Pontifícia Universidade Católica
do Rio de Janeiro

Roberto Pereira Garcia Junior

Portfolio selection using robust optimization and support vector machine (SVM)

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Engenharia Elétrica of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre in Engenharia Elétrica.

Advisor : Alexandre Street de Aguiar
Co-advisor: Davi Michel Valadão

Rio de Janeiro
November 2019

Pontifícia Universidade Católica do Rio de Janeiro

Roberto Pereira Garcia Junior

Portfolio selection using robust optimization and support vector machine (SVM)

Dissertation presented to the Programa de Pós–graduação em Engenharia Elétrica of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre in Engenharia Elétrica. Approved by the Examination Committee.

Alexandre Street de Aguiar
Advisor
Departamento de Engenharia Elétrica – PUC-Rio

Davi Michel Valadão
Co-advisor
Departamento de Engenharia Industrial – PUC-Rio

Marcus Vinicius Soledade Poggi de Aragão
Departamento de Informática – PUC-Rio

Betina Dodsworth Martins Froment Fernandes
IAG – PUC-Rio

Rio de Janeiro, November the 18th, 2019

Roberto Pereira Garcia Junior

He graduated in Electronic Engineering from the Military Institute of Engineering (IME-RJ). Is Investment Fund Analyst since 2014 at Unifinance Consultoria. Has worked as a consultant at MoningStar Consulting.

Bibliographic data

Garcia Junior, Roberto Pereira

Portfolio selection using robust optimization and support vector machine (SVM) / Roberto Pereira Garcia Junior; advisor: Alexandre Street de Aguiar; co-advisor: Davi Michel Valadão. – Rio de janeiro: PUC-Rio, Departamento de Engenharia Elétrica, 2019.

v., 62 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui bibliografia

1. Electrical engineering – Teses. 2. Teoria da Otimização;. 3. Teoria do aprendizado Estatístico;. 4. Aprendizado de Máquina;. 5. Análise Técnica;. 6. Classificação binária;. 7. Otimização robusta;. I. Street, Alexandre. II. Valadão, Davi Michel. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

To my parents, for their support
and encouragement.

Acknowledgments

Firstly, I would like to especially thank my advisor Dr. Alexandre Street and co-adviser Dr. Davi Valladão Gomes I have no words to express my sincere gratitude to them but I consider myself fortunate to have worked with them for the past two years.

I thank my parents and siblings for their support even many miles away, which has been essential in finding strength and keep fighting for my goals. Words cannot express the immensity of gratitude I have for them. To my family in general for strength and trust.

To all the teachers who took part in this journey, always helpful, even outside the course schedule, because without them there would be no enriching ideas. My sincere thanks.

To my friend Marcos Macedo, for the criticism, suggestions and support that helped to transform some of the ideas in this work into words. My eternal affection and gratitude.

To all co-workers at Unifinance Consultoria, without distinction.

My dear and super beloved girlfriend Daiana Dias, because she knew how to tolerate and understand my strange bad mood at certain times of this research, with wisdom. My thanks with great affection.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Garcia Junior, Roberto Pereira; Street, Alexandre (Advisor); Valadão, Davi Michel (Co-Advisor). **Portfolio selection using robust optimization and support vector machine (SVM)**. Rio de Janeiro, 2019. 62p. Dissertação de mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The difficulty of predicting the movement of financial assets is the subject of study by several authors. In order to obtain gains, it is necessary to estimate the direction (rise or fall) and the magnitude of the return on the asset in which it is intended to be bought or sold. The purpose of this work is to develop a mathematical optimization model with binary variables capable of predicting up and down movements of financial assets and using a portfolio optimization model to evaluate the results obtained. The prediction model will be based on the textit Support Vector Machine (SVM), in which we will make modifications in the regularization of the traditional model. For the portfolio management will be used robust optimization. The robust optimization techniques are being increasingly applied in portfolio management, since they are able to deal with the problems of the uncertainties introduced in the estimation of the parameters. It is noteworthy that the developed model is data-driven, i.e., the predictions are made using nonlinear signals based on past historical price / return data without any human intervention.

As prices depend on many factors it is to be expected that a set of parameters can only describe the dynamics of the prices of financial assets for a small interval of days. In order to more accurately capture this change in dynamics, the estimation of model parameters is done in a moving window To test the accuracy of the models and the gains obtained, a case study was made using 6 financial assets of the currencies, fixed income, variable income and commodities classes. The data cover the period from 01/01/2004 until 05/30/2018 totaling a total of 3623 daily quotations. Considering the transaction costs and out-of-sample results obtained in the analyzed period, it can be seen that the investment portfolio developed in this work shows higher results than the traditional indexes with limited risk.

Keywords

Theory of Optimization; Theory of Statistical Learning; Machine Learning; Technical analysis;; Binary classification; Robust optimization;

Resumo

Garcia Junior, Roberto Pereira; Street, Alexandre; Valadão, Davi Michel. **Seleção de portfólio usando otimização robusta e máquinas de suporte vetorial**. Rio de Janeiro, 2019. 62p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A dificuldade de se prever movimento de ativos financeiros é objeto de estudo de diversos autores. A fim de se obter ganhos, se faz necessário estimar a direção (subida ou descida) e a magnitude do retorno do ativo no qual pretende-se comprar ou vender. A proposta desse trabalho consiste em desenvolver um modelo de otimização matemática com variáveis binárias capaz de prever movimentos de subidas e descidas de ativos financeiros e utilizar um modelo de otimização de portfólio para avaliar os resultados obtidos. O modelo de previsão será baseado no *Support Vector Machine* (SVM), no qual faremos modificações na regularização do modelo tradicional. Para o gerenciamento de portfólio será utilizada otimização robusta. As técnicas de otimização estão sendo cada vez mais aplicadas no gerenciamento de portfólio, pois são capazes de lidar com os problemas das incertezas introduzidas na estimativa dos parâmetros. Vale ressaltar que o modelo desenvolvido é *data-driven*, i.e, as previsões são feitas utilizando sinais não-lineares baseados em dados de retorno/preço histórico passado sem ter nenhum tipo de intervenção humana.

Como os preços dependem de muitos fatores é de se esperar que um conjunto de parâmetros só consiga descrever a dinâmica dos preços dos ativos financeiros por um pequeno intervalo de dias. Para capturar de forma mais precisa essa mudança na dinâmica, a estimação dos parâmetros dos modelos é feita em janela móvel.

Para testar a acurácia dos modelos e os ganhos obtidos foi feito um estudo de caso utilizando 6 ativos financeiros das classes de moedas, renda fixa, renda variável e commodities. Os dados abrangem o período de 01/01/2004 até 30/05/2018 totalizando um total de 3623 cotações diárias. Considerando os custos de transações e os resultados *out-of-sample* obtidos no período analisado percebe-se que a carteira de investimentos desenvolvida neste trabalho exibe resultados superiores aos dos índices tradicionais com risco limitado.

Palavras-chave

Teoria da Otimização; Teoria do aprendizado Estatístico; Aprendizado de Máquina; Análise Técnica; Classificação binária; Otimização robusta;

Table of contents

List of figures

List of tables

List of Abreviations

TAP – Statistical Learning Theory
SVM – Support Vector Machine
AM – Machine Learning
IA – Artificial intelligence
BH – Buy and Hold
MDD – Max Drawdown
MILP – Mixed Integer Linear programming

1 Introduction

The Efficient Markets Hypothesis (EMH) is one of the most influential theories in the field of finance, being the basis for a number of asset pricing models (2). The basis of the EMH is the assertion that the price of an asset reflects the information available on the issuing institution, making it impossible for investors to gain abnormal returns. Despite the wide acceptance of EMH, many researchers did not give up examining the predictability of returns.

Initially, the most convenient assumption for financial theory and empirical methods is that distribution of security rates of return be multivariate normal with parameters that are stationary over time. However, Fama in (17) tested the normality hypothesis on daily returns for the Dow Jones Industrial stocks and found an excess kurtosis than that predicted from a sample of independent and identically distributed normal variables.

After these studies, many other authors attempted to describe the probability distribution of returns (7, 33, 30). There is no consensus on the actual distribution of probabilities of returns and due to the complexity of this task most of the researches focus on approach more data-driven. That is, currently the forecast models are no longer focused on the generating process, but on the recognition of hidden patterns.

Moreover, different from the traditional models that focused exclusively on predicting the conditional mean, the prediction of the return signal has been increasingly discussed. Diebold and Christoffersen in (13), for example, argue that although they do not find predictability in the conditional average of returns there is a dependency structure in the return signal. In (36) a new type of trend strategy is developed based on the past return signal. This strategy is driven mainly by signal dependence, which is positively related to the average return and negatively related to volatility.

Forecasting returns is a complex task, and it has not yet been possible to generate models with a high hit rate[1]. The relationship between past data and future data might be non-linear (24) and therefore traditional linear statistical

[1]Hit rate was defined as the ratio between the number of correct forecasts divided by the total number of forecasts.

models such as auto-regressive, integrated and moving average have not proved satisfactory for this task.

In view of this new paradigm, models such as Support Vector Machine (SVM), Neural Networks and Decision Trees are increasingly being used for time series forecasting. These models are totally data oriented and do not make any assumptions about the data generating process, that is, it adapts to the set of data received over time. A broad review of the nonlinear models applied to the prediction of actions can be found in (3).

Economic factors (market indices such as GDP and current account deficit), Institutional factors (information of companies as price on profit and distribution of dividends) were not addressed in this research. In this work a mathematical optimization model was developed for binary classification based on the SVM where the inputs are the historical prices of the assets.

The choice of SVM is based on some attractive features such as good generalization capability, robustness in large dimensions, convexity of objective function and well definied theory.

Among the characteristics cited, the most prominent is the generalization capacity. The first results were presented by Vapnik and Chernovenkis through the Theory of statistical learning, proposed by these authors in the 60s and 70s (Vapnik, 1995). The SVMs emerged as a direct result of using the principles presented in this study.

It is worth mentioning that the good generalization capacity of the SVM in the context of financial series is document in several studies. In (1), for example, an SVM adapted to show the flexibility of the support machines in the middle of chaotic series was used. The model considered the chaos intrinsic to the time series to make predictions about the USD / BRL exchange rates, reaching very satisfactory results and achieving a better performance than traditional models.

Also (37) analyzes the behavior of models generated with SVM in the American market focusing on two technical indicators: RSI[2] and MACD[3]. In addition to obtaining a good accuracy rate in the market index forecast, it concludes that SVM performs better in markets of high and low ballast than in markets of tediousness.

[2]The relative strength index (RSI) is a momentum indicator that measures the magnitude of recent price changes to evaluate overbought or oversold conditions in the price of a stock or other asset. The RSI is displayed as an oscillator (a line graph that moves between two extremes) and can have a reading from 0 to 100.

[3]Moving Average Convergence Divergence (MACD) is a trend-following momentum indicator that shows the relationship between two moving averages of a security's price. The MACD is calculated by subtracting the 26-period Exponential Moving Average (EMA) from the 12-period EMA.

However, to make a profit on the financial markets, not only is a good forecasting model needed, since most investors have a portfolio of investments. We need an allocation model that is capable of using the best forecasts possible, maximizing return and reducing losses.

The pioneering work in the area of portfolio optimization was the proposition of the mean-variance model by Harry Markowitz [4]. Which assumes that for the investor must maximize the expected return and minimize its volatility of returns to define define his optimal portfolio. In spite of the theoretical success of the mean-variance model, practitioners have shied away from this model. The solution of optimization problems is often very sensitive to perturbations in the parameters of the problem. Since the estimates of the market parameters are subject to statistical errors, they are very sensitive to the perturbations in the inputs. The results of the optimization problems may not be very reliable.

In this context, robust optimization (RO) techniques have received significant interest by the investment management community, as they allow portfolio managers to incorporate the uncertainty introduced by estimation errors directly into the optimization process. Its goal is to compute solutions with a prior ensured feasibility when the problem parameters are assumed to be unknown but confined within a prescribed uncertainty set. In (22) we have statistical methods for constructing uncertainty sets for factor models of asset returns. In that same line in (38) is proposed a model with box uncertainty sets for mean and covariance and show the arising model can be reduced to a smooth saddle-point problem subject to semi definite constraints.

Given optimization problems with uncertain parameters, RO finds the best decision in view of the worst-case parameter.

1.1
Literature review

The study of asset prices in financial time series using machine learning algorithms began in the early 2000s in the work of Fan and Palaniswami (18). The accounting information of shares traded on the Australian Stock Exchange for the period of 1992 and 2000 was used as input to the SVM. The results were compared with a benchmark model consisting of a uniformly weighted investment portfolio composed of all stocks available for ranking. It is noteworthy that only annual reports were considered and reports with more than one missing variable were discarded.

[4] For more details on Modern Portfolio Theory see (16, 32).

In 2001, (11) presented a study on the parameters of the SVM formulation in comparison to artificial neural networks - RNA. Although it presented unsatisfactory results in the classification of a new sample from the point of view of an investor, was superior to the use of RNA. In this work, in addition to highlighting the use of SVM in financial time series, they are shown that the use of external variables the financial time series, such as technical, fundamentalist or inter-market indicators, in most cases, does not improve the prediction power of the model when compared to the use of past returns.

In the same way, (28) follows the same line of reasoning in a similar work, ratifying cite 40 about the sensitivity of the C and δ^2 parameters of the SVM implementation, but focusing on distinct characteristics for the problem resolution. The tests are focused on the implementation of the kernel polynomial and RBF compared to a model generated by neural networks.

In HUANG (2005) (26) the SVM is used to predict the weekly movement of the NIKKEI 225 index, comparing its performance with linear discriminant analysis and quadratic discriminant analysis. The input data of the models are interest rate, short-term interest rate, long-term interest rate, industrial production, government consumption, private consumption, gross national product, gross domestic product. To evaluate the performance they used a random walk model as benchmark. The authors obtained results superior to linear discriminant and quadratic discriminant analysis, suggesting that this path can lead to capital gains.

In PHICHHANG (2009) (4) a group of algorithms are used to predict the daily trend of Hong Kong stock exchange assets. The results obtained were promising and the attributes used were the S&P 500 exchange rate between HK and US, plus the addition of the opening value, closing value, minimum value and maximum value in the day.

In ZHONGYUAN (2012) (39) the SVM is used to make forecasts in the Chinese stock market. As in similar works, it faces the problem of obtaining the optimal parameters for the algorithm, choosing them based on results obtained in cross-validation experiments.

In HUANG (2012) (25) proposed hybrid methodology using both GA and SVM for stock prediction. Genetic Algorithm (GA) is mainly used for parameter optimization of the model and to peprform feature selection to achieve optimal parameter as an input to the SVM model. The use of GA for feature selection is vital and helps to significantly outperform the benchmark schemes

In NAYAK (2015) (35) proposed a hybrid framework utilizing SVM with KNN. The proposed methodology was used to predict the Indian stock exchange market. SVM was utilized to predict future or loss. It also estimated the stock value over a time for one day, week and month. The model performed well for high dimensional feature vector and handled the error and the performance of the classification methods. The SVM-KNN model outperformed the mentioned models by removing the need to tune multiple parameters for ANN and fuzzy-based model.

Literature that connects the ideas of risk measurement or portfolio optimization with robust optimization is rather limited since this is a relatively new field. However, some works explore this field. Below we present some pioneering works in this area of robust optimization applied to portfolio selection.

In GOLDFARB and IYENGAR (2003) (22) analyze the problem of portfolio optimization, concentrating on the mean return, variance, and the Sharpe ratio. They develop a factor model with normality assumptions on the return distribution that reduce the size of the problem. Uncertainty is assumed to be present in the mean returns, residual variance of assets, and their factor exposure. It is shown that optimization of some portfolio characteristics, while keeping the others constrained, can be reduced to a second-order cone programming problem. Also, procedures for the construction of uncertainty sets are proposed.

In CERIA and STUBBS (2006) (12) analyze the standard Markowitz mean-variance optimization problem with uncertainty in the mean returns of assets. They construct ellipsoidal uncertainty region for them and prove that the robust optimal solution is equivalent to a stardard solution for an agent with higher risk aversion.

In BERTSIMAS and PACHAMANOVA (2008) (6) study the case of optimal multiperiod investment under uncertainty and transaction costs. They construct optimal investment strategies as a result of extensive linear programs. In the end they conduct a simulation study where they compare standard mean-variance multiperiod asset allocation to several variants of their robust decisions.

In GOH and SIM (2010) (21) analyze the problem of expected utility maximization under distributional uncertainty (which is defined in terms of supports and moments of the random variables). Since the problems they consider are highly nonlinear, they develop several approximative schemes based on linear and quadratic decision rules. Results are demonstrated on an example with multiperiod inventory management.

In WOZABAL (2012) (40) considers the problem of deriving the worst-case form of several risk measures under non parametric distributional uncertainty, defined by means of the Kantorovich distance. He combines a so-called subdifferential representation of common risk measures with a suitable form of the Kantorovich distance to derive tractable worst-case values of given risk measures.

1.2 Purpose of dissertation

The focus of this work is to create an optimal investment portfolio using forecasts based on an SVM to feed the robust optimization model. The forecast model, unlike the traditional SVM, will use binary variables to control the number of signals used in predictions. This form of regularization is still unusual in the financial market, mainly due to its high computational cost. However, we believe that the benefits obtained are worth the price paid for using the integer variables. In addition to the best results in terms of accuracy, it is easier to understand which are the variables that most influence predictions.

The model will be data-driven. For the distribution of the resources we use a robust optimization model based on the allocation model developed in (20) with some adaptations. Figure 1.1 below shows all steps of the process. Note that the model training was done using the Julia programming language (JuMP package (14)) that was chosen for having superior performance in the main programming languages such as Python, Matlab and R.

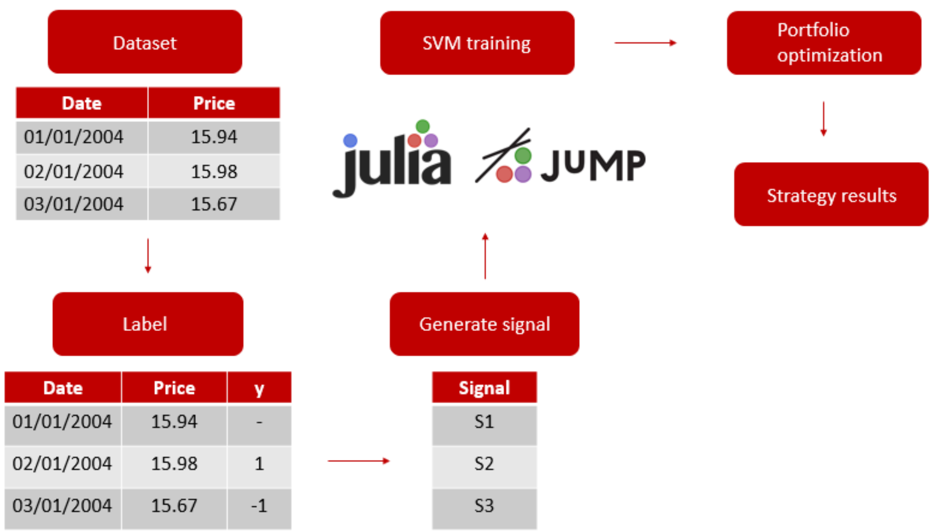

Figure 1.1: Schematic proposed in this work

The use of such a model has become increasingly popular since they do not make assumptions about the process generating the financial series.

In addition, they can be used in the most diverse markets: currencies, fixed income, variable income and commodities.

To obtain superior results is necessary that the signals used as inputs have good predictive power, therefore a broad review of the quantitative finance literature was necessary to understand which signals are the most used and with best performances. In addition to classical signals as a moving average, signals were still used in the Brazilian environment and have a good predictive power. In this work, we did not focus on any particular signal, since as the return distributions change over time the idea was to have a signal mix that adapts to the most varied market conditions.

1.3 Theoretical justification and contributions

To make a profit in the financial market it is very important to have subsidies for decision making. Such subsidies can be translated into information from financial analysts or even generated by mathematical models capable of anticipating the movement of asset prices. The risk involved in decision making is large and can cause considerable financial losses, so it is important to study and develop robust and reliable models that can serve as the basis for investors to make their decisions.

Over the last few decades, the amount of data generated has grown exponentially and it is becoming increasingly important for investment firms to process and analyze data faster and faster. In the face of this new scenario, systematic investments have gained increasing prominence in Brazil, since this type of strategy benefits from this large amount of available data and technological advances. The list of applications is varied and ranges from risk management, option pricing, portfolio optimization and robot advisors.

In addition, the theoretical relevance of this dissertation is related to the generation of scientific knowledge in optimization models with integer variables, quantitative finance and statistical learning theory applied to the Brazilian financial market.

Finally, our main contributions of work to literature are:

1. We propose a prediction model based on the SVM that uses integer variables to control the number of signals used to make the predictions on future returns.

2. We propose a methodology with a portfolio of signals, which in the literature are generally used and studied separately, in a combined manner.

3. We Consider that optimal portfolio losses are modeled using a robust adaptive approach. Also, unlike traditional portfolio optimization models that consider maximizing the expected return, we maximize a metric that is related to the chance our forecast is correct. The amount allocated in the assets will be distributed according to the confidence we have in the forecasts.

4. The model incorporates transactions costs, covering all fees structures typically observed in the market, to provide a more rigorous result for practical purposes.

1.4 Structure of dissertation

This dissertation was structured in 7 chapters, and the following is a brief description of each one.

In this chapter, general considerations were presented regarding the Brazilian financial market, the models that will be developed and the theoretical relevance of the topic addressed in this dissertation.

In Chapter 2 we present the two most popular types of support vector machines in the literature: Hard margin SVM and Soft Margin SVM. We will see that the former is used only when the data are linearly separable and therefore, it was necessary to extend the concept to the non-linearly separable case and hence the soft version appears.

Chapter 3 presents a quick review on robust optimization models. The objective is to present the model that will be used as the basis for this work.

Chapter 4 presents the signals that were used as predictors. The idea is to use signals with high predictive power and that are adaptive to capture the various dynamics of the market.

Chapter 5 presents the proposed model for the creation of investment portfolios. It shows the adaptations made in the traditional SVM standard. In addition, we present the robust portfolio optimization module that takes into account the forecasts made by SVM.

Chapter 6 presents the numerical results obtained with the application of the approaches proposed in chapter 5.

Finally, chapter 7 presents the conclusions obtained in this work, and suggestions for future improvements.

2
Support vector machine

In this section it is presented the use of SVMs to obtain linear boundaries to separate data belonging to two classes. The first case to be presented deals with linearly separable data. Since in real problem this situation hardly ever occurs, the formulation has been extended to the case where data are not separable. However, before entering SVM itself, let's present how the data set is organized and some considerations about the set of classifiers used throughout this work.

We assume training data are given consisting of observations $S_t^i = (s_{t,1}^i, s_{t,2}^i, \cdots, s_{t,N_s}^i)^T$ each having associated class label $y_t^i \in \{-1, 1\}$ where $i \in \mathcal{I} = \{1, 2, 3, \cdots, n\}$ refers to certain financial asset, $t \in \mathcal{T} = \{t_0, t_1, ..., t_m\}$ refers to a particular day in dataset. Already N_s refers to the total number of attributes used in this work. In addition, the set $\mathcal{K} = \{1, 2, \cdots, N_s\}$ is used to refer to all possible attributes. For example, $s_{t,k}^i$ refers to the attribute $k \in \mathcal{K}$ of asset $i \in \mathcal{I}$ in time $t \in \mathcal{T}$. The figure 2.1 below shows the matrix form of the dataset for a given asset.

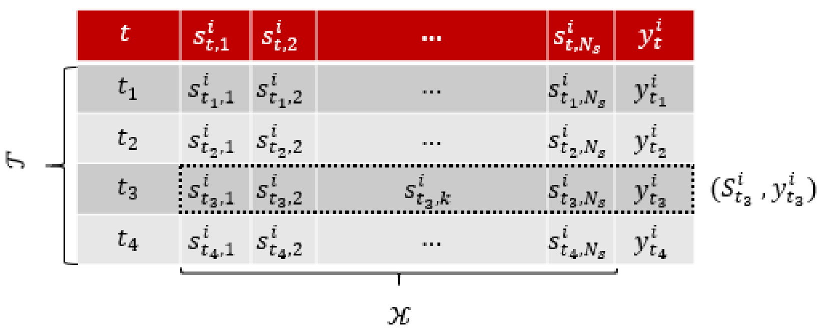

Figure 2.1: Matrix representation of the data set for a given asset. Matrix lines are time indexed since the data is time series.

For the remainder of the chapter it is not necessary to distinguish the asset being considered since the same model will be used for all and so let's consider our dataset as $S = \{(S_t, y_t) \ \forall t \in \mathcal{T}\}$.

It is typically assumed that there is a probability distribution $P(S_t, y_t)$ from which theses data are drawn. The set of classifiers shall be restricted to those of the form:

$$y_t = h_{\boldsymbol{w},b}(S_t) = I(f(S_t)) \tag{2-1}$$

$$f(S_t) = \langle S_t, \boldsymbol{w} \rangle + b \tag{2-2}$$

Where, I(z) = 1, if $z \geq 0$ and $I(z) = -1$ otherwise, $\boldsymbol{w} \in \mathbb{R}^{N_s}$ and $b \in \mathbb{R}$. Classifiers that separate the data through hyper-plane are called linear. Note that according to the $I(z)$ definition our classifier will return +1 or -1 without estimating probabilities (other than how logistic regression works, for example).

2.1 Functional and geometric margins

Given an example of the training set S of the form (S_t, y_t) the functional margin of $(\boldsymbol{w}, b)$ with respect to this training example is defined as:

$$\rho_t = y_t f(S_t) = y_t(\langle S_t, \boldsymbol{w} \rangle + b).$$

Note that if $\rho(f(S_t), y_t) > 0$ we have a correct sort because the signs of y_t and $f(S_t)$ are the same. For a linear classifier as shown in equations 2-1 we have that the functional margin is not a good measure for the confidence of our forecast, since it is sensitive to the rescaling of the parameters.For example, if we make the following transformation $(\boldsymbol{w}, b) \rightarrow (2\boldsymbol{w}, 2b)$ we have that the margin value is twice as large, but $h_{2\boldsymbol{w},2b}(S_t) = h_{\boldsymbol{w},b}(S_t)$.

To get around this problem we will introduce the concept of geometric margin. Consider the picture 2.2 below:

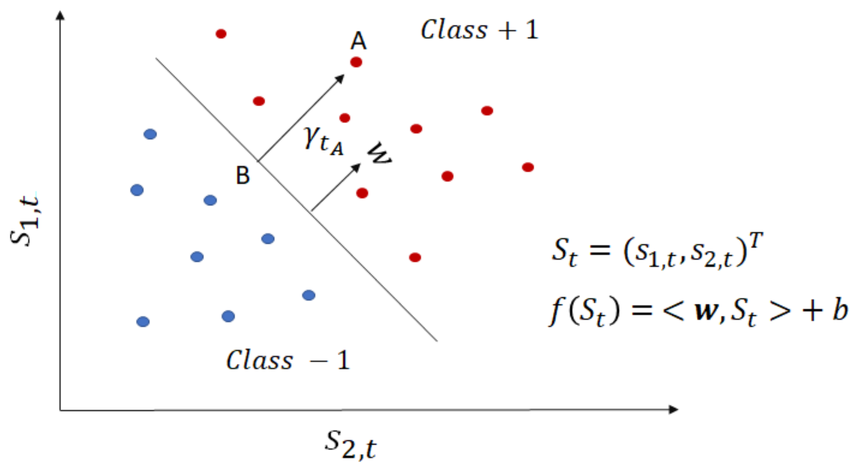

Figure 2.2: Example of a binary classification problem using a linear classifier.

We want to find the distance between $\overrightarrow{AB}$ defined by γ_t. Let S_t be the coordinate of point A, then the coordinate of point B is given by $S^B = S_t - \gamma_t \frac{\boldsymbol{w}}{||\boldsymbol{w}||}$. How $S^B \in \langle S_t, \boldsymbol{w} \rangle + b = 0$ we have,

$$\langle \boldsymbol{w}, S_t - \gamma_t \frac{\boldsymbol{w}}{||\boldsymbol{w}||} \rangle + b = 0 \tag{2-3}$$

Solving for γ_t,

$$\gamma_t = \frac{\langle S_t, \boldsymbol{w} \rangle + b}{||\boldsymbol{w}||} \tag{2-4}$$

This was worked out for the case of a positive training example A in the figure. More generally, we define the geometric margin of $(\boldsymbol{w}, b)$ with respect to a training example (S_t, y_t) to be:

$$\gamma_t = \frac{y_t(\langle S_t, \boldsymbol{w} \rangle + b)}{||\boldsymbol{w}||} \quad \forall t \in \mathcal{T} \tag{2-5}$$

Note that if $||\boldsymbol{w}|| = 1$ then $\rho_t = \gamma_t$. In addition we see that the geometric margin is not affected by the rescaling of (w, b). Given these properties we see that it makes sense to relate the geometric margin to the confidence of our forecast. For example, a high γ_t value indicates a point far from the decision boudary and as a consequence the chance of having made a wrong prediction is lower.

However, when we use the model to make predictions on unseen data, we do not have the label y_t and therefore we can not calculate γ_t directly. As a measure of the confidence of new forecasts we will use only the value given by:

$$d_t = \frac{(\langle S_t, \boldsymbol{w} \rangle + b)}{||\boldsymbol{w}||} \tag{2-6}$$

The magnitude of d_t is associated with confidence in the prediction and $I(d_t)$ the predicted class.

Finally, given a training set $S = \{(S_t, y_t), \forall t \in \mathcal{T}\}$, we also define the geometric margin of $(\boldsymbol{w}, b)$ with respect to S to be smallest of the geometric margins on the individual training examples:

$$\gamma = \min_{t \in \mathcal{T}} \gamma_t$$

2.2 Maximum margin classifiers

Given a set of training $S = \{(S_t, y_t), \forall t \in \mathcal{T}\}$ and what was presented in the previous section it is natural to try to find a classifier that presents maximum margin, since this would imply in forecasts with a high degree of confidence. This will result in a predictor separating the positive and negative examples with a large gap (geometric margin).

As mentioned earlier, we are assuming in this first moment that the data are linearly separable, i.e., it is possible to separate the positive and negative examples using a hyperplane of the form 2-2. The following optimization problem provides the hyperplane with maximum geometric margin:

$$\underset{\boldsymbol{w},b,\gamma}{\text{Maximize}} \quad \gamma \tag{2-7}$$
$$s.t.$$
$$||\boldsymbol{w}|| = 1 \tag{2-8}$$
$$y_t(\langle S_t, \boldsymbol{w}\rangle + b) \geq \gamma, \qquad \forall t \in \mathcal{T} \tag{2-9}$$

Note that the constraint $||\boldsymbol{w}|| = 1$ ensures that the margin coincides with the geometric margin and with this we have that all examples are at a distance of at least γ. At optimum, the result will be $(\boldsymbol{w}^*, b^*)$ which guarantees the largest possible geometric margin.

Although the model proposed in the equation 2-7 gives us the linear classifier with the maximum margin, a constraint $||\boldsymbol{w}|| = 1$ makes it non-convex. You can make some modifications to the model in order to make it convex. Consider the model,

$$\underset{\boldsymbol{w},b,\rho}{\text{Maximize}} \quad \frac{\rho}{||\boldsymbol{w}||} \tag{2-10}$$
$$s.t.$$
$$y_t(\langle S_t, \boldsymbol{w}\rangle + b) \geq \rho, \qquad \forall t \in \mathcal{T} \tag{2-11}$$

In the equation 2-10 we have the maximization of the geometric margin in the objective function and a constraint that guarantees that the marginal error is zero. We have succeeded in eliminating the constraint $||\boldsymbol{w}|| = 1$, but the problem is still non-convex due to $\frac{\rho}{||\boldsymbol{w}||}$. Let us argue that there is no loss of generality in assuming that $\rho = 1$. As previously discussed, the parameter rescheduling has no influence on the problem, so we can solve the problem with the $(\rho\boldsymbol{w}, \rho b)$ parameters. The problem becomes,

$$\underset{\boldsymbol{w},b}{\text{Maximize}} \quad \frac{1}{||\boldsymbol{w}||} \tag{2-12}$$
$$s.t.$$
$$y_t(\langle S_t, \boldsymbol{w}\rangle + b) \geq 1, \qquad \forall t \in \mathcal{T} \tag{2-13}$$

The problem 2-12 is mathematically equivalent to:

$$\underset{\boldsymbol{w},b}{\text{Minimize}} \quad ||\boldsymbol{w}||^2 \tag{2-14}$$

$$s.t.$$

$$y_t(\langle S_t, \boldsymbol{w}\rangle + b) \geq 1, \qquad \forall t \in \mathcal{T} \tag{2-15}$$

The model of the equation 2-14 is convex and is known in the literature as Support Vector Machine with rigid margin. The vectors that are closest to the optimal hyperplane (smaller geometric margin) are the support vectors as shown in the following figure 2.3.

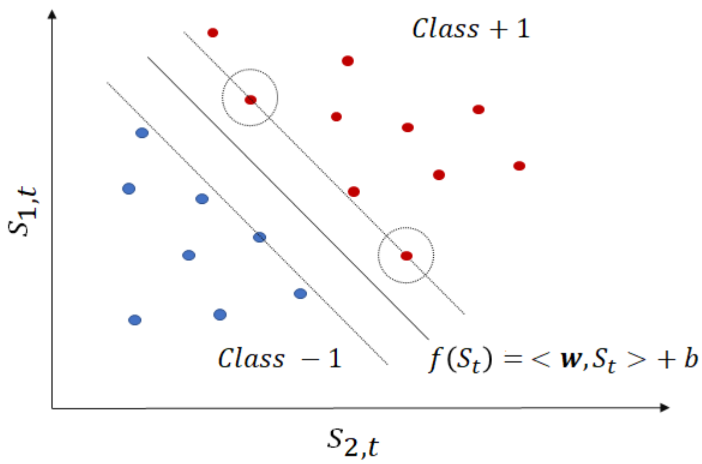

Figure 2.3: Optimum hyperplane and support vectors.

2.3 Soft margin SVM

The derivation of the SVM from the previous section assumed that the data are linearly separable. However, in most real problems this is not the case. This is due to several factors, among them the presence of noises and *outliers* in the data or to the very nature of the problem that may be nonlinear.

To accomplish this classification task, therefore it is allowed that some data may violate the constraint of the equation 2-14. This is done by introducing clearance variables z_t, for $t \in \{t_0, t_1, \cdots, t_m\}$. In this way, the optimization problem becomes:

$$\underset{\boldsymbol{w},b,\boldsymbol{z}}{\text{Minimize}} \quad \sum_{t=t_0}^{t_m} C_t z_t + \frac{1}{2}||\boldsymbol{w}||^2 \tag{2-16}$$

$$s.t.$$

$$y_t(\langle S_t, \boldsymbol{w}\rangle + b) \geq 1 - z_t, \qquad \forall t \in \mathcal{T} \tag{2-17}$$

$$z_t \geq 0, \qquad \forall t \in \mathcal{T} \tag{2-18}$$

The application of this procedure smoothes the edges of the linear classifier, allowing some data to remain between the dotted hyperplanes shown in figure 2.4 and also the occurrence of some classification errors. Note that the term $\sum_{t=t_0}^{t_m} z_t$ is nothing more than the marginal error and $||\boldsymbol{w}||^2$ is related to complexity of the model. The vector $\boldsymbol{C} = [C_{t_0}, C_{t_1}, ..., C_{t_m}]$ is a regularization term that imposes a weight on minimizing the complexity of the model, ie, it is the trade-off between complexity of the model and minimization of errors.

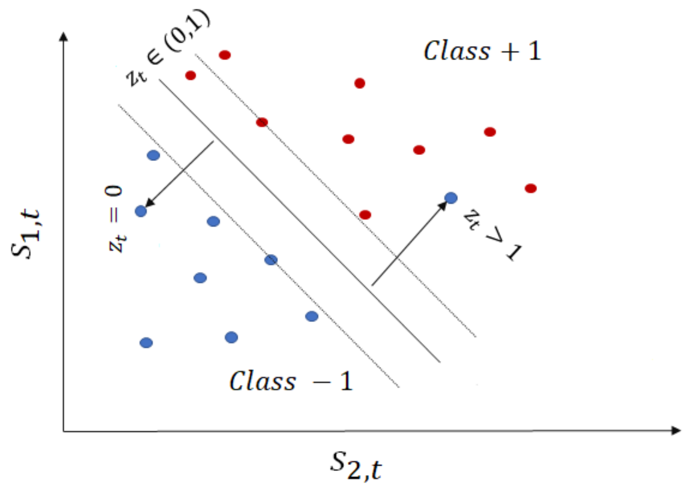

Figure 2.4: Binary classification problem where non-linearly separable data. The values of z_t serve as a counter for the number of errors, since $z_t > 1$ indicates erroneous predictions. Already values between 0 and 1 indicate that the forecast is correct, however, the point is close to the decision border.

3
Robust optimization and portfolio selection

Throughout this section we will present the basics concepts in robust optimization and the model developed in (20).

Given an objective function $f(\boldsymbol{x})$ to optimize subject to constraints $L(\boldsymbol{x}, \boldsymbol{R}_t) \leq \epsilon_1$ with uncertain parameters, $\{\boldsymbol{R}_t\}$, the general RO formulation is:

$$\underset{\boldsymbol{x}_t}{\text{Maximize}} \quad f(\boldsymbol{x}_t) \tag{3-1}$$

$$s.t.$$

$$L(\boldsymbol{x}_t, \boldsymbol{R}_t) \leq \epsilon_1, \qquad \forall \boldsymbol{R}_t \in \Xi_t \tag{3-2}$$

Here $\boldsymbol{x}_t \in \mathbb{R}^n$ is a vector of decision variables, $f, L \colon \mathbb{R}^n \to \mathbb{R}$ are functions, and the uncertainty parameters $\boldsymbol{R}_t \in \mathbb{R}^n$ are assumed to take arbitrary values in the uncertainty sets $\Xi_t \subseteq \mathbb{R}^n$. The goal of 3-1 is to compute minimum cost solution $\boldsymbol{x}^*$ among all those solutions which are feasible realizations of the disturbances $\boldsymbol{R}_t$ within Ξ_t.

It is worth mentioning that in the context of this paper $\boldsymbol{R}_t^T = (R_t^1, R_t^2, \cdots, R_t^n)$ is the vector of asset returns and $\boldsymbol{x}_t^T = (x_t^1, x_t^2, \cdots, x_t^n)$ is vector of allocation and $f(\boldsymbol{x}_t) = \hat{\boldsymbol{R}}_t^T \boldsymbol{x}_t$ with $\hat{\boldsymbol{R}}_t$ being the prediction of the return given a set of past information. Following what was done at (20) the loss restriction will be of the form $L(\boldsymbol{x}, \boldsymbol{R}_t) \leq \epsilon_1$ is described generally as:

$$\boldsymbol{R}_t^T \boldsymbol{x}_t \geq \gamma W_{t-1}, \quad \forall \boldsymbol{R}_t \in \Xi_t \tag{3-3}$$

Here W_{t-1} is the total wealth and γ the parameter that defines a percentage loss of the total wealth at time period t-1.

At this point we need to determine the structure for Ξ_t. There are several possible frameworks for the uncertainty set, however in finance the most commonly used are those listed below. (see (5) for more details)

$$\Xi_t(\sigma) = \{\tilde{\boldsymbol{R}}_t \in \mathbb{R}^n : (\tilde{\boldsymbol{R}}_t - \boldsymbol{\mu}_t)\Sigma^{-1}(\tilde{\boldsymbol{R}}_t - \boldsymbol{\mu}_t) \leq \sigma^2\} \tag{3-4}$$

$$\Xi_t(\Gamma) = \{\tilde{\boldsymbol{R}}_t \in \mathbb{R}^n : \exists \boldsymbol{q} \in \mathbb{R}^n_+ \quad s.t. \quad \tilde{\boldsymbol{R}}_t = \boldsymbol{\mu}_t + (\boldsymbol{r}_{low} - \boldsymbol{\mu}_t) \circ \boldsymbol{q}, \quad q_i \leq 1, \quad \sum_{i=1}^{n} q_i \leq \Gamma\} \tag{3-5}$$

$$\Xi_t(\alpha) = \{\tilde{\boldsymbol{R}}_t \in \mathbb{R}^n : \exists \boldsymbol{q} \in \mathbb{R}^n_+ \quad s.t. \quad \tilde{\boldsymbol{R}}_t = \sum_{j=1}^{J} q_j \boldsymbol{r}_{t-j}, \quad q_j \leq \frac{1}{N(1-\alpha)}, \quad \sum_{j=1}^{J} q_j = 1\} \tag{3-6}$$

The set in 3-4 is a quadratic or ellipsoidal uncertainty set, i.e, considers all returns within a radius σ from the mean return vector, where the ellipsoid is tilted by the covariance. When $\sigma = 0$ then $\tilde{\boldsymbol{R}}_t = \boldsymbol{\mu}_t$. The set 3-5 considers all returns such that each component of the return is in the interval $[r^i_{low}, \mu^i_t]$, with the restriction that the total weight of deviation from μ^i_t, summed across all assets, may be no more than Γ. Finally, 3-6 is the tail uncertainty set and considers the convex hull of all possible $N(1-\alpha)$ point average of the N returns.

However, in this work we will use the uncertainty set $\Xi_t(\gamma)$ defined below in equation 3-7, since it was applied to a similar problem to this work and the results were promising as presented in (20).

$$\Xi_t(\gamma) = \{\tilde{\boldsymbol{R}}_t \in \mathbb{R}^n : \exists \boldsymbol{\zeta} \in Z \quad s.t \quad \tilde{\boldsymbol{R}}_t = \sum_{j=1}^{J} \zeta_j \overline{\boldsymbol{R}}_{t-j}\ \} \tag{3-7}$$

where $\overline{\boldsymbol{R}}_{t-j}$ are return sample and ζ_j is defined in the set

$$Z = \{\boldsymbol{\zeta} \in [0,1]^J : \sum_{j=1}^{J} \zeta_j = 1\ \} \tag{3-8}$$

Therefore, the problem 3-1 can be written as follows

$$\underset{\boldsymbol{x}_t}{\text{Maximize}} \quad \hat{\boldsymbol{R}}_t^T \boldsymbol{x}_t \tag{3-9}$$

$$s.t.$$

$$\underset{\boldsymbol{\zeta} \in Z}{\text{Minimize}} \quad \sum_{j=1}^{J} \zeta_j (\overline{\boldsymbol{R}}_j^T \boldsymbol{x}_t) \geq \gamma W_{t-1} \tag{3-10}$$

It can be shown (see (20) for a complete and accurate demonstration) that problem 3-9 is equivalent to

$$\underset{\boldsymbol{x}_t, x_t^c, \boldsymbol{u}^+, \boldsymbol{u}^-}{\text{Maximize}} \quad \sum_{i=1}^{n} \hat{R}_t^i x_t^i + \hat{R}_{t,c} x_t^c \tag{3-11}$$

$$s.t.$$

$$x_t^c = W_{t-1} - \sum_{i=1}^{n} c_i (u_+^i + u_-^i) \tag{3-12}$$

$$\sum_{i=1}^{n} R_{t-j} x_t^i \geq \gamma W_{t-1}, \qquad \forall j = 1, 2..., J \tag{3-13}$$

$$x_t^i = x_{t-1}^i (1 + R_{t-1}^i) + (u_+^i - u_-^i), \qquad \forall i \in \mathcal{I} \tag{3-14}$$

$$u_+^i, u_-^i \in \mathbb{R}_+, \qquad \forall i \in \mathcal{I} \tag{3-15}$$

$$x_t^c \in \mathbb{R}_+ \tag{3-16}$$

Here $\hat{R}_t^i$ and $\hat{R}_{t,c}$ are the forecasts for financial assets and for the free risk asset, respectively. The values of $\boldsymbol{x}_t$ are the values allocated to each asset. In addition, the first restriction imposes the maximum loss (through the γ variable) that the portfolio may suffer each day for a period of J days. The other restrictions are related to the balance of inflows and outflows of the assets of the portfolio.

It is worth mentioning that there is no capital inflow during the analyzed period and $\boldsymbol{c}$ is the transaction cost vector of assets to buy and sell. For simplicity we are assuming that the cost of the transaction is independent of the volume operated.

4
Attributes and predictors

In this section we will present the attributes used to predict the trend of financial time series. One of the most used attributes as a trend predictor is the set of past returns $\boldsymbol{R_{t,J}} = (r_{t-1}, \cdots, r_{t-J})$, where J represents the number of past returns that will be used and t the day on which we will make the forecast. However, there are several critiques for this choice, since several studies point to the nonexistence of a linear correlation between the past and the future of financial series.

Thus, in addition to using the past returns we will use a set of $\boldsymbol{S^i_{t,N}} = (s^{i,N}_{t,1}, \cdots, s^{i,N}_{t,\frac{N_s}{2}})$[1], where each $s^{i,N}_{t,k}$, k[2] $\in \frac{\mathcal{K}}{2}$, $i \in \mathcal{I}$ and $t \in \mathcal{T}$ is a nonlinear function of the past data, i.e., $s^{i,N}_{t,k} = f_k(\boldsymbol{R}^i_{t,N})$. The models presented in this paper have the structure presented next.

1. **Model 1**: Model where we do not directly use past returns as predictors. In addition, the set of signals is divided into two: short-term and long-term signals.

$$\boldsymbol{S}^i_t = [\boldsymbol{S^i_{t,N_{short}}} \quad | \quad \boldsymbol{S^i_{t,N_{long}}}]$$

$$\boldsymbol{w} = [\boldsymbol{w_{short}} \quad | \quad \boldsymbol{w_{long}}]$$

   Where N_{long} and N_{short} are natural numbers with $N_{long} > N_{short}$ and $i \in \mathcal{I}$.

2. **Model 2**: Similar to model 1 with the inclusion of past returns.

$$\boldsymbol{S}^i_t = [\boldsymbol{S^i_{t,N_{short}}} \quad | \quad \boldsymbol{S^i_{t,N_{long}}} \quad | \quad \boldsymbol{R_{t,N_0}}]$$

$$\boldsymbol{w} = [\boldsymbol{w_{curto}} \quad | \quad \boldsymbol{w_{longo}} \quad | \quad \boldsymbol{w_0}]$$

   Where N_{long}, N_{short} and N are natural numbers with $N_{long} > N_{short}$ and $i \in \mathcal{I}$.

[1]This notation is in line with what was presented in the chapter 2. Here we only include the explicit dependence on the number of days spent (N) used in constructing the signals.

[2]$\frac{\mathcal{K}}{2} = \{1, 2, \cdots, \frac{N_s}{2}\}$

The total number of signals used in this work was 7 ($\frac{N_s}{2} = 7$)[3] and the detailed mathematical description of each of them is presented below[4]. The signal mix used to predict the ups and downs of returns must be flexible and dynamic to achieve good out-of-sample performance. The great majority of the signals used in this work are of type trend-followers, whose main characteristic is to be bought if the market is bullish and sold if the market is bass. That is, trend-follower models assume that if an asset is performing well, it will maintain that good performance.

1. **Signal 1:**
$$s_{t,1}^{N} = sign(R_{t-1}) * (\alpha + \beta * |R_{t-1}|) \tag{4-1}$$
Where the coefficients α and β are estimated using the training window that goes from R_{t-N} to R_{t-1}.

2. **Signal 2:**
$$s_{t,2}^{N} = sign(P_{t-1} - P_{t-N}) \tag{4-2}$$
Where P_t is the price of the financial asset at time t.

3. **Signal 3:**
$$s_{t,3}^{N} = \frac{(P_{t-1} - \mu_P)}{\sigma_P} \tag{4-3}$$
Where $\mu_P = \frac{1}{N}\sum_{i=t-N}^{t-1} P_i$ and $\sigma_P = \sqrt{\frac{\sum_{i=t-N}^{t-1}(P_i - \mu_P)^2}{N-1}}$.

4. **Signal 4:**
$$s_{t,4}^{N} = \frac{\sum_{i=t-N}^{t-1} R_i * I(R_i > 0)}{\sum_{i=t-N}^{t-1} -R_i * I(R_i < 0)} \tag{4-4}$$
Where $I(.)$ is a function such that $I(q) = 1$ if q is true and $I(q) = 0$ otherwise.

5. **Signal 5:**
$$s_{t,5}^{N} = \frac{\frac{max(P_{[t-N:t-1]})}{P_{t-1}} - \frac{P_{t-1}}{min(P_{[t-N:t-1]})}}{100 * \sigma_P} \tag{4-5}$$
Where $\mu_P = \frac{1}{N}\sum_{i=t-N}^{t-1} P_i$, $\sigma_P = \sqrt{\frac{\sum_{i=t-N}^{t-1}(P_i - \mu_P)^2}{N-1}}$ and $P_{[t-N:t-1]} = \{P_{t-N}, \cdots, P_{t-1}\}$.

6. **Signal 6:**
$$s_{t,6}^{N} = \frac{SMA_{\lfloor \frac{N}{10} \rfloor}}{SMA_N} - 1 \tag{4-6}$$
$$SMA_X = \frac{1}{X}\sum_{i=t-X}^{t-1} P_i$$

[3]It is worth mentioning that in model 1 we have N_S attributes and in model 2 we have $N_s + N_0$.

[4]Since we will use the same structure for all assets we will omit variable i in the definition of the predictors.

7. **Signal 7:**
$$s_{t,7}^{N} = max(R_{[t-N:t-1]}) \tag{4-7}$$
Where $R_{[t-N:t-1]} = \{R_{t-N}, \cdots, R_{t-1}\}$.

For a more in-depth discussion of the momentum strategy and sign constructions, we recommend references (34, 27, 15).

5
Proposed framework

Note that the model proposed in equation 2-16 above can be written as follows:

$$\underset{\boldsymbol{w},b}{\text{Minimize}} \quad \sum_{t=t_0}^{t_m} (1 - y_t * (\langle S_t, \boldsymbol{w} \rangle + b))_+ + \lambda * ||\boldsymbol{w}||^2 \tag{5-1}$$

Where $(1-v)_+ = max\{(1-v), 0\}$ is often called the hinge loss function, λ is tuning parameter which varies in range $(0, \infty)$.

The term $||\boldsymbol{w}||^2$ controls the complexity of the model. This term is known as regularized L_2 and is one of the most used. A disadvantage of this rule is that it does not lead to sparse solutions and sometimes it is difficult to interpret which variables are most relevant to the problem.

One of the most common ways to solve the problems of non-sparse solutions is to introduce the L_1 norm (see (31, 29) for more details) defined below.

Given a vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_{N_s})$ the norm L_1, denoted by $||\boldsymbol{w}||_1$, is defined as:

$$||\boldsymbol{w}||_1 = \sum_{k=1}^{N_s} |w_k| \tag{5-2}$$

The problem 5-1 became:

$$\underset{\boldsymbol{w},b}{\text{Minimize}} \quad \sum_{t=t_0}^{t_m} (1 - y_t * (\langle S_t, \boldsymbol{w} \rangle + b))_+ + \lambda * ||\boldsymbol{w}||_1 \tag{5-3}$$

The above formulation is convex and solves the problem of non-sparse solutions. However, it has two drawbacks. First, the λ value is continuous and the optimal value search process may be slow due to the large number of possibilities. Second, the relation between the λ value and the number of variables that do not assume zero is not immediate.

In order to solve the problem of the λ search space and the relation between the nonzero variables we can analogically to what was done in equation 5-3 changes the norm.

The norm that will be used throughout this paper will be based on the norm L_0 defined below.

Given a vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_{N_s})$ the norm L_0, denoted by $||w||_0$, is defined as:

$$||\boldsymbol{x}||_0 = \sum_{k=1}^{N_s} \frac{I(w_k \neq 0)}{N_s} \tag{5-4}$$

Where I (q) = 1 if q is true and I (q) = 0 otherwise.

Using this norm the problem 5-1 becomes:

$$\underset{\boldsymbol{w},b,\boldsymbol{z}}{\text{Minimize}} \quad \sum_{t=t_0}^{t_m} C_t z_t + ||\boldsymbol{w}||_0 \tag{5-5}$$

$$s.t.$$

$$y_t(\langle S_t, \boldsymbol{w}\rangle + b) \geq 1 - z_t, \qquad \forall t \in \mathcal{T} \tag{5-6}$$

$$z_t \geq 0, \qquad \forall t \in \mathcal{T} \tag{5-7}$$

In order to minimize the objective function of the problem 5-5 the norm L_0 will force the largest possible number of components of the vector $\boldsymbol{w}$ to be 0. However, the way the problem is formulated in 5-5 can not be solved by traditional solvers.

As the main purpose of using this norm is to control the number of variables that do not assume zero value, we introduce binary variables $\boldsymbol{I}$ to control the weights that are non-zero and K to restrict the total number of model variables. The problem 5-5 can be rewritten as follows:

$$\underset{\boldsymbol{w},b,\boldsymbol{z},\boldsymbol{I}}{\text{Minimize}} \quad \sum_{i=t_0}^{t_m} C_t z_t \tag{5-8}$$

$$s.t.$$

$$\sum_{k=1}^{N_s} I_k \leq K \tag{5-9}$$

$$y_t(\langle S_t, \boldsymbol{w}\rangle + b) \geq 1 - z_t, \qquad \forall t \in \mathcal{T} \tag{5-10}$$

$$-MI_k \leq w_k \leq MI_k, \qquad \forall k \in \mathcal{K} \tag{5-11}$$

$$I_k \in \{0,1\}, \qquad \forall k \in \mathcal{K} \tag{5-12}$$

$$z_t \geq 0, \qquad \forall t \in \mathcal{T} \tag{5-13}$$

In 5-8, M is a very large number and K is the maximum number of components of $\boldsymbol{w}$ that will not be 0. Also, note that if $I_k = 0$ implies $w_k = 0$ which was desired one.

Note that the equations 5-5 and 5-8 are equivalent (see (19) for more details) and solves the problem of controlling the number of nonzero variables. However, 5-8 is a linear convex problem with integer and non-integer variables (MILP) that can be easily solved with solvers as the Gurobi solver (see (23)).

Note that with this new formulation the number of variables is now controlled by the variable K which is integer and so the search of the optimal parameter is easily solved by doing a grid search in all p possibilities. Although there are N_s possibilities for K, some empirical studies suggest (see (10)) that the best choice for the maximum number of attributes is $\sqrt{N_s}$ and this is the value that we will use in this paper, i.e, $1 \leq K \leq \lfloor\sqrt{N_s}\rfloor$.

After finishing the modifications in the traditional SVM we need to generate the portfolios. This will be done based on what was presented in chapter-3, however we need to make some adaptations discussed below.

Note that in this work the predictions do not refer to the return but to the sign of R_t ($+1$ or -1) so adaptations will be made in the equation 3-11.

As mentioned in the section 2.1 when the signal from example $t \in \mathcal{T}$ to asset $i \in \mathcal{I}$ is positive , it is expected that $d_t^i = \frac{\langle \boldsymbol{w}_t^i, \boldsymbol{S}_t^i \rangle + b_t^i}{||w||}$ ($\boldsymbol{w}_t^i$ and b_t^i are the estimated values from the training data set and S_t^i will be the signal used in the return forecast) is a positive large number so that the point in question is distant of the plan that separates the positive and negative examples.

So that number can be looked at with a confidence we have for our forecast. Unlike the traditional optimization models that usually use the return on the objective function, here we will use the d_t^i in the objective function in order to place more emphasis on the predictions that we have the most confidence. Therefore, the model becomes:

$$\underset{x \in R^n, x_c \in R^+}{\text{Maximize}} \quad \sum_{i=1}^{n} d_t^i x_t^i + d_t^c x_t^c \tag{5-14}$$

$$s.t.$$

$$x_t^c = W_{t-1} - \sum_{i=1}^{n} c_i (u_+^i + u_-^i) \tag{5-15}$$

$$\sum_{i=1}^{n} R_{t-j} x_t^i \geq \gamma W_{t-1}, \qquad \forall j = 1, 2..., J \tag{5-16}$$

$$x_t^i = x_{t-1}^i (1 + R_{t-1}^i) + (u_+^i - u_-^i), \qquad \forall i \in \mathcal{I} \tag{5-17}$$

$$u_+^i \geq 0, \quad u_-^i \geq 0, \qquad \forall i \in \mathcal{I} \tag{5-18}$$

$$d_t^i = \frac{\langle \boldsymbol{w}_t^i, S_t^i \rangle + b_t^i}{||\boldsymbol{w}||}, \qquad \forall i \in \mathcal{I} \tag{5-19}$$

In addition, two more adaptations will be made to the model. First, the value of R_t^c has little variability and so d_t^c is always greater than the other d_t^i. This makes a large part of the portfolio to always be allocated to risk-free assets and to prevents this we will withdraw from the objective function and only include in the flow of purchases and sales. It will serve to finance the other assets.

Second, we include a parameter of regularization in transaction costs. This is a translation of a future cost function, which only exists because of transaction costs, so it is being used as an approximation of the forward view of transaction costs to generate implicit regularization. With this we managed to control the number of movements made and the idea is that in a regime of high costs the model does not make many moves.

$$
\begin{aligned}
\underset{x \in R^n}{\text{Maximize}} \quad & \sum_{i=2}^{n} d_t^i x_t^i - \lambda_t \sum_{i=1}^{n} c_i (u_+^i + u_-^i) && \text{(5-20)} \\
& s.t. \\
& \sum_{i=1}^{n} R_{t-j} x_t^i \geq \gamma W_{t-1}, \qquad \forall j = 1, 2..., J && \text{(5-21)} \\
& \sum_{i=1}^{n} c_i (u_+^i + u_-^i) + \sum_{i=1}^{n} (u_+^i - u_-^i) = 0, \qquad \forall i \in \mathcal{I} && \text{(5-22)} \\
& x_t^i = x_{t-1}^i (1 + R_{t-1}^i) + (u_+^i - u_-^i), \qquad \forall i \in \mathcal{I} && \text{(5-23)} \\
& u_+^i \geq 0, \quad u_+^i \geq 0, \qquad \forall i \in \mathcal{I} && \text{(5-24)} \\
& d_t^i = \frac{\langle \boldsymbol{w}_t^i, S_t^i \rangle + b_t^i}{||\boldsymbol{w}||}, \qquad \forall i \in \mathcal{I} && \text{(5-25)}
\end{aligned}
$$

Note that the 5-20 model is very similar to that proposed in (20) with the main differential being an inclusion of the regularizer λ_t. The regularization of costs is important in obtaining good results, since costs directly affect the result obtained as presented in the equation below, since the portfolio return can be broken down into (see (9) for more details):

$$R_t^P = \sum_{i=1}^{n} x_{t-1}^i R_t^i + \sum_{i=1}^{n} R_t^i (u_+^i - u_-^i) - \sum_{i=1}^{n} c_i (u_+^i + u_-^i) \qquad \text{(5-26)}$$

The first term of the equation represents the performance of the portfolio if no move had been made, the second can be regarded as the return obtained by having made the drives and the third is the price you pay for turning the wallet.

However, putting too high values for λ_t should lead to poor results, since in this case the algorithm will not do moves and the portfolio will not benefit from the predictions made.

Below we present the algorithm used to generate the results that will be presented in the next section. Initially we have to make the choice of the best K to be used by SVM. With this value we make the forecasts and the optimization model feeds. It is noteworthy that the parameters J and λ are updated every year as discussed in the next section.

Algorithm 1: Robust optimization model with predictions being made by an SVM - L0.

```
Result: x_t, W_t, ret_strategy
initialization: J_0, W_0, λ_0, X_0, 𝒦 = {1, 2, ···, √N_s} ;
while t ∈ 𝒯 do
    t_train = [t − 60, t − 20];
    t_val = [t − 20, t − 1];
    while i ∈ ℐ do
        while k ∈ 𝒦 do
            w_t^k = TrainSVM(S^i_train, Y^i_train, k);
            sharpe ← SharpeRatio(S^i_val, Y^i_val, ret^i_val, w_t^k);
            k = k + 1
        end
        IndexMax = argmax(sharpe);
        K* = 𝒦[IndexMax];
        w_t^{i,*} = TrainSVM(S^i_train, Y^i_train, K*);
        d_t ← (⟨w_t^{i,*}, S_t^i⟩ + b_t^i) / ||w||;
        i = i + 1;
    end
    if t == EndOfYear then
        J_new, λ_new = BestParameterRO([d_1, d_2, ···, d_{t−1}], ret_{1:t−1}, c);
    else
        J_new, λ_new = J_old, λ_old;
    end
    x_t, W_t, ret_strategy = RO(d_t, ret_t, x_{t−1}, J_new, λ_new, W_{t−1}, c);
    t = t + 1
end
```

6
Empirical results

For each portfolio financial asset we tested the 1-2 models for the different C_t functions and transaction cost given in the table 6.1 below.

Asset	Cost
USD/BRL	0.07 %
IBX-100	0.10%
IDKA 5	0.10%
IMA-B 5+	0.12%
GOLD	0.07%
S&P 500	0.10%

Table 6.1: Values used as transaction cost of assets

As mentioned in the previous section the period used for testing the models was 01/01/2004 a 30/05/2018 comprising 3623 daily quotes. For the forecast of each day, we use the set of parameters with the best performance in the validation window. For example, if we are on *trading* t, the validation window goes from $t - n_{validation}$ up to $t - 1$.

For each training window, we calculate the optimal parameter of the model for a given set of hyper-parameters, that is, Θ, where Θ is the set of hyper-parameters. With this value we calculate Sharpe's strategy in a validation window. We repeat this procedure until we sweep all possible values of the hyper-parameters (which here are discrete values).

Finally, we chose that set of parameters and hyper-parameters that lead to a strategy with higher sharpe. Mathematically we have,

$$w^*(\Theta^*) = \underset{\Theta}{\operatorname{argmax}} \quad Sharpe(w^*(\Theta), t_1, t_2) \tag{6-1}$$

In the equation 6-1 above, $[t_1, t_2]$ is the validation interval with $n_{validation}$ days and Θ is the set of hyper-parameters given by $\Theta = [N_{long}, N_{short}, N, K]$. The value $w^*(\Theta^*)$ is used to make the prediction in the test window given by $[t_2 + 1, t_2 + \tau]$. Here we will make $\tau = 1$, that is, we will use $w^*(\Theta^*)$ just to make the forecast on the day immediately after the training window. Once the forecast has been made the validation window is updated to $[t_1 + 1, t_2 + 1]$. In the table 6.2 we present the models and the respective values used for the hyper-parameters.

Model	N_{long}	N_{short}	N	K
1	{126,252}	$\{21, 63\}$	-	$\{1, 2, 3, 4\}$
2	$\{126, 252\}$	$\{21, 63\}$	5	$\{1, 2, 3, 4\}$

Table 6.2: Values used for the hyper-parameters of each model

Although the total number of daily quotes is 3623 we only have $3623 - (n_{training} + n_{validation} + max(N_{long}))$ days of *trading*, since we need those days to calculate the values of the signs, train the model and validate the results.

In addition, the value used for C_t will be of two forms: constant function and quadratic function. In the figure 6.1 below we have the form of the functions used. The constant function will penalize all errors in the same way and the quadratic the penalty will increase with the return. Note that for return with a module smaller than 1% the constant function $C_t = 1$ will be more punitive than the function $C_t = (100 * R_t)^2$. When the return module exceeds 1%, the quadratic function will become more punitive.

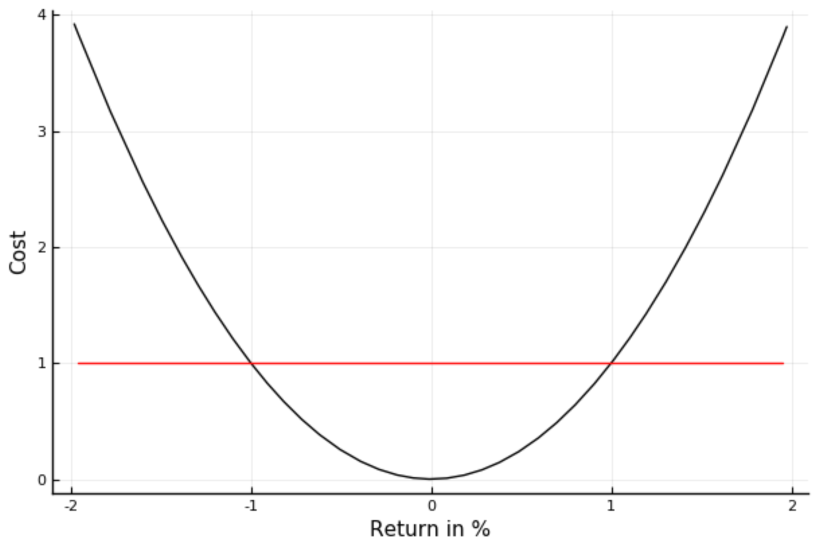

Figure 6.1: Forms of cost functions C_t. In black we have $C_t = (100 * R_t)^2$ and in red $C_t = 1$.

Finally, as mentioned in the previous chapter, we removed from the objective function the term $d_t^c x_t^c$, since it was always far superior to the others and this skewed allocations. However, without this term in the objective function we are somewhat restricted to allocations in risk free assets and as Brazil is one of the countries with the highest interest rates this strategy may mean a worse performance of the sharpe strategy.

To circumvent this situation predictions will be made using excess return over risk free rate[1]. That is, when the forecast for a given financial asset is positive it means that the return on this asset is higher than the risk free asset. Thus the model will only have incentives to allocate to risky assets if this return is at least that of risk-free assets.

In the next section we will present the results obtained for the accuracy of the models. It is worth mentioning that in this first phase of tests we will not take into account the costs in the transactions, since the main concern will be if the models have more accuracy than the random model.

6.1 Accuracy of Models

The figures 6.2 and 6.3 below show the accuracy obtained for all the variations tested. The values marked in green indicate that we have evidence to reject $H_0 : T_s = 0.5$ with a significance level of 5%. Note that for both the constant and the quadratic cost functions the results in terms of accuracy were similar.

Ativo	Modelo 1	Modelo 2
Dólar	0.5414	0.5282
IBX – 100	0.5183	0.5098
IDKA 5	0.5463	0.5373
IMA-B 5 +	0.5987	0.5801
OURO	0.5112	0.5341
S&P 500	0.5189	0.5286

Figure 6.2: Accuracy obtained for function $C_t = 1$.

Ativo	Modelo 1	Modelo 2
Dólar	0.5334	0.5298
IBX – 100	0.5201	0.5101
IDKA 5	0.5479	0.5345
IMA-B 5 +	0.5856	0.5680
OURO	0.5132	0.5295
S&P 500	0.5166	0.5201

Figure 6.3: Accuracy obtained for function $C_t = (100 * R_t)^2$.

The figure 6.4 presents the results of the statistical test for the comparison between the proportions ($H_0 : T_s^{constant} = T_s^{quadratic}$ versus $H_a : T_s^{constant} \neq$

[1] $\hat{R}_t^i = R_t^i - R_t^c$

$T_s^{quadratic}$) of two models. In it we can verify that there is no statistical evidence to affirm that the results are different.

Asset	Model 1	Model 2
USD/BRL	0	0
IBX – 100	0	0
IDKA 5	0	0
IMA-B 5 +	0	0
GOLD	0	0
S&P 500	0	0

Figure 6.4: Binary matrix where 0 indicates that we have no evidence to reject the null hypothesis.

6.2 Financial results of the models

In this section we will evaluate the results from the financial point of view. To do this, we will set up an investment portfolio using the best forecasts made by models 1 and 2. The 6.3 table below shows the assets and their forecasts.

Asset	Model
USD/BRL	1
IBX-100	1
IDKA 5	2
IMA-B 5+	1
GOLD	2
S&P 500	2

Table 6.3: Assets and their forecasting models. The forecast model was chosen according to the accuracy.

The search space of the hyper parameters γ, J and λ_t used to construct the portfolio were:

1. $\gamma \in \{-1.5\%\}$

2. $J \in \{10, 20, 30, 40, 50\}$

3. $\lambda_t \in \{100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$

In addition, to control the volatility and loss of the strategy we place a restriction on the maximum leverage allowed in 2 equity.

We initially searched for the optimal values of J^* and λ_t^* using all available predictions. The purpose of the search was to obtain the portfolio with the greatest excess of wealth when compared to the risk-free asset. The results obtained in this *gridsearch* are presented in the figure 6.5.

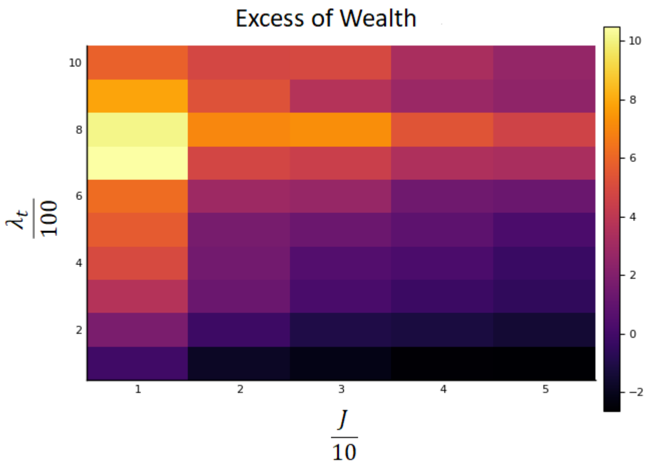

Figure 6.5: Map of indicated heat of wealth accomplishment when we varied λ_t and J.

The optimal value found was $(J^*, \lambda_t^*) = (10, 700)$. The value of J^* can be seen as a process memory, that is, the maximum number of days where we have information relevant to the portfolio creation process. In addition, a high value of λ_t^* indicates that controlling the number of moves is important.

Another point of attention would be in the value of γ, because we want that somehow this parameter is connected to the maximum loss that we can suffer. To see if it is in fact related to model losses, we calculated *Value At Risk* violations with a significance level of 5% in the search space. In black on figure 6.6 we have cases where losses were greater than 1.5% by more than 5% of days.

Note that for values of $\lambda_t > 200$ and $J > 10$ o *Value at Risk* $< |\gamma|$.

Since we use all the available quotes in the search for the best J and γ values we have no guarantee that these values will be the best for new quotes. That said, we will do the following consistency test on the sample: divide the period into 1-year sub-periods and calculate the parameter values in alternate years to see if the values are close to those obtained using the whole sample.

The results obtained showed that in some periods the results were far from the values $(10, 700)$ as can be seen in the table 6.4.

Given this fact, we decided to change the methodology for choosing J and λ. Initially we will start the portfolio optimization with arbitrary values for $(J_t, \lambda_t) = (J_0, \lambda_0)$ and every 1 year we reevaluate these values looking at

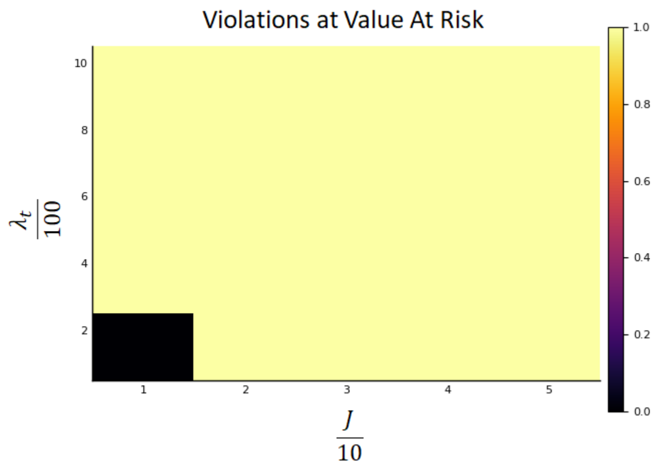

Figure 6.6: Map of indicated heat where we violate the value at risk when we change λ_t and J. In blacks we have the region where we violate Value at risk with a level of significance of 5%

$Year$	(J^*, λ_t^*)
Year 1	(10,700)
Year 3	(10,300)
Year 5	(30,800)
Year 7	(50,100)
Year 9	(30,700)
Year 11	(20,100)

Table 6.4: Values obtained for the pair (J, λ_t) in each sub-period.

all past quotes available. That is, for the calculation of (J_{t+1}, λ_{t+1}) we will do the search using all information until time t and we will keep these values for the next 1 year.

Looking at the results presented in table 6.5 we see that the value of J_t was 10 in practically all periods and the value of λ_t converged to 700 after the eighth year. In the figure 6.7 below we present the differences in terms of wealth between the best possible result (use (10,700) throughout the period) versus using the estimated values every 1 year.

The results presented here were obtained using the new methodology, as we found it more appropriate.

To evaluate financial results fairly, you need to create benchmark's appropriate. As the volatility of the strategy is around 15% it is natural to compare it with the Brazilian stock index IBX-100. In addition, since interest in Brazil is historically high (almost always above 10%) it is common to make comparisons with the risk free rate that is nothing more than the opportunity cost of the investor. Another benchmark commonly used is the equally weighted

Estimation	*Usage*	(J^*, λ_t^*)
-	02/05/2005 - 05/04/2006	(10,500)
02/05/2005 - 05/04/2006	05/04/2006 - 16/03/2007	(10,700)
02/05/2005 - 16/03/2007	16/03/2007 - 25/02/2008	(10,600)
02/05/2005 - 25/02/2008	25/02/2008 - 27/01/2009	(10,700)
02/05/2005 - 27/01/2009	27/01/2009 - 06/01/2010	(20,1000)
02/05/2005 - 06/01/2010	06/01/2010 - 14/12/2010	(20,800)
02/05/2005 - 14/12/2010	14/12/2010 - 21/11/2011	(10,700)
02/05/2005 - 21/11/2011	21/11/2011 - 24/10/2012	(10,700)
02/05/2005 - 24/10/2012	24/10/2012 - 01/10/2013	(20,900)
02/05/2005 - 01/10/2013	01/10/2013 - 08/09/2014	(10,700)
02/05/2005 - 08/09/2014	08/09/2014 - 13/08/2015	(10,700)
02/05/2005 - 13/08/2015	13/08/2015 - 21/07/2016	(10,700)
02/05/2005 - 21/07/2016	21/07/2016 - 29/06/2017	(10,700)
02/05/2005 - 29/06/2017	29/06/2017 - 08/06/2018	(10,800)

Table 6.5: Values obtained for the pair (J, λ_t) using the new methodology.

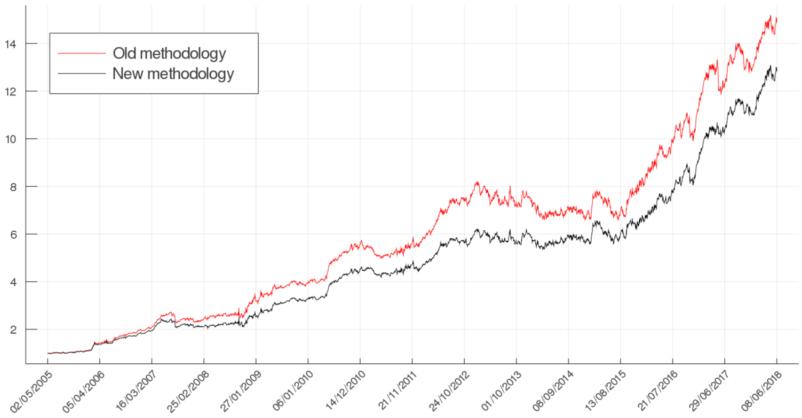

Figure 6.7: Result obtained using the new methodology.

(EW) portfolio to show that there are indeed benefits in making the moves suggested by the model developed. Finally, we will use the same methodology used in (20) to create a linear regression model (LR Model). The idea is to show that there are benefits to exchanging regression models for classification models when working with financial returns.

The figure 6.8 below shows the evolution of the wealth of a hypothetical investor who has applied R$ 1 real in the strategies mentioned above.

It is worth noting that during the 2008 crisis while the IBX-100 fell by around -60 % the strategy was almost unaffected falling only -3.91 %. In addition, in the period 2016 - 2018 we saw a significant increase in the Brazilian stock index (81 %), the strategy was able to capture this movement and also had a very significant result (94 %).

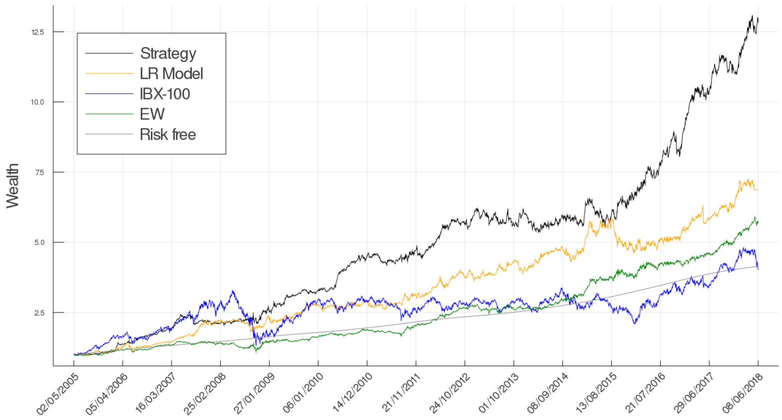

Figure 6.8: Evolution of the strategy wealth and the benchmarks in the period 2005-2018.

The table below 6.6 shows the volatility and sharpe of the strategies.

Asset	*Sharpe*	Volatility
Risk free	-	0.01%
Strategy	0.81	15.83%
LR Model	0.45	14.51%
EW	0.41	13.51%
IBX-100	0.12	23.03%

Table 6.6: Sharpe and volatility

Looking at the table, it can be seen that the portfolio has a sharpe well above those of benchmarks.

However, it is important to also note the consistency of the model over time. The ideal is that the gains are not concentrated in a single period but well distributed, because in this way as the investors apply their resources they will have good results regardless of the date of application.

To evaluate the consistency we will use sliding window of return, that is, we will calculate the returns in fixed windows over time. For example, imagine two investors who applied their funds at t and t + 1 and remained k invested days. Although both of them spend k days invested their results may have been quite different, but if the strategy is consistent, although these values must exceed the traditional indexes (risk free rate, IBX-100 related).

One way to measure consistency is to calculate how many windows the results exceed the risk free rate. This can be interpreted as the chance for an investor to buy that asset and to have a result higher than the opportunity cost.

The figure 6.9 below shows the evolution of the moving window considering $k = 252$. Note that the number of times the result was better than the risk free rate is 64.7% of the time.

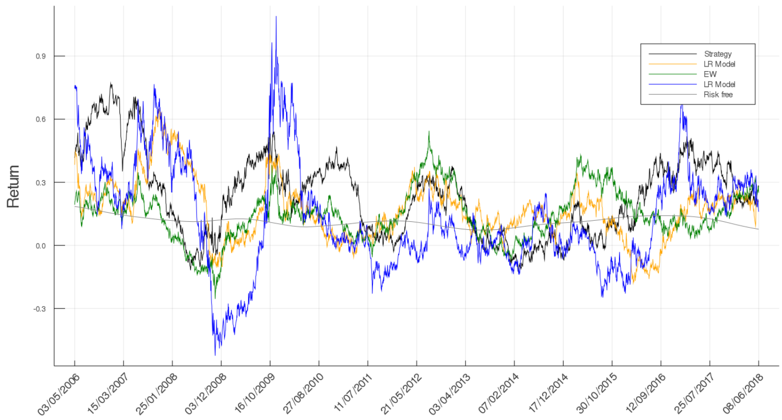

Figure 6.9: The moving window series when $k = 252$ for the period 2006 - 2018

In the figure 6.10 we have the histogram of the historical series of moving window. The average window for the Strategy is 22.30% against the risk free of 11.40%. The 6.7 table shows the mean, the quantile of 5% (Q1), the quantile of 95% (Q2) and the number of windows on the strategies exceeds the risk free.

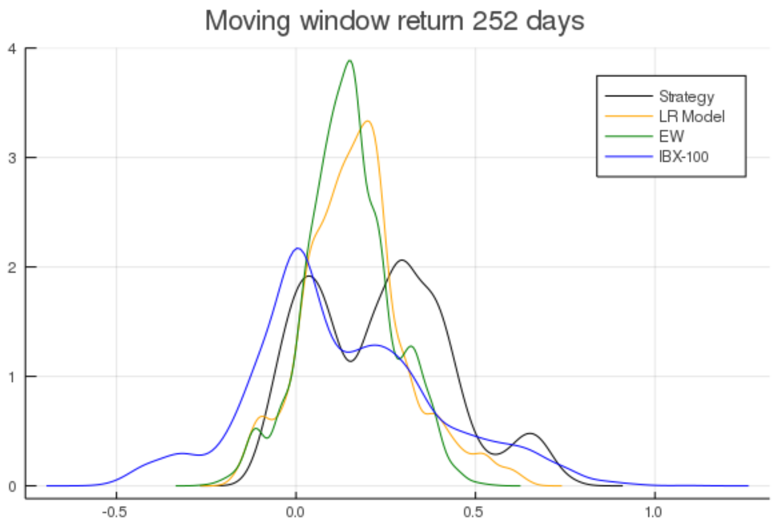

Figure 6.10: Histogram of the moving window series when $k = 252$ for the period 2006 - 2018

It can be seen in the table that both the 5% and 95% quantile are much higher than the IBX-100, ie in both bad and good results the investors who invested in the portfolio developed in this work will have better results.(in losses the strategy loses less and in gains it surpasses the IBX-100)

Asset	% > Risk free	Mean	Q1	Q2
Strategy	64.70	22.30%	-9.12%	72.03%
LR Model	63.75	15.65%	-6.59%	42.15%
IBX-100	47.77	14.40%	-28,30%	61.75%
EW	57.46	14.65%	-5.15%	35.73%
Risk free	-	11.40%	7.64%	15.55%

Table 6.7: Descriptive statistics for the mobile window when k = 252. The first column shows the number of times that the strategies exceed the risk free.

It is worth mentioning that the model developed in (20) had the 5% quantile with a less negative value, indicating that the losses suffered by this model are less severe.

Since the portfolio has a high standard deviation, it is also recommended to evaluate longer return windows, since we can have some periods where the portfolio had some more pronounced loss (which is normal and expected since the high standard deviation implies high variation both positive and negative). Having said that, we will also present the results for the value of $k = 504$.The result is shown in the figure 6.11.

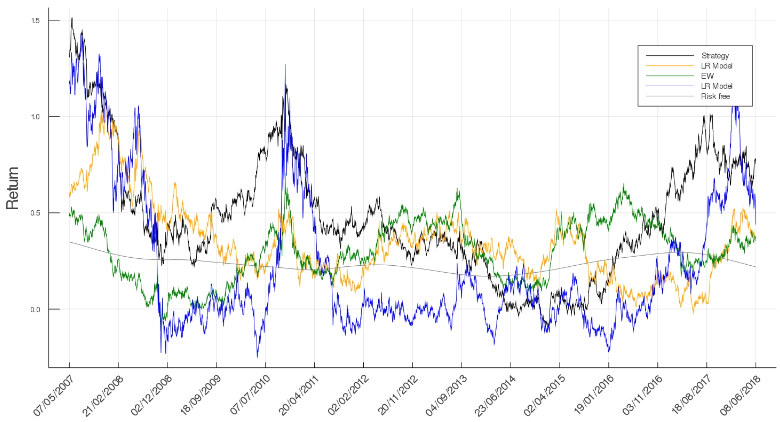

Figure 6.11: The moving window series when $k = 504$ for the period 2007 - 2018

As expected, increasing k to 504 the fraction of windows where the portfolio exceeds the risk free rises to 73.20%. In addition, the number of times EW exceeds the risk free is 63.37% and the LR Model is 66.77%. In the figure 6.12 we present the histogram of the series when $k = 504$ and in table 6.8 we present the average statistics, Q1 and Q2. The results are in line with those obtained for $k = 252$, with the main difference being 5% quantil of EW and the LR Model.

Finally, from the investor's point of view it is important to evaluate the actual losses that the strategy suffers over time. For this we will use the measure

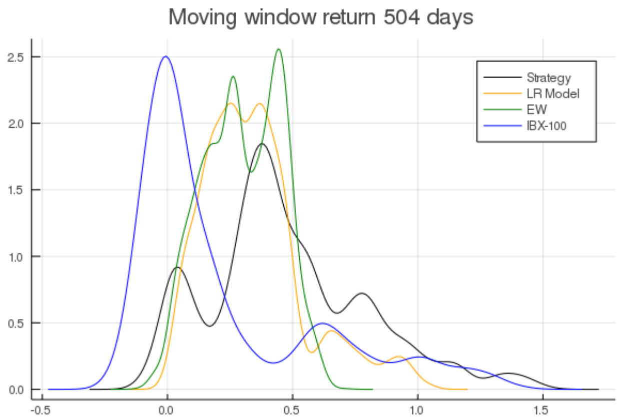

Figure 6.12: Histogram of the moving window series when $k = 504$ for the period 2007 - 2018

Asset	% > Risk free	Mean	Q1	Q2
Strategy	73.20%	44.50%	-4.38%	128.13%
LR Model	66.77%	32.64%	6.55%	75.28%
IBX-100	30.12%	22.37%	-11,85%	105.71%
EW	63.37%	29.99%	4.71%	52.26%
Risk free	-	23.81%	17.44%	29.83%

Table 6.8: Descriptive statistics for the mobile window when k = 504. The first column shows the number of times that the strategies exceed the risk free.

known in the literature as maximum drawdown (MDD). It measures the losses incurred between the highest quota and the lowest quota of the model. The evolution of the MDD is presented in the figure 6.13.

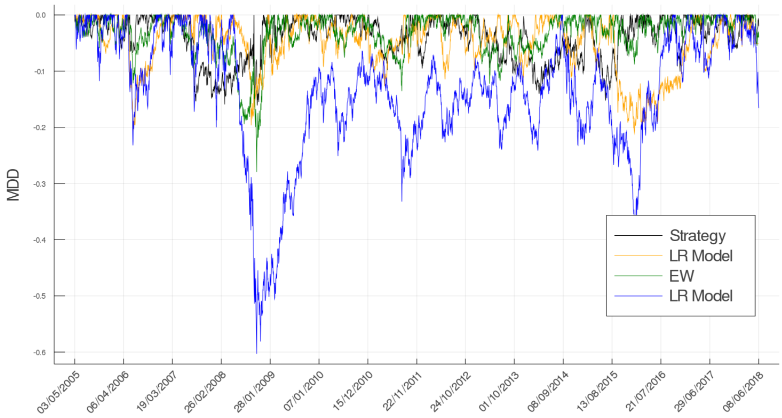

Figure 6.13: Evolution of the MDD of the strategies in the period 2005-2018

It should be noted that although the Strategy lost only -3.91% in the period of the 2008 crisis, losses for the previous 2 years totaled around 17% (the portfolio had a poor performance in the period before the crisis). LR

Model, on the other hand, had much less pronounced losses before the crisis and at the peak of the crisis reached -19%.

As can be seen from the viewpoint of the losses in the periods prior to the crisis of 2008, the EW and IBX-100 strategies were better, but in the subsequent periods the portfolio had MDD similar to EW and well below the IBX-100.

6.3 Sensitivity analysis

The results presented in the previous section are for γ and L fixed in -1.5% and 2, respectively. These values were chosen based on what the multimarket funds of the Brazilian industry use.

Since we have not done any optimization process in choosing these values it is natural to evaluate what happens when we change one of them while keeping the other fixed, that is, we will set $L = 2$ and vary the values of γ and vice verse. The 6.9 table below shows the values that will be used for γ and L.

γ	L
-1%	1
$\mathbf{-1.5\%}$	1.5
-2.0%	**2**
-2.5%	2.5
-3.0%	3.0
-3.5%	3.5
-4.0%	4.0
-4.5%	4.5
-5.0%	5.0

Table 6.9: Values used for γ and L. In black we highlight the pair used in this work.

The metrics used to evaluate the sensitivity results will be: accumulated wealth, ratio between the mean of the return and the standard deviation and the maximum drawdown.

In the figure 6.14 below we have the results when we set γ in -1.5% and we vary the value of L.

The results indicate that increasing the leverage value up to 4 is beneficial to accumulated wealth. However, the graph of the figure 6.14 (b) indicates that even in situations where the value of wealth is increasing with L the increment of risk that is generated is not fully translated into return. The MDD chart behaves as expected, since losses generally get worse as leverage increases.

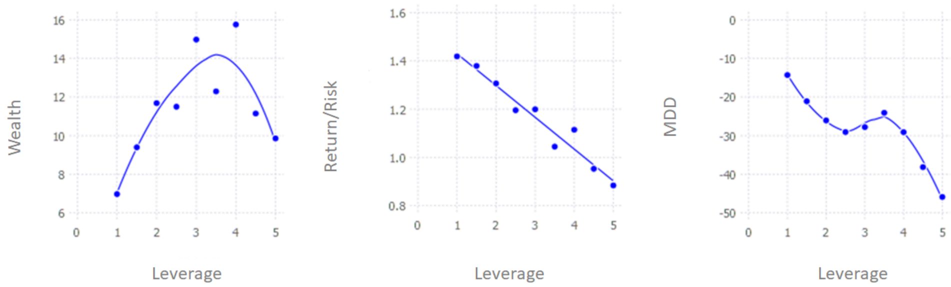

Figure 6.14: (a) Accumulated wealth. (b) Return to Risk Ratio. (c) MDD

In the figure 6.15 we present the results when we fix the value of L in 2 and vary the γ. For the graphs 6.15 (a) and (b) the results were similar to those shown in figure 5.16. There is an optimal value to where increasing $|\gamma|$ causes wealth to increase. However, the Return to Risk ratio worsens even in the case where wealth increases. In the case of the graph 6.15 (c) the results were divergent. Increasing the $|\gamma|$ did not lead to greater MDD as expected.

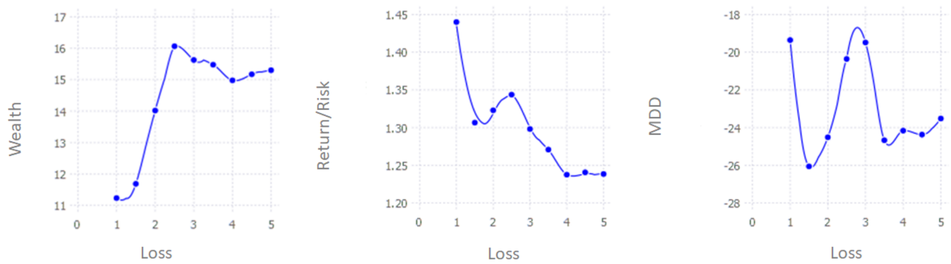

Figure 6.15: (a) Accumulated wealth. (b) Return to Risk Ratio. (c) MDD

7
Concluding Remarks

The central objective of this work was the creation of optimization that combines two elements: the prediction of trends in financial series and a model for the creation of investment portfolios.

For the prediction of financial series, the artificial intelligence model support vector machine was used as a basis, whose theory is widely diffused in the academic world. However, the work was not limited in just using a black blox form the already existing model, since several changes were proposed in the original model.

Of the proposed changes it is worth mentioning the change made in the regularization that went from norm L_2 to norm pseudo L_0. Changing the norm helps to prevent overfiting and makes it clear which (and how many) variables are relevant for the series forecast.

Already for the determination of the portfolio an adaptation of the model proposed in (20) was used. The main change was in the objective function that in the original work sought to maximize the expected return while in this work we maximized the confidence to make the correct forecast. That is, the financial amounts invested in financial assets are linked to our confidence in the forecast.

The empirical results of the investigation reported in this paper suggest that it is possible to obtain higher returns with this strategy when compared with the benchmark strategies (e.g., equally weighted and buy-and-hold, IBX-100 and betina's model) while considering transaction costs. Out-of-sample results indicate that the applied robust optimization specification showed an en-hanced portfolio performance while successfully constraining for significant losses.

Bibliography

[1] ALBUQUERQUE, P.; MORETTIN, P.. **Forecast of financial time series by means of vector support machines and wavelets.** PhD thesis, 08 2014.

[2] ALLAN G., CLIVE, TUMMERMANN. **Efficient Market Hypothesis and Forecasting**. CEPR Discussion Papers 3593, C.E.P.R. Discussion Papers, Oct. 2002.

[3] ATSALAKIS, G. S; VALAVANIS , K. P. **Surveying stock market forecasting techniques - part ii: sof comput methods**. Expert Systems with Applications, 3:5932–5941, 2009.

[4] BAFANDEH IMANDOUST, S.; BOLANDRAFTAR, M.. **Forecasting the direction of stock market index movement using three data mining techniques: the case of tehran stock exchange**. international Journal of Engineering Research and Applications, 4:2248–9622106, 06 2014.

[5] BERTSIMAS, D.; BROWN, D. B. ; CARAMANIS, C.. **Theory and applications of robust optimization**. SIAM Rev., 53(3):464–501, Aug. 2011.

[6] BERTSIMAS, D.; PACHAMANOVA, D.. **Robust multiperiod portfolio management in the presence of transaction costs**, 2008.

[7] BLATTBERG, R. C.; GONEDES, N. J.. **A comparison of the stable and student distributions as statistical models for stock prices**. The Journal of Business, 47(2):244–80, 1974.

[8] BOURGUIGNON, S.; NININ, J.; CARFANTAN, H. ; MONGEAU, M.. **Exact Sparse Approximation Problems via Mixed-Integer Programming: Formulations and Computational Performance**. IEEE Transactions on Signal Processing, 64(6):1405–1419, Mar. 2016.

[9] BOYD, S.; BUSSETI, E.; DIAMOND, S.; KAHN, R.; NYSTRUP, P. ; SPETH, J.. **Multi-period trading via convex optimization**. Foundations and Trends in Optimization, 3(1):1–76, 2017.

[10] BREIMAN, L.. **Random forests**. Machine Learning, 45(1):5–32, Oct 2001.

[11] CAO, L.; TAY, F.. **Financial forecasting using support vector machines**. Neural Computing and Applications, 10:184–192, 05 2001.

[12] CERIA, S.; STUBBS, R.. **Incorporating estimation errors into portfolio selection: Robust portfolio construction**. Journal of Asset Management, 7, 04 2006.

[13] CHRISTOFFERSEN, P. F; DIEBOLD,F. X. **Finanial asset returns, direction-of-change forecasting, and volatility dynamics**. Management Science, 52(8):1273–1287, 2006.

[14] DUNNING, I.; HUCHETTE, J. ; LUBIN, M.. **Jump: A modeling language for mathematical optimization**. SIAM Review, 59(2):295–320, 2017.

[15] DU PLESSIS, JOHAN. **Demystifying Momentum: Time-series and cross-sectional momentum, volatility and dispersion.** Master's thesis, University of Amsterdam, 2013.

[16] ELTON, E.; GRUBER, M.; BROWN, S. ; GOETZMANN, W.. **Modern Portfolio Theory and Investment Analysis**. John Wiley & Sons, 2009.

[17] FAMA, E. F.. **The behaviour of stock market prices**. Journal of Business, p. 34–105, 1965.

[18] FAN, A.; PALANISWAMI, M.. **Stock selection using support vector machines**. volumen 3, p. 1793 – 1798 vol.3, 02 2001.

[19] FENG, M.; MITCHELL, J. E.; PANG, J.-S.; SHEN, X. ; WÄCHTER, A.. **Complementarity formulations of ‘ 0-norm optimization problems**. 2013.

[20] FERNANDES, B. D. **Essays on Asset Allocation Optmization Problems Under Uncertainty**. Tese de doutorado em engenharia elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2014.

[21] GOH, J.; SIM, M.. **Distributionally robust optimization and its tractable approximations**. Oper. Res., 58(4-Part-1):902–917, July 2010.

[22] GOLDFARB, D.; IYENGAR, G.. **Robust portfolio selection problems**. Math. Oper. Res., 28(1):1–38, Feb. 2003.

[23] GUROBI OPTIMIZATION, L.. **Gurobi optimizer reference manual**, 2018.

[24] HSIEH, D. A. **Testing for nonlinear dependence in daily exchange rates**. The Journal of Business, 62(10):339–368, 1989.

[25] HUANG, C.-F.. **A hybrid stock selection model using genetic algorithms and support vector regression**. Appl. Soft Comput., 12(2):807–818, Feb. 2012.

[26] HUANG, W.; NAKAMORI, Y. ; WANG, S.. **Forecasting stock market movement direction with support vector machine.** Computers OR, 32:2513–2522, 01 2005.

[27] JEGADEESH, N.; TITMAN, S.. **Returns to buying winners and selling losers: Implications for stock market efficiency**. Journal of Finance, 48(1):65–91, 1993.

[28] KIM, K.-J.. **Financial time series forecasting using support vector machines**. Neurocomputing, 55:307–319, 09 2003.

[29] KOLLECK, A.; VYBÍRAL, J.. **Non-asymptotic analysis of l_1-norm support vector machines**. CoRR, abs/1509.08083, 2015.

[30] LINDEN, M.. **A Model for Stock Return Distribution**. International Journal of Finance & Economics, 6(2):159–169, April 2001.

[31] LIU, Y.; HELEN ZHANG, H.; PARK, C. ; AHN, J.. **Support vector machines with adaptive lq penalty**. Comput. Stat. Data Anal., 51(12):6380–6394, Aug. 2007.

[32] LUENBERGER, D.. **Investment Science**. Oxford University Press, 2009.

[33] MADAN, D.; SENETA, E.. **The variance gamma (vg) model for share market returns**. The Journal of Business, 63:511–24, 02 1990.

[34] MOSKOWITZ, T. J.; OOI, Y. H. ; PEDERSEN, L.. **Time series momentum**. Journal of Financial Economics, 104(2):228–250, 2012.

[35] NAYAK, R. K.; MISHRA, D. ; RATH, A. K.. **A naïve svm-knn based stock market trend reversal analysis for indian benchmark indices**. Appl. Soft Comput., 35(C):670–680, Oct. 2015.

[36] PAPAILIAS, F; LIU, J; THOMAKOS, D. D. **Return signal momentum**. May 2017.

[37] ROSILLO, R.; FUENTE, D. ; BRUGOS, J. A.. **Forecasting sp500 index movement with support vector machines**. Proceedings of the 2011 International Conference on Artificial Intelligence, ICAI 2011, 2, 04 2012.

[38] TÜTÜNCÜ, R.; KOENIG, M.. **Robust asset allocation**. Annals of Operations Research, 132(1):157–187, Nov 2004.

[39] WEI, Z.. **A svm approach in forecasting the moving direction of chinese stock indices**. 2012.

[40] WOZABAL, D.. **Robustifying convex risk measures for linear portfolios: A nonparametric approach**. Operations Research, 62:1302, 10 2014.

A
Equivalences between the problems 5-5 and 5-8

The purpose of this appendix is to present the reformulations of the L_0 norm based constraints. We will show,

$$(P_1): ||\boldsymbol{w}||_0 \leq K \Leftrightarrow (P_2): \begin{cases} \exists \boldsymbol{b} \in \{0,1\}^m \quad \text{such that} \\ \sum_{j=1}^m b_j \leq K, \quad \text{(i)} \\ -M\boldsymbol{b} \leq \boldsymbol{w} \leq M\boldsymbol{b}. \quad \text{(ii)} \end{cases}$$

The ($\Rightarrow$) implication is straightforward since $b_j = 0 \Leftrightarrow w_j = 0$. Now, let $\boldsymbol{b} \in \{0,1\}^m$ satisfy (i) and (ii), and suppose $||\boldsymbol{w}||_0 \geq K$. From (ii), one has $(b_j = 0) \Rightarrow (w_j = 0)$, that is, $(w_j \neq 0) \Rightarrow (b_j = 1)$. Hence $b_j = 1$ for at least $K+1$ indices j, which contradicts (i). Therefore, $||\boldsymbol{w}||_0 \leq K$.

For more details in exact sparse approximation problems see (8).

B
Allocations (x_t)

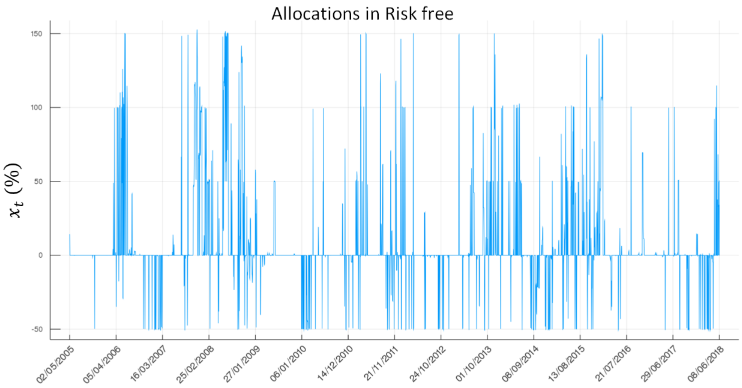

Figure B.1: Temporal evolution of risk-free rate allocations in the period 2005 - 2019.

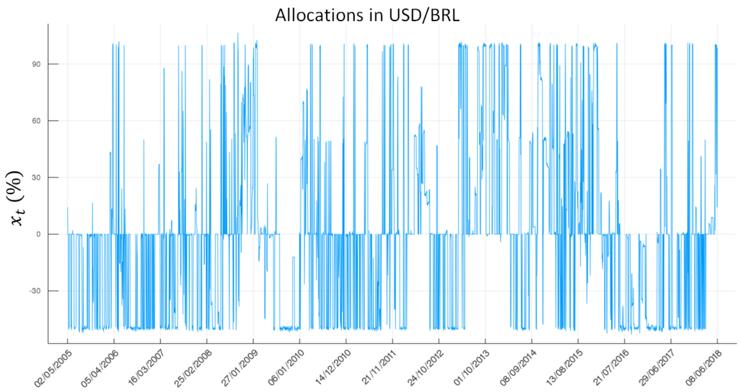

Figure B.2: Temporal evolution of USD/BRL allocations in the period 2005 - 2019.

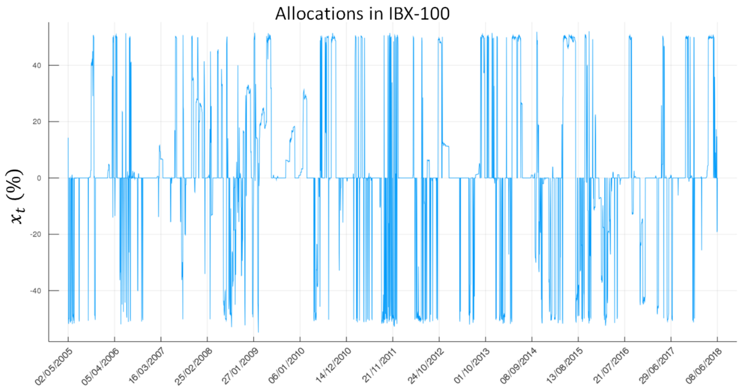

Figure B.3: Temporal evolution of IBX-100 allocations in the period 2005 - 2019.

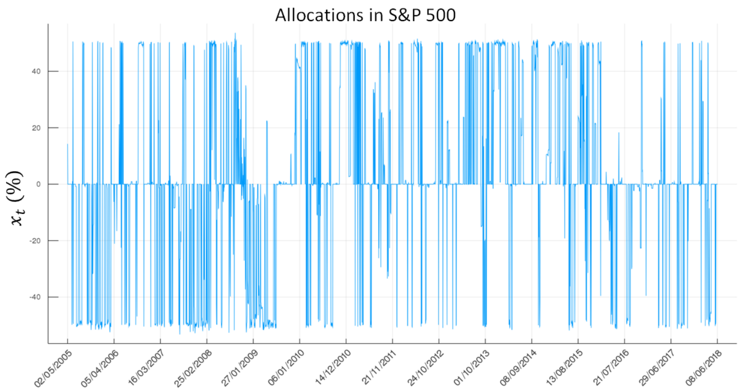

Figure B.4: Temporal evolution of S&P 500 allocations in the period 2005 - 2019.

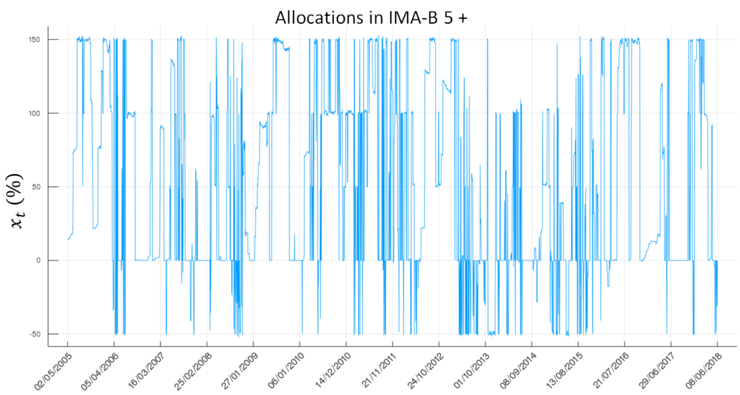

Figure B.5: Temporal evolution of IMA-B 5+ allocations in the period 2005 - 2019.

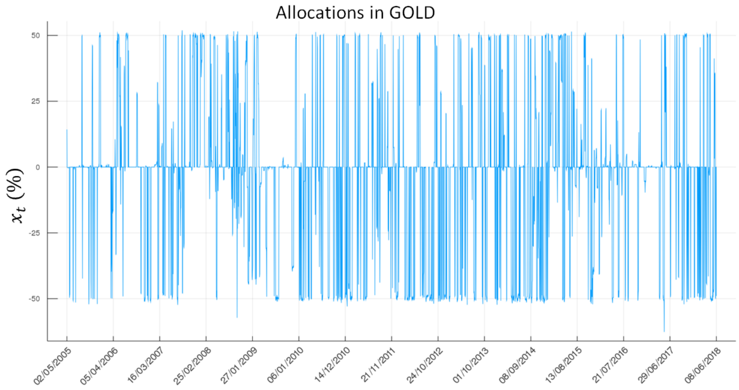

Figure B.6: Temporal evolution of GOLD allocations in the period 2005 - 2019.

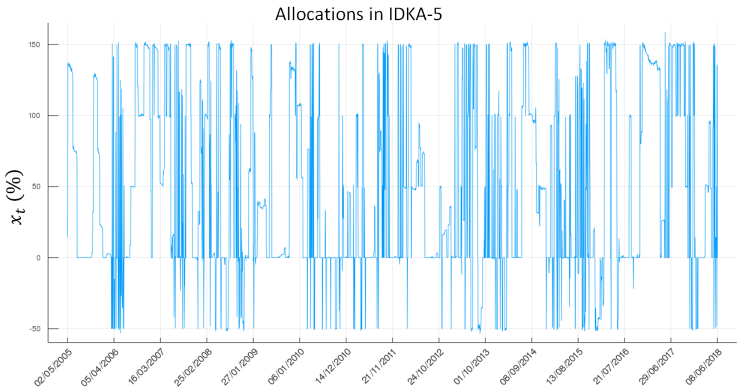

Figure B.7: Temporal evolution of IDKA-5 allocations in the period 2005 - 2019.

C
Forecast confidence (d_t)

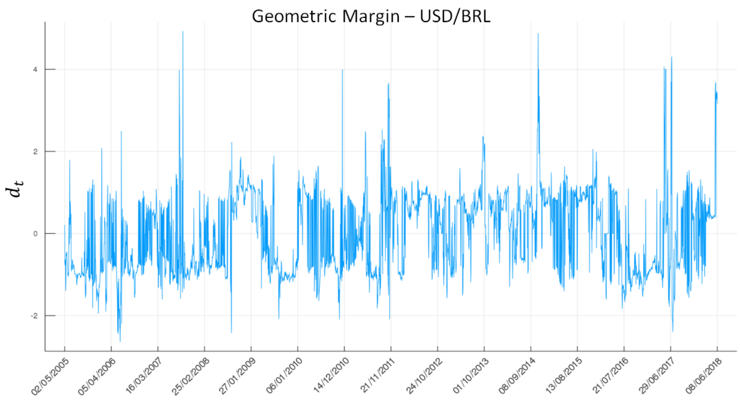

Figure C.1: Temporal evolution of USD/BRL geometric margin in the period 2005 - 2019.

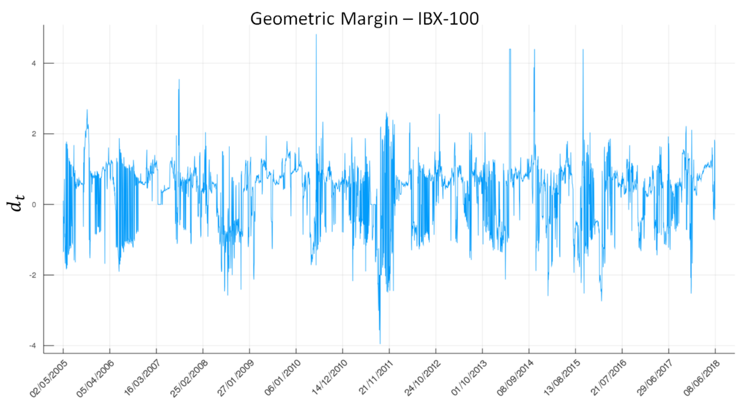

Figure C.2: Temporal evolution of IBX-100 geometric margin in the period 2005 - 2019.

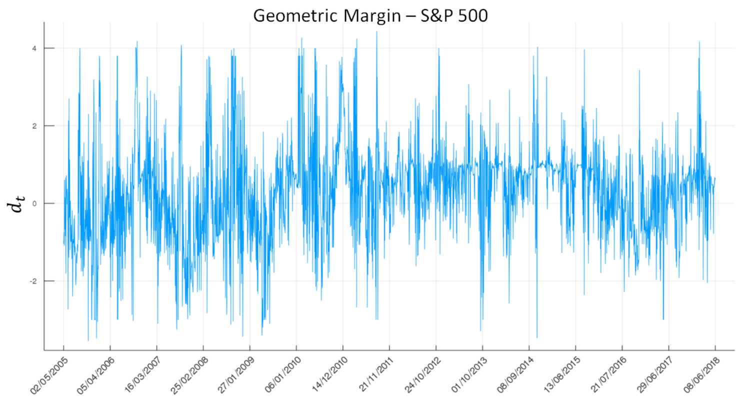

Figure C.3: Temporal evolution of S&P 500 geometric margin in the period 2005 - 2019.

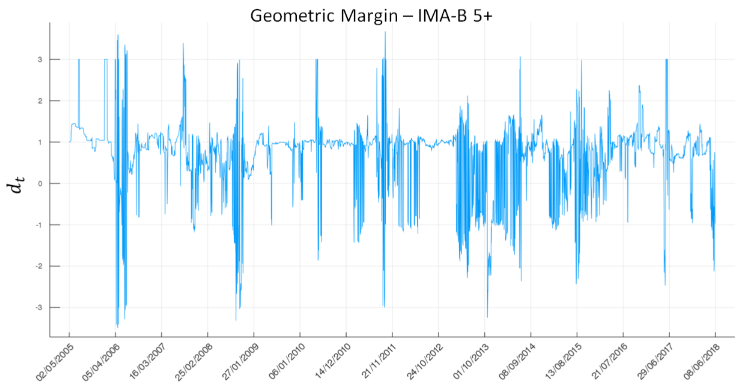

Figure C.4: Temporal evolution of IMA-B 5+ 500 geometric margin in the period 2005 - 2019.

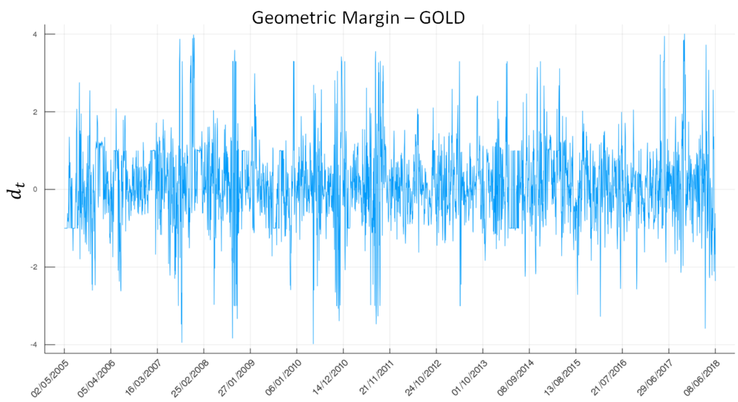

Figure C.5: Temporal evolution of GOLD geometric margin in the period 2005 - 2019.

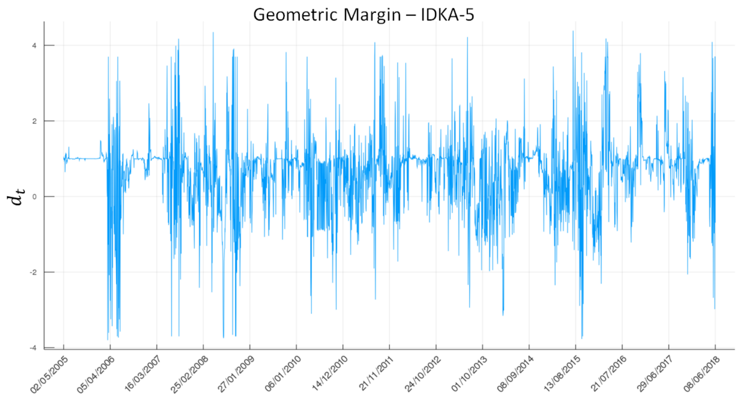

Figure C.6: Temporal evolution of IDKA-5 geometric margin in the period 2005 - 2019.

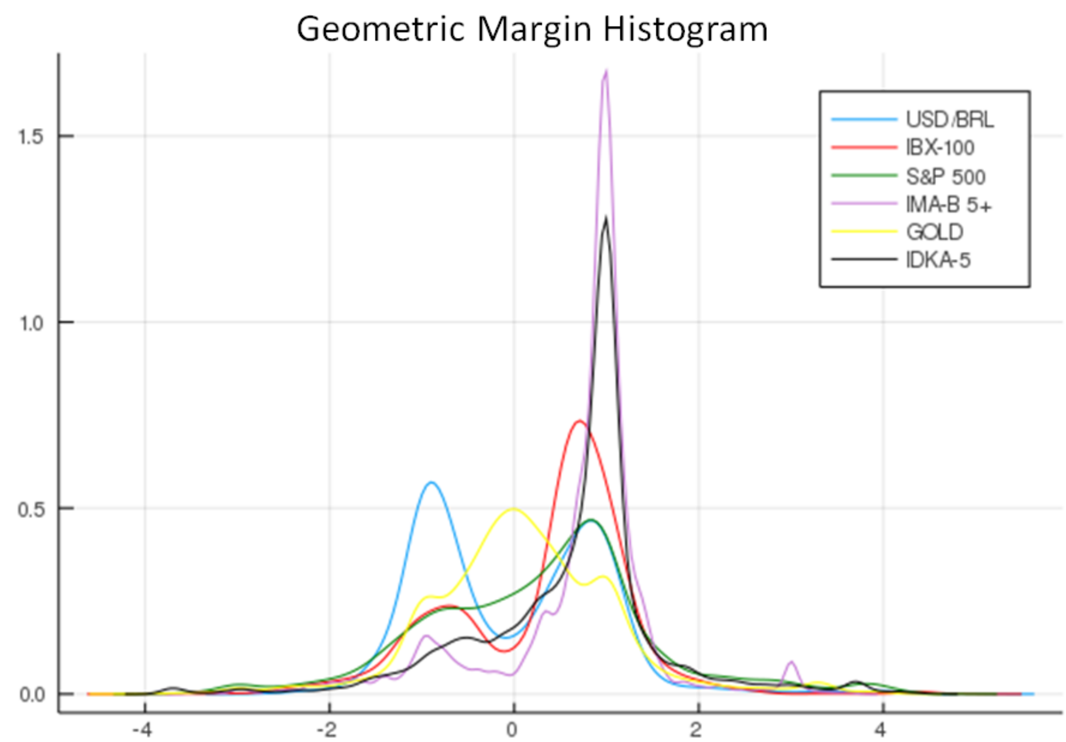

Figure C.7: Geometric Margin Histogram for all assets.