical Sharpe	0.60	0.81	0.52
MDD	-48.57%	-48.57%	-48.57%
Historical VaR	-2.58%	-2.60%	-2.60%
Nonparametric VaR	-2.59%	-2.58%	-2.58%
Historical CVaR	-4.32%	-4.34%	-4.29%
Nonparametric CVaR	-4.59%	-4.60%	-4.58%
Omega	1.21	1.28	1.20
Returns > 0	53.36%	53.66%	52.60%
Returns < 0	46.64%	46.34%	47.40%
Exc. Returns > 0	52.67%	52.90%	51.91%
Exc. Returns < 0	47.33%	47.10%	48.09%
Beta	0.99	0.97	1.01

5 CONCLUSION

The present work focused on designing a portfolio optimization model that extracts trading signals from financial ratios to determine which stocks are negotiating with a discount with respect to their fundamentals and, thus, tend to outperform in the future. For this purpose, it was chosen the parameterization approach developed by Brandt, Santa-Clara and Valkanov (2009) to write the asset weights of a portfolio in terms of assets' characteristics. In addition, supported by the literature on stock price direction predictions, machine learning models such as Random Forests, k-NN, and SVC are employed to convert financial ratios into the desired signals. By merging these distinct techniques and incorporating financial ratios into an innovative model, the work contributes to the literature on portfolio optimization.

With the objective of testing the optimization process proposed, a numerical exercise has been conceived, applying the model to the Brazilian stock market, and comparing its performance against appropriate benchmarks. The results confirm that the combination of the parameterization by Brandt, Santa-Clara and Valkanov (2009) with machine learning models for signal generation yields compelling out-of-sample risk-adjusted returns, beating all the benchmarks selected. These findings are robust, for they stem from a study covering a long and turbulent period for stock markets. The sensitivity analysis reinforces this conclusion by proving that changing the parameters of the model does not compromise the superiority of the optimized portfolio nor the value aggregated parameterization by Brandt, Santa-Clara and Valkanov (2009) and the signals produces by the machine learning models.

Despite the contributions of the present work to the literature, there are many opportunities for improvement and extension. Initially, the rebalancing strategy implemented here could be replaced by a stochastic dynamic optimization approach in which multiple periods ahead and the expected evolution of asset prices and volatilities are considered when deciding the allocation today; see Zhang, Li and Guo (2018). Through this adaptation, the optimization would be not only more realistic, but more precise, since it would capture more accurately the effects of transaction costs, for instance.

In an effort to attenuate the impacts of input uncertainty, robust optimization is another feasible and promising extension. As defined by Gabrel, Murat and Thiele (2014), this category of mathematical optimization encompasses several approaches designed to protect the decision-maker against parameter ambiguity and stochastic uncertainty. In the context contemplated in this dissertation, many sources of uncertainty can be identified. For instance, the financial ratios computed are prone to errors due to questionable (overly aggressive or conservative) or unstable accounting practices followed by firms. Expected returns are also vulnerable since they are not observable and, thus, must be estimated.

For illustration, an example of application of robust optimization to multiperiod portfolio management in the presence of transaction costs is developed by Bertsimas and

Pachamanova (2008). The results corroborate that robust portfolios are more resilient to misspecifications in the parameters of the model, clearly dominating other configurations in situations that most resemble real life.

With respect to the machine learning models used to generate trading signals from financial ratios, there is also room for enhancements. An alternative to be explored involves neural networks. Indeed, Giunduz et al. (2015) predict daily movement directions using deep networks. Inputs are comprised by technical indicators obtained from individual stock prices and by dollar and gold prices. Forecasts are performed by convolutional networks. The authors obtain satisfactory accuracy rates out-of-sample. In a similar manner, Oliveira, Nobre and Zárate (2013) use neural networks for stock price prediction in the Brazilian stock market, reaching a percentage of correct predictions of 87.50% for the validation set.

BIBLIOGRAPHY

ABARBANELL, J. S.; BUSHEE, B. J. Abnormal returns to a fundamental analysis strategy. *The Accounting Review*, v. 73, n. 1, p. 19–45, Jan. 1998. Cited on page 27.

ABROKWA, J.; NKANSAH, P. Predictors of stock returns: Some evidencefrom an emerging market. *Academy of Accounting and Financial Studies Journal*, v. 19, n. 3, p. 1–8, Mar. 2015. Cited 2 times on pages 25 and 26.

ARTZNER, P. et al. Coherent measures of risk. *Mathematical Finance*, v. 9, n. 3, p. 203–228, Jul. 1999. Cited on page 59.

AVRAMOV, D.; KAPLANSKI, G.; SUBRAHMANYAM, A. Moving average distance as a predictor of equity returns. Review of Financial Economics, Forthcoming. 2020. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3111334>. Last access: 2 Out. 2020. Cited on page 26.

BALLINGS, M. et al. Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, v. 42, n. 20, p. 7046–7056, Nov. 2015. Cited 4 times on pages 13, 21, 30, and 62.

BARESA, S.; BOGDAN, S.; IVANOVIC, Z. Strategy of stock valuation by fundamental analysis. *UTMS Journal of Economics*, v. 4, n. 1, p. 45–51, Jan. 2013. Cited 3 times on pages 13, 21, and 25.

BERNSTEIN, A. L. In defense of fundamental investment analysis. *Financial Analysts Journal*, v. 31, n. 01, p. 57–61, Feb. 1975. Cited on page 21.

BERTRAND, P.; PRIGENT, J. L. Omega performance measure and portfolio insurance. *Journal of Banking & Finance*, v. 34, n. 12, p. 1–18, Dec. 2010. Cited on page 60.

BERTSIMAS, D.; LAUPRETE, G. J.; SAMAROV, A. Shortfall as a risk measure: properties, optimization and applications. *Journal of Economic Dynamics and Control*, v. 28, n. 7, p. 1353–1381, Apr. 2004. Cited on page 59.

BERTSIMAS, D.; PACHAMANOVA, D. Robust multiperiod portfolio management in the presence of transaction costs. *Computer & Operations Research*, v. 35, n. 1, p. 3–17, Jan. 2008. Cited on page 74.

BODIE, Z.; KANE, A.; MARCUS, A. J. *Investments*. 10. ed. New York: McGraw-Hill Education, 2014. Cited 2 times on pages 21 and 58.

BONDT, W. F. M. D.; THALER, R. Does the stock market overreact? *The Journal of Finance*, v. 40, n. 2, p. 28–30, Jul. 1985. Cited on page 54.

BRANDT, M. W.; SANTA-CLARA, P.; VALKANOV, R. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, v. 22, n. 09, p. 3411–3447, Sep. 2009. Cited 17 times on pages 9, 11, 13, 14, 22, 23, 42, 43, 44, 45, 48, 49, 50, 52, 55, 66, and 73.

CESSIE, S. L.; HOUWELINGEN, J. C. V. Ridge estimators in logistic regression. *Journal of the Royal Statistical Society*, v. 41, n. 1, p. 191–201, Jul. 2007. Cited on page 34.

CHEKHLOV, A.; URYASEV, S.; ZABARANKIN, M. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance*, v. 8, n. 01, p. 13–58, Jan. 2005. Cited on page 59.

CHEN, Y.; HAO, Y. A feature weighted support vector machine and k-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, v. 80, n. 1, p. 340–355, Sep. 2017. Cited 2 times on pages 30 and 31.

CONT, R. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, v. 1, n. 2, p. 223–236, Feb. 2001. Cited 3 times on pages 57, 58, and 63.

DEMIGUEL, V. et al. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, v. 55, n. 5, p. 798–812, May 2009. Cited on page 58.

DEMIGUEL, V. et al. A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, v. 33, n. 5, p. 2180–2222, Apr. 2020. Cited 3 times on pages 44, 45, and 48.

DOMINGOS, P.; PAZZANI, M. On the optimality of the simple bayesian classifier under zero-one loss: Improving performance by constraining portfolio norms. *Machine Learning*, v. 29, n. 1, p. 103–130, Nov. 1997. Cited on page 40.

DUTTA, A.; BANDOPADHYAY, G.; SENGUPTA, S. Prediction of stock performance in the indian stock market using logistic regression. *The International Journal of Business and Information*, v. 7, n. 1, p. 105–136, Jun. 2012. Cited on page 30.

ELY, R. A. Returns predictability and stock market efficiency in brazil. *Revista Brasileira de Finanças*, v. 9, n. 4, p. 571–584, Jan. 2011. Cited 3 times on pages 21, 26, and 28.

FABOZZI, F. J. et al. *Robust portfolio optimization and management*. 1. ed. New Jersey: John Wiley & Sons, 2007. Cited 2 times on pages 41 and 59.

FAMA, E. F.; MACBETH, J. D. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, v. 81, n. 3, p. 607–636, May. 1973. Cited 2 times on pages 25 and 26.

FAVRE-BULLE, A.; PACHE, S. *The omega measure:Hedge fund portfolio optimization*. 45 f. Dissertação (Mestrado em Finanças) — Université de Lausanne, Lausanne, 2003. Cited on page 60.

FERREIRA, L. A. S.; RIBEIRO, C. O. Uma contribuição ao problema de composição de carteiras de mínimo valor em risco. *Gestão & Produção*, v. 12, n. 2, p. 295–304, May 2005. Cited on page 59.

FONG, W. M.; WONG, W. K.; LEAN, H. H. International momentum strategies: A stochastic dominance approach. *Journal of Financial Markets*, v. 8, n. 1, p. 89–109, Feb. 2005. Cited on page 51.

GABREL, V.; MURAT, C.; THIELE, A. Recent advances in robust optimization: An overview. *European Journal of Operational Research*, v. 235, n. 3, p. 471–483, Jun. 2014. Cited on page 73.

GHAOUI, L. E.; OKS, M.; OUSTRY, F. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operations Research*, v. 51, n. 4, p. 543–556, Jul. 2003. Cited on page 59.

GHAZANFAR, M. A. et al. Using machine learning classifiers to predict stock exchange index. *International Journal of Machine Learning and Computing*, v. 7, n. 2, p. 24–29, Apr. 2017. Cited on page 30.

GIUNDUZ, H. et al. Stock market prediction using deep neural networks. In: SIGNAL PROCESSING AND COMMUNICATIONS APPLICATIONS CONFERENCE, 25., 2017, Antalya, Turkey. *Proceedings...* [S.l.]: IEEE, 2015. p. 1–4. Cited on page 74.

GRAHAM, B. *O investidor inteligente*. Rio de Janeiro: Nova Fronteira, 2007. Cited on page 44.

GROSSMAN, S. J.; STIGLITZ, J. E. On the impossibility of informationally efficient markets. *The American Economic Review*, v. 70, n. 3, p. 393–408, Jun. 1980. Cited 2 times on pages 26 and 27.

HAND, D. J.; YU, K. Idiot's bayes: Not so stupid after all? *International Statistical Review*, v. 69, n. 03, p. 385–398, May 2001. Cited on page 40.

HAND, J. R. M.; GREEN, M. The importance of accounting information in portfolio optimization. *Journal of Accounting, Auditing & Finance*, v. 26, n. 01, p. 01–34, Jan. 2011. Cited 4 times on pages 22, 43, 49, and 55.

HANSEN, P. R.; LUNDE, A.; NASON, J. M. The model confidence set. *Econometrica*, v. 79, n. 2, p. 453–497, Mar. 2011. Cited on page 48.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning*: Data mining, inference, and prediction. 2. ed. California: Springer, 2008. Cited 6 times on pages 15, 31, 32, 34, 35, and 36.

HORRIGAN, J. O. Some empirical bases of financial ratio analysis. *The Accounting Review*, v. 40, n. 03, p. 558–568, Jul. 1965. Cited on page 25.

HSU, J.; KALESNIK, V.; KOSE, E. What is quality? *Financial Analysts Journal*, v. 75, n. 2, p. 44–61, Mar. 2019. Cited 2 times on pages 26 and 29.

HUANG, W.; NAKAMORI, Y.; WANG, S. Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*, v. 32, n. 10, p. 2513–2522, Oct. 2005. Cited 2 times on pages 13 and 21.

HULL, J. C. *Options, futures and other derivatives*. 7. ed. New Jersey: Pearson Prentice-Hall, 2008. Cited on page 59.

JAMES, G. et al. *An introduction to statistical learning with applications in R*. 1. ed. New York: Springer, 2013. Cited 8 times on pages 15, 31, 32, 33, 36, 37, 38, and 39.

JEGADEESH, N.; TITMAN, S. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, v. 48, n. 1, p. 65–91, Mar. 1993. Cited on page 54.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, v. 349, n. 6245, p. 255–260, Jul. 2015. Cited 3 times on pages 13, 22, and 29.

KARA, Y.; BOYACIOGLU, M. A.; BAYKAN, O. K. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. *Expert Systems with Applications*, v. 38, n. 5, p. 5311–5319, May 2011. Cited 2 times on pages 13 and 21.

KEATING, C.; SHADWICK, W. F. A universal performance measure. *The Journal of Performance Measurement*, v. 6, n. 3, p. 1–42, May 2002. Cited on page 60.

KHERADYAR, S.; IBRAHIM, I.; NOR, F. M. Stock return predictability with financial ratios. *International Journal of Trade, Economics and Finance*, v. 2, n. 5, p. 391–396, Oct. 2011. Cited on page 26.

KOH, K.; KIM, S.; BOYD, S. An interio-point method for large-scale l1-regularized logistic regression. *Expert Systems with Applications*, v. 8, n. 54, p. 1519–1555, Jul. 2007. Cited on page 34.

KOLM, P. N.; TUTUNCU, R.; FABOZZI, F. J. 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, v. 234, n. 2, p. 356–371, Apr. 2014. Cited 3 times on pages 40, 41, and 42.

KONNO, H.; YAMAZAKI, H. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management Science*, v. 37, n. 5, p. 519–531, May 1991. Cited on page 58.

KOSOWSKI, R.; NAIK, N. Y.; TEO, M. Do hedge funds deliver alpha?: A Bayesian and bootstrap analysis. *Journal of Financial Economics*, v. 84, n. 1, p. 229–264, Jan. 2007. Cited 4 times on pages 13, 21, 26, and 28.

KROKHMAL, P.; PALMQUIST, J.; URYASEV, S. Portfolio optimization with conditional value-at-risk objective and constraints. *The Journal of Risk*, v. 4, n. 2, p. 11–27, Jan. 2002. Cited on page 59.

KUMAR, N. Importance of technical and fundamental analysis and other strategic factors in the indian stock market. *Management Review: An International Journal*, v. 8, n. 1, p. 38–120, 2013. Cited 2 times on pages 26 and 29.

KUNCHEVA, L. I. On the optimality of naive bayes with dependent binary features. *Pattern Recognition Letters*, v. 27, n. 7, p. 830–837, May 2006. Cited 2 times on pages 39 and 40.

LAPORTA, R. Expectations and the cross-section of stock returns. *The Journal of Finance*, v. 51, n. 5, p. 1715–1742, Dec. 1996. Cited on page 26.

LEWELLEN, J. Predicting returns with financial ratios. *Journal of Financial Economics*, v. 74, n. 2, p. 209–235, Nov. 2004. Cited on page 26.

LIM, K.; BROOKS, R. The evolution of stock market efficiency over time: A survey of the empirical literature. *Journal of Economic Surveys*, v. 25, n. 1, p. 69–108, Feb. 2011. Cited 3 times on pages 26, 27, and 28.

LO, A. W. The adaptive markets hypothesis. *The Journal of Portfolio Management 30th Anniversary Issue*, v. 30, n. 5, p. 15–29, 2004. Cited 3 times on pages 26, 27, and 28.

LUENBERGER, D. G. *Investment science*. 1. ed. New York: Oxford University Press, 1998. Cited on page 41.

LYLE, M. R.; YOHN, T. L. Fundamental analysis and mean-variance optimal portfolios. Kelley School of Business Research Paper Forthcoming. 2020. Available at: <https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3300976>. Last access: 21 Sep. 2020. Cited 3 times on pages 22, 43, and 55.

MARKOWITZ, H. Portfolio selection. *The Journal of Finance*, v. 7, n. 1, p. 77–91, Mar. 1952. Cited 3 times on pages 40, 41, and 42.

MEDEIROS, M. C.; PASSOS, A. M.; VASCONCELOS, G. F. R. Parametric portfolio selection: Evaluating and comparing to markowitz portfolios. *Revista Brasileira de Finanças*, v. 12, n. 2, p. 257–284, Ju. 2014. Cited 6 times on pages 22, 43, 48, 49, 50, and 55.

MURESAN, E.; WOLITZER, P. Organize your financial ratios analysis with palms. EMuresan Working Paper No. 02-01. 2004. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=375880>. Last access: 20 Oct. 2020. Cited on page 25.

OLIVEIRA, F. A.; NOBRE, C. N.; ZáRATE, L. E. Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index: Case study of PETR4, Petrobras, Brazil. *Computer & Operations Research*, v. 40, n. 18, p. 7596–7606, Dec. 2013. Cited on page 74.

PALAZZO, V. et al. Analysis of value portfolios in the brazilian market. *Revista de Contabilidade e Finanças*, v. 29, n. 78, p. 452–468, Dec. 2018. Cited 3 times on pages 44, 53, and 54.

PANAGIOTIDIS, T.; STENGOS, T.; VRAVOSINOS, O. On the determinants of bitcoin returns: A lasso approach. *Finance Research Letters*, v. 27, n. 1, p. 235–240, Dec. 2018. Cited on page 30.

PATEL, J. et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, v. 42, n. 1, p. 259–268, Jan. 2015. Cited on page 30.

PEREIRA, J. M.; BASTO, M.; SILVA, A. F. The logistic lasso and ridge regression in predicting corporate failure. In: GLOBAL CONFERENCE ON BUSINESS, ECONOMICS, MANAGEMENT AND TOURISM, 39., 2015, Rome, Italy. *Proceedings...* [S.l.]: Procedia Economics and Finance, 2016. p. 634–641. Cited 2 times on pages 30 and 35.

PLYAKHA, Y.; UPPAL, R.; VILKOV, G. Why does an equal-weighted portfolio outperform value- and price-weighted portfolios. Edhec-Risk Institute, Working Paper. 2012. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2724535>. Last access: 20 Feb. 2020. Cited on page 54.

RAPACH, D. E.; WOHAR, M. E. Valuation ratios and long-horizon stock price predictability. *Journal of Applied Econometrics*, v. 20, n. 3, p. 327–344, Mar. 2005. Cited on page 26.

ROCKAFELLAR, R. T.; URYASEV, S. Optimization of conditional value-atrisk. *The Journal of Risk*, v. 2, n. 3, p. 21–41, Apr. 2000. Cited 2 times on pages 59 and 60.

SEHGAL, S. Equity valuation using price multiples: A comparative study for bricks. *Asian Journal of Finance & Accounting*, v. 2, n. 1, p. 68–91, Dec. 2010. Cited 2 times on pages 26 and 27.

SHEN, K.; TZENG, G. Combined soft computing model for value stock selection based on fundamental analysis. *Applied Soft Computing*, v. 37, p. 142–155, Aug. 2015. Cited 2 times on pages 26 and 27.

SHIHAVUDDIN, A. S. M. et al. Prediction of stock price analyzing the online financial news using naive bayes classifier and local economic trends. In: INTERNATIONAL CONFERENCE ON ADVANCED COMPUTER THEORY AND ENGINEERING, 1., 2010, Washington, DC, USA. *Proceedings...* [S.l.]: IEEE, 2010. p. 22–26. Cited on page 30.

SILVA, A. D. V. *Portfolio optimization using fundamental indicators based on multi-objective EA*. 108 f. Dissertação (Mestrado em Engenharia e Gestão Industrial) — Técnico Lisboa, Lisboa, 2014. Cited on page 43.

SONG, Y. et al. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing*, v. 251, n. 3, p. 26–34, Aug. 2017. Cited on page 36.

SPOONER, M. C. Origin of fundamental analysis. *Financial Analysts Journal*, v. 40, n. 04, p. 79–80, Apr. 1984. Cited on page 21.

TOLA, V. et al. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, v. 32, n. 1, p. 235–258, Jan. 2008. Cited on page 58.

TSAI, C. et al. Predicting stock returns by classifier ensembles. *Applied Soft Computing*, v. 11, n. 2, p. 2452–2459, Mar. 2011. Cited 2 times on pages 13 and 21.

TSAY, R. S. *Analysis of financial time series*. 3. ed. New Jersey: Wileyr, 2010. Cited 2 times on pages 57 and 58.

VASCONCELOS, G. F. R. *Forecasting in high-dimension*: Inflation and other economic variables. 169 p. Tese (Doutorado em Engenharia Elétrica) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2018. Cited 2 times on pages 35 and 48.

WADHWA, B. Financial ratios: The precarious core of fundamental analysis. *Frontiers Journal of Accounting and Business Research*, v. 1, n. 1, p. 33–35, Apr. 2019. Cited 2 times on pages 21 and 25.

WHITTINGTON, G. Some basic properties of accounting ratios. *Journal of Business Finance and Accounting*, p. 219–223, 1980. Cited on page 25.

WOLFF, D.; ECHTERLING, F. Stock picking with machine learning. Forthcoming. 2020. Available at: <https://ssrn.com/abstract=3607845>. Last access: 2 Out. 2020. Cited on page 30.

WU, L.; YANG, Y.; LIU, H. Nonnegative-lasso and application in index tracking. *Computational Statistics and Data Analysis*, v. 70, n. 1, p. 116–126, Feb. 2014. Cited on page 30.

YANG, H.; LIU, X.; WU, Q. A practical machine learning approach for dynamic stock recommendation. In: INTERNATIONAL CONFERENCE ON TRUST, SECURITY AND PRIVACY IN COMPUTING AND COMMUNICATIONS, 17., 2018, New York, USA. *Proceedings...* [S.l.]: IEEE, 2018. p. 1693–1697. Cited 2 times on pages 29 and 30.

ZHANG, N.; LIN, A.; SHANG, P. Multidimensional k-nearest neighbor model based on eemd for financial time series forecasting. *Physica A: Statistical Mechanics and its Applications*, v. 477, n. 1, p. 161–173, Jul. 2017. Cited 2 times on pages 30 and 31.

ZHANG, Y.; LI, X.; GUO, S. Portfolio selection problems with markowitz's mean-variance framework: A review of literature. *Fuzzy Optimization and Decision Making*, v. 17, n. 1, p. 125–158, Jun. 2018. Cited on page 73.

APPENDIX A – STOCK DATASET

Company	Ticker
Ambev S.A.	ABEV3 BS Equity
Banco Bradesco S.A.	BBDC4 BS Equity
Localiza Rent a Car S.A.	RENT3 BS Equity
B3 S.A. - Brasil, Bolsa, Balcão	B3SA3 BS Equity
Banco Bradesco S.A.	BBDC3 BS Equity
Banco do Brasil S.A.	BBAS3 BS Equity
BR Malls Participacoes S.A.	BRML3 BS Equity
Braskem S.A.	BRKM5 BS Equity
BRF S.A.	BRFS3 BS Equity
CCR S.A.	CCRO3 BS Equity
Cia de Saneamento Basico do Estado de Sao Paulo S.A.	SBSP3 BS Equity
Cia Energetica de Minas Gerais S.A.	CMIG4 BS Equity
Cia Siderurgica Nacional S.A.	CSNA3 BS Equity
Cogna Educacao S.A.	COGN3 BS Equity
Companhia Hering S.A.	HGTX3 BS Equity
Cosan S.A.	CSAN3 BS Equity
CPFL Energia S.A.	CPFE3 BS Equity
Cyrela Brazil Realty S.A.	CYRE3 BS Equity
Embraer S.A.	EMBR3 BS Equity
Energias do Brasil S.A.	ENBR3 BS Equity
Energisa S.A.	ENGI11 BS Equity
Engie Brasil Energia S.A.	EGIE3 BS Equity
Equatorial Energia S.A.	EQTL3 BS Equity
Gerdau S.A.	GGBR4 BS Equity
Hypera S.A.	HYPE3 BS Equity
Itau Unibanco Holding S.A.	ITUB4 BS Equity
Itausa S.A.	ITSA4 BS Equity
JBS S.A.	JBSS3 BS Equity
Lojas Americanas S.A.	LAME4 BS Equity
Lojas Renner S.A.	LREN3 BS Equity
Metalurgica Gerdau S.A.	GOAU4 BS Equity
MRV Engenharia e Participacoes S.A.	MRVE3 BS Equity
Multiplan Empreendimentos Imobiliarios S.A.	MULT3 BS Equity
Natura & Co Holding S.A.	NTCO3 BS Equity
Petroleo Brasileiro S.A.	PETR4 BS Equity
Petroleo Brasileiro S.A.	PETR3 BS Equity

Raia Drogasil S.A.	RADL3 BS Equity
Transmissora Alianca de Energia Eletrica S.A.	TAEE11 BS Equity
Sul America S.A.	SULA11 BS Equity
Telefonica Brasil S.A.	VIVT4 BS Equity
TOTVS S.A.	TOTS3 BS Equity
Ultrapar Participacoes S.A.	UGPA3 BS Equity
Usinas Siderurgicas de Minas Gerais S.A.	USIM5 BS Equity
Vale S.A	VALE3 BS Equity
WEG S.A.	WEGE3 BS Equity
YDUQS Participacoes S.A.	YDUQ3 BS Equity

APPENDIX B – FINANCIAL RATIOS

Indicator	Description
Asset Turnover	Amount of sales or revenues generated per dollar of assets
Cash Coversion Cycle	Metric which expresses the length of time, in days, that it takes for a company to convert resource inputs into cash flows
Current Ratio	Ratio to indicate the company's ability to pay back its short-term liabilities with its short-term assets. Unit: Actual
Degree of Financial Leverage	Leverage ratio summarizing the affect a particular amount of financial leverage has on a company's earnings
Dividend 12 Month Yield	The sum of net dividend per share amounts that have gone ex-dividend over the prior 12 months, divided by the current stock price
Dividend Payout Ratio	Fraction of net income a firm pays to its shareholders in dividends, in percentage
Dividend Yield	Ratio calculated by dividing the dividend per share (DPS) estimate based on the current fiscal year provided by the requested firm/broker by the current price of the security
EBIT Margin	Ratio which measures the company's profitability
EBITDA Margin	Percentage margin of trailing 12 month Earnings Before Interest Taxes Depreciation and Amortization (EBITDA) divided by the trailing 12 month Sales
EV to Trailing 12M Sales	Current Enterprise Value / Trailing 12 Month Sales
Interest Coverage Ratio	Net income, less interest expense, real estate depreciation and amortization of wholly owned and joint venture communities, other depreciation and amortization, minority interests, net gain on the sale of depreciable property, excluding income tax, divided by total interest
Leverage Ratio - Tangible Capital Ratio	Leverage Ratio or Tangible Cap Ratio
Market Capitalization	Total current market value of all of a company's outstanding shares stated in the pricing currency
Net Income to Common Margin	Ratio of net income available to common shareholders to total revenue, expressed in percentage. This alternative calculation of profit margin makes the ratio more relevant to common shareholders
Operating Margin	Ratio used to measure a company's pricing strategy and operating efficiency, in percentage

Operating Return on Total Invested Capital	Indication of how effectively a company uses the sources of capital through its operations
Periodic EV to Trailing 12M EBIT	Periodic enterprise value as a multiple of earnings before interest and taxes (EBIT)
Periodic EV to Trailing 12M EBITDA	Periodic enterprise value as a multiple of earnings before interest, taxes, depreciation and amortization (EBITDA)
Pre-Tax Return on Capital Employed	Trailing 12 month (T12M) pretax earnings and interest expense, divided by the average of capital employed, as a percentage
Price Cash Flow	The price/cash flow ratio is the ratio of a stock's price divided by the cash flow per share
Price Earnings Growth Ratio	A company's Bloomberg Estimated P/E ratio for the specified period divided by the company's Bloomberg Estimated Long Term Growth (LTG) rate
Price Earnings Ratio	Ratio of the price of a stock and the company's earnings per share
Price to Book Ratio	Ratio of the stock price to the book value per share
Price to Free Cash Flow	Valuation metric that compares a company's market price to its level of trailing 12 month free cash flow per share
Price to Sales Ratio	The price to sales ratio is the ratio of a stock's last price divided by sales per share
Quick Ratio	Ratio to indicate the company's ability to pay back its short-term liabilities with its liquid assets
Return on Asset (ROA)	Indicator of how profitable a company is relative to its total assets, in percentage. Return on assets gives an idea as to how efficient management is at using its assets to generate earnings
Return on Capital	Metric that measures the return that an investment generates for capital contributors, in percentage. It indicates how effective a company is turning capital into profits
Return on Equity (ROE)	Measure of a corporation's profitability by revealing how much profit a company generates with the money shareholders have invested, in percentage
Return on Invested Capital (ROIC)	Indicates how effectively a company uses the sources of capital (equity and debt) invested in its operations. Average Invested Capital is the average of the beginning and ending balance of Total Invested