GABRIEL DE SOUZA OLIVA
VINICIUS AKIRA IMAIZUMI

INVESTMENT STRATEGIES BASED ON FORECASTING DIRECTION OF CHANGE

São Paulo

GABRIEL DE SOUZA OLIVA
VINICIUS AKIRA IMAIZUMI

INVESTMENT STRATEGIES BASED ON FORECASTING DIRECTION OF CHANGE

Work presented to the Escola Politécnica da Universidade de São Paulo in order to obtain the title of Engenheiro de Computação e Engenheiro Eletricista com Ênfase em Computação.

São Paulo

GABRIEL DE SOUZA OLIVA
VINICIUS AKIRA IMAIZUMI

INVESTMENT STRATEGIES BASED ON FORECASTING DIRECTION OF CHANGE

Work presented to the Escola Politécnica da Universidade de São Paulo in order to obtain the title of Engenheiro de Computação e Engenheiro Eletricista com Ênfase em Computação.

Advisor:

Pedro Luiz Pizzigatti Corrêa

São Paulo

ABSTRACT

This paper describes the process of developing a quantitative investment strategy from scratch. Our goal was to implement it by predicting the direction of market changes. We start with extensive research on predicting market direction, different prediction models and work on their implementation. We focus our work on exploring the advantages of polymodels, an ensemble model composed of univariate forecasting models. It is used to predict the volatility of major American and Brazilian stock indices. This predicted volatility is then used to estimate the probability that the market will have a positive return over a given time horizon. Once the forecasting model is implemented, we build an investment strategy using the inputs from the model's prediction and test whether it has the potential to outperform the market. Our results are quite satisfactory, as we were able to find profitable out-of-sample strategies that outperformed the market benchmark.

Keywords – quantitative investments, polymodel, volatility forecasting, market direction, investment strategies

RESUMO

Este trabalho descreve o processo de desenvolvimento de uma estratégia quantitativa de investimento a partir do zero. Nosso objetivo era implementá-la prevendo a direção das variações do mercado. Começamos fazendo uma extensa pesquisa sobre a previsão da direção do mercado, diferentes modelos de previsão e trabalhamos em sua implementação. Concentramos nosso trabalho na exploração das vantagens de um modelo conhecido como *polymodels*, que consiste numa composição de modelos de previsão univariados. Ele é usado para prever a volatilidade dos principais índices de ações americano e brasileiro. Essa volatilidade prevista é então usada para estimar a probabilidade de o mercado ter um retorno positivo em um determinado horizonte de tempo. Depois que o modelo de previsão é implementado, criamos uma estratégia de investimento utilizando os dados da previsão do modelo e testamos se ela tem o potencial de superar o desempenho do indice de referencia do mercado. Nossos resultados são bastante satisfatórios, conseguimos encontrar estratégias lucrativas *out of sample* que superaram o desempenho do benchmark de mercado.

Palavras-chave – investimentos quantitativos, polymodel, previsões de volatilidade, direção de mercado, estratégias de investimento

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 Motivation

In the dynamic and challenging world of financial markets, the search for profitable investment strategies remains a primary objective for both individual and institutional investors. With the advancement of computational techniques and the availability of vast amounts of financial data, quantitative investment strategies have gained significant popularity in recent years. In this dynamic context, this work embarks on a comprehensive journey, delving into the implementation process of a quantitative investment strategy aimed at predicting the direction of market changes, from initial theory to final backtesting. Our research seeks to answer the key question: Can we build a successful investment strategy based on predicting the direction of market change?

1.2 Objectives and literature

We start by analyzing the predictability of the market. A substantial body of research has explored various quantitative investment strategies, ranging from simple moving averages to sophisticated machine learning algorithms. Researchers have primarily focused on predicting directional trends in excess returns, as highlighted by Nyberg's [1] study on directional forecasting. Many of them have documented that only the direction of excess returns is predictable. According to Christoffersen and Diebold [2], while there's an extensive attempt to predict market returns, their random behavior makes them very difficult and perhaps impossible to predict. It's said that conditional mean independence is considered a good approximation to the dynamics of asset returns, and therefore their

unpredictability is assumed. However, when asset returns are decomposed into the product of return sign and return module, as in the equation below, it is shown that these terms are individually predictable. Later, the same authors attempt, with some success, to produce direction-of-change forecasts useful for market timing [3].

$$r_t = sign(r_t) \cdot |r_t|$$

Their approach had the goal of estimating the probability of future positive returns - the market direction - and had expected market volatility and expected conditional returns as key inputs. In order to extend this model, we will focus our research on volatility forecasting.

According to Campbell [4], changing market volatility affects the expected returns of all assets, and measuring variations in volatility can help explain changes in expected returns. In his work, he rejects the hypothesis that volatility changes are unpredictable on a daily basis, but he doesn't succeed in finding a trading strategy that generates economic profits. Hansen and Lunde [5] compare GARCH(1,1) to other ARCH model variations for predicting daily volatility. Their conclusion is that few models outperform GARCH and the difference in performance is very small.

In this paper, we will predict volatility by implementing a new model known as polymodels. The theory of polymodels was first introduced by Raphael Douady in his paper on StressVaR [6], a new risk indicator. The polymodel consists of a group of so-called elementary models - univariate models - that share the same objective. Most of its applications so far have been related to risk. In another article, Douady uses nonlinear polymodels to manage the downside of strategies [7]. Thomas Barrau and Douady [8] provide details on the conception and the main advantages of using polymodels, among them the absence of multicollinearity problems and the possibility of having different time series with different lengths. The flexibility of polymodels allows them to be used in different ways, especially when calculating the final output of the model and allowing more interpretability of the prediction process. Different approaches can lead to different

aggregation methods for different interpretations [8]. Wang et al. [9] provided an updated review of the extensive literature on forecast combinations, such as some averaging and regression methods that we will use in this paper.

In developing an investment strategy, it is critical to have an allocation system that determines how much to invest in a given asset. Many studies have been done on calibration and position sizing for portfolio strategies. Markowitz's 1952 paper on portfolio selection [10] introduced the concept of mean-variance optimization, which has become a cornerstone of modern portfolio theory and influenced the development of position sizing techniques. Markowitz's work has paved the way for many practical applications of position sizing, such as risk parity, minimum variance, and maximum diversification strategies, which aim to achieve optimal portfolio allocations while effectively managing risk. Meyer et al. [11] provide a study that analyzes and compares various methods commonly used by portfolio managers, including the Kelly criterion, all-or-nothing betting, and other sizing methods. In this paper, we will use these sizing techniques to build an investment strategy from our predictions.

Another important step in implementing our investment strategy was backtesting, which involves testing a trading strategy against historical market data to evaluate its potential performance. When working with financial time series, the data length isn't usually very long, so in order to get the most out of it, it's important to cross-validate when training a model. However, we must be careful to avoid data leakage during this process. Leakage occurs when the training set contains information that also appears in the test set. One reason why k-fold CV fails in finance is that it cannot be assumed that the observations come from an IID process [12]. Another potential problem in backtesting is data snooping, which refers to any statistical procedure that uses sample data with the with the intention of improving the reported results. Therefore, we use the cleaned cross-validation [13].

The main contributions of this paper lie in the development of the prediction model and the predictability of the direction of return changes. First, we extend the model

proposed by Christoffersen and Diebold [2] by implementing a new volatility prediction model from scratch. While the authors use GARCH models to predict volatility, we will use other techniques to better predict volatility. Second, this work will be the first to use the polymodels theory [8] to predict the direction of market changes. The choice of using polymodels theory to predict market direction represents a departure from traditional models that rely on single forecasting techniques. By using polymodels, we introduce a more versatile and adaptive approach that can capture intricate patterns in market dynamics.

This work is organized as follows. In Section 2, we focus on the main theory of our model to estimate the probability of the direction of market change. In section 3, we go deeper into the theoretical concepts of polymodels and propose a new approach to it. We also have a review in other volatility prediction models and compare them with the polymodels. The next section details the methodology of this work, how we manage the data, implementation of the polymodels and backtesting. Section 5 presents and discusses the main results of our strategy. Finally, we present the conclusions of this work.

2   MARKET DIRECTION AND STRATEGY

In this work, we propose to extend the implementation proposed by Christoffersen and Diebold [2]. In their work, it was revealed a direct connection between asset return volatility and asset return sign, which we call direction of change. While the authors use GARCH models to predict volatility, we will implement and make use of the polymodels with the same goal.

2.1   Probability of positive returns

Let R_t be a series of returns and Ω_t the information set available at time t. We define the probability of a positive return as $Pr[R_t > 0]$. The conditional mean and variance are denoted respectively as $\mu_{t+1|t}$ and $\sigma^2_{t+1|t}$. If $Pr[R_{t+1} > 0|\Omega_t]$ varies with Ω_t, exists conditional dependence and the return series is sign predictable. We can denote a generic distribution of the conditional returns as:

$$R_{t+1}|\Omega_t \sim (\mu, \sigma^2_{t+1|t})$$

Then the conditional probability of positive return can be written as:

$$Pr[R_{t+1} > 0|\Omega_t] = 1 - Pr(R_{t+1} \leq 0|\Omega_t)$$

$$= 1 - F\left(\frac{-\mu}{\sigma_{t+1|t}}\right)$$

where F is the distribution function of the standardized return $(R_{t+1} - \mu)/\sigma_{t+1|t}$.

In this paper, we are going to use the same way:

$$Pr(R_{t+1} > 0|\Omega_t) = 1 - F\left(\frac{-\mu_{t+1|t}}{\sigma_{t+1|t}}\right)$$

We can see that with the equation above, we can now calculate the probability of having positive returns in function of the conditional mean of expected returns in $t+1|t$ and of the expected future volatility at the same time defined by $\sigma_{t+1|t}$. The next step is to model and predict these two components of the equation. Using the historical data, the expected conditional return can be computed on a regression in function of the volatility. For that, we are going to fit the following estimator for our direction model:

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 log(\hat{\sigma}_t) + \hat{\beta}_2[log(\hat{\sigma}_t)]^2 + \hat{\beta}_3[log(\hat{\sigma}_t)]^3, t = 1, ..., k$$

where k is the length of available data at time k. With this estimator we can estimate $\hat{\mu}_{t+1|t}$ in function of $\hat{\sigma}_{t+1|t}$. Following that, we are going to estimate the future volatility $(\hat{\sigma}_{t+1|t})$ in order to estimate the probability of positive returns, $Pr(R_{t+1} > 0|\Omega_t)$. The polymodels are going to be used in order to predict volatility in different time frames. This process is going to be explained in detail in the methodology section.

2.2   Position sizing

Given that now we have the probability of having positive returns for a given market, it is important to have a strategy in order to consistently profit from this information. Our strategy will be executed in function of different position sizing methods. Their output corresponds to the percentage in which the strategy will be allocated to the correspondent asset. These methods are going to be explained below:

Model confidence

In this strategy we will be positioned in the selected asset only if the probability of positive

returns is bigger than 50%.

$$\begin{cases} f(p) = p, if\, p > 0.5 \\ f(p) = 0, otherwise \end{cases}$$

where $f(p)$ is the position size and p is the probability of positive returns.

All or nothing

In this strategy we will be fully positioned in the selected asset if the probability of positive returns is bigger than 50%.

$$\begin{cases} f(p) = 1, if\, p > 0.5 \\ f(p) = 0, otherwise \end{cases}$$

Full long and short

In this strategy we take a binary position: fully long on the asset if the positive returns probability is over 50%, otherwise the position is fully short.

$$\begin{cases} f(p) = 1, if\, p > 0.5 \\ f(p) = -1, otherwise \end{cases}$$

NCDF

Lopez de Prado [13] described this position normal cumulative distribution function sizing method. In this methodology the following H_0 hypothesis is tested: $p[x = 1] = \frac{1}{2}$. In case of probability of positive returns smaller than 50%, we are full short position.

$$\begin{cases} f(p) = 2Z[z] - 1, if\, p > 0.5 \\ f(p) = -1, otherwise \end{cases}$$

$$z = \frac{p - 0.5}{\sqrt{p(1 - p)}} \sim Z$$

Kelly criterion

Presented by Kelly in 1956 [14], is a mathematical formula relating to the long-term growth of capital developed.

$$f(p) = 2p - 1$$

3   PREDICTION MODELS

In this section, we are going to dive into the main prediction model used in this paper: the polymodel. We are also going to have a short review on the random forest algorithm, since it's going to be a reference model in which we are going to compare the performance along with the polymodels. To assure the superiority of the polymodels, we are going to make some tests comparing its performance against other known models for volatility prediction.

3.1   Random Forest

The Random Forest algorithm, introduced by Leo Breiman and Adele Cutler in 2001 [15], is a versatile and widely used ensemble learning method for both classification and regression tasks. It belongs to the family of decision tree-based algorithms and is known for its robustness, accuracy, and resistance to overfitting. This section provides an overview of the Random Forest algorithm, its key features, and its applications within prediction models.

3.1.1   Algorithm Overview

The Random Forest algorithm starts with a training set. It creates several subsets of this data using bootstrap sampling, and constructs a decision tree for each subset. These trees are diversified by randomly selecting features at each node. During prediction, each tree provides its output through voting (for classification) or averaging (for regression). This ensemble of trees reduces overfitting, increases accuracy, and makes Random Forest

a robust machine learning algorithm. This process is illustrated in the figure below.
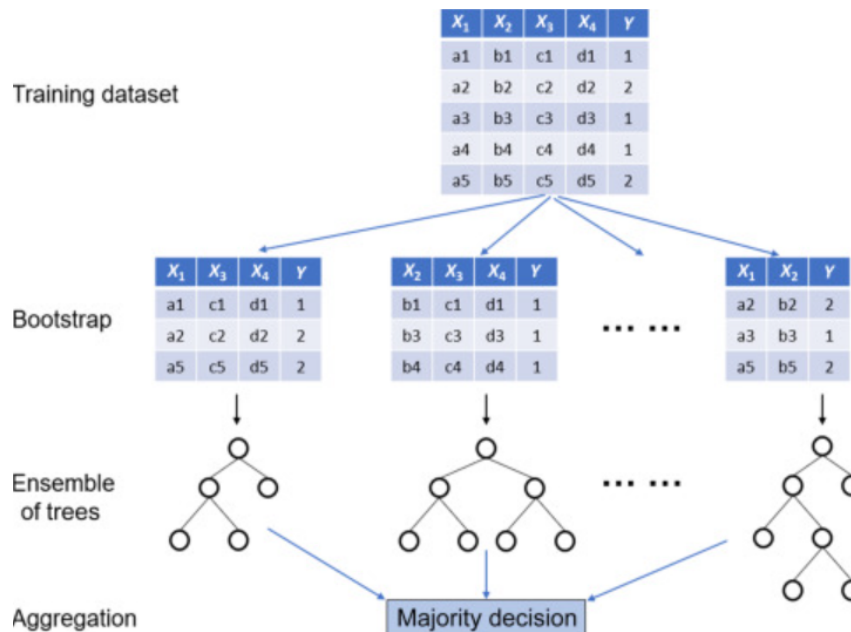
Figure 1: Random forest functioning diagram

3.1.2 Key Features

Random Forests offer several key features that contribute to their popularity and effectiveness:

- **Ensemble Diversity:** By using bootstrapped samples and random feature selection, Random Forests create a diverse set of decision trees, which reduces overfitting and improves generalization.

- **Robustness:** Random Forests are less sensitive to noisy or irrelevant features in the dataset, making them suitable for handling high-dimensional data.

- **Out-of-Bag (OOB) Estimation:** The bootstrapped samples mean that around one-third of the data is not used to train each tree. This "out-of-bag" data can be used for validation and estimating the model's performance without the need for a separate validation set.

- **Variable Importance:** The algorithm provides a measure of feature importance based on how much a feature improves the predictive performance of the model.

This information can be used for feature selection and understanding the underlying relationships in the data.

3.1.3 Applications

Random Forests have found applications in finance, including:

- Predicting stock prices

- Credit risk assessment

- Fraud detection

In conclusion, the Random Forest algorithm's ability to create diverse decision trees through bootstrapped sampling and random feature selection makes it a powerful tool for prediction modeling. Its robustness, accuracy, and applicability to a wide range of domains have contributed to its popularity in both academic research and practical applications.

3.2 Polymodels

The polymodels theory was first introduced by Raphael Douady [6] and it consists in a collection of nonlinear single factor models that represents the same target variable. These univariate models are called "elementary models" and they compose the polymodel providing different perspectives from different sets of data towards a single target. The equations set below represents a general structure of a polymodel.

$$
\begin{cases}
Y = \varphi_1(X_1) \\
Y = \varphi_2(X_2) \\
... \\
Y = \varphi_n(X_n)
\end{cases}
$$

For a dataset composed of n features, the polymodel is going to be composed of n elementary models φ with the same target Y. For the final output of the model, it's necessary

to implement an aggregation method.

3.2.1 Main advantages

The polymodel presents some advantages in relation to other models in general. They are consequence of its univariate structure. Some of these advantages are described below.

Overfitting reduction

As the model is composed of univariate models, they are independent between themselves and do not influence each other. It avoids the common overfitting present in multivariated models.

Precision increase

When working with univariate models, it is possible to work with non-linear models for each predictor. As many times the relation between the variable and the target is non-linear it promotes an increase in the precision.

Robustness

Usually multivariate models like linear regression present low stability in its coefficients in cause of multicollinearity. It is most noticeable when a new feature is added and usually there's a large change in the value of these coefficients.

Flexibility

One of the main advantage of the polymodel is its flexibility concerning the elementary models. The implementation of these models can be done according to the data. They can be build from any other model from linear regressions to neural networks.

Different series lengths

As each of the components of the polymodel is an elementary model, during the fitting process, each model is fitted using only its corresponding feature. As a consequence, the

series are independent and they can have different lengths. Therefore, the whole model can be fitted without losing data even if a feature has more samples than other.

3.2.2   The original elementary model

The elementary model aims to represent the target in function of a single variable. Thomas Barrau and Raphael Douady [8] implement this model in what they call LNLM - Linear non-linear model.

$$Y = LNLM(X)$$

This model is composed in two parts, the first one being a non-linear model and the second a linear model. The representation below depicts the structure of the LNLM model.

$$LNLM(X) = \bar{y} + \mu \sum_{h=1}^{u} \hat{\beta}_h^{NonLin} H_h(X) + (1 - \mu)\hat{\beta}^{Lin} X + \epsilon$$

μ is a parameter to control the linearity and non-linearity weights. And in their methodology, as well as in [16], $H(X)$ is an Hermite polynomial.

3.2.3   The hybrid elementary model

In this work, it was developed a new version of the Polymodel, specially concerning the elementary model. During the implementation process, which will be better explained in the Methodology section, the elementary model had 3 possible methods: linear, non-linear and hybrid.

Linear Model

In this case, the elementary model is composed simply by a linear model. In this paper, the only linear implemented method was a linear regression. As it's an univariate model, the model is simply defined as

$$Y = \beta_0 + \beta(X)$$

and the Polymodel is composed by n linear regressions.

Non-linear model

The non-linear approach consisted in setting the elementary models as non-linear models. Two models were tested: Polynomial regression and extreme learning machines. The former consisted in a simple 5 degrees polynomial regression and the latter corresponds to a single-hidden layer feedforward neural networks that provide good generalization performance at extremely fast learning speed [17]. Unlike traditional methods that involve iterative parameter tuning, ELMs adopt a feedforward neural network architecture with a single hidden layer, where the weights connecting the input layer to the hidden layer are randomly initialized and fixed. The crucial innovation lies in the training process, as only the weights connecting the hidden layer to the output layer are adjusted using a single-pass learning algorithm. In literature, it also shows that these models can outperform support vector machines in both classification and regression applications [18].

Hybrid model

The hybrid model was implemented in order to capture both the linear and non-linear behavior of the data. Therefore, the following univariate model was developed:

$$Y = f(X) + \epsilon_1$$

$$\epsilon_1 = g(X) + \epsilon_2$$

In the first equation, we have the $f(X)$ is a linear model fitting the feature with the Polymodel target Y. Following that, a non-linear model $g(X)$ fits the residuals of $f(X)$.

3.2.4   Aggregating predictions

For a dataset composed of n features, the corresponding Polymodel will also have n elementary models. As consequence, when predicting the target, there will be n different predictions coming from each of these individual models. One may wonder what would

be the final prediction. As mentioned by Barrau in [8], the Polymodel offers a large representation of the links that each variable has with our target. Many approaches are proposed like maxima, minima or percentiles.

For this paper, our goal was to find the best aggregation method for the output of the elementary models of the polymodel, giving us a final and unique volatility prediction value. Initially, some approaches were tested like weighting each elementary model forecast by the mean squared error of the model. Other approaches using AIC (Akaike's Information Criterion) and BIC (Bayesian Information Criterion) as in the work of Kolassa [19] were tested. To aggregate polymodels predictions, we also tested an entropy method, using Shannon's information measure and innovations from the models as in [20]. However, these methods were not forward used because their results were not satisfactory and their complexity.

The work of Wang et al.[9] mentioned before provided an extensive review of the literature on forecast combinations. We decided to implement several aggregation methods within the polymodel. The simplest one was the median value of the predictions. Other methods tested to aggregate were a linear regression and a random forest algorithm.

Implementing an aggregation model inside the polymodel required the elaboration of a more complex structure to not have any overfitting or data leakage. Since we had to fit another model while fitting the polymodel, the solution was to implement a nested cross-validation. It is a technique used in machine learning to assess a model's performance and hyperparameter tuning simultaneously. It involves two layers of cross-validation: the outer loop splits the data into training and testing sets multiple times to estimate the model's overall performance, while the inner loop is used for hyperparameter optimization (which in this case was the aggregation model fitting) by splitting the training data into training and validation sets. This process was used to avoid overfitting the aggregation process and it's shown below.
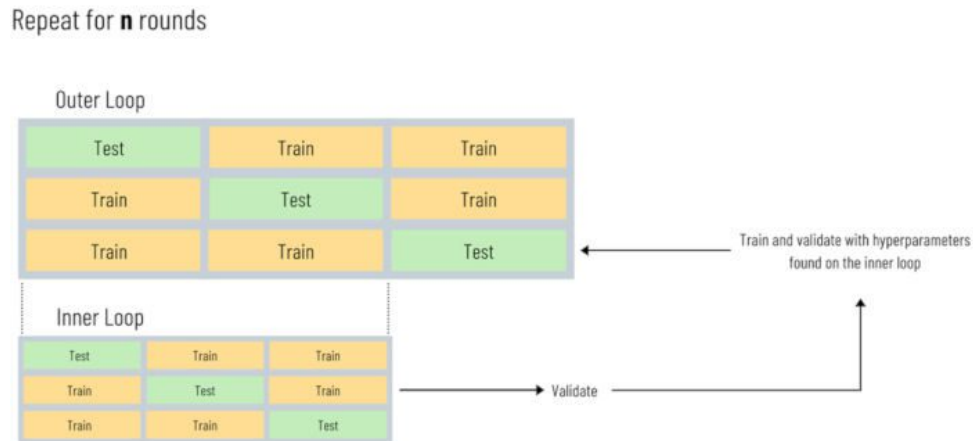
Figure 2: Nested cross-validation scheme

3.3   Polymodels versus other models

In order to use the polymodels in this paper for volatility forecasting, it was important to assure its superiority against other models. Therefore, we compared the volatility forecast among some known models for volatility prediction. Based on the work of Karasan [21] that implemented various models with the same goal, we tested the following models: ARCH, GARCH, SVR-GARCH, Polymodel and Random Forest.

3.3.1   VIX

The VIX, or CBOE Volatility Index [22], is a widely recognized financial gauge that reflects market participants' expectations of future volatility in stock prices. Commonly referred to as the "fear index," the VIX is calculated based on the prices of options contracts for the S&P 500 index. A higher VIX value suggests heightened market uncertainty and potential for larger price swings, while a lower value indicates calmer market conditions. As a valuable tool for investors and analysts, the VIX provides insights into market sentiment and risk perception, aiding in decision-making and risk management strategies. In this case, we use the VIX as the predicted volatility.

3.3.2 ARCH

One of the first attempts to model volatility by Engle in 1982 [23], the ARCH(p) model has the form:

$$\sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_k (r_{t-k})^2$$

where $r_t = \sigma_t \epsilon_t$.

ARCH model is a univariate model and it is based on the historical asset returns.

3.3.3 GARCH

One of the most known volatility prediction models [5], it's an extension of the ARCH model, incorporating lagged conditional variance. In this enhancement, ARCH is elevated by the inclusion of "p" delayed conditional variance values. Consequently, the GARCH model transforms into a multivariate entity, representing autoregressive moving average structures for conditional variance. This structure is defined by incorporating "p" lagged squared returns and "q" lagged conditional variance components.

$$\sigma_t^2 = \omega + \sum_{k=1}^{q} \alpha_k r_{t-k}^2 + \sum_{k=1}^{p} \beta_k \sigma_{t-k}^2$$

3.3.4 SVR-GARCH

Support Vector Machines (SVMs) are powerful supervised machine learning algorithms primarily used for classification and regression tasks. SVMs work by finding a hyperplane that best separates different classes of data points while maximizing the margin between them.

The core idea of SVMs involves transforming the original data into a higher-dimensional space using a mathematical function called a kernel. In this higher-dimensional space, the SVM seeks to find the hyperplane that maximizes the distance between the closest

data points of different classes, known as support vectors.

The idea of the SVR-GARCH proposed by Karasan [24] is to fit a SVR model with the components of a GARCH model. This means that the SVR is fitted with squared lagged returns and squared lagged variance.

3.3.5   Comparison metric: Mean Absolute Scaled Error (MASE)

In order to score and compare different models we make use of an useful metric proposed by Hyndman in 2006 [25] called Mean Absolute Scaled Error. It serves as an evaluation tool for assessing the accuracy of a forecasting model by comparing its performance to that of a simple naive forecasting approach. MASE addresses some of the limitations of other metrics like Mean Absolute Error (MAE) by providing a normalized measure of error that is not affected by the scale of the data or the magnitude of errors.

The formula for calculating MASE involves dividing the mean absolute error of the forecasted values by the mean absolute error of the naive forecast (typically the one-step ahead forecast). The result is a ratio that indicates how much better or worse the forecasting model performs in comparison to the simple naive approach. A MASE value less than 1 suggests that the model's predictions are better than the naive method, whereas a value greater than 1 indicates that the model is performing worse than the naive approach.

3.3.6   Results comparison

The mentioned models were implemented in addition to the polymodels and random forest. Following that, these models were fitted and generated volatility predictions for different horizons of time (5,10,22,66 days). As these two models have a bigger flexibility concerning the input variables, they had as inputs the same features that are going to be mentioned in the methodology section.

In order to verify if the polymodels would perform any better than the other models,

the MASE score was calculated as well as the mean squared error of the predictions. The only methods that presented a MASE score lower than 1, besides having a lower MSE, were the polymodels and the random forest model, meaning that they beat a naive forecast model. The other models presented results indicating that their performance for the selected forecast horizons weren't better than a naive model. Therefore, this paper is going to be continued using the polymodels, as expected and the random forest for comparison.

4   METHODOLOGY

The previous sections provided an overview of some theory that is going to be implemented in this paper. The goal is to construct a profitable strategy by forecasting the direction of change.

To achieve our ultimate goal, we have a two step identification strategy. The first one involves a comprehensive analysis and testing of the polymodels structure to predict the market volatility and subsequently estimate the probability of positive returns. Our predicting architecture is shown in the following diagram.
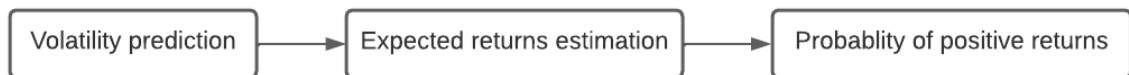
Figure 3: Market direction prediction architecture

In the second step, we will apply our forecasts to formulate an optimal strategy that surpasses the benchmark's returns. This step involves testing different sizing strategies to measure our exposition to the selected index. We then backtest the strategy and test if it is statistically better than the selected benchmark.

Our methodology follow the sequence: initially we collect and treat the data, then it is inserted in a polymodel in order to predict the expected volatility. The next step is to apply this expected volatility to calculate the probability of positive returns. These probabilities are then used by a sizing function that will set a portfolio position along the time. Once the strategy is settled, we will backtest it and execute statistical tests to

confirm its validity and robustness.

4.1   Data

The data used in this paper consists in a group of financial indicators. Some of the features are derived from other time series based on economical concepts. As the focus of this work is the implementation of a model and a strategy, we will not cover the chosen indicators individually. Most of this data corresponds to tradable instruments and therefore are available on a daily frequency. These indicators are mostly related to market surprises and uncertainties, credit spreads and volatility signals.

This data was obtained mostly from Bloomberg terminal and other financial data platforms. Only stationary series were selected to be used in the model. The final dataset is composed by 66 features and they have a date range from 2004 to 2021 with over 4000 rows. In the appendix, we present plots to depict the different features that compose the used dataset.

4.2   Volatility prediction

The first step in our methodology to obtain a direction prediction is to forecast the market volatility. In order to achieve that, this paper is going to make direct use of the polymodels.

We are going to make volatility forecasts for the S&P500 index - one of the most commonly followed US equity indices - and the Ibovespa, the Brazilian stocks index. Besides that, for each index, the predictions are going to be made for different time horizons: 5, 10, 22 and 66 days. The model's target volatility is defined as the log of the standard deviation of returns calculated in a rolling window of the horizon we are predicting. We also annualize this value and shift it by the horizon size. This means that

at time t, the target will be the annualized volatility of the returns H days ahead.

$$target_t = log(std([r_{t+H}, r_t]) \cdot \sqrt{252})$$
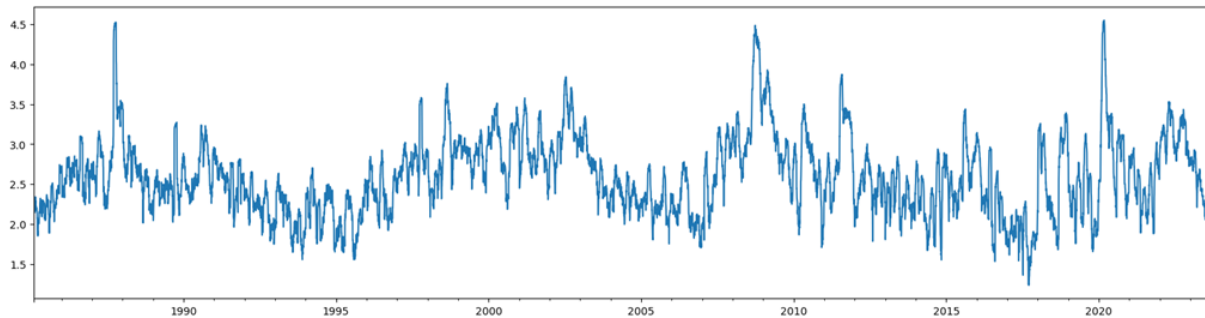
Figure 4: Target visualization for S&P500 22 days volatility

In the previous section, it was presented the hybrid version of the polymodel. In order to find the best useful volatility prediction, several possible combinations of linear and non linear models were used to compose the elementary model. Through this work, 11 different combinations were created and they represent the 3 types of elementary models: linear, nonlinear and hybrid. For the linear model combinations the Linear Regression model was tested both with and without intercept fitting. Concerning the nonlinear model, tests were made using the Extreme Learning Machine (ELM) model and polynomial regressions of 5 degrees. The different compositions are detailed in the table below.

Combination	Elementary type	Linear Model	Non Linear Model	Combiner Model
comb1	linear	LinearRegression()	None	Median
comb2	linear	LinearRegression(fit_intercept=False)	None	Median
comb3	linear	LinearRegression()	None	Random Forest
comb4	nonlinear	None	ELM	Median
comb5	nonlinear	None	ELM	Random Forest
comb6	hybrid	LinearRegression()	ELM	Median
comb7	hybrid	LinearRegression()	ELM	Random Forest
comb8	nonlinear	None	Polynomial Regression	Median
comb9	nonlinear	None	Polynomial Regression	Random Forest
comb10	hybrid	LinearRegression()	Polynomial Regression	Median
comb11	hybrid	LinearRegression()	Polynomial Regression	Random Forest

Table 1: Polymodels tested combinations

4.3   Data snooping and cross-validation

Once set the combinations to be tested, the next step is to run the models and generate the volatility predictions. When developing a quantitative model, one must be careful with the dangers of data snooping. It refers to the practice of repeatedly adjusting or optimizing a trading strategy based on historical data until it appears to yield impressive results, without considering whether those results are statistically significant or likely to hold up in future market conditions. The danger of data snooping in backtesting lies in the potential to create strategies that appear highly profitable solely due to chance patterns in historical data. Performing multiple tests increases the likelihood that some variables exceed the predetermined significance level just by luck. In general, the more inferences are made, the more likely false inferences occur [26].

The cross-validation is a relevant technique applied during this process to improve the reliability of the backtest. It is used in machine learning to assess the performance of a model by dividing the dataset into multiple subsets. The model is trained on some subsets and tested on others, allowing for a more reliable estimation of its generalization ability than a single train-test split [27]. To generate predictions out-of-sample for the whole time series and reduce the risk of over fitting and data leakage, the model fitting and prediction is done with the use of purged cross-validation. The purging removes from the training set those samples that are build with information that overlaps samples in the testing set [13].
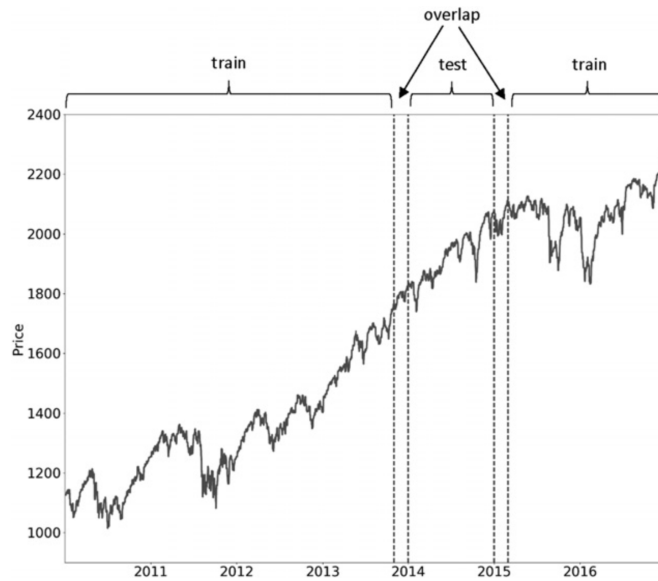
Figure 5: Purging illustration, from [13]

At each iteration of the cross validation process, the current training set passed through a transformation named Quantile Transform. It is a statistical method employed to transform a dataset's probability distribution so that the transformed values have a specified desired distribution, usually uniform or Gaussian (normal) distribution [28].

The models outputs are then re-transformed and undone the log operation by exponentializing the values. The final output is the predicted volatility on a daily frequency.

4.4   Direction prediction and backtesting

Once the volatility predictions are generated, a backtest framework is implemented as depicted below.
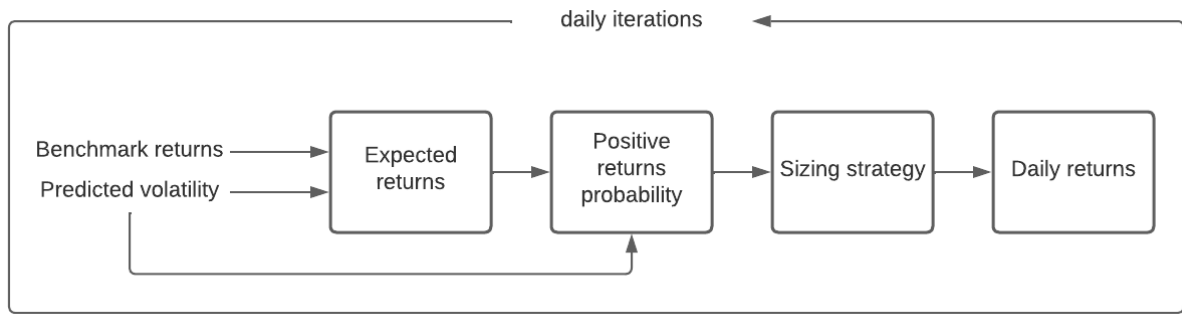
Figure 6: Backtest framework

In this process, the benchmark - S&P500 or Ibovespa - as well as the volatility forecasts are given as inputs. For each calendar day, in an iterative process, they are used to fit a linear regression as stated previously in the section 2.1:

$$\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 log(\hat{\sigma}_t) + \hat{\beta}_2 [log(\hat{\sigma}_t)]^2 + \hat{\beta}_3 [log(\hat{\sigma}_t)]^3$$

With the expected returns and the predicted volatility, the probability of positive returns is obtained from the cumulative distribution function of the standardized returns. The next step in the backtesting process is to apply these probabilities into the different sizing strategies. Their output corresponds to the percentage in which the strategy will be long or short in the correspondent index. The remaining percentage is considered as cash, with no returns.

$$allocation_\% = f(p)$$

Where $f(p)$ is the sizing function. In order to calculate the daily returns of the different strategies, this work simulates a real environment by adding costs and a delay between the investment signal and the positioning date, it represents the time necessary to investment managers to achieve the desired position in the market. This delay is set as two days, which means that an allocation defined at time t will have its returns reflected in $t + 2$. For the costs, the estimated value for trading futures of indexes was of 5 bps (0.05%) over

the turnover, which is the variation of the position size.

$$costs = \frac{0.05}{100} \cdot (allocation_t - allocation_{t-1})$$

For the matter of analyzing if this whole volatility and direction predictions, as well as the different sizing strategies have some real potential to bring profits, a simulation was made supposing that the models predict the real volatility, which means predicting the future with no error.
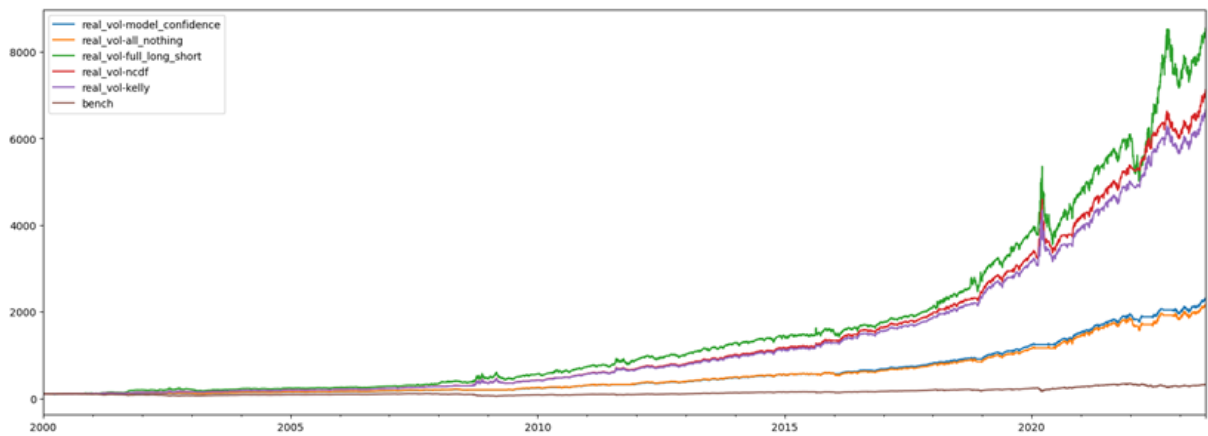
Figure 7: Real volatility prediction simulation of cumulative returns, S&P500 22 days

The simulation above depicts that the formulated strategy combining probabilities of positive returns along with a sizing method can achieve great returns potential as they perform significantly above the benchmark. We can also check the behavior of the different sizing methods through time, as shown below.
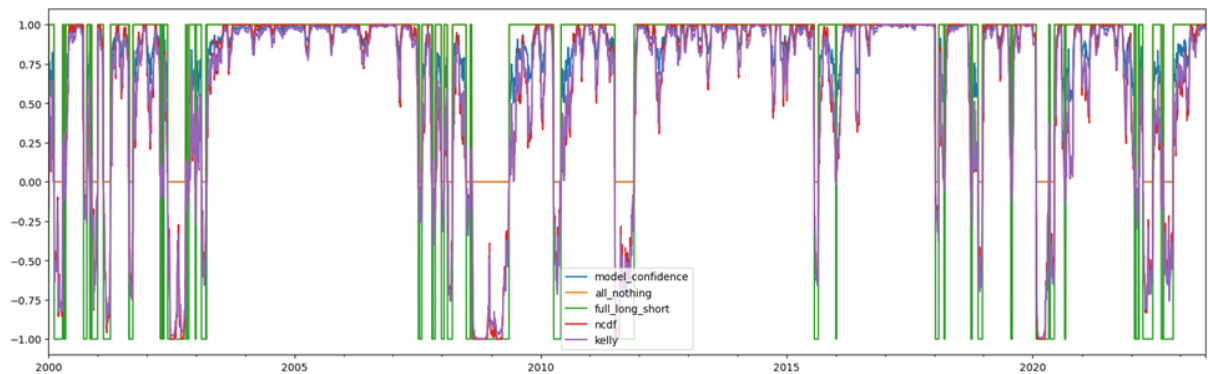
Figure 8: Real volatility prediction simulation of sizing methods, S&P500 22 days

Even though it is shown the potential of this strategy, it is totally dependent on the quality of the volatility forecasts. In this work this backtest process is run for the 11 different combinations of polymodels and for the 5 sizing strategies, totalling 55 series of returns to be analyzed.

4.5   Robustness tests - MCS

During the process of creating an investment strategy and backtesting, many results can be difficult to be compared among themselves and a consistent method to find the best strategy is required. Besides that, it is important to have trustful results and to check if the results are statistically better than the selected benchmark. Therefore, this paper makes use of a method called Model Confidence Set (MCS).

The inception of the Model Confidence Set (MCS) can be attributed to the work of Hansen et al. [29], who introduced this methodology with the aim of refining the process of selecting forecasting models. At its core, the Model Confidence Set methodology seeks to streamline the extensive array of potential models into a more manageable subset — known as a model confidence set — which is strategically designed to encompass the most accurate forecasting model with a designated level of confidence. The defining aspect of this approach lies in its adaptability to various scenarios, allowing users to tailor their criteria for the selection of the 'best' model to align with the specific demands of the given context.

The Model Confidence Set methodology's fundamental philosophy is rooted in the concept of parsimony. The goal is not to inundate decision-makers with an overwhelming assortment of models, but rather to present a concise collection that captures the most promising candidates for optimal forecasting. By employing a pre-specified level of confidence, MCS introduces a level of statistical rigor, ensuring that the selected model confidence set not only contains the best-performing model but also offers quantifiable certainty in its efficacy.

For our strategies selection, our goal was to select the best combination of polymodels based on the strategy performance from its predictions. Therefore, the MCS methodology is going to be applied for every set of combinations for each sizing methods backtests. This means that as previously mentioned, we had 5 sizing methods with 11 polymodels combinations, then we are going to create 5 different groups. Each group will contain 11 different backtests and strategies performances corresponding to each polymodel combination. The MCS will then be applied to each one of these groups in order to find the best strategy inside this group.

The criterion defined in this paper for the MCS to find the best strategy was the sortino ratio (risk-adjusted performance measure that evaluates an investment's return relative to its downside volatility). The MCS output is a value between 0 and 1 meaning the probability that the respective backtest belongs to the model confidence set of the best performing strategies.

5 RESULTS

After running the entire backtesting process, we end up with the performance statistics of 55 simulations for each selected period and index. The performance metrics of the best performing combinations are shown in the appendix of this paper.
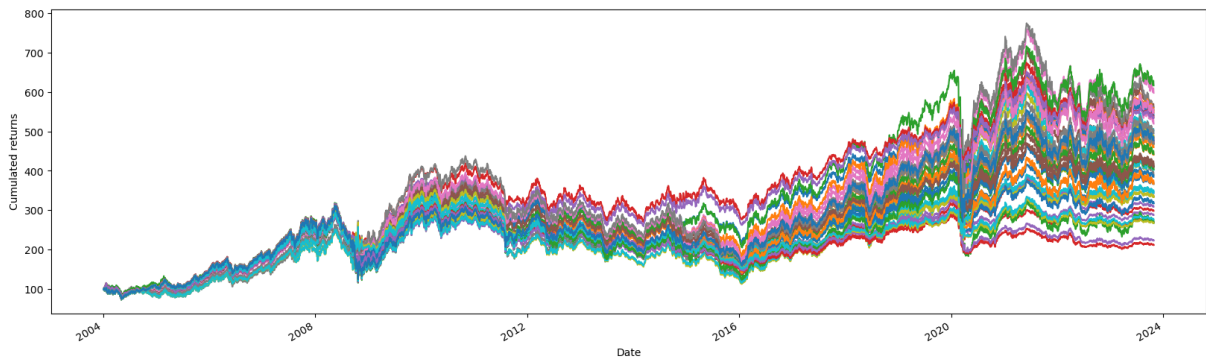
Figure 9: Cumulative returns of example simulation, Ibovespa 22 days

Some of the key metrics calculated are described below:

Volatility (ann.)

Measure of the variability of an asset's returns over time, annualized.

Skewness (ann.)

Indicates the asymmetry of a distribution of returns around its mean, annualized.

Kurtosis (ann.)

Quantifies the tail-heaviness of a distribution of returns, annualized.

Sharpe Adjusted (ann.)

Similar to the Sharpe ratio but adjusts for the skewness and kurtosis of returns, providing a more accurate risk assessment, annualized.

Tail Ratio

Measures the ratio of the average of the extreme values to the average of the non-extreme values in a distribution.

Sortino (ann.)

Evaluates the risk-adjusted return by considering only downside volatility, annualized.

Calmar (ann.)

Computes the ratio of the average annual return to the maximum drawdown, providing insight into risk and return, annualized.

Max Drawdown

Represents the maximum percentage loss from a peak to a trough in an investment's value.

Historical VaR (Value at Risk)

A quantile-based measure that estimates the potential loss at a specified confidence level, based on historical data.

Historical Expected Shortfall

Also known as Conditional Value at Risk (CVaR), it estimates the average loss beyond the VaR level, providing a more comprehensive view of potential losses based on historical data.

MCS 3Y

Model confidence set, probability of this model to belong to the model confidence set of

best models. The criterion is the rolling sortino ratio of 3 years.

We will now dive into an analysis of the results obtained for each index and period. In each analysis, we will show a graph of the best performing strategies in terms of returns and Sharpe. We will also show a table with the top 5 models (including the benchmark) in terms of the average MCS of the sizing methods for each combination. The idea of obtaining the average MCS is to find the most consistent polymodel combination where it has superior performance across most of the sizing methods. The detailed results tables with each sizing method and combination are available in the appendix of this paper.

5.1 S&P 500

5.1.1 S&P 500, 5 days

Figure 10: Best results backtest, S&P500 5 days

The table below compares the average Model Confidence Set (MCS) scores of various investment models to a benchmark, which has the highest MCS score at 0.94. This means that while we could find specific combinations of models and sizing strategies that outperformed the benchmark, on average they weren't more consistent. The benchmark emerges as the model with the most confidence in its performance. In particular, *comb8* and *comb4* show robust performance with MCS scores of 0.8 and 0.6, respectively, demonstrating substantial confidence but falling slightly short of the benchmark. In contrast, *comb1* and *comb2* lag behind with MCS scores of 0.13 and 0.1, indicating a significant divergence in confidence compared to the benchmark.

Model	MCS 3Y
bench	0,94
comb8	0,8
comb4	0,6
comb1	0,13
comb2	0,1

Table 2: Average MCS per model, S&P500 5 days

5.1.2   S&P 500, 10 days

Figure 11: Best results backtest, S&P500 10 days

The table compares the Model Confidence Set (MCS) scores of various investment models to a benchmark with an MCS score of 0.74. Interestingly, *comb4* outperforms the benchmark with an MCS score of 0.93, indicating a high level of confidence in its performance. This suggests that *comb4* - a nonlinear elementary model composed by an Extreme Learning Machine and the individual predictions aggregated by the median - may represent a strong and reliable investment strategy compared to the benchmark. In contrast, *bench* itself, despite having the second-highest MCS score at 0.74, falls short of *comb4*, implying that other models may offer more consistent and confident performance.

Furthermore, *comb5*, *comb8*, and *comb7* exhibit lower MCS scores of 0.27, 0.26, and 0.17, respectively, signaling a notable divergence in confidence compared to the benchmark.

Model	MCS 3Y
comb4	0,93
bench	0,74
comb5	0,27
comb8	0,26
comb7	0,17

Table 3: Average MCS per model, S&P500 10 days

5.1.3 S&P 500, 22 days

Figure 12: Best results backtest, S&P500 22 days

For a horizon of one month (22 days), combination 4 stands out with the best results for all sizing methods. In particular, *comb4* stands out with a perfect MCS score of 1, indicating the highest possible level of confidence in its performance. This implies that *comb4* is exceptionally reliable, potentially making it a strong candidate for consideration in the investment strategy.

In contrast, the benchmark (*bench*) has a relatively lower MCS score of 0.23, indicating a lower level of confidence compared to *comb4*. In addition, *comb5*, *comb8*, and *comb6* have MCS scores of 0.03, 0.01, and 0, respectively, indicating a significant divergence in confidence levels from the benchmark.

Model	MCS 3Y
comb4	1
bench	0,23
comb5	0,03
comb8	0,01
comb6	0

Table 4: Average MCS per model, S&P500 22 days

5.1.4   S&P 500, 66 days
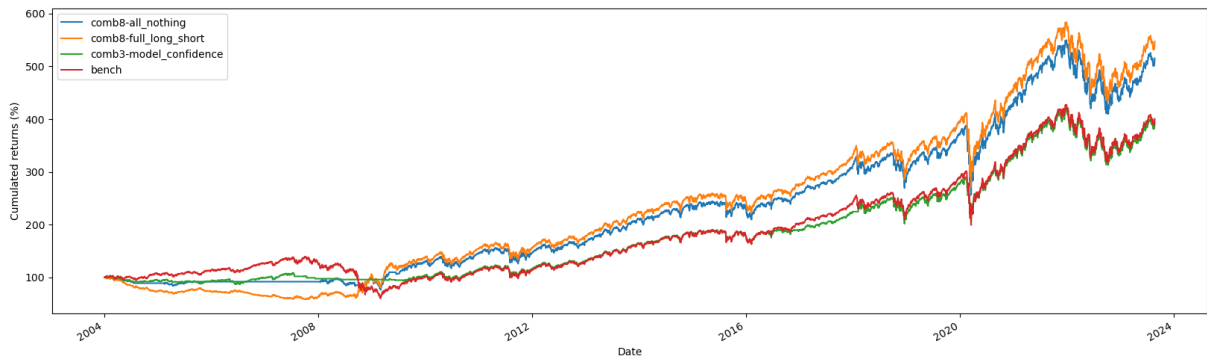
Figure 13: Best results backtest, S&P500 66 days

When working with a longer forecast horizon, the results are less consistent. While for the other periods combination 4 was consistently the best model choice, for this period we can see that the benchmark has the best score. *comb8* follows closely with an MCS score of 0.68, indicating a high level of confidence in its performance and making it a strong contender relative to the benchmark. *comb3*, with an MCS score of 0.53, shows a moderate level of confidence.

In contrast, *comb4* and *comb2* have lower MCS scores of 0.23 and 0.12 respectively, indicating a significant divergence in confidence compared to the benchmark.

Model	MCS 3Y
bench	0,7
comb8	0,68
comb3	0,53
comb4	0,23
comb2	0,12

Table 5: Average MCS per model, S&P500 66 days

5.2   Ibovespa

5.2.1   Ibovespa, 5 days
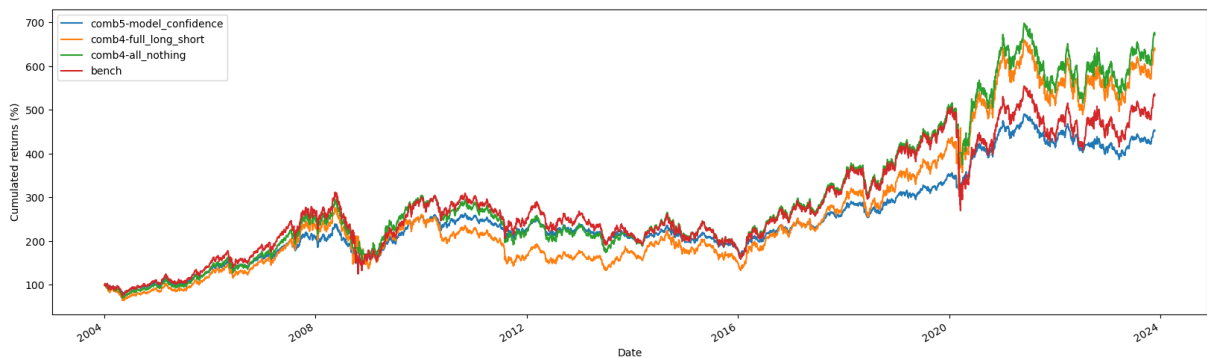
Figure 14: Best results backtest, Ibovespa 5 days

For the Brazilian index with a forecast horizon of 5 days, *comb5* emerges as the top performer with an MCS score of 0.95. This high level of confidence positions *comb5* as a standout model in the set. *comb5* is the polymodel in which the elementary model also consists of an Extreme Learning Machine but with a Random Forest as a combiner model of the individual predictions. *comb4* follows with an MCS score of 0.67, indicating a respectable level of confidence in its performance.

In contrast, the benchmark has a lower MCS score of 0.09, indicating a significant difference in confidence compared to *comb5* and *comb4*. In addition, *comb7* and *random-forest* have MCS scores of 0.08, indicating comparable confidence levels, but still lower than *comb5* and *comb4*. This analysis highlights *comb5* as a highly reliable model, out-

performing both the benchmark and other models in terms of confidence. *comb4* also stands out as a robust performer.

Model	MCS 3Y
comb5	0,95
comb4	0,67
bench	0,09
comb7	0,08
randomforest	0,08

Table 6: Average MCS per model, IBOV 5 days

5.2.2 Ibovespa, 10 days
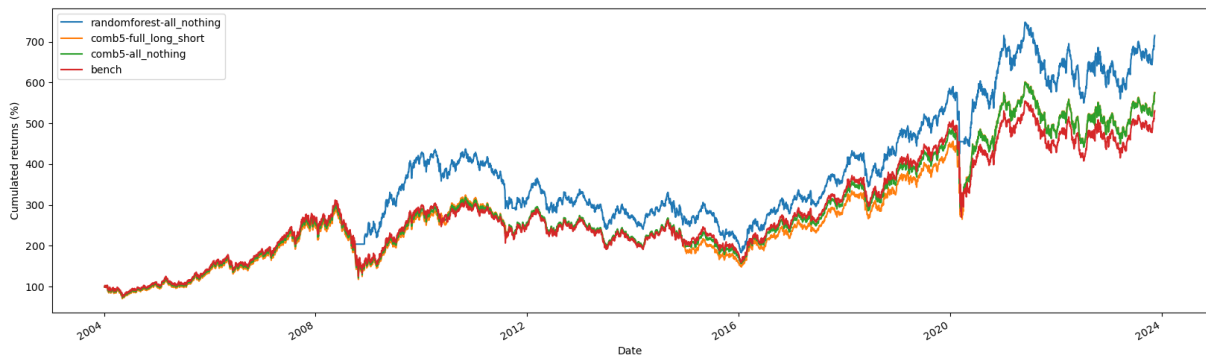
Figure 15: Best results backtest, Ibovespa 10 days

While the benchmark leads the set with an MCS score of 0.69, *randomforest* follows closely with an MCS score of 0.6, indicating a significant level of confidence in its performance. This was the only prediction horizon where we found that the Random Forest model outperformed the different polymodels.

In contrast, *comb5* has a lower MCS score of 0.37, indicating a significant difference in confidence compared to the benchmark and *randomforest*. *comb7* and *comb3* both have an MCS score of 0.19, indicating comparable confidence levels, but lower than the benchmark and *randomforest*.

This analysis highlights *randomforest* as a strong performer, closely following the

benchmark.

Model	MCS 3Y
bench	0,69
randomforest	0,6
comb5	0,37
comb7	0,19
comb3	0,19

Table 7: Average MCS per model, IBOV 10 days

5.2.3 Ibovespa, 22 days
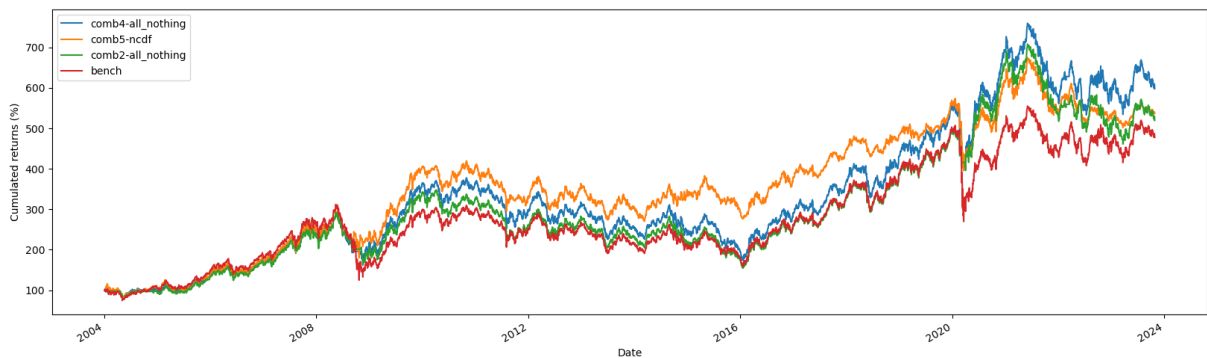
Figure 16: Best results backtest, Ibovespa 22 days

For the one month forecast horizon, *comb5* again leads the set with an MCS score of 0.82. This high confidence level suggests that *comb5* is a strong performer compared to other models and the benchmark. *comb4* follows with an MCS score of 0.65, also indicating a respectable level of confidence in its performance.

On the other hand, *comb2* has a lower MCS score of 0.42, indicating a significant difference in confidence compared to *comb5* and *comb4*. The benchmark has a much lower MCS score of 0.03, indicating a significant divergence in confidence compared to the top performing models. Additionally, *randomforest* has an MCS score of 0, indicating little to no confidence in its performance based on the model comparison.

This analysis highlights *comb5* and *comb4* as robust models that outperform both

the benchmark and other models in terms of confidence.

Model	MCS 3Y
comb5	0,82
comb4	0,65
comb2	0,42
bench	0,03
randomforest	0

Table 8: Average MCS per model, IBOV 22 days

5.2.4   Ibovespa, 66 days
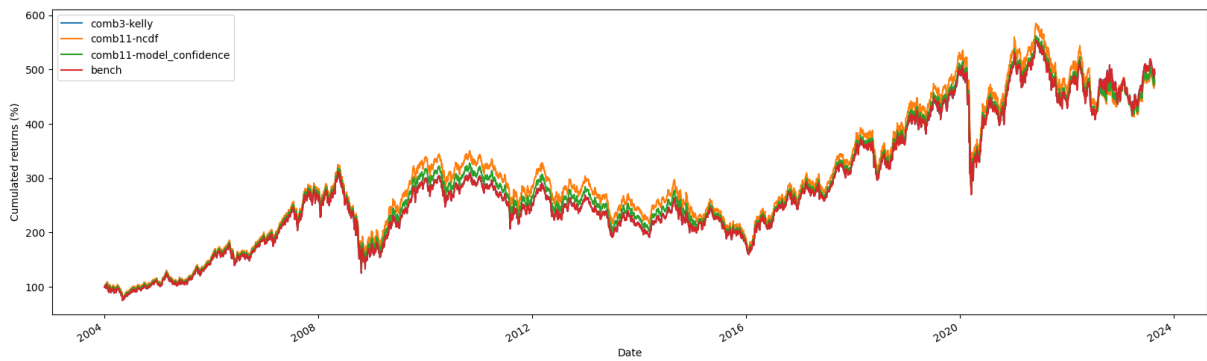
Figure 17: Best results backtest, Ibovespa 66 days

For the longer prediction horizon, we can see that the *comb3* is very close to the benchmark. In fact, when analyzing the performance table, it's remarkable that this combination has the highest MCS score for the *model confidence* and *kelly* sizing methods, while it underperforms the benchmark for the *all or nothing, full long short and NCFD* methods.

Model	MCS 3Y
comb3	0,68
bench	0,6
comb8	0,24
comb11	0,12
comb1	0,01

Table 9: Average MCS per model, IBOV 66 days

6 CONCLUSION

This thesis presents a detailed approach to developing a profitable investment strategy that predicts the direction of market changes. The research covered every stage of the process, including comprehensive analysis, strategy construction, and backtesting to ensure superior statistical results compared to the chosen benchmark. We achieved reasonably satisfactory results, as we were able to identify profitable out-of-sample strategies that outperformed the market.

We explored the approach in the academic literature that relates market volatility to future market direction. We also proposed a new version of the Polymodels ensemble model that was used to predict the volatility of stock indexes, and we focused on exploring the benefits it provided. The predicted volatility was also used to estimate the probability that the market will have a positive return over a given time horizon, improving on the framework proposed by Campbell. We also tested different sizing strategies to measure the exposure to the selected index.

Our results show that we were indeed able to identify good strategies based on our predictions. One relevant observation we were able to make was the power of the extreme learning machines. In all the horizons tested, the combinations 4 and 5 of the polymodels were the highlight. In both cases, the elementary model consisted only of an ELM and the difference between the two was the method of aggregation of the predictions, the former being the median and the latter a Random Forest model.

We could also see that our volatility predictions led to strategies that improved the Sharpe Ratio relative to the benchmark, meaning that we could get a better relationship

between returns and volatility, including costs. Since we didn't have a leveraged strategy, the main advantage of the proposed methods is a large reduction in drawdown.

This work succeeds in answering the question "Can we build a successful investment strategy based on investment strategy based on predicting the direction of market changes". We have demonstrated a systematic and rigorous process for achieving profitable results. To extend this work, several approaches are possible. One could try to improve volatility forecasting with new models or an improvement of the polymodels. On the sizing side, one could also improve it with new and more active positioning systems.

BIBLIOGRAPHY

[1] Henri Nyberg. "Forecasting the direction of the US stock market with dynamic binary probit models". In: *International Journal of Forecasting* 27.2 (2011), pp. 561–578. ISSN: 0169-2070. DOI: `https://doi.org/10.1016/j.ijforecast.2010.02.008`.

[2] Peter F. Christoffersen and Francis X. Diebold. "Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics". In: *Management Science* 52.8 (2006), pp. 1273–1287. URL: `http://www.jstor.org/stable/20110599`.

[3] Peter F. Christoffersen; Francis X. Diebold; Roberto S. Mariano; Anthony S. Tay; Yiu Kuen Tse. "Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence". In: *Journal of Financial Forecasting* 1 (2007), pp. 1–22.

[4] Campbell R. Harvey and Robert E. Whaley. "Market volatility prediction and the efficiency of the SP100 index option market". In: *Journal of Financial Econometrics* 31 (1992), pp. 43–73.

[5] Peter R. Hansen and Asger Lunde. "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?" In: *Journal of Applied Econometrics* 20.7 (2005), pp. 873–889. DOI: `https://doi.org/10.1002/jae.800`.

[6] Ilija I. Zovko Cyril Coste Raphael Douady. "The StressVaR: A New Risk Concept for Extreme Risk and Fund Allocation". In: (2009). DOI: `https://doi.org/10.3905/jai.2011.13.3.010`.

[7] Raphaël Douady. "Managing the Downside of Active and Passive Strategies: Convexity and Fragilities". In: *Journal of portfolio management* 46.1 (Nov. 2019), pp. 25–37. DOI: `10.3905/jpm.2019.1.112`. URL: `https://hal.science/hal-02488589`.

[8] Raphael Douady Thomas Barrau. *Artificial intelligence for financial markets: the polymodel approach.* Springer, 2022.

[9] Xiaoqian Wang et al. "Forecast combinations: An over 50-year review". In: *International Journal of Forecasting* (2022). ISSN: 0169-2070. DOI: `https://doi.org/10.1016/j.ijforecast.2022.11.005`.

[10] Harry Markowitz. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91. ISSN: 00221082, 15406261.

[11] Michael Meyer; Illya Barziy; Jacques Francois Joubert. "Meta-Labeling: Calibration and Position Sizing". In: *Portfolio Management Research* (2023). DOI: `https://doi.org/10.3905/jfds.2023.1.119`.

[12] Marcos López de Prado. "The 10 Reasons Most Machine Learning Funds Fail". In: *The Journal of Portfolio Management* (2018), pp. 120–133. DOI: `10.3905/jpm.2018.44.6.120`.

[13] Marcos Lopez de Prado. *Advances in Financial Machine Learning*. John Wiley & Sons, 2018.

[14] John L. Kelly. "A new interpretation of information rate". In: *IRE Trans. Inf. Theory* 2 (1956), pp. 185–189.

[15] Leo Breiman. "Random Forests". In: (2001). DOI: `10.1023/A:1010933404324`.

[16] Raphael Douady Xingxing Ye. "Systemic Risk Indicators Based onNonlinear Poly-Model". In: *Journal of Risk and Financial Management* 12 (2019). DOI: `https://doi.org/10.3390/jrfm12010002`.

[17] Chee-Kheong Siew Guang-Bin Huang Qin-Yu Zhu. "Extreme learning machine: Theory and applications". In: *Neurocomputing* (2006). DOI: `https://doi.org/10.1016/j.neucom.2005.12.126`.

[18] Guang-Bin Huang. "An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels". In: *Cognitive Computation* 6 (Sept. 2014), pp. 376–390. DOI: `10.1007/s12559-014-9255-2`.

[19] Stephan Kolassa. "Combining exponential smoothing forecasts using Akaike weights". In: *International Journal of Forecasting* 27.2 (2011), pp. 238–251. ISSN: 0169-2070. DOI: `https://doi.org/10.1016/j.ijforecast.2010.04.006`.

[20] In Jae Myung, Sridhar Ramamoorti, and Andrew D. Bailey. "Maximum Entropy Aggregation of Expert Predictions". In: *Management Science* 42.10 (1996), pp. 1420–1436. ISSN: 00251909, 15265501.

[21] A. Karasan. *Machine Learning for Financial Risk Management with Python: Algorithms for Modeling Risk*. O'Reilly, 2022. ISBN: 9781492085256.

[22] CBOE. *VIX Index*. `https://www.cboe.com/tradable_products/vix/`. 2023. (Visited on 08/21/2023).

[23] Robert F. Engle. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4 (1982), pp. 987–1007. ISSN: 00129682, 14680262.

[24] Abdullah Karasan and Esma Gaygısız. "Volatility Prediction and Risk Management: An SVR-GARCH Approach". In: *The Journal of Financial Data Science*. 2020.

[25] Rob J. Hyndman and Anne B. Koehler. "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4 (2006), pp. 679–688. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001.

[26] Halbert White. "A Reality Check for Data Snooping". In: *Econometrica* 68.5 (2000), pp. 1097–1126. ISSN: 00129682, 14680262.

[27] Payam Refaeilzadeh, Lei Tang, and Huan Liu. "Cross-Validation". In: *Encyclopedia of Database Systems*. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pp. 532–538. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_565. URL: https://doi.org/10.1007/978-0-387-39940-9_565.

[28] Scikit Learn. *Quantile Transform*. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.QuantileTransformer.html. 2023. (Visited on 08/10/2023).

[29] Peter R. Hansen, Asger Lunde, and James M. Nason. "THE MODEL CONFIDENCE SET". In: *Econometrica* 79.2 (2011), pp. 453–497. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/41057463.

APPENDIX A – FEATURES

The plots below depict the some of different features that compose the used dataset. They all correspond to stationary series with a daily frequency.
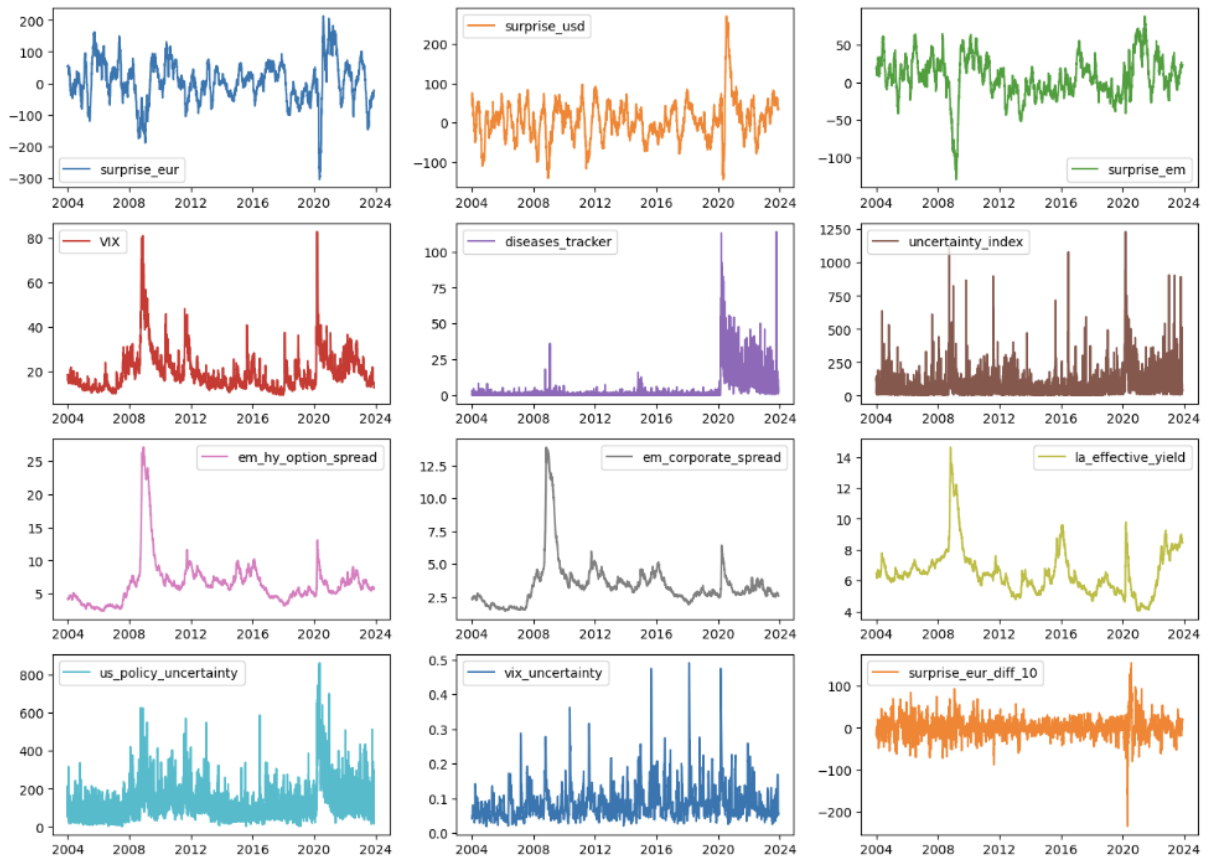
Figure 18: Features plot

APPENDIX B – RESULTS PERFORMANCE METRICS

B.1 S&P 500

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	comb8	1	5,82	11,5	-0,41	5,93	0,55	0,93	0,77	0,25	-25,37	-1,18	-1,79
model_confidence	bench	0,956	7,3	19,16	-0,25	12,86	0,46	0,92	0,65	0,16	-56,78	-1,79	-2,97
model_confidence	comb4	0,956	6,62	14,38	-0,34	7,36	0,52	0,94	0,72	0,2	-38,05	-1,42	-2,25
model_confidence	comb1	0,064	4,86	12,63	-0,26	11,41	0,44	0,92	0,61	0,15	-38	-1,2	-1,98
model_confidence	comb2	0,033	5,74	13,66	-0,46	7,89	0,48	0,95	0,66	0,16	-39,49	-1,36	-2,16
all_nothing	comb8	1	8,15	17,22	-0,45	9,55	0,54	0,95	0,75	0,24	-38,28	-1,69	-2,66
all_nothing	bench	0,827	7,3	19,16	-0,25	12,86	0,46	0,92	0,65	0,16	-56,78	-1,79	-2,97
all_nothing	comb4	0,827	7,55	16,63	-0,36	6,99	0,52	0,94	0,73	0,2	-43,77	-1,65	-2,6
all_nothing	comb2	0,397	6,55	15,98	-0,43	6,71	0,48	0,94	0,66	0,16	-48,59	-1,61	-2,53
all_nothing	comb11	0,382	4,72	11,29	-0,49	4,17	0,47	0,96	0,64	0,14	-37,62	-1,21	-1,81
full_long_short	comb8	1	8,24	19,16	-0,43	12,88	0,51	0,92	0,71	0,27	-36,06	-1,78	-2,92
full_long_short	bench	0,911	7,3	19,16	-0,25	12,86	0,46	0,92	0,65	0,16	-56,78	-1,79	-2,97
full_long_short	comb4	0,204	6,83	19,17	-0,27	12,86	0,44	0,91	0,62	0,23	-36,34	-1,79	-2,96
full_long_short	comb1	0,119	5,17	19,16	-0,25	12,84	0,36	0,9	0,5	0,11	-60,38	-1,81	-2,97
full_long_short	comb6	0,119	5,38	19,17	-0,25	12,84	0,37	0,9	0,51	0,12	-59,18	-1,81	-2,97
ncdf	bench	1	7,3	19,16	-0,25	12,86	0,46	0,92	0,65	0,16	-56,78	-1,79	-2,97
ncdf	comb4	0,463	5,33	13,16	-0,39	9,72	0,46	0,97	0,64	0,18	-33,78	-1,27	-2,08
ncdf	comb8	0,463	2,62	5,41	-0,44	4,32	0,51	0,96	0,7	0,2	-13,86	-0,56	-0,87
ncdf	comb1	0,209	2,52	5,62	-0,32	9,6	0,47	0,96	0,66	0,19	-14,09	-0,56	-0,89
ncdf	comb2	0,041	4,68	12,56	-0,52	10,78	0,43	0,99	0,59	0,21	-25,9	-1,19	-1,99
kelly	bench	1	7,3	19,16	-0,25	12,86	0,46	0,92	0,65	0,16	-56,78	-1,79	-2,97
kelly	comb4	0,529	5,29	13,01	-0,4	9	0,46	0,99	0,64	0,18	-33,09	-1,25	-2,04
kelly	comb8	0,529	3,11	6,31	-0,44	4,2	0,52	0,98	0,71	0,21	-15,37	-0,64	-1,01
kelly	comb1	0,13	2,96	6,61	-0,32	9,82	0,47	0,98	0,66	0,19	-16,77	-0,65	-1,05
kelly	comb2	0,005	4,56	12,51	-0,53	10,39	0,42	0,99	0,58	0,2	-26,72	-1,19	-1,98

Table 10: Performances of top 5 models for each sizing method, S&P 500 5 days

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	comb4	1	7,29	15,64	-0,37	7,35	0,53	0,93	0,74	0,21	-40,27	-1,57	-2,47
model_confidence	bench	0,342	7,25	19,17	-0,25	12,85	0,46	0,92	0,64	0,16	-56,78	-1,79	-2,97
model_confidence	comb5	0,342	4,52	9,51	-0,61	2,97	0,51	0,93	0,7	0,19	-26,33	-1,04	-1,51
model_confidence	comb1	0,069	5,56	13,89	-0,27	12,16	0,46	0,92	0,64	0,16	-40,39	-1,34	-2,19
model_confidence	comb8	0,069	5,72	12,98	-0,44	6,2	0,49	0,94	0,68	0,17	-36,7	-1,32	-2,04
all_nothing	comb5	1	6,21	12,67	-0,57	5,17	0,54	0,96	0,74	0,18	-37,11	-1,34	-2,02
all_nothing	comb4	0,872	7,75	17,09	-0,42	8,02	0,52	0,93	0,73	0,2	-44,24	-1,68	-2,7
all_nothing	comb7	0,872	5,84	11,85	-0,68	3,98	0,54	0,98	0,74	0,22	-28,76	-1,26	-1,92
all_nothing	bench	0,361	7,25	19,17	-0,25	12,85	0,46	0,92	0,64	0,16	-56,78	-1,79	-2,97
all_nothing	randomforest	0,248	5,7	12,45	-0,55	4,39	0,51	0,95	0,7	0,18	-35,04	-1,32	-2
full_long_short	bench	1	7,25	19,17	-0,25	12,85	0,46	0,92	0,64	0,16	-56,78	-1,79	-2,97
full_long_short	comb4	0,907	7,43	19,17	-0,38	12,87	0,47	0,92	0,66	0,24	-37,18	-1,79	-2,95
full_long_short	comb8	0,775	6,06	19,17	-0,62	12,85	0,4	0,92	0,56	0,14	-55,61	-1,79	-2,98
full_long_short	comb1	0,024	6	19,17	-0,25	12,84	0,4	0,9	0,56	0,14	-56,78	-1,81	-2,98
full_long_short	comb5	0,005	3,02	19,18	-0,16	12,8	0,25	0,9	0,35	0,09	-51,75	-1,81	-2,94
ncdf	bench	1	7,25	19,17	-0,25	12,85	0,46	0,92	0,64	0,16	-56,78	-1,79	-2,97
ncdf	comb4	0,872	6,61	15,59	-0,38	8,61	0,49	0,96	0,68	0,25	-30,18	-1,51	-2,45
ncdf	comb8	0,063	3,6	7,8	-0,42	4,47	0,49	0,96	0,68	0,2	-19,36	-0,8	-1,25
ncdf	comb1	0,023	3,48	8,36	-0,35	10,83	0,45	0,98	0,63	0,17	-21,59	-0,82	-1,32
ncdf	comb6	0,016	3,39	8,42	-0,3	9,68	0,44	0,98	0,61	0,16	-23,16	-0,83	-1,34
kelly	comb4	1	6,69	15,39	-0,35	8,54	0,5	0,93	0,7	0,27	-28,12	-1,52	-2,4
kelly	bench	0,98	7,25	19,17	-0,25	12,85	0,46	0,92	0,64	0,16	-56,78	-1,79	-2,97
kelly	comb8	0,173	4,06	8,59	-0,41	4,28	0,51	0,97	0,7	0,22	-20,05	-0,89	-1,36
kelly	comb1	0,027	3,88	9,3	-0,34	11,13	0,46	0,97	0,64	0,17	-24,59	-0,91	-1,47
kelly	comb6	0,018	3,77	9,37	-0,28	10,12	0,44	0,96	0,62	0,16	-26,43	-0,93	-1,48

Table 11: Performances of top 5 models for each sizing method, S&P 500 10 days

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	comb4	1	7,82	15,91	-0,3	7,28	0,55	0,94	0,78	0,22	-39,88	-1,58	-2,5
model_confidence	bench	0,013	6,88	19,18	-0,25	12,86	0,44	0,92	0,62	0,15	-56,78	-1,79	-2,97
model_confidence	comb5	0,004	4,37	10,37	-0,62	6,6	0,46	0,94	0,63	0,22	-22,1	-1,11	-1,65
model_confidence	comb8	0,004	5,8	13,29	-0,49	8,04	0,49	0,95	0,68	0,27	-24,3	-1,36	-2,1
model_confidence	comb1	0,003	4,66	14,5	-0,33	13,67	0,39	0,93	0,54	0,13	-42,1	-1,41	-2,29
all_nothing	comb4	1	8,16	16,69	-0,26	7,2	0,55	0,93	0,78	0,22	-42,77	-1,66	-2,61
all_nothing	comb5	0,153	6,13	13,32	-0,52	11,58	0,51	0,94	0,71	0,25	-27,39	-1,37	-2,12
all_nothing	bench	0,012	6,88	19,18	-0,25	12,86	0,44	0,92	0,62	0,15	-56,78	-1,79	-2,97
all_nothing	comb7	0,009	4,91	13,32	-0,18	8,46	0,43	0,95	0,59	0,23	-24,34	-1,36	-2,12
all_nothing	randomforest	0,007	3,88	13,17	-0,6	5,33	0,35	0,93	0,48	0,12	-39,41	-1,4	-2,14
full_long_short	comb4	1	8,49	19,17	-0,14	12,85	0,52	0,92	0,74	0,22	-45,13	-1,79	-2,88
full_long_short	bench	0,299	6,88	19,18	-0,25	12,86	0,44	0,92	0,62	0,15	-56,78	-1,79	-2,97
full_long_short	comb6	0,006	3,11	19,18	-0,19	12,86	0,26	0,91	0,36	0,07	-70,62	-1,79	-2,96
full_long_short	comb8	0,006	5,32	19,17	-0,41	12,85	0,37	0,92	0,51	0,17	-40,56	-1,79	-2,93
full_long_short	comb1	0,003	1,88	19,18	-0,2	12,82	0,19	0,9	0,27	0,05	-73,8	-1,81	-2,97
ncdf	comb4	1	7,67	17,66	-0,28	15,1	0,51	0,95	0,71	0,29	-31,09	-1,61	-2,69
ncdf	bench	0,419	6,88	19,18	-0,25	12,86	0,44	0,92	0,62	0,15	-56,78	-1,79	-2,97
ncdf	comb8	0,008	3,99	10,82	-0,48	5,24	0,42	0,96	0,57	0,18	-24,83	-1,13	-1,74
ncdf	comb1	0,006	3,4	11,43	-0,44	13,56	0,35	0,97	0,49	0,13	-31,1	-1,1	-1,78
ncdf	comb6	0,006	3,44	11,4	-0,37	12,01	0,35	0,98	0,49	0,13	-30,26	-1,1	-1,78
kelly	comb4	1	7,59	17,4	-0,31	14,44	0,51	0,94	0,71	0,28	-31,7	-1,6	-2,65
kelly	bench	0,422	6,88	19,18	-0,25	12,86	0,44	0,92	0,62	0,15	-56,78	-1,79	-2,97
kelly	comb8	0,01	4,22	10,99	-0,48	5,74	0,43	0,96	0,6	0,19	-24,56	-1,13	-1,75
kelly	comb1	0,003	3,45	11,64	-0,42	13,34	0,35	0,97	0,49	0,12	-33,91	-1,14	-1,8
kelly	comb6	0,003	3,5	11,61	-0,37	12,2	0,35	0,97	0,49	0,12	-32,98	-1,13	-1,8

Table 12: Performances of top 5 models for each sizing method, S&P 500 22 days

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	comb8	1	7,76	15,2	-0,51	12,83	0,57	0,96	0,79	0,27	-31,96	-1,48	-2,38
model_confidence	bench	0,715	7,32	19,22	-0,25	12,84	0,46	0,91	0,65	0,16	-56,78	-1,79	-2,98
model_confidence	comb3	0,715	7,21	14,27	-0,29	8,48	0,56	0,98	0,78	0,27	-29,05	-1,42	-2,25
model_confidence	comb4	0,585	7,43	16,46	-0,35	6,95	0,52	0,93	0,72	0,19	-45,01	-1,65	-2,61
model_confidence	comb2	0,336	6,92	15,6	-0,42	7,05	0,51	0,94	0,7	0,19	-42,41	-1,59	-2,49
all_nothing	comb8	1	8,71	16,49	-0,33	13,15	0,59	0,97	0,83	0,29	-33,92	-1,58	-2,58
all_nothing	bench	0,256	7,32	19,22	-0,25	12,84	0,46	0,91	0,65	0,16	-56,78	-1,79	-2,98
all_nothing	comb3	0,253	7,32	15,39	-0,48	15,32	0,54	0,96	0,75	0,24	-33,92	-1,49	-2,39
all_nothing	comb4	0,253	7,41	16,73	-0,36	7,1	0,51	0,93	0,71	0,19	-45,96	-1,66	-2,65
all_nothing	comb2	0,221	7,12	16,02	-0,43	7,44	0,51	0,94	0,71	0,19	-43,78	-1,62	-2,56
full_long_short	comb8	1	9,05	19,22	-0,21	12,86	0,55	0,99	0,78	0,25	-42,05	-1,7	-2,87
full_long_short	bench	0,537	7,32	19,22	-0,25	12,84	0,46	0,91	0,65	0,16	-56,78	-1,79	-2,98
full_long_short	comb3	0,084	5,9	19,23	-0,29	12,82	0,39	0,96	0,56	0,19	-40,81	-1,73	-2,91
full_long_short	comb4	0,066	6,54	19,23	-0,27	12,8	0,43	0,92	0,6	0,18	-45,31	-1,79	-2,96
full_long_short	comb7	0,066	5,32	19,23	-0,33	12,81	0,37	0,95	0,51	0,19	-36,97	-1,75	-2,93
ncdf	bench	1	7,32	19,22	-0,25	12,84	0,46	0,91	0,65	0,16	-56,78	-1,79	-2,98
ncdf	comb3	0,854	6,42	17,49	-0,18	12,81	0,44	0,95	0,63	0,24	-31,86	-1,61	-2,66
ncdf	comb4	0,12	6,77	18,62	-0,26	12,67	0,45	0,94	0,63	0,19	-43,77	-1,72	-2,86
ncdf	comb8	0,12	5,66	16,46	-0,58	12,56	0,42	0,96	0,58	0,16	-42,51	-1,56	-2,51
ncdf	comb1	0,08	5,52	16,04	-0,39	12,74	0,42	0,96	0,58	0,16	-40,58	-1,54	-2,42
kelly	bench	1	7,32	19,22	-0,25	12,84	0,46	0,91	0,65	0,16	-56,78	-1,79	-2,98
kelly	comb3	0,757	6,14	17,21	-0,2	13,21	0,43	0,95	0,61	0,24	-31,26	-1,59	-2,61
kelly	comb8	0,281	5,56	16	-0,53	11,76	0,42	0,95	0,58	0,16	-41,98	-1,52	-2,43
kelly	comb1	0,132	5,28	15,6	-0,38	12,18	0,41	0,95	0,57	0,15	-41,55	-1,51	-2,35
kelly	comb4	0,132	6,65	18,53	-0,27	12,68	0,44	0,94	0,62	0,19	-43,94	-1,71	-2,85

Table 13: Performances of top 5 models for each sizing method, S&P 500 66 days

B.2 Ibovespa

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	comb5	1	8,03	17,55	0,03	4,04	0,53	1	0,76	0,22	-42,76	-1,73	-2,49
model_confidence	comb4	0,592	9,31	20,94	-0,14	2,58	0,53	1	0,76	0,24	-47,03	-2,09	-2,96
model_confidence	bench	0,027	8,94	27,13	-0,14	8,14	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,85
model_confidence	comb1	0,024	6,94	19,16	-0,04	8,13	0,45	0,98	0,64	0,17	-50,13	-1,8	-2,78
model_confidence	comb6	0,024	6,93	19,24	-0,03	8,42	0,44	0,99	0,63	0,17	-50,59	-1,8	-2,79
all_nothing	comb4	1	10,24	24,3	-0,2	3,07	0,52	0,97	0,75	0,27	-47,97	-2,46	-3,44
all_nothing	comb5	0,732	9,96	26,08	0,06	7,43	0,49	0,96	0,71	0,22	-59,96	-2,51	-3,66
all_nothing	randomforest	0,078	9,92	25,1	-0,19	5,54	0,5	0,96	0,72	0,23	-55,96	-2,49	-3,53
all_nothing	comb7	0,044	9,3	26,9	-0,14	8,43	0,47	0,95	0,66	0,21	-59,96	-2,55	-3,82
all_nothing	bench	0,033	8,94	27,13	-0,14	8,14	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,85
full_long_short	comb5	1	10,37	27,13	0,23	8,09	0,5	0,96	0,73	0,23	-59,96	-2,58	-3,75
full_long_short	comb4	0,848	9,94	27,13	-0,18	8,15	0,49	0,97	0,69	0,25	-51,83	-2,55	-3,83
full_long_short	comb7	0,356	9,52	27,13	-0,14	8,14	0,47	0,96	0,67	0,21	-59,96	-2,58	-3,85
full_long_short	bench	0,308	8,94	27,13	-0,14	8,14	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,85
full_long_short	comb1	0,308	8,71	27,14	-0,14	8,14	0,44	0,96	0,63	0,2	-59,96	-2,58	-3,85
ncdf	comb5	1	4,26	8,88	0,04	5,64	0,51	1,05	0,74	0,22	-20,36	-0,84	-1,37
ncdf	comb4	0,409	7,34	19,07	-0,22	4,07	0,47	1,05	0,67	0,16	-54,74	-1,85	-2,8
ncdf	bench	0,043	8,94	27,13	-0,14	8,14	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,85
ncdf	comb2	0,017	6,22	18,36	-0,3	4,28	0,42	1,08	0,59	0,13	-58,35	-1,77	-2,74
ncdf	comb1	0	3,64	10,47	0,1	10,2	0,39	1,04	0,56	0,12	-35,04	-0,95	-1,65
kelly	comb5	1	4,88	10,01	0,02	4,71	0,53	1,05	0,76	0,23	-22,81	-0,96	-1,52
kelly	comb4	0,482	7,39	19,09	-0,22	3,76	0,47	1,03	0,67	0,17	-53,68	-1,85	-2,78
kelly	bench	0,028	8,94	27,13	-0,14	8,14	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,85
kelly	comb2	0,003	6,23	18,42	-0,3	3,99	0,42	1,06	0,59	0,14	-57,27	-1,77	-2,72
kelly	comb6	0,001	4,14	11,98	0,18	11,55	0,4	1,03	0,57	0,12	-39,51	-1,08	-1,86

Table 14: Performances of top 5 models for each sizing method, Ibovespa 5 days

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	randomforest	1	7,88	19,46	-0,17	2,22	0,49	0,99	0,69	0,18	-51,46	-1,97	-2,79
model_confidence	bench	0,524	8,93	27,15	-0,14	8,13	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,86
model_confidence	comb5	0,524	8,04	21,19	-0,13	4,97	0,47	0,99	0,67	0,21	-47,33	-2,06	-3,08
model_confidence	comb4	0,473	7,38	21,81	-0,16	3,03	0,44	0,98	0,62	0,17	-57,11	-2,25	-3,15
model_confidence	comb2	0,22	6,22	21,27	-0,2	2,76	0,39	0,97	0,55	0,12	-67,69	-2,21	-3,06
all_nothing	randomforest	1	10,61	24,34	-0,15	3,18	0,54	0,96	0,77	0,22	-58,16	-2,46	-3,41
all_nothing	comb3	0,208	8,56	26,78	-0,14	8,68	0,44	0,95	0,63	0,2	-59,96	-2,55	-3,84
all_nothing	comb5	0,208	9,37	26,4	-0,13	8,54	0,47	0,95	0,67	0,21	-59,96	-2,54	-3,75
all_nothing	comb7	0,208	8,82	27,07	-0,14	8,23	0,45	0,95	0,64	0,2	-59,96	-2,57	-3,84
all_nothing	bench	0,193	8,93	27,15	-0,14	8,13	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,86
full_long_short	randomforest	1	10,7	27,14	-0,11	8,13	0,51	0,96	0,73	0,21	-66,73	-2,57	-3,79
full_long_short	comb5	0,732	9,38	27,14	-0,11	8,12	0,47	0,95	0,67	0,21	-59,96	-2,59	-3,84
full_long_short	comb7	0,727	8,66	27,15	-0,13	8,13	0,44	0,95	0,63	0,2	-59,96	-2,59	-3,85
full_long_short	bench	0,714	8,93	27,15	-0,14	8,13	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,86
full_long_short	comb3	0,714	7,98	27,15	-0,14	8,13	0,42	0,95	0,6	0,19	-59,96	-2,58	-3,86
ncdf	bench	1	8,93	27,15	-0,14	8,13	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,86
ncdf	comb5	0,177	6,2	17,21	-0,06	5,74	0,44	1,02	0,62	0,2	-36,89	-1,67	-2,65
ncdf	comb4	0,079	4,08	23,27	-0,33	5,76	0,29	0,99	0,4	0,09	-76,09	-2,26	-3,41
ncdf	comb8	0,044	5,92	18,46	-0,02	8,29	0,4	1,07	0,58	0,14	-53,95	-1,65	-2,87
ncdf	comb2	0,011	1,87	23,16	-0,41	7,29	0,2	0,98	0,27	0,05	-84,56	-2,22	-3,4
kelly	bench	1	8,93	27,15	-0,14	8,13	0,45	0,96	0,64	0,2	-59,96	-2,58	-3,86
kelly	comb5	0,224	6,12	17,33	-0,11	5,02	0,43	1,01	0,61	0,2	-37,05	-1,7	-2,63
kelly	comb4	0,067	4	23,07	-0,32	5,37	0,29	0,99	0,4	0,09	-75,85	-2,24	-3,37
kelly	comb8	0,033	6,02	18,49	-0,02	7,59	0,41	1,03	0,58	0,14	-52,67	-1,72	-2,82
kelly	comb2	0,011	1,96	22,95	-0,37	6,87	0,2	0,99	0,28	0,05	-83,92	-2,18	-3,36

Table 15: Performances of top 5 models for each sizing method, Ibovespa 10 days

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	comb4	1	9,27	23,69	-0,11	2,88	0,49	0,96	0,71	0,21	-55,04	-2,41	-3,35
model_confidence	comb2	0,751	8,43	22,78	-0,1	2,45	0,47	0,96	0,67	0,19	-56,7	-2,35	-3,22
model_confidence	comb5	0,751	9,06	23,99	-0,05	6,73	0,48	0,97	0,69	0,23	-50,49	-2,33	-3,43
model_confidence	bench	0,003	8,36	27,16	-0,14	8,13	0,43	0,95	0,61	0,2	-59,96	-2,58	-3,86
model_confidence	comb1	0	7,51	24,81	-0,03	7,95	0,42	0,97	0,59	0,17	-59,8	-2,33	-3,58
all_nothing	comb2	1	8,83	23,4	-0,1	2,29	0,48	0,97	0,69	0,2	-55,5	-2,41	-3,32
all_nothing	comb4	0,964	9,62	24,16	-0,14	3,09	0,5	0,96	0,72	0,22	-54	-2,46	-3,41
all_nothing	comb5	0,371	9,14	26,94	-0,14	8,47	0,46	0,95	0,65	0,21	-59,18	-2,55	-3,84
all_nothing	bench	0,001	8,36	27,16	-0,14	8,13	0,43	0,95	0,61	0,2	-59,96	-2,58	-3,86
all_nothing	comb1	0,001	8,31	27,16	-0,14	8,13	0,43	0,95	0,61	0,19	-59,96	-2,58	-3,86
full_long_short	comb5	1	9,79	27,15	-0,14	8,15	0,48	0,96	0,68	0,22	-58,4	-2,57	-3,85
full_long_short	comb4	0,835	9,19	27,16	-0,09	8,12	0,46	0,97	0,66	0,21	-58,94	-2,56	-3,79
full_long_short	comb2	0,179	7,24	27,17	-0,04	8,12	0,39	0,97	0,56	0,17	-62,67	-2,56	-3,8
full_long_short	bench	0,007	8,36	27,16	-0,14	8,13	0,43	0,95	0,61	0,2	-59,96	-2,58	-3,86
full_long_short	comb1	0,007	8,26	27,16	-0,14	8,13	0,43	0,95	0,61	0,19	-59,96	-2,58	-3,86
ncdf	comb5	1	9,01	21,81	0,06	5,45	0,5	1,01	0,73	0,28	-38,93	-2,12	-3,14
ncdf	comb4	0,202	8,46	25,61	-0,26	6,07	0,45	0,97	0,63	0,2	-57,82	-2,48	-3,65
ncdf	comb2	0,107	7,41	25,66	-0,06	9,37	0,41	0,99	0,58	0,16	-64,31	-2,39	-3,58
ncdf	bench	0,099	8,36	27,16	-0,14	8,13	0,43	0,95	0,61	0,2	-59,96	-2,58	-3,86
ncdf	comb1	0,003	6,41	22,88	0,12	9,24	0,39	1,02	0,55	0,15	-59,96	-2,11	-3,4
kelly	comb5	1	8,99	21,66	0,02	5,46	0,51	1,02	0,73	0,27	-40,34	-2,09	-3,12
kelly	comb4	0,268	8,64	25,52	-0,26	6,05	0,45	0,98	0,64	0,2	-57,47	-2,48	-3,62
kelly	comb2	0,05	7,41	25,5	-0,09	8,84	0,41	0,98	0,59	0,16	-63,47	-2,38	-3,56
kelly	bench	0,048	8,36	27,16	-0,14	8,13	0,43	0,95	0,61	0,2	-59,96	-2,58	-3,86
kelly	comb8	0,002	6,6	22,71	0,03	7,64	0,4	1,01	0,56	0,15	-59,29	-2,13	-3,35

Table 16: Performances of top 5 models for each sizing method, Ibovespa 22 days

Sizing	Model	MCS 3Y	Ann. Returns	Volatility	Skewness	Kurtosis	Sharpe	Tail Ratio	Sortino	Calmar	Max Drawdown	Historical VaR	Expected Shortfall
model_confidence	comb3	1	8,6	27,14	-0,13	7,9	0,44	0,95	0,63	0,2	-59,91	-2,59	-3,86
model_confidence	comb11	0,09	8,44	26,98	-0,12	7,75	0,44	0,95	0,62	0,2	-58,09	-2,59	-3,84
model_confidence	comb2	0,01	2	18,42	-0,23	5,73	0,2	0,98	0,28	0,07	-50,66	-2	-3
model_confidence	comb4	0,01	0,76	19,3	-0,32	7,56	0,14	0,96	0,19	0,04	-68,32	-2,07	-3,14
model_confidence	comb9	0,01	8,36	27,01	-0,14	7,88	0,43	0,95	0,61	0,2	-58,83	-2,59	-3,85
all_nothing	bench	1	8,6	27,24	-0,14	8,1	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87
all_nothing	comb2	0,01	2,01	18,82	-0,34	7,49	0,2	0,97	0,28	0,07	-51,11	-2,02	-3,05
all_nothing	comb4	0,01	0,84	19,55	-0,35	7,71	0,14	0,96	0,2	0,04	-69,01	-2,1	-3,18
all_nothing	comb9	0,01	8,48	27,21	-0,14	8,15	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87
all_nothing	comb11	0,008	8,39	27,22	-0,14	8,14	0,43	0,95	0,61	0,2	-59,96	-2,59	-3,87
full_long_short	bench	1	8,6	27,24	-0,14	8,1	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87
full_long_short	comb11	0,004	8,15	27,24	-0,14	8,1	0,42	0,95	0,6	0,19	-59,96	-2,6	-3,87
full_long_short	comb4	0,004	-9,68	27,25	-0,15	8,06	-0,24	0,9	-0,33	-0,07	-95,58	-2,69	-3,95
full_long_short	comb7	0,004	7,37	27,25	-0,16	8,09	0,4	0,93	0,56	0,18	-59,96	-2,62	-3,89
full_long_short	comb9	0,004	8,33	27,24	-0,14	8,1	0,43	0,95	0,61	0,2	-59,96	-2,6	-3,87
ncdf	bench	1	8,6	27,24	-0,14	8,1	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87
ncdf	comb8	0,717	8,6	27,24	-0,14	8,1	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87
ncdf	comb11	0,407	8,37	26,98	-0,12	7,65	0,43	0,95	0,62	0,21	-55,91	-2,6	-3,83
ncdf	comb1	0,033	8,6	27,24	-0,14	8,1	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87
ncdf	comb3	0,033	8,6	27,24	-0,14	8,1	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87
kelly	comb3	1	8,6	27,03	-0,12	7,71	0,44	0,95	0,63	0,2	-59,86	-2,58	-3,85
kelly	comb11	0,084	8,29	26,8	-0,12	7,47	0,43	0,95	0,61	0,21	-56,17	-2,57	-3,81
kelly	comb4	0,001	-9,73	26,27	-0,19	6,6	-0,26	0,9	-0,35	-0,07	-94,88	-2,62	-3,81
kelly	comb7	0,001	7,4	26,92	-0,15	8,17	0,4	0,94	0,57	0,18	-59,95	-2,57	-3,85
kelly	bench	0	8,6	27,24	-0,14	8,1	0,44	0,95	0,62	0,2	-59,96	-2,59	-3,87

Table 17: Performances of top 5 models for each sizing method, Ibovespa 66 days