Group part

A Work Project, presented as part of the requirements for the Award of a Master's degree in

Finance from the Nova School of Business and Economics.

A DEEP LEARNING ALGORITHM TO OPTIMIZE MULTI-ASSET PORTFOLIO RETURNS BASED ON MACROECONOMIC VARIABLES
-
APPLYING DIFFERENT PORTFOLIO OPTIMIZERS

MARGARIDA CARVALHOSA

Work project carried out under the supervision of:

Professor Gonçalo Sommer Ribeiro

20/12/2023

Group part

Abstract

The individual investor does not usually take advantage of quantitative methods to invest, such as Machine Learning algorithms. Additionally, the individual investor is often tempted to time the market. This paper proposes an investment strategy that dynamically allocates weights based on macroeconomic variables, aimed at facilitating the access to advanced methods for the individual investor, diminish the influence of emotional behavior of the investor in his personal investments, and corresponding attempts to time the market. It does so by studying past relationships between macroeconomic variables and the optimal asset weights that would have led to a good performance in the past. These relationships are examined by means of an Artificial Neural Network. Finally, to better take advantage of the different economic scenarios, and to potentially benefit from diversification gains, different asset classes were considered, beyond the most adopted bonds and equity.
Moreover, this dissertation comprehends the study of different portfolio optimization methods.

Keywords

Machine Learning, Multi-Asset, Macro, Portfolio Optimization

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. INTRODUCTION

This paper aims to create a systematic and macro-based investment strategy and the corresponding algorithm that dynamically adapts the allocation weights to each asset according to the changes in macroeconomic variables, as well as the changes in the relationships between these variables and the assets themselves. The most important feature is for the study to provide practical value so that an investor can implement it.

The investment strategy incorporates a machine learning approach, employing an Artificial Neural Network (ANN). The algorithm, which will be further discussed in the subsequent sections, focuses on capturing and adapting to the evolving relationships between the macroeconomic variables and the assets in the portfolio. By using a machine learning approach that can consider the interactive effects among the inputs, it allows for a more concrete understanding of real-life problems, whereas in academic statistical methods, such as regressions, because it relies on predefined models and assumptions, its adaptability to more complex problems is more questionable. The investment philosophy adopted throughout the model is the risk-parity meaning the model attempts to predict weights that would result in each asset contributing equally to the portfolio's total risk.

This strategy lies within the scope of global macro, as both share a common foundation in economic data. Indeed, global macro strategies analyse macroeconomic indicators to make investment decisions, and as the strategy mentioned above is a systematic and macro-based approach that employs an algorithm for asset allocation purposes, one can say that it emphasizes its alignment with global macroeconomic strategies. Global macro is a type of strategy known for being adopted by hedge funds and for some historically well-known trades such as George Soros' breaking of the British pound, which profited him over £1 billion. Regarding the motivations for this paper, we believe that investors should consistently consider the macroeconomic environment when making investment decisions to achieve a sustained long

term financial gain, i.e., by adopting a top-down approach, which serves as the primary motivation for this research paper.

The choice of a multi-asset strategy derives from the intention of capturing the different relationships between each asset and the macroeconomic fundamentals, as well as benefiting from potential diversification gains by investing in multiple asset classes, namely equity, commodities and fixed income. Finally, we found that there was little research on a trading strategy encompassing all the features we were looking for: practical implementation, multi-asset investment universe, macro-based strategy, and using a ML algorithm.

The research goes beyond just analysing and predicting security premiums based on macroeconomic variables, as the algorithm dynamically adapts to changes in relationships, tackling the "coefficient instability" problem, and the fact that the predictive power of these variables is not stationary. It also directly optimizes the portfolio allocation weights using the systematic macro-based model mentioned above, rather than just predicting the main drivers of security premiums. In summary, this research is an attempt to solve the practical problems that usually come hand in hand with applying theoretical frameworks to produce outperforming strategies.

2. LITERATURE REVIEW

To implement a multi-asset macro-investing strategy through the construction of a Machine Learning algorithm that dynamically adjusts the weights according to the macroeconomic scenario, it is then necessary to understand which variables have a prediction power on excess returns. Several papers have attempted (and claimed) to find variables that would predict returns. This research is usually focused on predicting the security premium or the excess return over the risk-free rate. Goyal and Welch 2008, examined the in-sample (IS) and out-of-sample (OOS) performance of different independent variables in predicting the equity premium, which was researched by different authors throughout the years, and

concluded it is yet to find variables with "meaningful robust empirical equity premium forecasting power". Moreover, these models could not be practically implemented by an investor with access only to the information available at the time. Another important finding was that variables with lower frequency tended to be better predictors than higher-frequency variables. It is important to notice, however, that this paper examined only the performance of linear regressions in predicting the equity premium. The problem with this approach is that the relationship between asset returns and the explanatory variables may not be linear (Desai and Bharati 1998). A potential contributor to nonlinearity, pointed out by Boer et al. 2018, may be the implementation constraints that don't allow investors to fully realize the potential of a factor strategy. This suggests there is still much research to do on finding variables with predictive power of security premiums.

Buncic and Tischhauser 2017 added to Goyal and Welch's research by showing that OOS forecasts of the equity premium can be considerably enhanced by including macroeconomic data. Thus, to narrow down the focus of the paper, one decided to use only macroeconomic variables to assess its effectiveness in predictive modelling. To define the macro factors, we followed BlackRock's research in 2020, which defines six macro factors - real rates, inflation, credit, economic growth, commodities, and emerging markets - that explain over 90% of returns across different asset classes (Ang et al. 2020), Ana Simões replicates in her paper these factors using 87 U.S. based macro variables (Simões 2021). Since these variables worked well in this model, we incorporated the same in the model and developed in this research.

Identifying excess return predictors for stocks can be a challenging task. In the following paragraphs, we will delve into the variables that show the premise of predicting returns of several asset classes, including Equity, Fixed Income and Commodities. Indeed, research suggests that returns are more predictable for corporate bonds than for stock returns (Lin et al.

2014). Ludvigson and Ng 2009 also show that macroeconomic fundamentals have a strong prediction power for the bond risk premia of US government bonds. Moreover, there is a lack of research for multi-asset investing based on macroeconomic variables and not much use of ML models, which should be studied due to the potential for model improvement as pointed out previously, that may explain the difficulty finding variables with predictive power over returns. Moreover, a multi-asset strategy allows for diversification gains when the strategy is macro-based, and the stock-bond correlation changes dynamically with the macroeconomic scenario (AQR 2023). There is also evidence that as commodities are exposed to market forces such as unexpected inflation and potential positive roll return in futures-based commodity investment in periods of high spot volatility, adding a commodity component to a diversified portfolio results in enhanced risk-adjusted performance (Georgiev 2001). For these reasons, we decided to broaden the range of investable asset classes in our study.

The aim of our research is a bit more practical than these studies by not restricting the research to finding risk premium predictors but rather constructing an algorithm that optimizes dynamically the portfolio allocation accordingly, with the information available at the time. The dynamic construction is crucial since many variables cannot sustain their predictive power throughout time, and hence the poor out-of-sample performance verified in many papers. As such, an algorithm was developed. As mentioned previously, the one common statistical method to predict asset returns based on different variables is through regression as described by Adrian et al., 2015, where the authors introduce a framework for estimating beta representations of dynamic asset pricing models, using a three-step regression for the estimators that efficiently generates a slope and risk betas for each test asset. Alternatively, numerous nonlinear regression models (e.g., logarithmic, and exponential functions) can be used to explore nonlinear relationships. However, these require transformations of the independent variables, which need to be specified before identifying the appropriate model.

Thus, Machine Learning positions itself as an alternative approach that has been increasingly adopted to predict asset returns. When compared with statistical modelling, machine learning places more emphasis on prediction power and not on explaining the outcome (Daul et al. 2022). It is important to notice that Machine Learning is not all the same and includes different methods or subsets. Many authors have found that regression trees and neural networks (ANN), which are two of the most traditionally used ML methods, perform better than linear models in predicting stock returns (Rasekhschaffe and Jones 2019; Gu et al. 2020; Choi et al. 2019; Leung et al. 2021). Daul et al. also found that regression trees and neural networks' superiority in predicting returns can be explained by their robust regularisation methods and their ability to capture interaction effects. Between these two ML models, research suggests that for this paper, predicting asset returns, neural networks are more appropriate given their higher prediction accuracy and capability to achieve a higher return on investment when compared with other classifier ensembles - both homogeneous (similar classifiers) and heterogeneous (diverse classifiers) – and single classifiers such as decision trees and logistic regressions (Tsai et al. 2011). Hence, an artificial neural network was the model chosen to implement the algorithm. It is important to notice, however, that research on the comparison of the out-of-sample performance between ML models and traditional methods, as well as between different ML models for this specific purpose is limited, mostly focusing on the stock market, and results differ between studies. Given that this evidence of higher prediction accuracy and return on investment provided in the previously mentioned paper was found particularly in a sub-type of ANN – multilayer perceptron (MLP) – and given that MLP allows for nonlinearity, we opted for using the same architecture.

Artificial neural networks are particularly interesting as they attempt to simulate the human brain. They are called "neural networks" since they are organised in nodes, which function like neurons and are connected to each other. These nodes are organised in layers: the

input layer; the hidden layers, which can go from 0 up to any number of layers; and the output layer. The model must be trained, to understand the relationships between the inputs and the output. Thus, a training data set is used, which should be representative of the underlying model. When the network is properly achieving the output, the training is completed, and the network can be implemented (Zou et al. 2009). Since the network was constructed with more than 3 layers, it is included in a particular subclass within Machine Learning, which is Deep Learning.

The construction of the algorithm is not complete without defining the goal of the algorithm. Risk-adjusted returns are generally accepted to be more comprehensive than other goals, such as total return solely. Initially, the investment strategy aimed to maximize risk-adjusted returns (i.e. mean-variance optimization) by calculating the asset allocation that achieved the highest Sharpe Ratio, the most adopted measure for risk-adjusted returns. However, optimizing for the Sharpe directly has numerous complications, mainly caused by a high sensitivity to estimate errors. This often leads to an unstable model and a weak performance out-of-sample (Chakrabarti 2021). Considering other options, research points to risk parity consistently outperforming optimal allocation strategies like mean-variance and minimum-variance efficient portfolios (Shakernia et al. 2010). In line with this, we decided to adopt a risk-parity allocation for the target weights, being the Sharpe Ratio a key metric to assess the strategy´s performance compared to a benchmark that will be further detailed in the paper.

3. METHODOLOGY

The decision on the characteristics, features, inputs, and model architecture must be carefully thought out. ML-based strategies are usually reserved for hedge funds and other institutional investors, but this does not have to necessarily be the case.

3.1. GOALS AND FEATURES ALIGNMENT

This strategy chases the Sharpe using advanced methods, which the individual investor does not usually make use of, but it does so by considering the constraints and preferences of the typical retail investor. These include a preference for a passive approach, a low to medium risk tolerance, and resources far more limited than institutional investors. Considering this profile, the strategy is rebalanced monthly, a frequency that is not too high for most investors and that makes sense since personal investors usually rely on their monthly salary to apply their savings. It also uses a risk-parity allocation, as this allocation is built on risk contributions of each asset to the overall portfolio and is known for usually achieving higher Sharpe ratios and being more robust, i.e. avoiding drawdowns, than traditional portfolios, which tend to focus on dollar allocation (Hurst et al. 2010). This allocation also allows for shorting and leveraging, but we opted for an unleveraged portfolio and no-shorting policy since we are tailoring the strategy to the typical profile of a retail investor. Moreover, research suggests that imposing weight constraints can improve portfolio performance, as described by Behr et al.,2012 as it was demonstrated that a constrained minimum-variance portfolio yields significantly lower out-of-sample variances when compared with many established minimum-variance portfolio strategies. This may seem counterintuitive, as in the end, it is an optimization problem, and with constraints, one may not be able to achieve optimal solutions. Still, it is also a prediction problem, so the actual risk and return associated with securities are unknown, meaning these constraints can be reasonable (Frost and Savarino 1988). Lastly, considering the resources' constraint, the model although developed on ML, does not require demanding computational resources. Personal investors usually are less informed and have limited access to research, so leaving the allocation decisions to an algorithm comes in handy. All used data can be publicly accessed.

3.2. DATA TREATMENT

Data collection was the first step of the process. Given the literature review on related work that was conducted beforehand, the decision on which macroeconomic variables to use was based on research, particularly Ana Simões's (2021) paper which uses 87 macro variables as inputs for her algorithm (Appendix – Figure 1). As for the investment universe, different securities from different asset classes were considered, all ETFs or futures contracts, again, for real-life implementation. The comprehensive list of the considered securities can be found in the Appendix – Figure 2 (**Note:** only the securities corresponding to tickers "DIA", "QQQ", "GSG", "SPY", and "AGG" were considered in the final model). The prices of each security and the information on the 87 macroeconomic indicators were extracted. For dates in which not all the securities were active yet, they were entirely disregarded. Thus, the investment universe returns' data starts in August 2006.

However, one cannot use the information of a specific month unless it is already available. Thus, to account for the data that would be available each month, one should lag the data with the appropriate lag, which depends on how long it takes to release that information. Therefore, for the macroeconomic variables that need to be lagged, one, two, or three months of lag was applied, depending on release dates (applied lags in the Appendix – Figure 3). Additionally, the indicators do not have the same frequency, comprehending daily, weekly, monthly, and quarterly data. All the indicators were then adjusted so that monthly data would be available. To do this, for quarterly data, the last available data would be the one used (forward-fill), e.g., if in May the last available value was the one of March, that would be the one used. As for indicators with a frequency higher than monthly, the average of the values of the corresponding month was calculated.

Different variables take different scales and dimensions, e.g. considering the macroeconomic variables data since October 2007, the average value of the variable "Building Permits" was 1245.82, while the average value of the variable "Building Permits MoM" was

0.01. This may be because variables can take different representations, like in the example provided, in which the former represents the absolute value of instances verified on the corresponding period, while the latter represents the growth rate of the same variable. In other cases, the variables may be accounted for in the same way but simply take different dimensions. As such, variables are not on a comparable basis, and variables of a higher magnitude can "dominate" the objective function (Alshaher 2021, 4). Hence, so that the different indicators could be on the same level and to avoid model distortion, all data was subject to feature scaling, transforming data into a comparable scale, through the normalization method, also known as Min-Max scaling. This method works by shifting and re-scaling methods so that they end up ranging between 0 and 1, where 0 corresponds to the minimum value, and 1 to the maximum value in each column. This was applied not only to macroeconomic data but also to weights data, for the same reason, as these were used as input for the model. To understand this, it is important to understand the model first.

3.3. MODEL CONSTRUCTION

The model is a means towards an end, in this case, to make investment decisions: the output of the model is the weight to invest in each security for the upcoming month. It is then a problem with a predictive nature: to predict the weights that would lead to equal risk contributions based on the macroeconomic environment.

Two important points should then be noted: (1) the input of the model should not be confused with the concept of independent variables. The input is the data that is used by the model. Since we are trying to predict weights, the weights that would be optimal in the past, according to the objective – equal risk contributions - were used as input data for the model. If the model only used macroeconomic data, the independent variables, it could not establish any relationship between them and returns. Thus, both the macroeconomic data as well as the optimal weights were used as input data. (2) The 'optimal weights' that the model studies

depend on the objective function. For example, one can aim just to maximize returns, independently of the risk; one can desire to minimize the volatility; and so on, and the 'optimal weights' would vary accordingly. In the constructed model, the objective function was a risk-parity function, so the optimal weights are the ones that result in approximately equal risk contributions of each asset to the overall portfolio risk. To obtain these weights, 14 months of data were used, meaning that the input data did not start in August 2006, the first observation point. Instead, the input data starts 14 months later in October 2007.

The step-by-step procedure to find the optimal weights can then be summed up into the following steps: (1) for each date starting in October 2007, select the returns of the previous 14 months; (2) calculate the standard deviation for each asset over the same period; (3) set the initial weights to the inverse of the individual asset's standard deviation; (4) divide each weight by the sum of the weights, so that they sum up to 1, meaning the portfolio is fully invested; (5) define the risk-parity objective function, which first calculates the portfolio variance based on past returns and weights for each month, using equation (1); calculates the risk contributions of each security for the total portfolio risk over the same period; and then calculates the square difference of the sum between the individual risk contributions and the target that is desired for them: 1/number of securities.

$$\sigma^2 = w^T * \sum * w \qquad (1)$$

After defining the objective function, (2) this is minimized. Essentially, we are obtaining the weights that minimize the deviation from the individual risk contribution's target. Additionally, the objective function considers the constraints that weights should add up to 1 (portfolio fully invested) and no individual weight should be negative or above 1 (no shorting or leveraging).

So that the strategy could be dynamic and applicable in real life, the model was constructed on a rolling basis: it makes predictions only based on the available data up to that

point. As such, for the weight's predictions for October 2013, for example, the model only used data available in September 2013, meaning the model was not all trained on the same data. As for the start date of the training periods, there were two possibilities: it could be fixed, starting from the earliest date in which data is available, or it could also be rolling so that only more recent data would be used to train the model and the training period would hence be fixed, e.g., always use the past 1 years to train data. We opted for the latter, with a training period of three years, with the reasoning later explained in the paper.

3.4. MODEL CONSTRUCTION – TECHNICAL OVERVIEW

Neural networks have built-in layers containing nodes or neurons. The first layer is the input layer, and the last is the output layer. Intuitively, one can understand that the former includes the data, while the latter outputs the produced result. There may exist hidden layers in between, which are the ones that perform the transformations and computations, by allocating weights to the inputs and applying an "activation function", which is a function that calculates the output.

The used neural network for this research was built with a multilayer perceptron (MLP) architecture, which has some features that are valuable for our purpose, such as the neurons being fully connected, meaning every neuron in one layer is connected to every neuron in the other layer (see architecture in the Appendix – Figure 4). This allows the network to learn complex relationships about the data.

MLPs also use the backpropagation method, which fine-tunes the weights by repeatedly adjusting them according to the error rate or loss observed throughout the flow of the network, following the processing of each piece of data.

With the type of architecture settled, certain parameters in its construction must be decided as well. Given the complexity of our dataset and its patterns, a deeper network, i.e. with an increased number of layers, is more appropriate. However, this can also lead to capturing

noise instead of meaningful patterns. On the other hand, a larger dataset allows to increase the number of layers. Instead of trying to find the best architecture, we opted for using the randomized search method.

Randomized search is a method to tune models. Essentially, since it is quite hard to find the optimal parameters to train a neural network and achieve predictive power, one can perform a randomized search that builds different models with different combinations of hyperparameters and evaluates them, enabling one to choose the optimal hyperparameters combination. After doing this, the model was tuned taking the optimal combination into account: batch sizes (i.e., number of samples used per process of model optimisation) of 16 samples; 4 hidden layers with 256, 128, 128, and 64 neurons nodes per hidden layer, respectively; a ReLU (Rectified Linear Unit) activation function, and an alpha of 0.01 (learning rate parameter).

The choice of the activation function is crucial and influences the algorithm's performance. To avoid two very common complications in training models, the vanishing gradient, and the exploding gradient problems, the Rectified Linear Unit (ReLu) function was used:

$$f(x) = \max(0, x) \tag{1}$$

A gradient is the derivative of a function with multiple input variables. When this derivative gets lower and lower as it propagates between layers, in the case of the vanishing gradient problem, or too high, in the case of the exploding gradient problem, the feedback to optimize the model is distorted. This means that the weights could either not be updated anymore, preventing the model from being trained, or the weights would be overcorrected, resulting in an unstable network. Other characteristics, such as time and computational resources demanded are also important to consider when deciding on the activation function. Therefore, the ReLU function is often used to avoid these complications, although in some

cases it can still result in these problems (when there are too many neurons, or the learning rate is set too high).

Data should be split into a training dataset, from which the algorithm learns, and a test dataset, that corresponds to the objective evaluation of the model. It is common practice to split randomly between both datasets. Instead, since this is a time-series prediction problem, and to ensure the robustness and dynamic feature of the model, it uses rolling training data. Thus, to split the data into training and test sets, the algorithm works on a loop. For each different month, it uses past data available up to that month as the training dataset. Then, it is given the macroeconomic variables and weights of that month (independent variables from the test dataset) and predicts the weights, considering these new data, and applying the learned patterns from the training. This prediction is then compared to the dependent variable values in the test dataset, also known as the target variables. This method is known as the time or temporal split of the dataset. Often, a validation set is also used to tune the model, but since the training set is continuously extended on each iteration, as new data becomes available (walk-forward validation), there is no need for this additional data set.

To evaluate the performance of the model, one can use different metrics, but due to the nature of the model and the problem it aims to solve, using the Mean Squared Error (MSE), which measures the average squared difference between the predicted and actual values was the most logical choice. The Root Mean Squared Error (RMSE) was also used since it is expressed in the same units as the original data, and thus, more intuitive. The final model registered an MSE of 1.29 and a RMSE of 1.13. This RMSE value of 1.13, implies that the predicted weights deviate 1.13% from the actual weights of the initial risk parity optimizer that was used as a comparison when calculating this metric. On a practical view, a 1.13% error rate

in an asset allocation problem is quite low, and it implies that the model is robust in its predictions.

4. RESULTS AND CONCLUSIONS

In this section, the results of our final strategy will be discussed, as well as the conducted analysis and process it took to achieve this final model.

4.1. PREDICTION ALGORITHM AND STRATEGY RESULTS

As explained previously, the investment strategy itself starts in October 2010, while the model's input, which includes the training dataset in addition, starts in October 2007. The end date is November 2022. Within this investment period, it was possible to include out-of-the-normal circumstances, such as the COVID-19 Pandemic, and assess the robustness of the strategy throughout this period.

Considering it is a multi-asset investment strategy, the benchmark for comparison was built on a 60/10/10/10/10 split, consisting of AGG US Equity, DIA US Equity, GSG US Equity, SPY US Equity, and QQQ US Equity, respectively. The benchmark yielded a 5.30% average annualized return, with a 6.37% annualized volatility, meaning its Sharpe Ratio was 0.74. On comparison, our strategy would have yielded a 4.60% average annualized return, 5.96% annualized volatility, and a Sharpe Ratio of 0.67, underperforming the benchmark when considering the portfolio returns and Sharpe Ratio, but with significantly less volatility. Indeed, even though the strategy's volatility is lower than the benchmark, it couldn't offset the difference in returns, therefore leading to a lower Sharpe Ratio.

For this comparison between the constructed portfolio and the benchmark, the Info Sharpe might be another metric that one can look at as it also reflects the risk-adjusted returns of a strategy, but without considering the risk-free rate, thus not being influenced by variations of this variable. The strategy's Info Sharpe was 0.77 while the benchmark one was 0.83, thus reinforcing the outperformance of the benchmark.

Looking at the cumulative returns (Appendix – Figure 5), one can observe that until the beginning of the Covid-19 pandemic, the strategy was outperforming the benchmark, but since then, the strategy has been underperforming the benchmark. One of the possible reasons for this underperformance could be related to the weight allocation given to a particular asset class, commodities, which is represented by the GSG US Equity. Indeed, looking at the Appendix - Figure 6, where the weights of the indexes since 2020 are displayed, one can see that the model's weight allocation has been different from the benchmark one, as there is an initial overweight trend for fixed income and an underweight trend in the remaining asset classes. However, this trend started to reverse around January 2021, as fixed income started to have lower weights and commodities began to have higher allocations, with equities remaining stable. Moreover, looking at the Appendix – Figure 7, where the returns of the indexes are displayed since 2020, one can see that the fixed income returns were low, almost zero, therefore, as the model gave a higher allocation to this asset class than the benchmark, one might expect the performance of the benchmark to be better in this case. Moreover, since January 2021, one can see that commodities returns have been quite volatile, having long periods of negative returns. With this in mind, one might expect that as the benchmark's exposure to commodities is lower than the model one, the model's performance will be more severely affected by the volatility of the commodity index (GSG US Equity) than the benchmark.

Still, these results do not account for trading costs, which can make a difference big enough to make a strategy unviable. Thus, trading costs of 20bps were accounted for, and the strategy still yielded a Sharpe of 0.63 (4.33% annualized return and 5.95% standard deviation). The comprehensive results before and after accounting for trading costs can be found in the Appendix – Figure 8.

The deterioration of the strategy was not significant for two reasons: (1) the portfolio is only rebalanced every month, and (2) the weights were somewhat stable, avoiding extreme values and constant buying and selling in large relative amounts.

The strategy's solid results were only possible due to the high predictive power of the model. When assessing the portfolio's accuracy, as mentioned in Section 3.4., the MSE of the model was only 1.28, and the RMSE was 1.13, i.e., on average, the model only deviated 1.13% from the test dataset.

4.2. FEATURE IMPORTANCE ANALYSIS

It is common practice in Machine Learning to perform a feature importance analysis, which attributes scores to each feature to understand its importance for the model. This analysis is essential to not implement the findings of the model without understanding it. Additionally, it is also useful to understand what features contribute negatively to the model and could be excluded. Contributing negatively to the model means reducing the model's accuracy, not reducing the strategy's performance.

The permutation importance technique was used to assess feature importance. The way this technique works is by measuring the change in the model's performance when a particular feature's values are randomly permuted, therefore disrupting the feature's link to the target variable. It was implemented, again, using a rolling window approach, as otherwise, it would not accurately represent the impact on the real model. It was then calculated the average of the importance of each feature. The features that are the most important for the model and the least important ones can be seen in Appendix – Figures 9 and 10, respectively. Going into more detail in the first five most important variables, one can see that these correspond to the AGG US Equity, the US Corporate BAA 10 Year Spread (BICLB10Y Index), the S&P CoreLogic CS US HPI NSA Index (SPCSUSA Index), the New Home Sales index (NHSLTOT Index) and the CPI Core Index SA (CPUPAXFE Index). Starting with the first variable, one can understand

its importance in the model as it is the asset class that has the highest asset allocation in each period, thus having a big impact on the remaining weights that need to be allocated to Equities and Commodities. Moreover, historically, the correlation between stocks and bonds has been negative, therefore, it is expectable that the weight allocation given to either asset class might be intertwined. In fact, in an economic expansion where equity prices are going up, investors will tend to capitalize on that growth in the prices, therefore leading to a lower demand for fixed income instruments, thus decreasing their prices. Regarding commodities, one can say that it has a negative correlation with fixed income, as when commodity prices go up, it creates an inflationary pressure that can lead to an increase in interest rates, which will decrease bond prices, thus, one can understand the reasoning behind the importance of fixed income market instruments when calculating commodity weights. Regarding the second variable, the BICLB10Y Index, one can also understand its importance in an asset allocation problem as it is through changes in spreads that there are changes in the way financial assets are priced, as interest rates are one of the crucial factors when calculating the price of an asset. Thus, it comes as no surprise that this variable is one of the most important features in the model. Regarding the third variable and fourth variable, SPCSUSA Index and NHSLTOT Index, respectively, one can say that there is no direct relationship between these two macroeconomic indicators and the asset classes considered as both these indicators are related to the housing market. However, it is important to note that housing is a very important sector when analysing a country's economic performance. Consequently, the economic performance of a country will affect how financial assets behave and how they are perceived by investors. For instance, when deciding what monetary policies to implement, central banks consider a variety of indicators in their decisions, being housing one of them, and depending on the monetary policies adopted by central banks, financial assets will behave accordingly, which in turn will influence an investor's decision-making process. For the final variable, the CPUPAXFE Index, its

importance in this asset allocation problem shouldn't be questioned. Inflation is one of the main pillars of an economy and it is an indicator that is closely monitored by central banks and investors as high inflation data creates future uncertainty for future interest rates, thus often leading to market volatility. Regarding the macroeconomic variables, one knows that there are variables with a lower impact in the model, therefore the least important ones were excluded from the dataset and the model was applied again. As a result of this analysis, a model excluding the 10 worst features was tested, and even though it led to the final Sharpe ratio of 0.69, in comparison to 0.67 before removing these features, the RMSE increased to 2.17% in comparison to the previous value of 1.13%. A possible explanation behind this counterintuitive result is the existence of non-linearity relationships between features, hence the model performing better when including the less important features (Hölvold & Skenderovic, 2023)

4.3. SECURITIES ANALYSIS

Initially, 9 different securities were considered for the investment universe. The objective was to broaden our scope since it is a multi-asset strategy. Thus, the universe ranged from the traditional asset classes (e.g. bonds and equity) to the more 'exotic', such as Private Equity (comprehensive list of the securities in the Appendix – Figure 2). Nevertheless, even though 9 securities were considered, this was just a starting point. It is generally not recommended to invest in all possible securities, it is necessary a previous analysis – not only to understand securities individually, but also how they perform when combined. Thus, an analysis of the best combinations of securities was performed. Still, if this analysis was to be performed today, it would give rise to the look-ahead bias – one could not decide beforehand, with data not available at the time. Still, it would be very hard to make this backtest analysis before implementing the strategy. The problem arises from the fact that data is only available for all the 9 securities starting in December 2007. The optimal weights using the risk-parity objective function need 14 months of data to be calculated. Additionally, the training period

takes another 3 years, so one would only be available to perform this backtest afterwards, and still needs to wait a few years to have a reasonable period to backtest. The solution we found to the problem was to make the decision based only on the backtest of the optimal weights – target variables – since, if the model performs well, the target variables are the ones that influence the strategy's performance. As such, we positioned ourselves in September 2011, backtesting 30 months of optimal weights obtained through the risk-parity objective function, from the previous years, since March 2009.

It was also tested whether it was preferred to have a smaller or larger number of securities to invest in. Again, we positioned in September 2011 to avoid being biased. All possible combinations of securities were tested, and an average Info Sharpe was calculated for each. In total, with 9 securities, one can invest in 510 different combinations. The combinations were grouped by the number of securities and the average Info Sharpe for each number of securities was calculated – results below.

Number of ETFs	**Average Info Sharpe**
1	0.44
2	0.54
3	0.60
4	0.64
5	0.66
6	0.68
7	0.70
8	0.73
9	0.75

Exhibit 1 – Average Info Sharpe obtained in optimal weights by number of ETFs invested in.

The conclusion is that with our risk-parity objective function, one would have performed better, on average, for the period between March 2009 and August 2011, by

investing in 9 different securities. There was a clear positive relationship between performance and number of securities. It was also tested which were the best-performing combinations overall. It was found that the best-performing strategy would have been invested only in Invesco QQQ Trust Series 1 and iShares Core U.S. Aggregate Bond ETF. Still, there was a preference for diversity of assets, and it is unlikely that the best performer in the past would remain the same in the future. As such, the chosen combination of securities for the strategy was the best within a 5-asset investment universe, which corresponded to the 19th best combination overall for the backtested period. This combination consisted of DIA US Equity, QQQ US Equity, GSG US Equity, SPY US Equity, and AGG US Equity (see full security names in Appendix – Figure 2). Despite the careful analysis, it is in the best interest of the research to include a broader investment period, so even though we simulated an unbiased decision by positioning ourselves in 2011, over 3 years of data would be removed from the analysis if the final model only started after this decision. Nevertheless, the positive/negative impact this bias may have in the final model can easily be tested by running the developed algorithm starting only after the decision.

4.4. TRAINING PERIODS

The model is trained on a rolling window period, i.e., using the last x years to train the model, and the chosen period can influence the model's performance. The different tested periods were 2, 3, and 4 years, and the corresponding RMSE were 1.91%, 1.13%, and 1.11%, respectively. The corresponding Sharpe Ratios were 0.74, 0.67 and 0.58. With these results, one can understand that there is an inverse relationship between the RMSE and Sharpe Ratio of the strategy if the training window is changed. Theoretically, it is expected that a larger window would lead to more accurate weight estimation, as the model considers more data when predicting the weights. When defining the optimal time window, we had to perform a cost- benefit analysis between the Sharpe Ratio and the RMSE. As a result, we didn’t consider 4 years as the optimal rolling window because, even though it presents the lowest RMSE, we

believed that the Sharpe Ratio could be improved significantly without increasing too much the RMSE value. Therefore, after having considered 3 years as the model's rolling window, the Sharpe ratio increased significantly (from 0.58 to 0.67), while the RMSE remained stable (from 1.13% to 1.11%). After these findings, we decided to increase the rolling window, even more, to see if the change in the time window would lead to more improvements, but this wasn't necessarily the case as despite the big increase in the Sharpe Ratio, from 0.67 to 0.74, the RMSE increased from 1.13% to 1.91%, therefore we considered that the increase in the Sharpe wasn't enough to offset the increase in RMSE, therefore, decided to use a 3-year rolling window in the model. More details on the respective performance of the model in each training window can be found in the Appendix – Figure 11.

4.5. DRAWDOWNS AND POOR PERFORMANCE ANALYSIS

When observing the cumulative returns of the strategy, the only major drawdown (-14.17%) started by the end of 2021. Since the COVID-19 Pandemic started in 2020 both the strategy and the benchmark's performance have been affected by the strategy being more severely impacted with the inflation and hawkish environment than the benchmark. The model, due to its risk-parity objective function, tends to overweight bonds, specifically, the AGG US Equity, the fixed-income ETF. One can assess this by testing for predictive causality of the macroeconomic variables, applying the Granger causality test. Indeed, it points towards 7 statistically significant macroeconomic variables considering a p-value of 0.05 (Appendix – Figure 12), of which 1 was an inflation indicator, 5 were interest rates (classified as 'Market'), and 1 was a labour market indicator. Unfortunately, since the strategy of implementing the final model was only possible from October 2010, it was not possible to test it during the Global Financial Crisis.

4.6. IMPLEMENTING THE MODEL WITH NO MACROECONOMIC FUNDAMENTALS

The model seems to work effectively and with high prediction accuracy. However, there is no evidence that the macroeconomic variables cause this. The feature importance analysis suggests that the macroeconomic variables do explain this, as from the 10 most important features, 8 are macroeconomic variables, but there is the possibility that the model can predict optimal weights based on their values in the past. To assess this hypothesis, the model was tested using the same dataset but excluding all macroeconomic variables. The Sharpe ratio remained the same at 0.67, and the model's RMSE changed from 1.13% to 0.65%. One possible reason for this finding could be the correlation between the indexes considered in the analysis. Indeed, out of the 5 indexes considered, 3 are equity indexes that have a big correlation among themselves, therefore, in an asset allocation problem, this indirect correlation could have a significant impact thus potentially mitigating the effect of macroeconomic indicators in the problem. If a certain macroeconomic variable affects one equity index, the presence of correlations among the indexes can imply that the impact might be spread among them, reducing the overall influence of macroeconomic factors on an asset allocation problem. This confirms that indeed it is not proved that the accuracy of the model and the strategy derive from its macroeconomic fundamentals. Nevertheless, it is worth noting that accuracy, alone, can be misleading. If the model was 100% accurate, the model would predict the optimal weights that were obtained using past data, and there would be no advantage. Some inaccuracy is therefore desired: the model should use target variables to learn from past relationships, not exactly to replicate them.

5. CONCLUSION

This research results on a Machine Learning algorithm that intends to bring together the benefits of passive and active investing, by allowing investors to rebalance their portfolio regularly, without the involvement typically demanded by active investing. Additionally, it also provides an opportunity for the individual investor to use quantitative methods, in this case, through Machine Learning. Still, above all, the algorithm is a tool for a strategy that until 2020

outperformed the considered benchmark, on a cumulative returns' comparison. This suggests that the weakening of the strategy and, therefore, the goal of outperforming the benchmark over the backtested period not being attained, may be explained by (1) a poor algorithm in unfamiliar macroeconomic environments, such as the Covid-19 pandemic, leading to a fragile strategy; or (2) the portfolio optimizer itself – Risk Parity strategies tend to overweight bonds, which perform poorly in inflationary environments. Thus, potential improvements of the strategy could include modifications such as making use of cosine similarity instead of rolling windows, to include the most similar past periods instead of the most recent periods, or attempt different optimizers, either on the entire strategy, or by means of a trading signal that would allow to switch to a more adequate optimizer.

Moreover, this project is an extension of other research and projects that have used ML or other Artificial Intelligence algorithms to add value to the finance and investments field, building upon past results seen using macroeconomic variables, extending these to different asset classes, and testing different ML methods and portfolio allocation strategies. Due to the broad range of possible characteristics, methods, and data one can make use of; future research can apply the model to different variables, instead of being macro-based, or different securities. Moreover, we opted for the most accessible securities to the public, even when opting for alternative asset classes, such as Real Estate, to which we used a Real Estate Investment Trust so that we could get exposure to the asset class. This results in a very limited approach to alternative assets. However, with more resources and capital, one could consider direct exposure to this or other asset classes, and test this model, or others, as a tool for different strategies.

Finally, when implementing the strategy, one must also account for other implicit costs that vary according to each investor, such as exchange rates and tax rates applied.

1. APPLICATION OF DIFFERENT PORTFOLIO OPTIMIZATION METHODS

Following the algorithm's construction, the fundamentals of the strategy were reviewed. The decision on the portfolio optimization approach to follow should be given utter importance. The problem adjacent to the portfolio optimization is essentially a constrained utility maximization problem, or minimization, depending on the objective (e.g. minimum variance portfolios). To further improve the strategy, and to complement existing research on portfolio optimization theory, by applying it to a ML-built model, different approaches were analysed.

The methods here studied were (1) Mean-Variance Optimization (MVO); (2) Minimum-Variance Portfolio; (3) Minimum-Semivariance; (4) Risk Parity; and (5) Hierarchical Risk Parity.

The mean-variance optimization is an extension of the Modern Portfolio Theory developed by Harry Markowitz. It attempts to maximize the expected return for a certain risk. It lies on the principle that an investor is risk-averse and, consequently, when faced with two assets with the same return but different risks, the least risky option would be preferred (or the same reasoning but targeting risk instead). Diversification is also a principle of this theory, as a portfolio's variance can be reduced by holding assets that are not perfectly correlated. This research indeed follows the principles of the Modern Portfolio Theory, in the sense that it attempts to build a diversified portfolio, by looking at a multi-asset universe, and recognizes the risk-aversion rationale, by assessing the strategy's performance through its risk-adjusted returns. Despite these resemblances, the MVO was not the method chosen for the researched algorithm, due to its pitfalls. There are some limitations in implementing a mean-variance optimizer: extreme and unstable weights, caused by estimation errors in input parameters. This often leads to poor out-of-sample performance. Some variants tackle this problem through the use of regularization methods (e.g. Lasso and Ridge regression).

The minimum-variance portfolio, as the name suggests, is the portfolio that minimizes overall portfolio risk. It is also built on the Modern Portfolio Theory, in particular, on the efficient frontier. It corresponds to the portfolio with the lowest variance within the efficient frontier (Appendix – Figure 13). Theoretically, it should be the most efficient portfolio for the risk-averse investor. Intuitively, considering the risk-return trade-off principle, this portfolio typically results in lower returns, but offers downside risk protection.

A large argument against this approach is that it measures risk through variance, a symmetric measure. Consequently, when penalizing variance, investors are also penalizing large positive returns, which, naturally, should be favoured. This led to the minimum-semivariance approach (MSVP), which only considers returns as risky when deviating below the mean or a certain threshold (detailed methodology in Section 3 – Minimum Semivariance Portfolio Methodology).

Risk Parity, introduced by Ray Dalio in his "All Weather Portfolio", can follow different approaches, but the most widely known and commonly adopted attempts to build a portfolio with equal risk contributions (ERCs). Mathematically, this problem can be expressed by Equation 2.

$$\arg\min_{w} \sum_{i=1}^{N} \left[w_i - \frac{\sigma(w)^2}{(\Sigma w)_i \mathrm{N}} \right]^2 \tag{2}$$

Where:

N = Number of assets

w_i = Weight allocated to asset i

w = Weights vector formed by w_i of each asset

Σ = Covariance matrix

$\sigma(w)$ = Standard deviation of the portfolio

This can be seen as a constrained minimum variance portfolio, subject to the constraint that the assets' risk contribution should be equal (Maillard et al. 2009, 1). The common objective

behind the different approaches is to allocate risk instead of capital, and its resilience to endure economic downturns is an argument in favour of this method. However, Risk Parity is highly conditional on the investment universe (Shakernia et al. 2010, 1). Hence, the selection of the investment universe is a crucial step for the strategy to perform well. Then again, the objective of the development of this strategy is to make investors' lives easier, and a careful selection of the investment universe can demand significant efforts, as one should consider different characteristics for each asset class, past performance, expense ratios, etc., to end up with a robust and "trust-worthy" investment universe. Since this is not the desired procedure, other options should be considered. The previously mentioned optimizers are some of the alternatives, but research has led to the development of a Risk Parity approach that, in theory, should not require such a careful selection of assets. This alternative is known as Hierarchical Risk Parity (HRP)(López de Prado 2016), and it attempts to avoid this pitfall by including a more careful analysis of the similarities between different securities, applying a hierarchy to the different assets through statistical procedures described in Section 2 – Understanding Hierarchical Risk Parity Optimization. Another recognized advantage is that it does not require the inversion of the covariance matrix, which can lead to unstable and extreme results.

The different optimizers were assessed and backtested on the entire investment universe (Exhibit 2). In practice, several modifications are usually applied to these optimizers, such as regularization, to ensure robustness and numerical stability, but we here applied the 'pure' optimizers, as considering all typical adaptations would be too broad and falls behind of the scope of this study. Still, better results can be achieved combining these optimizers with certain considerations. Price data has been available for all securities since January 2008. However, this is not the beginning of the backtest period since the optimizer requires past data to calculate future weights. The window sizes of past data here tested vary between 14 and 54, so the beginning of the backtest period varies accordingly. Using different backtest periods might not

be the best approach but the goal is to compare different optimizers overall and, as such, choosing one single window for the analysis would be of little added value, as some optimizers perform better on different window sizes than others. The spike in performance for the 14-months window indeed results from including the rebound after the 2008 crash, not due to a particular superiority of considering this window size over different sizes. Still, by excluding certain periods, the larger window sizes performed worse in general.

The optimizers are not directly included in the Machine Learning model, but rather in the definition of the target variables the model studies. Thus, the results mentioned in this section were not observed after running the model, but beforehand. It is an attempt to improve the model by increasing the quality of its input data. The production of the variables followed the same unbiased procedure as the target variables used in the model: implementing an optimizer that outputs weights based on information available at the time and calculated on a rolling basis.

The heatmap below illustrates the different risk-adjusted returns, through the Info Sharpe indicator, of the different optimizers. *Note: a line chart can also be seen in the Appendix – Figure 14, for easier interpretability.*

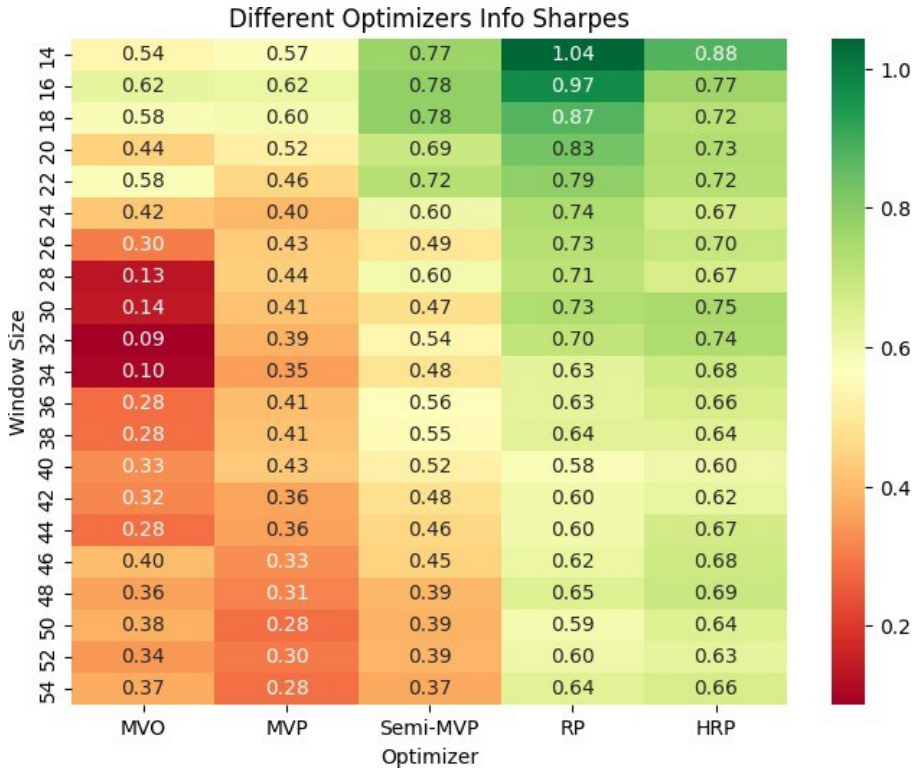

Exhibit 2 – Performance comparison (Info Sharpe) of different optimizers, considering different window sizes of past historical data (heatmap chart).

Risk Parity consistently outperforms all other methods for every window size until 28 months, while HRP outperforms for the larger window sizes considered, albeit with less pronounced differences. This suggests that despite its efforts to avoid Risk Parity's weaknesses, it is not clear that HRP is a more optimal approach. In particular, considering the resemblance among certain securities (e.g. ETFs on the SP&500, Dow Jones Industrial Average, and NASDAQ), HRP could distinguish itself through its hierarchical approach. Additionally, robust results obtained with smaller window sizes are an advantage for the ML model, since it already requires an extensive training period. The MVP also displays a tendency toward better performance with shorter window sizes, with a decrease in the Info Sharpe from 0.76, considering 10 months of historical data, to 0.30, considering 54 months. The Mean-Variance Optimizer exhibits unstable results, with Info Sharpe levels varying between 0.16 and 0.60. The results are in line with prior research on the limitations of this method.

Examining the different weights each strategy allocated to each security, on average (Appendix – Figure 15), a general trend emerges – most optimizers tend to allocate a significant portion to bonds. MVP, due to its inherent risk aversion, was the portfolio that allocated the largest weight to bonds (nearly 91%). In contrast, HRP typically allocates only around 2% to bonds, and most of the portfolio to equity. Within similar assets, such as the three equity ETFs considered, RP allocates nearly equal weights – around 6% each, while HRP, leveraging its hierarchical feature, concentrates most of its equity allocation on a single ETF. Plotting the dendrogram of the resulting clusters obtained reveals the similarity between the equity ETFs ("QQQ", "DIA", and "SPY" tickers) (Appendix – Figure 16).

It is important to notice that a different selection of securities could lead to different results. One could further investigate the optimizers on a different number of assets and combine this with a careful selection of the same. Still, the use of ETFs for this research already allows to indirectly reach a broad range of securities.

Considering the novelty of Hierarchical Risk Parity, despite it not being a clear "winner" based on the backtesting results here studied, and of Minimum-Semivariance, there is still considerable research to be done on these allocation methods. Thus, more emphasis was given to them in this paper than the other methods, which already count with decades of research and various modifications and extensions.

2. UNDERSTANDING HIERARCHICAL RISK PARITY OPTIMIZATION

The mathematical process of Risk Parity is somewhat straightforward, but Hierarchical Risk Parity is a bit more complex. López de Prado introduced the HRP framework in 2016. In his paper, "Building Diversified Portfolios that Outperform Out of Sample", he breaks down the algorithm into three major steps: (1) Tree clustering, (2) Quasi-diagonalization, and (3) Recursive bisection.

One of the drawbacks of Risk Parity is that it requires the inversion of the covariance matrix (see Equation 2), which can result in sensitive and unstable values. Thus, to understand the process below explained, it is key to understand that the main goal is to allocate weights optimally without inverting the covariance matrix, and to transmit a sense of hierarchy to the weights allocation.

The first step, tree clustering is essentially a more rigorous method to achieve what investors usually attempt to do when investing, for example, in companies of different sectors. The rationale behind this is to increase diversification and reduce the portfolio's risk by investing in less correlated assets. Thus, in this step, securities are clustered in a hierarchical structure. In the second step, the covariance matrix is reorganized in a way that the diagonal displays the largest values. In other words, the most similar assets are arranged next to each other, over the diagonal, while distinct assets are further away from the diagonal. In the final step, one of the most common Risk Parity approaches, the inverse variance, is applied. However, it is implemented on a quasi-diagonal covariance matrix, which is optimal (see proof on López de Prado's original paper). Moreover, the calculations are not conducted with the overall matrix. Instead, as the name "recursive bisection" suggests, the process iterates over pairs of adjacent subsets. This results in an important benefit of HRP: making use of a hierarchy. The detailed explanation of the entire algorithm can be found below:

1) Tree clustering
    a. Compute $N \times N$ correlation matrix;
    b. Convert to a correlation-distance matrix D, in which each (I,j) element is: $D(i,j) = \sqrt{\frac{1}{2} \times (1 - \rho(i,j))}$. The reason for this conversion is that correlation by itself does not hold certain mathematical properties that a distance measure does (non-negativity, coincidence, symmetry, and sub-additivity). Thus, in D we are essentially measuring the distance between different asset returns;

c. Compute Euclidean-distance matrix $\bar{D}$ between any two column-vectors of D in which each *(i,j)* element is $\bar{D}(i,j) = \sqrt{\sum_{k=1}^{N} (D(k,i) - - D(k,j))^2}$. The reasoning behind this step might be not so obvious. $\bar{D}$ is measuring the similarity of assets' returns through their overall distance from the other assets. If a pair of assets has a similar behaviour, their distance from other assets will also be similar and, thus, the Euclidean distance will tend to 0. This way, interactions between all assets are considered instead of simple pairs' correlations.

d. The first cluster is defined, by joining the pair of different assets ($i \neq j$) that are the most similar, i.e. with the lowest $\bar{D}(i,j)$ value ("linkage criterion");

e. Matrix $\bar{D}$ should be updated considering this cluster's distance from other assets, like this cluster was one single asset part of the matrix. This distance can be calculated through different methods since there are multiple elements in a cluster to consider the distance from. The method used here is to consider the distance between the closest elements' pair;

f. This new distance is updated in the $\bar{D}$ matrix, and the columns and rows of assets that belong to the cluster are dropped;

g. Repeat steps d to f until N-1 clusters have been appended. The result is a linkage matrix of dimensions (N-1)×4;

2) Quasi-diagonalization

a. "Decluster" the linkage matrix recursively, i.e. iteratively replace clusters with their constituents;

b. Sort and reindex the constituents based on similarity;

3) Recursive bisection

a. Initialize set of weights to 1;

b. Loop through the sorted list of assets outputted in step 2, and repeatedly bisect the list into halves, until each subset contains only a single asset;

c. Group the clusters into pairs and loop through each pair. For each pair:

d. Calculate the variance of each cluster, $\tilde{V}_i$, with Equation (2), in which the weights were obtained through the inverse-variance allocation and normalization of these weights;

e. Allocate weights to each cluster based on their contribution to the pair's volatility: $w_1 = 1 - \frac{V_1}{V_1 + V_2}$; $w_2 = 1 - w_1$. **Note:** this approach is top-down, meaning the clusters that are looped first and assigned their corresponding weights are the ones with the most assets, and the loop ends with the assignment of the individual weights.

3. MINIMUM SEMIVARIANCE PORTFOLIO METHODOLOGY

The minimum semivariance portfolio follows the same reasoning as the minimum variance portfolio – it is the portfolio with the least risk within the efficient frontier. The difference lies in the risk measure, which in this method would be semivariance. Other past attempts to avoid only downside risk over symmetric variance have used more complex methods and that led to this approach not gaining much adherence.

The approach here followed is the same as in the minimum variance portfolio approach, but by means of mean-semivariance optimization instead. Semivariance can be calculated with the below formula:

$$\text{Semivariance} = \frac{1}{n} \times \sum_{r_t < \text{Average}}^{n} (\mu - r_t)^2 \qquad (3)$$

Where:

n = Number of observations below the mean or target

r_t = Observed return

μ = Mean or target value of the dataset

Given the corresponding window's past returns, and a benchmark return, a convex optimization is performed, in this case, the minimization of the portfolio's semivariance. Although it is a process that can be easily implemented by means of an algorithm, it may result on some issues as numerical instability.

From Equation 3, one can understand that this method limits the number of observations, by considering only below the mean/target returns. This is one of the limitations of this approach, and Markowitz pointed out that, for this reason, it makes more sense to use semivariance instead of variance when assets' returns are significantly asymmetric (Markowitz 2020, 372). Considering this observation, the skewness of individual assets' returns was assessed. Five of the nine securities had a skew that did not fall between -0.5 and 0.5. From these, all were negatively skewed, meaning they have a longer tail on the negative side (Appendix – Figure 17). This may explain the clear outperformance of the minimum semivariance portfolio over the minimum variance portfolio and different assets might have led to different results. A contradictory observation, however, would be that larger window sizes resulted in worse Info Sharpes. Smaller window sizes, together with the below target/mean constraint, diminishes the number of observations even further. This should negatively impact the strategy, but that was not observed. It could be argued that the decline in performance is explained by the different backtest periods considered in each period. Indeed, in general, the optimizers performed better on narrower windows.

4. EXTENSIONS TO THE HIERARCHICAL RISK PARITY ALLOCATION

The HRP algorithm suggested by López de Prado follows an inverse-variance approach. This may lead to overweighting assets with lower volatilities. Equal Risk Contributions are an alternative approach that could reduce concentration risk. This can be implemented in HRP by adapting the step 3.d) described in Section 2. Another suggested modification to the algorithm is the replacement of the sample covariance by an exponentially weighted covariance matrix

with Ledoit-Wolf shrinkage (Molyboga 2020, 1). The underlying principle to this modification is that recent data should be given more importance than older observations and, thus, more weight should be assigned to recent observations. Following this step, a Ledoit-Wolf shrinkage is employed to improve robustness by diminishing estimation errors, through a structured estimator diag(*S*). This relies on the assumption that the covariance is close to a diagonal matrix, meaning that the assets are nearly independent. This is a bold assumption and might be unrealistic. To further study the feasibility and impact of these modifications, the corresponding variations were applied to the HRP algorithm and again, backtested.

Exhibit 3 – Performance comparison (Info Sharpe) of different modifications applied to the Hierarchical Risk Parity algorithm considering different window sizes.

The noticeable observation is the heightened instability evident in the Equal-Risk Contributions modification, with higher peaks and throughs in performance. Regarding the adjustments to the covariance matrix – replacement of the covariance matrix by a Ledoit-Wolf shrinkage of an exponentially weighted covariance matrix–- this modification resulted only in neglectable differences. Still, it is an interesting result, as despite the near independency assumption it relies on, it demonstrates comparable performance to the conventional Hierarchical Risk Parity. This analysis leads to the conclusion that it is not yet clear that suggested modifications benefit the Hierarchical Risk Parity allocation.

5. DISCUSSION

An overview of different optimizers is a central step for the construction of a systematic investment strategy. Yet, research evolves, and new suggestions come along. This study attempts to (1) shed light on different portfolio optimization methods, (2) constrast them with conventional portfolio optimization methods, (3) stress the backtest scenario, by testing different window sizes of past historical data, different asset classes, different numbers of constituents in the portfolio, and (4) review recent modifications and developments on the Hierarchical Risk Parity.

Key findings were the outperformance of Risk Parity against traditional allocation methods, HRP, and the Minimum Semivariance Portfolio, when using more recent data, i.e. smaller window sizes, until 28 months. For larger window sizes, the results are less clear, with Risk Parity and Hierarchical Risk Parity as the top 2 performers during the backtest, with marginal differences in performance. Mean-Variance Optimization displays inconsistent results, achieving a wide range of Info. Additionally, the Minimum Semivariance Portfolio mostly outperformed the Minimum Variance Portfolio in about 0.20 risk-adjusted returns, an important insight for the risk-averse investor. Most of the ETFs studied have asymmetric historical returns, so the choice between the two methods under circumstances in which the considered assets for the investment universe have historical symmetric returns should be carefully thought before. Finally, modifications to the Hierarchical Risk Parity appear to lack evident positive impacts on performance. Nevertheless, it is the opinion of the author that further developments of this allocation method, and exploitations of some of the inherent concepts, such as the sense of hierarchy, should continue to be studied. The geometric approach to financial problems, as López de Prado employed in the development of the Hierarchical Risk Parity through the replacement of the covariance matrix for the distance matrix, is also a tactic

with potential for continued research findings, due to its focus on the structure of the problems at hand.

Limitations exist within this studies' findings. Computational resources required to backtest the different optimizers and window sizes constrained the performance assessment, therefore it was opted to use only one performance indicator, the Info Sharpe. The investment universe was also limited for the same reason. A broader comprehension of the performance using standard metrics, such as annualized return and skewness would complement the results. Nevertheless, the Info Sharpe is still an insightful indicator for the fact that the investor can make use of leverage to achieve higher returns if desired.

6. REFERENCES

Ang, Andrew, David Chua, Katelyn Gallagher, and Stephen Hull. 2020. "BlackRock: Reserves Management with Factors and Reference Portfolios." In *Asset Management at Central Banks and Monetary Authorities*, edited by Jacob Bjorheim, 459–84. Cham: Springer.

Behr, Patrick, Guettler, Andre, and Felix Miebs. 2013." On portfolio optimization: Imposing the right constraints". *The Journal of Banking and Finance* 37, 4: 1232–1242. https://doi.org/10.1016/j.jbankfin.2012.11.020

Brixton, Alfie, Jordan Brooks, Peter Hecht, Antti Ilmanen, Thomas Maloney, and Nicholas McQuinn. 2023. "A Changing Stock-Bond Correlation.". *AQR Research* 49: 4. https://www.aqr.com/Insights/Research/Journal-Article/A-Changing-Stock-Bond-Correlation.

Buncic, Daniel, and Martin Tischhauser. 2017. "Macroeconomic Factors and Equity Premium Predictability." *International Review of Economics & Finance* 51: 621–44. https://doi.org/10.1016/j.iref.2017.07.006

Chaves, Denis, Jason Hsu, Feifei Li, and Omid Shakernia. 2011. "Risk Parity Portfolio vs. Other Asset Allocation Heuristic Portfolios." *The Journal of Investing* 20, 1: 108–18. https://doi.org/10.3905/joi.2011.20.1.108

Choi, Darwin, Wenxi Jiang, and Chao Zhang. 2019. "Alpha Go Everywhere: Machine Learning and International Stock Returns." https://doi.org/10.2139/ssrn.3489679.

Daul, Stéphane, Thibault Jaisson, and Alexandra Nagy. 2022. "Performance Attribution of Machine Learning Methods for Stock Returns Prediction." *The Journal of Finance and Data Science* 8, 8: 86–104. https://doi.org/10.1016/j.jfds.2022.04.002.

De Boer, Sanne, Julian Keuerleber, and Carsten Rother. 2018. "Performance Attribution through a Factor Lens." *Social Science Research Network*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3209424.

Desai, Vijay S, and Rakesh Bharati. 1998. “The Efficacy of Neural Networks in Predicting Returns on Stock and Bond Indices.” *Decision Sciences* 29, 2: 405–23. https://doi.org/10.1111/j.1540-5915.1998.tb01582.x.

Frost, Peter A., and James E. Savarino. 1988. “For Better Performance.” *The Journal of Portfolio Management* 15*,* 1: 29–34. https://doi.org/10.3905/jpm.1988.409181.

Georgiev, Georgi. 2001. “Benefits of Commodity Investment.” *The Journal of Alternative Investments* 4, 1: 40–48. https://doi.org/10.3905/jai.2001.318997.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. “Empirical Asset Pricing via Machine Learning.” *The Review of Financial Studies* 33, 5. https://doi.org/10.1093/rfs/hhaa009.

Leung, Edward, Harald Lohre, David Mischlich, Yifei Shea, and Maximilian Stroh. 2021. “The Promises and Pitfalls of Machine Learning for Predicting Stock Returns.” *The Journal of Financial Data Science* 3, 2: 21–50. https://doi.org/10.3905/jfds.2021.1.062.

Lin, Hai, Chunchi Wu, and Guofu Zhou. 2018. “Forecasting Corporate Bond Returns with a Large Set of Predictors: An Iterated Combination Approach.” *Management Science* 64, 9: 4218–38. https://doi.org/10.1287/mnsc.2017.2734.

López de Prado, Marcos. 2016. “Building Diversified Portfolios that Outperform Out of Sample.” *The Journal of Portfolio Management* 42*,* 4: 59–69. https://doi.org/10.3905/jpm.2016.42.4.059

Ludvigson, Sydney C., and Serena Ng. 2009. “Macro Factors in Bond Risk Premia.” *Review of Financial Studies* 22, 12: 5027–67. https://doi.org/10.1093/rfs/hhp081.

Maillard, Sébastien, Thierry Roncalli, and Jérôme Teïletche. 2010. “The Properties of Equally Weighted Risk Contribution Portfolios” The Journal of Portfolio Management 36, 4: 60–70. https://doi.org/10.3905/jpm.2010.36.4.060.

Markowitz, Harry M., David Starer, Harvey Fram, Sander Gerber. 2020. “Avoiding the Downside: A Practical Review of the Critical Line Algorithm for Mean-Semivariance Portfolio Optimization.” In *Handbook of Applied Investment Research,* edited by John B. Guerard and William T. Ziemba. World Scientific.

Molyboga, Marat. 2023. “A Modified Hierarchical Risk Parity Framework for Portfolio Management.” *The Journal of Financial Data Science* 2, 3: 128-139. https://doi.org/10.3905/jfds.2020.1.038

Rasekhschaffe, Keywan Christian, and Robert C. Jones. 2019. “Machine Learning for Stock Selection.” *Financial Analysts Journal* 75, 3: 70–88. https://doi.org/10.1080/0015198x.2019.1596678.

Simões, Ana Bárbara. 2022. “A Machine Learning Algorithm Applied to Macroeconomic Factor Investing.” Master’s thesis, Nova School of Business and Economics. http://hdl.handle.net/10362/144695.

T. Kwok, James, Zhi-Hua Zhou, and Lei Xu. 2015. “Machine Learning.” In *Handbook of Computational Intelligence*, edited by Witold Pedrycz and Janusz Kacprzyk, 495–522. Berlin Heildeberg: Springer.

Tsai, Chih-Fong, Yuah-Chiao Lin, David C. Yen, and Yan-Min Chen. 2011. “Predicting Stock Returns by Classifier Ensembles.” *Applied Soft Computing* 11, 2: 2452–59. https://doi.org/10.1016/j.asoc.2010.10.001.

Welch, Ivo, and Amit Goyal. 2007. “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction.” *Review of Financial Studies* 21, 4: 1455–1508. https://doi.org/10.1093/rfs/hhm014.

Zou, Jinming, Yi Han, and Sung-Sau So. 2009. “Overview of Artificial Neural Networks.” In *Artificial Neural Networks*, edited by David J. Livingstone. Totowa: Humana.

7. APPENDIX

Ticker	Name	Category	Frequency
SBOITOTL INDEX	NFIB Small Business Optimism	Consumer Confidence	M
CONSSENT Index	U. of Mich. Sentiment	Consumer Confidence	M
CICRTOT Index	Consumer Credit	Credit	M
MOODCAAA Index	Moody's Corporate Bond Spread AAA	Credit	D
MOODCBAA Index	Moody's Corporate Bond Spread BAA	Credit	D
BASPCAAA Index	US Corporate AAA 10 Yr Spread	Credit	D
BICLB10Y Index	US Corporate BAA 10 Year Spread	Credit	D
IP YOY Index	US Industrial Production YOY	Economic Activity	M
USTGTTCB Index	Advance Goods Trade Balance	Economic Activity	M
NHSPATOT Index	Building Permits	Economic Activity	M
NHCHATCH Index	Building Permits MoM	Economic Activity	M
MTIBCHNG Index	Business Inventories	Economic Activity	M
CGNOXAI% Index	Cap Goods Orders Nondef Ex Air	Economic Activity	M
CGSHXAI% Index	Cap Goods Ship Nondef Ex Air	Economic Activity	M
CPTICHNG Index	Capacity Utilization	Economic Activity	M
CFNAI Index	Chicago Fed Nat Activity Index	Economic Activity	M
CONCCONF Index	Conf. Board Consumer Confidence	Economic Activity	M
CONCEXP Index	Conf. Board Expectations	Economic Activity	M
CONCPSIT Index	Conf. Board Present Situation	Economic Activity	M
EXP1CMOM Index	Export Price Index MoM	Economic Activity	M
EXP1CYOY Index	Export Price Index YoY	Economic Activity	M
IMP1XPM% Index	Import Price Index ex Petroleum MoM	Economic Activity	M
IMP1CHNG Index	Import Price Index MoM	Economic Activity	M
IMP1YOY% Index	Import Price Index YoY	Economic Activity	M
IP CHNG Index	Industrial Production MoM	Economic Activity	M

FRNTTNET Index	Total Net TIC Flows	Economic Activity	M
CONSCURR Index	U. of Mich. Current Conditions	Economic Activity	M
SAARTOTL Index	Wards Total Vehicle Sales	Economic Activity	M
FLSLCHAN Index	Bureau of Economic Analysis Real Final Sales of GDP Dollars SA Chain 2012 Price	Entire economy	Q
FLSLCHA% Index	Bureau of Economic Analysis Real Final Sales of GDP Dollars SA Chain 2012 Price	Entire economy	Q
COMFBTWR Index	Langer Economic Expectations	Entire economy	M
LEI CHNG Index	Leading Index	Entire economy	M
FDDSSD Index	Monthly Budget Statement	Entire economy	M
CONSEXP Index	U. of Mich. Expectations	Entire economy	M
NHSPSTOT Index	Housing Starts	Housing	M
NHCHSTCH Index	Housing Starts MoM	Housing	M
USHBMIDX Index	NAHB Housing Market Index	Housing	M
NHSLTOT Index	New Home Sales	Housing	M
NHSLCHNG Index	New Home Sales MoM	Housing	M
SPCSUSA Index	S&P CoreLogic CS US HPI NSA Index	Housing	M
SPCSUSAY Index	S&P CoreLogic CS US HPI YoY NSA	Housing	M
CPUPAXFE Index	CPI Core Index SA	Inflation	M
CPUPXCHG Index	CPI Ex Food and Energy MoM	Inflation	M
CPI XYOY Index	CPI Ex Food and Energy YoY	Inflation	M
CPURNSA Index	CPI Index NSA	Inflation	M
CPI CHNG Index	CPI MoM	Inflation	M
CPI YOY Index	CPI YoY	Inflation	M
DGNOXTCH Index	Durables Ex Transportation	Inflation	M
PCE CMOM Index	PCE Core Deflator MoM	Inflation	M
PCE CYOY Index	PCE Core Deflator YoY	Inflation	M
PCE DEFM Index	PCE Deflator MoM	Inflation	M

PCE DEFY Index	PCE Deflator YoY	Inflation	M
PITLCHNG Index	Personal Income	Inflation	M
PCE CRCH Index	Personal Spending	Inflation	M
PCE CHNC Index	Real Personal Spending	Inflation	M
CONSPXMD Index	U. of Mich. 1 Yr Inflation	Inflation	M
CONSP5MD Index	U. of Mich. 5-10 Yr Inflation (Inflation Expectation)	Inflation	M
NFP TCH Index	Change in Nonfarm Payrolls	Labor market	M
NFP PCH Index	Change in Private Payrolls	Labor market	M
INJCJC Index	Initial Jobless Claims	Labor market	W
NAPMEMPL Index	ISM Employment	Labor market	M
PRUSTOT Index	Labor Force Participation Rate	Labor market	M
USURTOT Index	Unemployment Rate	Labor market	M
USMMMNCH Index	Change in Manufact. Payrolls	Manufacturing & Industrial	M
TMNOCHNG Index	Factory Orders	Manufacturing & Industrial	M
NAPMPMI Index	ISM Manufacturing	Manufacturing & Industrial	M
NAPMNEWO Index	ISM New Orders	Manufacturing & Industrial	M
NAPMPRIC Index	ISM Prices Paid	Manufacturing & Industrial	M
IPMGCHNG Index	Manufacturing (SIC) Production	Manufacturing & Industrial	M
CHPMINDX Index	MNI Chicago PMI	Manufacturing & Industrial	M
OUTFGAF Index	Philadelphia Fed Business Outlook	Manufacturing & Industrial	M

IP Index	US Industrial Production	Manufacturing & Industrial	M
H15T10Y Index	Constant Maturity Rate 10Y US Fixed Income	Market	D
H15T1Y Index	Constant Maturity Rate 1Y US Fixed Income	Market	D
H15T20Y Index	Constant Maturity Rate 20Y US Fixed Income	Market	D
H15T2Y Index	Constant Maturity Rate 2Y US Fixed Income	Market	D
H15T30Y Index	Constant Maturity Rate 30Y US Fixed Income	Market	D
H15T3M Index	Constant Maturity Rate 3M US Fixed Income	Market	D
H15T3Y Index	Constant Maturity Rate 3Y US Fixed Income	Market	D
H15T5Y Index	Constant Maturity Rate 5Y US Fixed Income	Market	D
H15T7Y Index	Constant Maturity Rate 7Y US Fixed Income	Market	D
EURUSD Curncy	EUR USD Exchange Rate	Market	D
GB12 Govt	Generic US 1Yr Government Bill	Market	D
US0003M Index	ICE LIBOR USD 3 Month	Market	D
FRNTTOTL Index	Net Long-term TIC Flows	Market	M
SMART MONEY FLOW INDEX	SMART MONEY INDEX	Market	D
USGG10YR Index	US Generic Govt 10 Yr	Market	D

Figure 1 – Macroeconomic variables information

Ticker	Name	Asset Class Focus
AGG US Equity	iShares Core U.S. Aggregate Bond ETF	Fixed Income
DIA US Equity	SPDR Dow Jones Industrial Average ETF Trust	Equity
EMB US Equity	iShares J.P. Morgan USD Emerging Markets Bond ETF	Fixed Income
GLD US Equity	SPDR Gold Shares	Commodities
GSG US Equity	iShares S&P GSCI Commodity-Indexed Trust	Commodities
PLD US Equity	Prologis Inc	Real Estate
PSP US Equity	Invesco Global Listed Private Equity ETF	Private Equity
QQQ US Equity	Invesco QQQ Trust Series 1	Equity
SPY US Equity	SPDR S&P 500 ETF Trust	Equity

Figure 2 – Securities studied and corresponding information

Ticker	Name	Lag
SBOITOTL INDEX	NFIB Small Business Optimism	1-month
CONSSENT Index	U. of Mich. Sentiment	No lag
CICRTOT Index	Consumer Credit	1-month
MOODCAAA Index	Moody's Corporate Bond Spread AAA	No lag
MOODCBAA Index	Moody's Corporate Bond Spread BAA	No lag
BASPCAAA Index	US Corporate AAA 10 Yr Spread	No lag
BICLB10Y Index	US Corporate BAA 10 Year Spread	No lag
IP YOY Index	US Industrial Production YOY	1-month
USTGTTCB Index	Advance Goods Trade Balance	1-month
NHSPATOT Index	Building Permits	1-month
NHCHATCH Index	Building Permits MoM	1-month
MTIBCHNG Index	Business Inventories	2-month
CGNOXAI% Index	Cap Goods Orders Nondef Ex Air	1-month

CGSHXAI% Index	Cap Goods Ship Nondef Ex Air	1-month
CPTICHNG Index	Capacity Utilization	1-month
CFNAI Index	Chicago Fed Nat Activity Index	1-month
CONCCONF Index	Conf. Board Consumer Confidence	No lag
CONCEXP Index	Conf. Board Expectations	No lag
CONCPSIT Index	Conf. Board Present Situation	No lag
EXP1CMOM Index	Export Price Index MoM	1-month
EXP1CYOY Index	Export Price Index YoY	1-month
IMP1XPM% Index	Import Price Index ex Petroleum MoM	1-month
IMP1CHNG Index	Import Price Index MoM	1-month
IMP1YOY% Index	Import Price Index YoY	1-month
IP CHNG Index	Industrial Production MoM	1-month
FRNTTNET Index	Total Net TIC Flows	2-month lag
CONSCURR Index	U. of Mich. Current Conditions	No lag
SAARTOTL Index	Wards Total Vehicle Sales	1-month
FLSLCHAN Index	Bureau of Economic Analysis Real Final Sales of GDP Dollars SA Chain 2012 Price	1-quarter
FLSLCHA% Index	Bureau of Economic Analysis Real Final Sales of GDP Dollars SA Chain 2012 Price	1-quarter
COMFBTWR Index	Langer Economic Expectations	1-month
LEI CHNG Index	Leading Index	1-month
FDDSSD Index	Monthly Budget Statement	1-month
CONSEXP Index	U. of Mich. Expectations	No lag
NHSPSTOT Index	Housing Starts	1-month
NHCHSTCH Index	Housing Starts MoM	1-month
USHBMIDX Index	NAHB Housing Market Index	No lag
NHSLTOT Index	New Home Sales	1-month
NHSLCHNG Index	New Home Sales MoM	1-month

SPCSUSA Index	S&P CoreLogic CS US HPI NSA Index	2 month
SPCSUSAY Index	S&P CoreLogic CS US HPI YoY NSA	2-month
CPUPAXFE Index	CPI Core Index SA	1-month
CPUPXCHG Index	CPI Ex Food and Energy MoM	1-month
CPI XYOY Index	CPI Ex Food and Energy YoY	1-month
CPURNSA Index	CPI Index NSA	1-month
CPI CHNG Index	CPI MoM	1-month
CPI YOY Index	CPI YoY	1-month
DGNOXTCH Index	Durables Ex Transportation	1-month
PCE CMOM Index	PCE Core Deflator MoM	1-month
PCE CYOY Index	PCE Core Deflator YoY	1-month
PCE DEFM Index	PCE Deflator MoM	1-month
PCE DEFY Index	PCE Deflator YoY	1-month
PITLCHNG Index	Personal Income	1-month
PCE CRCH Index	Personal Spending	1-month
PCE CHNC Index	Real Personal Spending	1-month
CONSPXMD Index	U. of Mich. 1 Yr Inflation	No lag
CONSP5MD Index	U. of Mich. 5-10 Yr Inflation (Inflation Expectation)	No lag
NFP TCH Index	Change in Nonfarm Payrolls	1-month
NFP PCH Index	Change in Private Payrolls	1-month
INJCJC Index	Initial Jobless Claims	No lag
NAPMEMPL Index	ISM Employment	1-month
PRUSTOT Index	Labor Force Participation Rate	1-month
USURTOT Index	Unemployment Rate	1-month
USMMMNCH Index	Change in Manufact. Payrolls	1-month
TMNOCHNG Index	Factory Orders	2-month

NAPMPMI Index	ISM Manufacturing	1-month
NAPMNEWO Index	ISM New Orders	1-month
NAPMPRIC Index	ISM Prices Paid	1-month
IPMGCHNG Index	Manufacturing (SIC) Production	1-month
CHPMINDX Index	MNI Chicago PMI	No lag
OUTFGAF Index	Philadelphia Fed Business Outlook	No lag
IP Index	US Industrial Production	1-month
H15T10Y Index	Constant Maturity Rate 10Y US Fixed Income	No lag
H15T1Y Index	Constant Maturity Rate 1Y US Fixed Income	No lag
H15T20Y Index	Constant Maturity Rate 20Y US Fixed Income	No lag
H15T2Y Index	Constant Maturity Rate 2Y US Fixed Income	No lag
H15T30Y Index	Constant Maturity Rate 30Y US Fixed Income	No lag
H15T3M Index	Constant Maturity Rate 3M US Fixed Income	No lag
H15T3Y Index	Constant Maturity Rate 3Y US Fixed Income	No lag
H15T5Y Index	Constant Maturity Rate 5Y US Fixed Income	No lag
H15T7Y Index	Constant Maturity Rate 7Y US Fixed Income	No lag
EURUSD Curncy	EUR USD Exchange Rate	No lag
GB12 Govt	Generic US 1Yr Government Bill	No lag
US0003M Index	ICE LIBOR USD 3 Month	No lag
FRNTTOTL Index	Net Long-term TIC Flows	2-month
SMART MONEY FLOW INDEX	SMART MONEY INDEX	No lag
USGG10YR Index	US Generic Govt 10 Yr	No lag

Figure 3 – Macroeconomic variables and corresponding lags

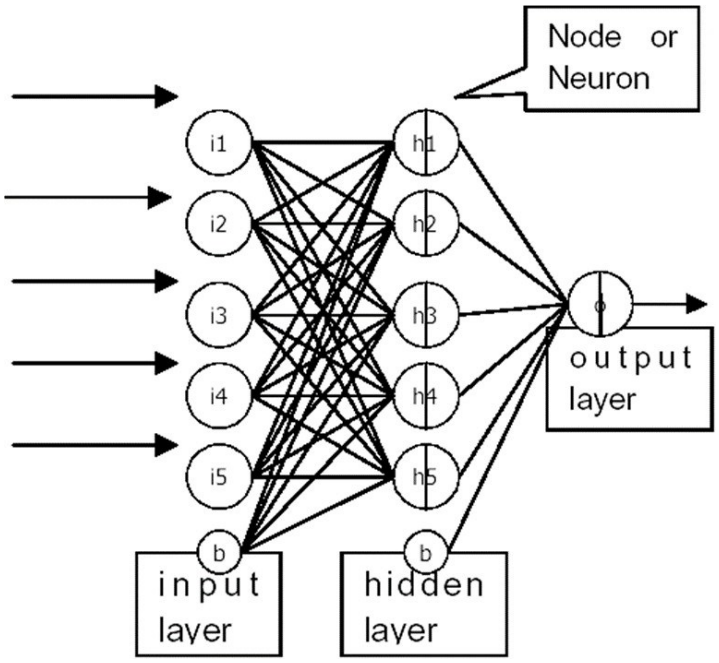

Figure 4 – Multilayer Perceptron (MLP) model scheme

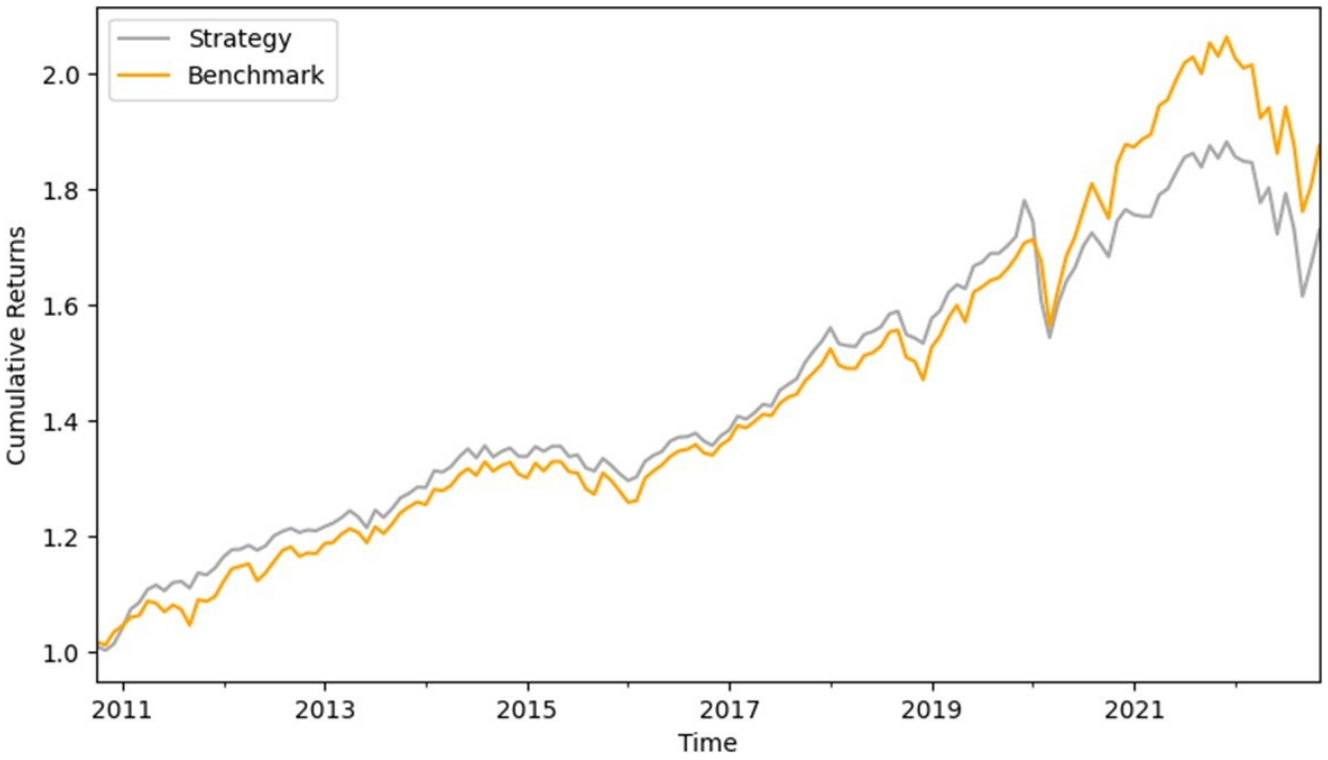

Figure 5 – Comparison between cumulative returns of the model and the benchmark

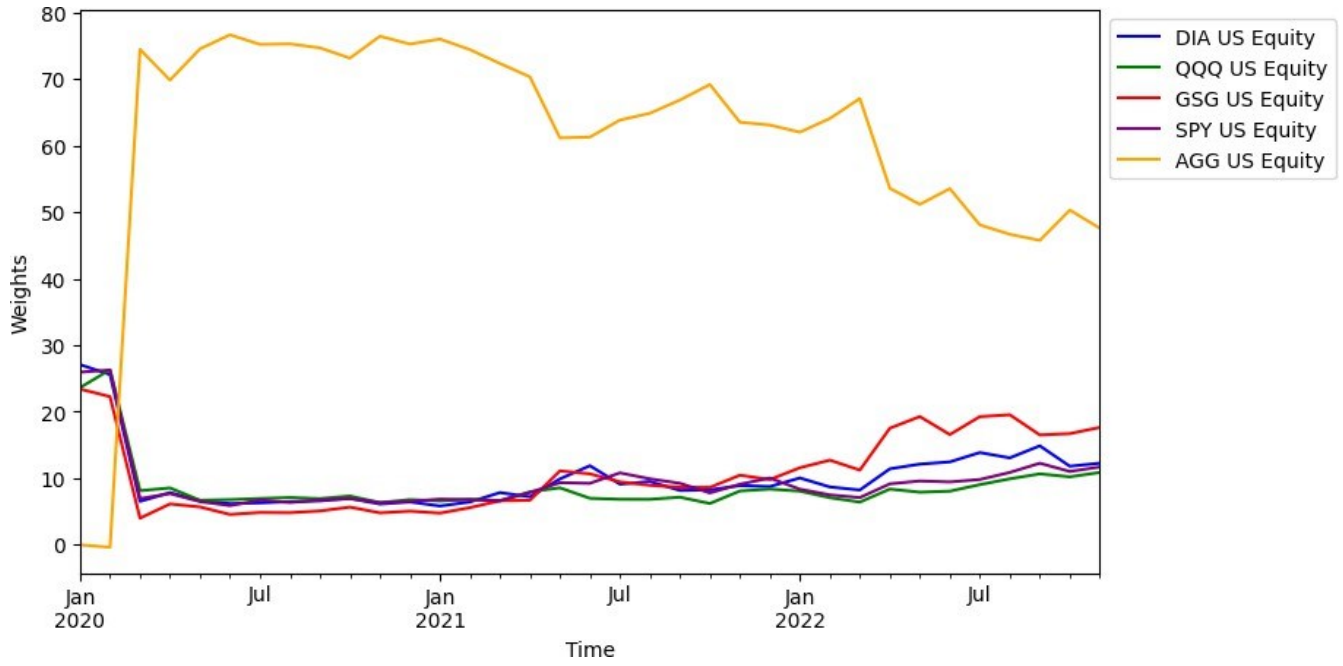

Figure 6 –Model's allocation weights since 2020

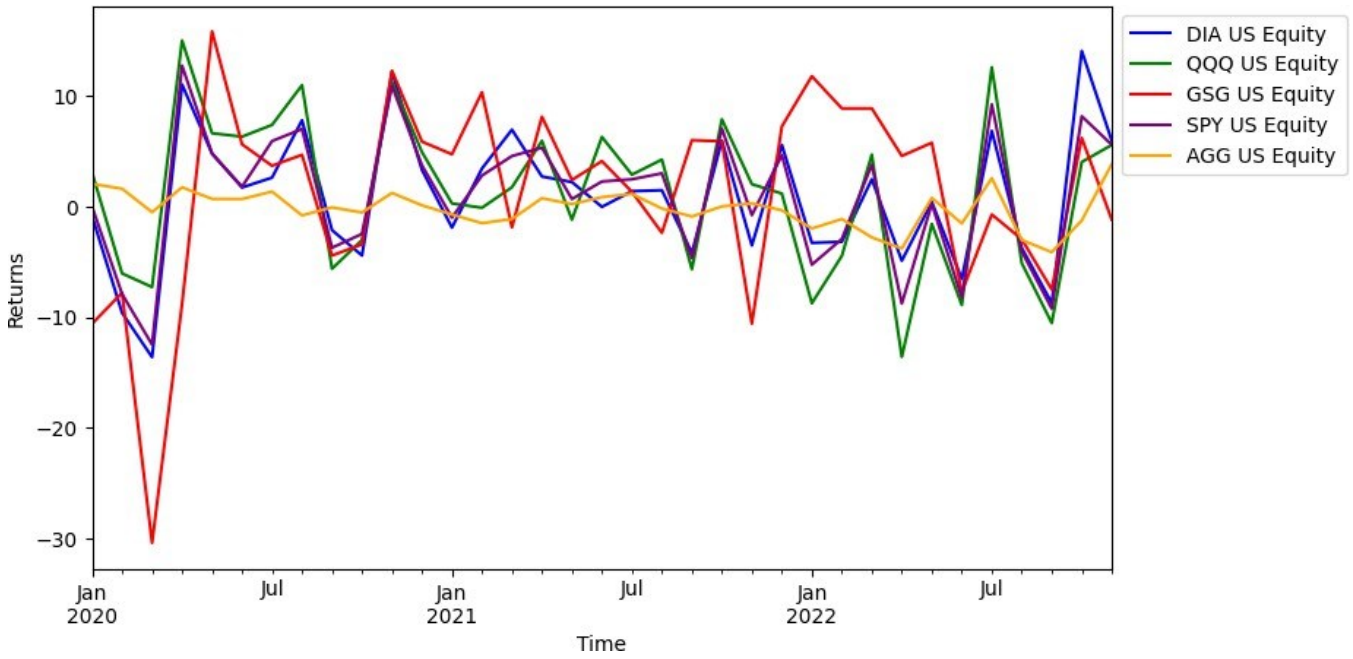

Figure 7 – Indexes returns since 2020

	Final Model	Final Model after trading costs	Benchmark	Benchmark after trading costs
Annualized return	4.60%	4.33%	5.30%	5.28%
Volatility	5.96%	5.95%	6.37%	6.37%
Sharpe Ratio	0.67	0.63	0.74	0.73
Info Sharpe	0.77	0.73	0.83	0.83
Skewness	-1.38	-1.42	-0.72	-0.72
Kurtosis	5.04	5.10	2.16	2.16
Maximum Drawdown	-14.17%	-14.31%	-14.62%	-14.62%
Positive Months	65.07%	64.38%	66.44%	66.44%
Sortino Ratio	0.73	0.67	0.95	0.94

Figure 8 – Performance overview

Model Variable	Feature Importance
AGG US Equity	0.1636
BICLB10Y	0.1574
SPCSUSA Index	0.1538
NHSLTOT Index	0.1187
CPUPAXFE Index	0.1126
CPURNSA Index	0.0703
FLSLCHAN Index	0.0658
PRUSTOT Index	0.0544
GSG US Equity	0.0536
FRNTTOTL Index	0.0498

Figure 9 – Top 10 most important features

Model Variable	Feature Importance
CONSEXP Index	-0.0722
CONSP5MD Index	-0.0773
CONCPSIT Index	-0.0971
CONSSENT Index	-0.0974
EURUSD Currency	-0.1033
CONCCONF Index	-0.1183
CONSCURR Index	-0.1189
SPCSUSAY Index	-0.1392
INJCJC Index	-0.1808
USURTOT Index	-0.2056

Figure 10 – Bottom 10 least important features

	Annualized Return	Volatility	Sharpe Ratio	Info Sharpe	Skewness	Kurtosis	Maximum Drawdown	Positive Months	Sortino Ratio
2 years	4.89%	5.87%	0.74	0.83	-1.35	4.85	-13.33%	65.82%	0.81
3 years	4.60%	5.96%	0.67	0.77	-1.38	5.04	-14.17%	65.07%	0.73
4 years	4.12%	6.01%	0.58	0.69	-1.44	5.12	-13.35%	64.93%	0.61

Figure 11 – Performance comparison between different training periods

Name	Soft/Hard	Category	Original frequency
ICE LIBOR USD 3 Month	Hard	Market	D
Constant Maturity Rate 3M US Fixed Income	Hard	Market	D
Generic US 1Yr Government Bill	Hard	Market	D
Constant Maturity Rate 1Y US Fixed Income	Hard	Market	D
Initial Jobless Claims	Hard	Labour market	W
U. of Mich. 1 Yr Inflation	Hard	Inflation	M
Constant Maturity Rate 2Y US Fixed Income	Hard	Market	D

Figure 12 – Statistically significant macroeconomic variables causing the major drawdown

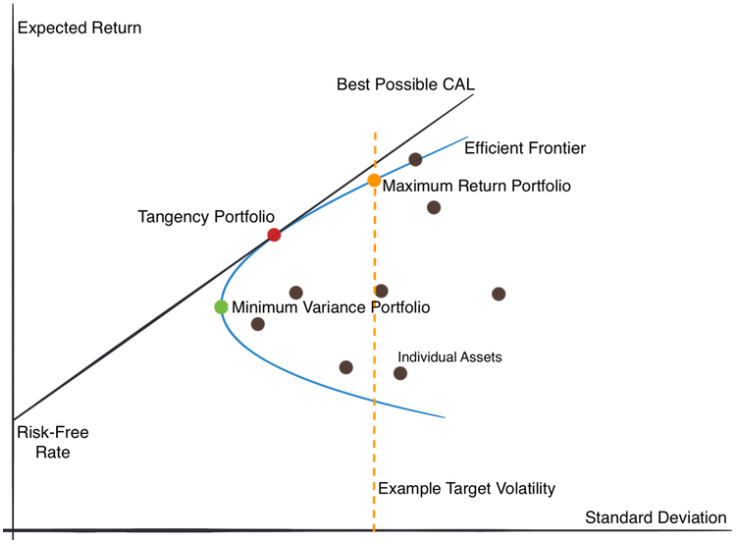

Figure 13 – Illustrative example of the Minimum Variance Portfolio and the Efficient Frontrier.

Source: Quantpedia. 2021. "Markowitz Model". Accessed December 12, 2023. https://quantpedia.com/markowitz-model/

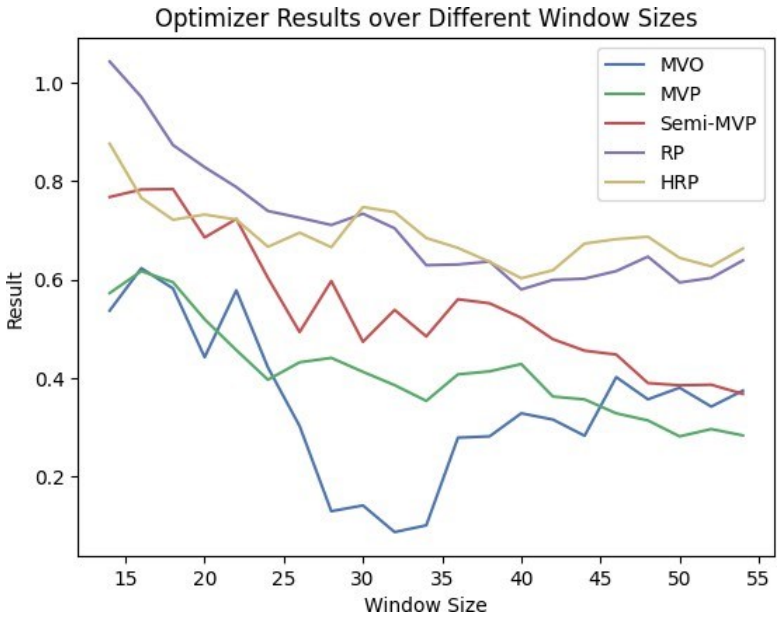

Figure 14 – Performance comparison (Info Sharpe) between different periods of past historical data used for different optimizers

	Mean-Variance Optimization	**Minimum Variance Portfolio**	**Minimum Semivariance Portfolio**	**Risk Parity**	**Hierarchical Risk Parity**
DIA US Equity	6.05	3.53	5.11	6.72	72.83
QQQ US Equity	13.12	0.21	2.69	5.58	2.32
GSG US Equity	5.93	3.74	1.89	6.57	9.75
SPY US Equity	3.96	0.94	1.52	6.26	4.80
AGG US Equity	59.90	90.81	84.86	46.42	2.32
EMB US Equity	1.48	0.00	0.12	9.82	1.46
GLD US Equity	4.22	0.27	2.52	9.75	1.51
PLD US Equity	4.44	0.25	1.06	4.47	2.00
PSP US Equity	0.90	0.26	0.22	4.40	3.00

Figure 15 – Average weights obtained by the different optimizers

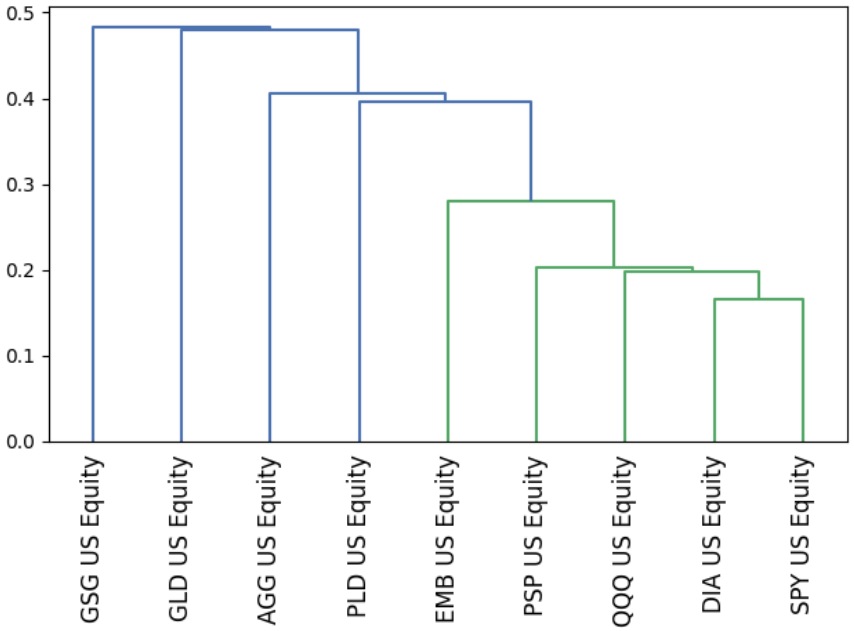

Figure 16 – Dendrogram illustrating the formed clusters resulting from Step 1) Tree clustering of HRP for the last month of the backtest period and considering a 54-months window size.

	DIA	QQQ	GSG	SPY	AGG	EMB	GLD	PLD	PSP
Skewness	-0.37	-0.43	-0.80	-0.57	0.28	-0.74	0.04	-0.68	-1.18

Figure 17 – Skewness of the individual ETFs' returns.