University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 21/2020 (327)

HRP Performance Comparison in Portfolio Optimization under Various Codependence and Distance Metrics

Illya Barziy
Marcin Chlebus

Warsaw 2020

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

HRP performance comparison in portfolio optimization under various codependence and distance metrics

Illya Barziy, Marcin Chlebus[a] *

[a] *Faculty of Economic Sciences, University of Warsaw*
* *Corresponding author: mchlebus@wne.uw.edu.pl*

Abstract: Problem of portfolio optimization was formulated almost 70 years ago in the works of Harry Markowitz. However, the studies of possible optimization methods are still being provided in order to obtain better results of asset allocation using the empirical approximations of codependences between assets. In this work various codependences and metrics are tested in the Hierarchical Risk Parity algorithm to determine whether the results obtained are superior to those of the standard Pearson correlation as a measure of codependence. In order to compare how HRP uses the information from alternative codependence metrics, the MV, IVP, and CLA optimization algorithms were used on the same data. Dataset used for comparison consisted of 32 ETFs representing equity of different regions and sectors as well as bonds and commodities. The time period tested was 01.01.2007-20.12.2019. Results show that alternative codependence metrics show worse results in terms of Sharpe ratios and maximum drawdowns in comparison to the standard Pearson correlation for each optimization method used. The added value of this work is using alternative codependence and distance metrics on real data, and including transaction costs to determine their impact on the result of each algorithm.

Keywords: Hierarchical Risk Parity, portfolio optimization, ETF, hierarchical structure, clustering, backtesting, distance metrics, risk management, machine learning

JEL codes: C32, C38, C44, C51, C52, C61, C65, G11, G15

1. Introduction

The modern approach of portfolio optimization has started from the works of Harry Markowitz (1952) and until now is a topic of continuous development. Works that contributed to the development of the topic include the works by Merton (1969), Magill (1976), Lopez de Prado (2016), Alipour (2016), and many other scientists who have studied the problem of optimal asset allocation both theoretically and empirically.

The variety of techniques and approaches for portfolio optimization include various programming methods, optimization methods, and machine learning approaches. Some of the popular methods are the equal weighting (EW), mean-variance (MV), inverse variance parity (IVP), critical line algorithm (CLA) and the hierarchical risk parity (HRP). Each of these approaches has its own drawbacks, as discussed in the works of Lopez de Prado (2016) and Jain (2019). Many papers on the topic don't describe how the performance of a portfolio optimization algorithm is affected by transaction costs, slippage, and other elements that can decrease returns and substantially affect the benchmarks of the algorithm, as shown in the work by Chavalle (2019).

The hierarchical risk parity algorithm was chosen as a basic algorithm in this work to be analyzed and modified as it possesses a unique characteristic of not requiring the covariance matrix of the portfolio assets to be invertible. This requirement must be fulfilled in order to use the methods created by Markowitz and their modifications, as stated in Lopez de Prado (2016). However, this is often impossible for real-life data, as the covariance matrices are estimated from empirical observations and pose bad numerical characteristics such as a high conditional number. This makes the practical use of these algorithms complicated, especially when the portfolio consists of a high number of financial instruments. From Lopez de Prado (2020), these problems are solved by clustering the instruments into sub-portfolios, de-noising, and de-toning the covariance matrix, using different covariance matrix estimators, or shrinking the covariance matrix.

The HRP algorithm was first presented in works by Marcos Lopez de Prado (2016) and Alipour (2016). In the original and following works of the authors, the basic algorithms are described in detail, however, the possible modifications are only briefly mentioned, as potential fields of study for future researchers. The possibilities of modifications are present in multiple steps of the algorithm and are described further in this work. After researching the literature no works were found that cover these particular modifications, their comparison, and how are they affecting the performance of the algorithm and how is it affected by the introduction of the transaction costs. The purpose of this work is to extend the results of authors and explore the ways the algorithm is applicable to optimizing the portfolio of ETFs over the course of the previous 13 years.

Algorithms developed for asset allocation may have different optimization goals, which go in line with the interests of a portfolio manager, such as maximization of the returns, minimization of the volatility, or maximization of the Sharpe ratio – a combination of both previous goals. The goal of the chosen HRP algorithm is the minimization of the portfolio variance, however, even with this mentioned, it can outperform other algorithms, which strive to achieve the highest Sharpe ratio, due to differences in the algorithms.

The way to measure the portfolio profitability relative to risk was introduced by William Sharpe in the year 1966 and described in the work of Sharpe (1994). The Sharpe ratio is measuring the average excess revenue over a benchmark in relation to the variance of portfolio profits. Choosing the portfolio that is positioned on the efficient frontier (set of portfolios that pose the highest expected return for a given level of risk) with the highest Sharpe ratio, is the way the optimal portfolios are chosen.

In this work, a multi-period problem of portfolio optimization is analyzed, as in the approaches by Magill (1976). The investor is constructing a portfolio at time t and after a short period e, the portfolio turns from optimal one to sub-optimal. This results in the need for reconstruction of the portfolio when it's no longer optimal.

Approaches to the timing of rebalancing the portfolio are different, they include – "time-only" (rebalancing at a specific time once every period), "threshold-only" (rebalancing after a certain threshold of revenue is reached, i.e. if a value of a portfolio has changed 1%), "time-and-threshold" which is a combination of the two mentioned approaches, as described in the work of Jaconetti (2010). In terms of time rebalancing, usual practices are applying rebalancing annually, quarterly, monthly, weekly, and daily. In terms of the threshold, the values can vary from 0-1% to 5-10%, as discussed in the work of Jaconetti (2010).

As stated in the above work, a choice of a rebalancing method depends on the individual characteristics of a portfolio, however, a too frequent rebalancing will lead to a significant increase in costs. For a portfolio consisting of diversified instruments, a 5% threshold or reasonable time frequency is assumed to be delivering an optimal balance between risk control and costs. In this particular work, the time-only rebalancing with monthly time frequency is used to test the modifications of the portfolio optimization algorithms.

This work assumes transaction costs to be equal to the Interactive Brokers rate, which is available in Appendix C. The sensitivity of profitability and other metrics to transaction costs increase for the chosen strategies are also analyzed, however, it's not the main goal of this work.

It is considered that the portfolio manager can only open long positions and not short ones (a long-only portfolio). This is the approach supported by the HRP algorithm, it simplifies the modeling and takes into account that the algorithm is minimizing the variance of the portfolio and not maximizing the returns. Short positions and leverage can be added to the HRP by creating the corresponding instruments from the currently available ones and adding them to the portfolio.

The structure of this work consists of the following parts: Methodology briefly specifies the problem being solved in mathematical terms as the steps of the HRP algorithm, Codependence matrices and distance metrics part describes the theory behind the modifications applied to the HRP algorithm and others used for comparison, Data set used describes the origin and the character of data used. Finally, the Results of the research are presented and the Conclusions are drawn.

2. Methodology

In this section the general problem of portfolio optimization will be described, the steps of the HRP algorithm are briefly covered as places where modifications will take place. Only the main ideas of other algorithms used in the comparison – MV, IVP, and CLA are mentioned, as investigating their structure is not the main goal of this work. Finally, this part includes the details on how the transaction costs are counted, how the portfolio is being rebalanced, and the way metrics used to compare the performances of the algorithms are calculated.

2.1. The optimization problem

As described in the work of Markowitz (1952), an investor is choosing a portfolio with a minimum level of risk for a given level of return, thus solving an optimization problem. According to Mitchell (2002), the portfolio optimization problem that takes transaction costs into account is described by the following set of conditions:

$$\min\{w^T \cdot C \cdot w\} \tag{1}$$

$$\text{Subject to: } E[R]^T \cdot w \ \geq E \tag{2}$$

$$e^T \cdot w = 1 \tag{3}$$

$$\overline{w} - \mathrm{b} + \mathrm{s} = \mathrm{w} \tag{4}$$

$$(C_B + e)^T \cdot b + (C_S + e)^T \cdot s = 0 \tag{5}$$

$$b, s, w \ \geq 0 \tag{6}$$

where:

w	- vector of weights for instruments,
e	- vector of ones,
C	- estimation of the covariance matrix based on the historical returns,
$E[R]$	- vector of expected returns of the instruments based on the historical returns,
E	- expected return of a portfolio,
$\overline{w}$	- vector of new weights for instruments in a new portfolio after rebalancing,
b, s	- vectors of amounts bought and sold of each instrument during rebalancing,
C_B, C_S	- vectors of transaction costs at rebalancing for each instrument.

In the above problem, conditions (1) – (3), (6) are representing a standard portfolio problem, as described by Markowitz. According to Mitchell (2002), this formulation of portfolio problem "minimizes a quadratic risk measurement with a set of linear constraints specifying the minimum expected portfolio return and enforcing a full investment of funds". The conditions (4) – (6) are showing the rebalancing problem. The proportional brokerage fees are assumed, the transaction costs used in this paper are following this requirement. The weights of the securities are non-negative, which means that the portfolio is long-only.

In the first period t, a basic optimization problem is being solved and the efficient frontier is constructed. From the efficient frontier, a portfolio with the highest Sharpe ratio should then be chosen.

For the next periods $t + \varepsilon$, the investor needs to solve a full optimization problem to obtain the new efficiency frontier. The goal is again to get the portfolio with the highest Sharpe ratio and repeat the steps above for all the moments the portfolio rebalances are needed.

2.2. The Hierarchical Risk Parity approach

The algorithm was originally developed by Lopez de Prado in 2016. This method allows the rebalancing of a portfolio in the conditions in which it would be impossible to do so under the MV method due to numerical issues – codependence matrix is impossible to invert. It consists of three stages, namely the tree clustering, quasi-diagonalization, and the recursive bisection. The stages are described below as well as points where modifications are introduced.

2.2.1. Tree clustering

The input of the algorithm is a series of returns with N observations for each of the securities M. From this data, an empirical codependence matrix is calculated. Based on the correlation matrix, a distance matrix is calculated using a defined distance metric.

On this first step, there is a modification to apply to the algorithm. There are multiple ways to calculate the codependence matrix and the distance to use. The classical and most widely used correlation metric is the Pearson correlation, however, there are also other metrics that are being used in this work – distance correlation, mutual information, variation of information. They are described in more detail later. As for the distance metric, there are also multiple options available.

To make the comparison fair – so each of the optimization algorithms has access to the same information regarding the codependences between the instruments in a portfolio, the same type of codependence metric will be used in each one of them. This modification only changes the codependence matrix provided to the MV, IVP, and CLA algorithms.

The distance matrix, calculated from the codependence matrix is showing how each of the elements in the portfolio is related to other elements. From the book of Arkhangel'skii (1990), the distance is a metric, in a sense that it obeys the three properties of a metric – non-negativity (8), identity of indiscernibles (9), symmetry (10), and these three metrics imply subadditivity (11):

$$d: X \times X \rightarrow [0, \infty);\ i, j, k \in X \tag{7}$$

$$d(i,j) \geq 0 \tag{8}$$

$$d(i,j) = 0 \Leftrightarrow i = j \tag{9}$$

$$d(i,j) = d(j,i) \tag{10}$$

$$d(i,j) \leq d(i,k) + d(k,j) \tag{11}$$

where:

X - set of elements,

$d(i, j)$ - distance between elements i and j,

From a distance matrix a hierarchical structure of the securities is drawn using the clustering algorithm. The default clustering algorithm used in the work of Lopez de Prado (2016) is the single-linkage clustering. As shown in Pfitznger (2019), other clustering algorithms can be applied such as Ward, DIANA, and Genetic Permutation. It was also shown that these clustering techniques outperform the standard single-linkage one when used in HRP on the Pearson correlation codependence metric in terms of the diagonalization measure of the resulting correlation matrix. In this work, the single-linkage clustering method is used to compare the performance of codependence metrics.

2.2.2. Quasi-diagonalization

In this step, order of the elements in a portfolio is changed according to the hierarchical structure discovered in the previous step. It is done by replacing each cluster consecutively with the elements included inside of it, the order of the clustering is preserved. This allows positioning the highest values in the correlation matrix along the main diagonal, thus reducing the condition number and making the portfolio more stable according to Lopez de Prado (2016).

The similar elements are also placed near each other after this step in terms of correlation close to each other. The change of the correlation matrix before and after the quasi-diagonalization can be seen in Appendix A.

2.2.3. Recursive bisection

In this step, the algorithm is assigning the weights to the securities in a portfolio. The main idea is to unpack each cluster (consisting of two elements – left and right) one by one while assigning weights to each part until all clusters are unpacked and each individual asset in a portfolio has an assigned weight. The weights are assigned based on the variance of left and right elements in a cluster unpacked. The more variance the element has, the less weight it gets.

Modifications to the algorithm can also be applied on this step. For example, if one of the securities in a portfolio possesses a substantially lower level of variance, a higher portion will be allocated to this security in comparison to others. One might want to limit this by setting a threshold of maximum allowed weight for a single asset and redistribute the excess.

Two ways of the redistribution of the excess weight may be applied. One is redistributing it among all other elements in direct proportion to their current weights. The other is to redistribute it only between the elements in the same cluster (using the built hierarchical structure) in direct proportion to weight already distributed by the HRP algorithm. These redistribution approaches weren't mentioned in other literature and are an interest for future research.

2.3. Alternative portfolio optimization methods

The allocation of weights for the MV algorithm is done by solving the equations (1) - (3) and is described in more detail in the notes by Bruke (2020).

The IVP method, as mentioned in Alipour (2016), takes into account only the variance of individual instruments and not the covariance between them. This is one of the popular frameworks used in financial literature.

The CLA, as well as the MV method, was developed by Markowitz. The main idea behind the approach is also setting inequality conditions to the weights of each individual instrument in a portfolio. The implementation of the algorithm used in this work is described in Bailey and de Prado (2013). The CLA outputs multiple weights sets that satisfy the optimization problem conditions. The set of weights that results in the highest Sharpe ratio is chosen from the possible solutions, as this is the main benchmark used for comparison.

When the HRP algorithm is tested using another dependence metric instead of the Pearson correlation, the MV, IVP, and CLA are also used with that same metric. This is done to test if it's the metric used is being more informative or the HRP actually uses this information better in comparison to other metrics. Also, this allows to test the effectiveness of each codependence metric by looking at the average results of the optimization algorithms.

2.4. Metrics for performance comparison

In the same way, as described in the work of Markowitz (1952), it is assumed that an investor is aiming to minimize the risk of a portfolio at a given level of return.

The portfolio is being initially constructed during the first period and is reconstructed at the end of each of the next rebalancing periods. Rebalancing used in this work is time-only and is set to monthly. To estimate a codependence matrix of elements a 6-month rolling window is chosen, as in Clarke (2006) and Alipour (2016). After each rebalancing, the transaction costs from Appendix C are recorded.

The transaction costs are assumed to be symmetric – equal for buying and selling an instrument. Taxes are not paid from the profit and can be added easily – they will be proportional to the returns and therefore the sensitivity of the algorithms to them is not calculated in this work. The constructed portfolio is long-only, it allows only buying and holding the assets and selling at a profit, short sales are not allowed. The price of a portfolio is recalculated after each observation – in our case daily.

Input data is a time series of prices that are being transformed into a series of returns in the following way:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{12}$$

where:

R_t - return value at the observation t,

P_t, P_{t-1} - price values at observations t and $t-1$.

The performance of the algorithms is compared using the following measures: The annualized Sharpe ratio, average daily return, a standard deviation of daily returns, annualized rate of return, maximum drawdown, total returns including cost. The costs incurred are analyzed by: average cost of rebalancing, standard deviation of costs, and total returns including costs.

It is also possible to test the significance of differences between the metrics such as annualized Sharpe ratios and annualized volatilities using the approach described in the works of Ledoit and Wolf (2008, 2011).

As for the risk-free rate used in the annualized Sharpe ratio formula, 0% was chosen as it has changed over the previous few decades. Changing the risk-free ratio won't affect the comparison of performances of actual algorithms, as it will decrease the Sharpe ratios with equal proportion. For comparison to the results of other researchers, the results of this work can be easily changed to annualized Sharpe ratios with fixed risk-free rates. For example, in the work of Zakamulin (2016), the risk-free rate for the period 1989-2017 was assumed to be fixed at 2% annually.

3. Codependence matrices and distance metrics

In this work, various dependence measures were used. The classical one is the Pearson correlation. The same metric used since the works of Markowitz. The measures used are described in detail in this part of work, the type of dependencies that they are measuring will be stated. The dependence metrics used are:

- Pearson correlation
- Distance correlation
- Mutual information
- Variation of information

And the distance metrics compared in the tree clustering step of the HRP algorithm:

- Angular distance
- Absolute angular distance
- Squared angular distance

3.1. Pearson correlation

Originally the correlation coefficient was developed by Pearson, the mathematical formula for the correlation was first published by Bravais (1844). It's a widely used statistic in theoretical and practical applications for codependence measurement between elements.

The Pearson correlation is a measure of linear codependence between the two random variables, but is not a metric since it doesn't satisfy the conditions of non-negativity (8) and subadditivity (11). As mentioned in Lopez de Prado (2020), keeping the metric conditions allows for inducing a topology on a set of data. Lack of such a topology can make the outcomes of the measurements being incoherent. The example stated in the above work is that the difference between correlations (0.9, 1.0) and (0.1, 0.2) is the same, but the second pair poses a higher difference of codependence. For this reason, in the HRP algorithm, a distance matrix is calculated from the correlation matrix, as the distance is a metric.

The Pearson correlation coefficient for the sample is calculated as:

$$\rho_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{13}$$

where:

ρ_{xy}	- Pearson correlation coefficient between variables x and y,
x_i, y_i	- i-th observation of variables x and y,
$\bar{x}, \bar{y}$	- average value of observations for x and y,
N	- number of observations.

The value of statistic falls in the range of $(-1, +1)$, where a coefficient of $+1$ means that the relationship between variables can be described using a linear equation and with an increase of one variable, the other increases too. With the coefficient of -1, the relationship is also linear, but with an increase of one variable, the other one decreases. The range of values is from $+1$ to -1, with 0 is no linear correlation, 1 is a truly positive, and -1 is a true negative. Correlation between 0 and $+1$ means that the observations of variables lie on the same side of distributions in relation to the means. As mentioned in Lopez de Prado (2020), this statistic is also sensitive to outliers.

3.2. Distance correlation

The distance correlation codependence measure was first introduced by Szekely in 2005 and is described in the work of Szekely (2007) as a generalization of Pearson correlation to take non-linear codependences into account. As mentioned in Lopez de Prado (2020), it is a computationally expensive statistic. In contrast to the Pearson correlation, the distance correlation equal to zero implies that the variables are independent.

The approach of Szekely also allows for a calculation of analogs for ordinary moments in the Pearson correlation, such as distance variance, distance standard deviation, and the distance covariance. The distance correlation statistic is calculated in the following way:

$$\rho_{dist}[X,Y] = \frac{dCov[X,Y]}{\sqrt{dCov[X,X]dCov[Y,Y]}} \tag{14}$$

$$dCov^2[X,Y] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} D(x_i,x_j) * D(y_i,y_j) \tag{15}$$

$$D(x_i,x_j) = a_{i,j} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}\, , D(y_i,y_j) = b_{i,j} - \bar{b}_{i.} - \bar{b}_{.j} + \bar{b}_{..} \tag{16}$$

$$a_{i,j} = \|X_i - X_j\|\, , b_{i,j} = \|Y_i - Y_j\|\, , i,j = 1,2, \dots N. \tag{17}$$

where:

$\rho_{dist}[X, Y]$ - distance correlation between random variables X and Y,

$dCov[X, Y]$ - sample distance covariance between variables X and Y,

$D(x_i, x_j)$ - doubly centered Euclidean matrices of variables X and Y,

a, b - Euclidean distance matrices between observations of X and Y respectively,

$a_{i,j}$ - element of a matrix that is placed on the i-th row and the j-th column,

$\bar{a}_{i.}$ - mean of i-th row of matrix a,

$\bar{a}_{.j}$ - mean of j-th column of matrix a,

$\bar{a}_{..}$ - grand mean of matrix a,

X_i - i-th observation of the X variable,

$||.||$ - Euclidean norm,

N - number of observations.

This statistic falls in the range of $(0, +1)$, where the value 0 is observed if and only if variables X and Y are independent, it also allows to measure the codependence between the two random vectors that not necessarily have equal dimension. If the statistic value 1 is observed and it is assumed that the subspaces spanned by X and Y are equal, then:

$$Y = A + b * C * X \tag{18}$$

where:

X, Y - vectors of observations of variables X and Y,

A - vector,

b - scalar,

C - orthonormal matrix.

As the distance correlation is always positive, it's impossible to use it to find negative codependences between variables. Due to the need for the calculation of the distance matrices for each of the variables, the space complexity of the calculation increases to $O(n^2)$ in comparison to $O(n)$ for Pearson correlation. This makes it harder to use this statistic for large samples of data. In this work a rolling window approach is used to estimate the codependence matrix, so the difficulties related to calculations are omitted.

3.3. Mutual information

Mutual information is a measure of mutual dependence between the two random variables. It arises from the information theory of Shannon, which was originally described in the work by Shannon and Weaver (1949) and later developed further in the work by Cover and Thomas (1991). Mutual information determines the difference between the joint distribution of the pair (X, Y) and the product of the marginal distributions X and Y, it allows quantifying how much information is obtained about one variable through only observing the second one.

From the Cover and Thomas (1991), the mutual information can be calculated as:

$$H[X] = -\sum_{x \in S_x} p[x] \log [p[x]] \tag{19}$$

$$H[X, Y] = -\sum_{x, y \in S_x \times S_y} p[x, y] \log [p[x, y]] \tag{20}$$

$$I[X, Y] = H[X] + H[Y] - H[X, Y] = \sum_{x \in S_x} \sum_{y \in S_y} p[x, y] log \left[\frac{p[x, y]}{p[x]p[y]}\right] \tag{21}$$

where:

$H[X]$	- entropy of random variable X,
$p[x]$	- probability of random variable X being equal to x,
S_x	- set of values that that the random variable X can be equal to,
$H[X, Y]$	- joint entropy of variables X and Y,
$p[x, y]$	- probability of pair of random variables (X, Y) being equal to (x, y),
$I[X, Y]$	- mutual information between variables X and Y.

The entropy is interpreted as the amount of uncertainty associated with variable X, it is equal to zero if the probability of one element from S_x is 1 and equal to maximum value $\log [\|S_x\|]$ when the probability is distributed uniformly across elements in S_x. The mutual information measurement falls in the range of $(0, \min\{H[X], H[Y]\})$, is non-negative (8) and symmetric (10), but it's not a metric since it doesn't satisfy the subadditivity property (11).

As the entropy is finite only in case of discretizing the random variables, the calculation of it includes limiting density to discrete points, the same way they are calculated in the work of Lopez de Prado (2020):

$$I[X,Y] = \sum_{i=1}^{|U|}\sum_{j=1}^{|V|} \frac{|U_i \cap V_j|}{N} log \frac{N|U_i \cap V_j|}{|U_i||V_j|} \tag{22}$$

where:

$\lvert U\rvert, \lvert V\rvert$	- number of bins in the discretization for variables X and Y respectively,
$\lvert U_i\rvert$	- number of elements in a i-th bin from total bins U,
$\lvert U_i \cap V_j\rvert$	- sum of elements in i-th bin from U and j-th bin from V,
N	- total number of elements in bins U and V.

As for the number of bins to use for the discretization of variables X and Y, the optimal binning formula for the joint entropy case from the work of Hacine-Gharbi and Ravier (2018), as discussed in Lopez de Prado (2020) was used:

$$|U| = |V| = round\left[\frac{1}{\sqrt{2}}\sqrt{1+\sqrt{1+\frac{24N}{1-\rho_{XY}}}}\right] \tag{23}$$

where:

$\lvert U\rvert , \lvert V\rvert$	- number of bins to use in the discretization for variables X and Y respectively,
N	- number of observations of variables X and Y,
ρ_{XY}	- Pearson correlation between observations of variables X and Y.

In order to use this measure in the chosen portfolio optimization algorithms, it is being normalized to the range of $(0\,,+1)$:

$$I_{normalized}[X,Y] = \frac{I[X,Y]}{\min\{H[X],H[Y]\}} \tag{24}$$

3.4. Variation of Information

The variation of information is the measure of the distance between two random variables and according to Lopez de Prado (2020) can be interpreted as the uncertainty that is expected from one variable if the other variable is known. The same as with the mutual information measure, it’s closely related to Shannon’s information theory.

The variation of information was originally presented in a work by Meila (2006). This is a true metric as it follows the subadditivity (11) rule. This measure is calculated as:

$$VI[X,Y] = H[X] + H[Y] - 2I[X,Y] \tag{25}$$

where:

$VI[X,Y]$ - variation of information between variables X and Y,

$H[X]$ - entropy of variable X,

$I[X,Y]$ - mutual information between variables X and Y.

This metric falls in the range of $(0\,,\ H[X\,,Y])$. If one wants to compare the variances of information across populations of different sizes, the problem occurs. As the $H[X\,,Y]$ doesn't have a firm upper bound. Alternative metrics with upper bounds are mentioned in Lopez de Prado (2020). As for the number of bins to use in the discretization of variables X and Y, the same approach as in the mutual information calculation was used. The entropy of discretized variables was calculated as:

$$H[X] = -\sum_{i=1}^{|U|} \frac{|U_i|}{N} log \frac{|U_i|}{N} \tag{26}$$

where:

$|U|$ - number of bins in the discretization for variable X,

$|U_i|$ - number of elements in a i-th bin from total bins U,

N - total number of elements in bins U.

In the same way as mutual information metric, the variation of information is being normalized to the range of $(0\,,+1)$:

$$VI_{normalized}[X,Y] = \frac{VI[X,Y]}{H[X\,,Y]} \tag{27}$$

3.5. Angular distance

The angular distance is the measure used to transform Pearson correlation into a metric as it's a linear multiple of the Euclidean distance between the vectors. It was introduced in the work of Lopez de Prado (2016).

According to the author, this metric is well suited in the tasks of building long-only portfolios, exactly the ones that the HRP algorithm builds, as well as other algorithms that are used for the comparison. Angular distance is calculated based on Pearson correlation coefficient as:

$$d_{\rho}[X,Y] = \sqrt{\frac{1}{2}(1-\rho_{XY})} \tag{28}$$

where:

$d_{\rho}[X,Y]$ - angular distance between variables X and Y,

ρ_{XY} - Pearson correlation coefficient between variables X and Y.

This metric transforms values from range $(-1\,,+1)$ to range of $(0,+1)$. To use this metric on other types of codependences, they have to be also scaled $(-1\,,+1)$. It makes the variables that have a negative correlation have a greater distance between them. In the long-only portfolio the instruments that have a negative correlation can only offset risk and it's not possible to use them to capture the same direction movements of the market, therefore they are treated as different for diversification purposes. The graph below shows how this metric transforms the range of Pearson correlation coefficient to $(0,+1)$.

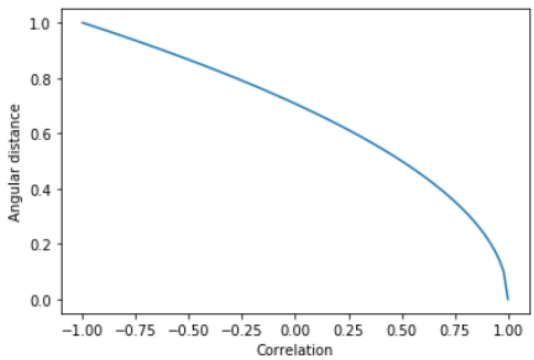

Figure 1: Transformation of Pearson correlation using the angular distance metric

Source: Own calculations.

3.6. Absolute angular distance

The absolute angular distance is a measure similar to the angular distance but is more suited for tasks of building the long-short portfolios, as the elements with high negative correlations are set to have smaller distances between them. In the long-short portfolios, the highly negatively-correlated securities can be treated as similar, as it's possible to invest in the second security with a negative sign. This measure was introduced in the work of Lopez de Prado (2016). Absolute angular distance is calculated based on Pearson correlation coefficient as:

$$d_{|\rho|}[X,Y] = \sqrt{\frac{1}{2}(1 - |\rho_{XY}|)} \tag{29}$$

where:

$d_{|\rho|}[X,Y]$ - absolute angular distance between variables X and Y,

ρ_{XY} - Pearson correlation coefficient between variables X and Y.

The graph below shows how this metric transforms the range of Pearson correlation coefficient to $(0, +1)$.

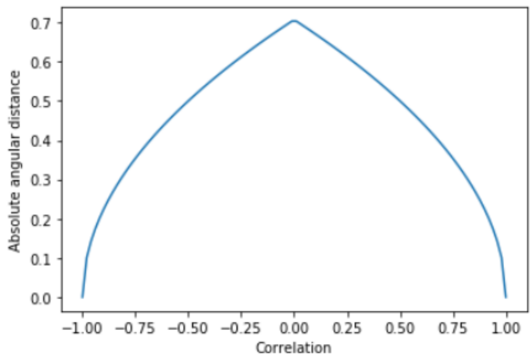

Figure 2: Transformation of Pearson correlation using the absolute angular distance metric
Source: Own calculations.

3.7. Squared angular distance

The squared angular distance is almost the same as the absolute angular distance in terms of properties and transformation of the correlation. However, it assigns slightly larger distances to correlations in comparison to the absolute angular distances. This measure was introduced in the work of Lopez de Prado (2016). Squared angular distance is calculated based on Pearson correlation coefficient as:

$$d_{\rho^2}[X,Y] = \sqrt{\frac{1}{2}(1 - \rho_{XY}{}^2)} \tag{30}$$

where:

$d_{\rho^2}[X,Y]$ - squared angular distance between variables X and Y,

ρ_{XY} - Pearson correlation coefficient between variables X and Y.

The graph below shows how this metric transforms the range of Pearson correlation coefficient to $(0, +1)$.

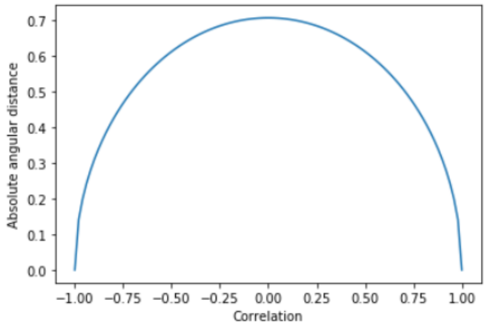

Figure 3: Transformation of Pearson correlation using the squared angular distance metric
Source: Own calculations.

4. Data set used

In order to test the performance of portfolio algorithms with different distance metrics, it was decided to use a diversified portfolio consisting of exchange-traded funds (ETFs). The picked ETFs are representing different asset types – stocks, bonds, and commodities. The stocks ETFs are chosen in a way to represent different regions of the world – USA, Asia, Europe, etc., and specific industries of the economy. The work of Schlanger (2018) describes how various ETFs can be used to construct portfolios with different investment goals in terms of risk-return balance.

The first criterion used when choosing a set of ETFs was the diversification of the investments, but also the liquidity of the ETFs was taken into account. This is done to ensure the simulations done will be less affected by biases of price slippage and low activity on the market, making it impossible to execute trades at the available prices. These biases can transform a profitable strategy during the backtesting process into an unprofitable one on real world-tests. However, the modeling of the slippages is not the goal of this work, as the comparison between algorithms and modifications is being done and these biases will affect the result of each algorithm in the same way. The threshold set to the liquidity of the ETF is 300000 average daily traded contracts based on the previous three months. This criterion has resulted in more than 600 ETFs to choose from in a portfolio. The instruments to be used in the portfolio are listed in Appendix B. The annualized Sharpe ratios and the time frame for which data is available for each ETF is also included to compare to the results of each portfolio optimization algorithm.

The data used was purchased from Kibot data provider and has a 30 minutes' interval of observations. The values representing the prices 30 minutes before market close were used as daily prices, as in practice it's impossible to execute orders exactly at close prices. This makes the modeling in this work closer to a bias-free one.

The period of time used in the simulation is from 01.01.2007 to 20.12.2019, as it reflects the recent changes in asset price movement and dependencies as well as a period of a crisis. This period also covers a full cycle of expansion and contraction according to the National Bureau of Economic Research. This will show how each of the algorithms with the corresponding codependence metrics can handle periods of recessions.

For some of the instruments in a chosen set data was not available from the start of the observation period – 01.01.2007, as they were created later. This is taken into account in a way that during each rebalancing only those instruments are used for which a codependence matrix can be estimated, which means that data should be available for at least 6 months before the rebalancing time.

As it was discussed previously, a rolling window approach was used for rebalancing with 6 months to estimate the codependence matrix and this matrix is used for the next month of daily data. It's being controlled that at the time of rebalancing the algorithm has no access to the data from the next month, in order to omit the look-ahead bias.

5. Results of the research

In this section first, the results regarding the distance metrics will be discussed, then the comparison of performances of each of the algorithms for different codependence metrics will be presented. Finally, the sensitivity of the algorithm performances to changes in transaction costs will be mentioned.

5.1. Distance metrics

For each of the codependence metrics, the HRP algorithm was used with different distance metrics. The results among the distance metrics were similar for each codependence metric.

This was expected as distance metrics have similar properties when used on a set of positively correlated instruments. As seen from the formulas described in the methodology part, the differences between the angular distance and the absolute/squared will be seen only if the assets are negatively correlated.

The results of portfolio optimizations are just slightly different, so a conclusion can be drawn that at some periods of time the dependencies between the elements were negative. Full table comparing the annualized Sharpe ratio, the average daily returns, and the standard deviation of returns is presented below.

Table 1: Comparison of distance metrics results

Codependence metric	Distance metric	Annualized Sharpe Ratio	Average daily return	Standard deviation of returns
Pearson correlation	Angular	0.644110	0.000253	0.006234
	Absolute angular	0.6524116	0.000256	0.006238
	Squared angular	0.6524116	0.000256	0.006238
Distance correlation	Angular	0.427612	0.000288	0.010702
	Absolute angular	0.427612	0.000288	0.010702
	Squared angular	0.427612	0.000288	0.010702
Mutual information	Angular	0.423383	0.000289	0.010702
	Absolute angular	0.423383	0.000289	0.010702
	Squared angular	0.423383	0.000289	0.010702
Variation of information	Angular	0.424777	0.000289	0.010786
	Absolute angular	0.424777	0.000289	0.010786
	Squared angular	0.424810	0.000289	0.010786

Source: Own calculations.

Note: The table above presents the results of running HRP portfolio optimization algorithm with multiple combinations of metrics. The total number of models created was 12.

The best results in terms of annualized Sharpe ratio are achieved using the Pearson correlation and the absolute or squared angular distance. The average daily returns are higher when using the alternative codependence metrics, but that is being offset with a substantially higher variation of returns.

The higher Sharpe ratio with the Pearson correlation using the squared or absolute angular distance can be explained by the fact that some of the instruments in a portfolio while having a negative correlation but clustered closer are offsetting the risk.

For the codependence metrics – distance correlation, mutual information, and the variation of information, the average returns are 12,5% higher, which is a significant increase in the observed period. But the variance of the returns is 71,5% higher, resulting in 34,5% lower annualized Sharpe ratio. The figure below shows side by side comparison of the portfolio rebalanced by HRP using various codependence metrics.

Figure 4: Accumulated returns for angular distance and various codependence metrics.

Source: Own calculations.

In the next comparisons the absolute angular distance is used as the distance metric for the HRP algorithm, as it's showing the best results with the Pearson correlation metric and the same results as other distance metrics when other codependence metrics are used.

5.2. Codependence metrics

The MV, IVP, CLA, and the HRP algorithms will be used on the data set with different codependence metrics to compare their performance. The results of the algorithms in terms of annualized Sharpe ratios, average daily returns, standard deviations of the daily returns, annualized rate of returns, and maximum drawdowns are shown in the table below. The way accumulated returns compare and how the weights are distributed among the instruments with each rebalancing for the optimization methods is also displayed. The results for the MV algorithm on the alternative codependence metrics is left blank as the obtained codependence matrices were impossible to use in the convex optimization problem being solved by the algorithm.

Table 2: Comparison of the optimization algorithms with different codependence metrics

Codependence metric	Optimiz-ation algorithm	Annualized Sharpe Ratio	Average daily return	Returns standard deviation	Annualized rate of return	Maximum drawdown
Pearson correlation	MV	0.395689	0.000231	0.009243	4.95%	47.38%
	IVP	0.512581	0.000274	0.008481	6.31%	40.69%
	CLA	0.446839	0.000248	0.008805	5.53%	43.14%
	HRP	0.652416	0.000256	0.006238	6.19%	26.16%
Distance correlation	MV	-	-	-	-	-
	IVP	0.418334	0.000298	0.011323	6.20%	52.19%
	CLA	0.328224	0.000348	0.016814	5.45%	68.48%
	HRP	0.427612	0.000288	0.010702	6.12%	48.56%
Mutual information	MV	-	-	-	-	-
	IVP	0.418334	0.000298	0.011323	6.20%	52.19%
	CLA	0.378348	0.000365	0.015297	6.33%	65.77%
	HRP	0.423399	0.000289	0.020831	6.09%	49.03%
Variation of information	MV	-	-	-	-	-
	IVP	0.418334	0.000298	0.011323	6.20%	52.19%
	CLA	0.377542	0.000337	0.014159	6.27%	61.49%
	HRP	0.424810	0.000289	0.010786	6.10%	48.83%

Source: Own calculations.

Note: The table above presents the results of running each portfolio optimization algorithm with multiple codependence metrics. The total number of models created was 13.

Figure 5: Accumulated returns for each algorithm using the Pearson correlation.

Source: Own calculations.

Figure 6: Accumulated returns for CLA algorithm using different codependence.

Source: Own calculations.

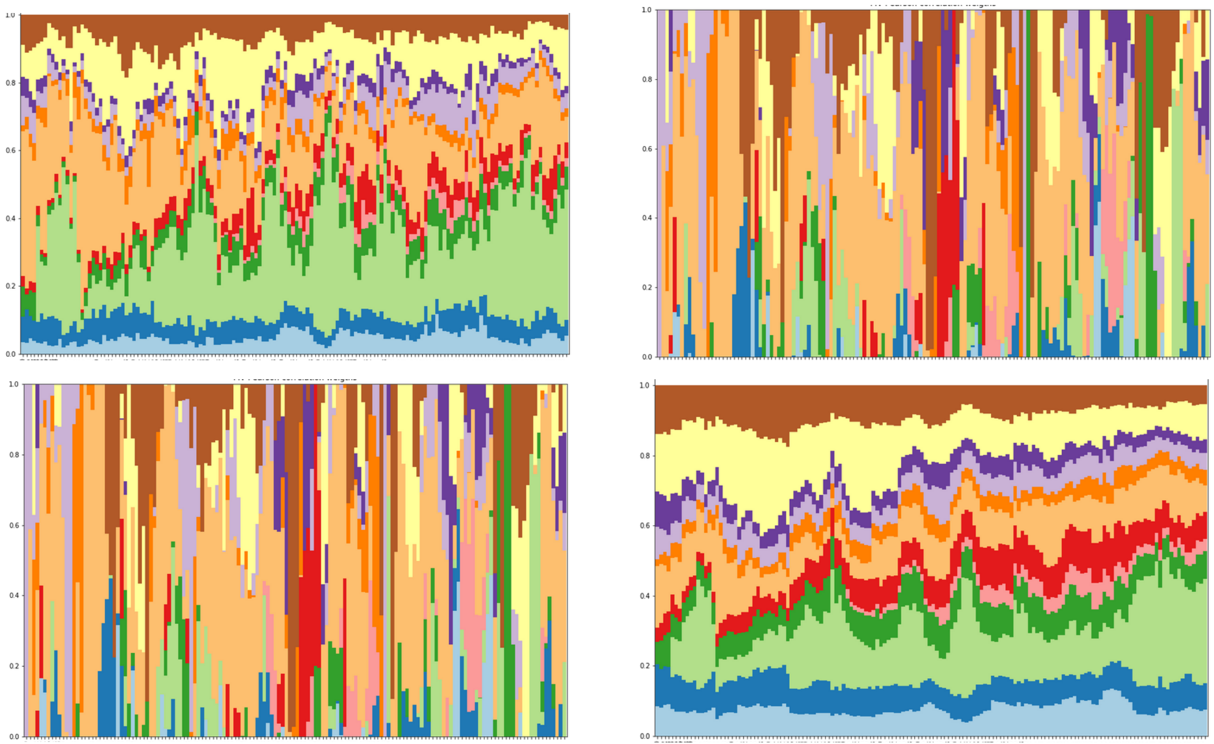

Figure 7: Bar plots of weights distribution for HRP (top left), CLA (top right), MV (bottom left), IVP (bottom right) algorithms using Pearson correlation.
Source: Own calculations.

For every codependence metric used the HRP algorithm has shown better results in terms of the annualized Sharpe ratios as a comparison measure. Unfortunately, the alternative codependence metrics as distance correlation, mutual information, and the variation of information show worse results than the standard Pearson correlation. On the other side, the average returns for alternative methods are higher.

From graphs of weights redistribution for different algorithms using the Pearson correlation, it can be concluded that the allocations of the IVP and the HRP are more robust than the MV and the CLA ones. This means that rebalancing the portfolio using the latter ones may result in more difficulties and higher rebalancing costs as the amounts of traded instruments are higher with each rebalancing. The CLA and the HRP have very similar average returns, but the CLA has a 41,1% higher standard deviation of returns.

5.3. Sensitivity to transaction costs

Figure 7 shows that algorithms have different robustness and as a result, different transaction costs. These transaction costs were calculated on a portfolio with a starting value of 1’000’000 USD and monthly rebalancing. The total number of rebalances is 148. Transaction costs for further analysis are presented in the table below.

Table 3: Comparison of transaction costs and total returns for different algorithms

Codependence metric	Optimization algorithm	Total transaction costs	Average transaction costs	Total returns including costs
Pearson correlation	MV	12`596.63 USD	85.11 USD	828`751.47 USD
	IVP	5`536.39 USD	37.41 USD	1`148`367.88 USD
	CLA	13`251.52 USD	89.54 USD	958`677.03 USD
	HRP	8`075.74 USD	54.57 USD	1`118`506.66 USD
Distance correlation	MV	-	-	-
	IVP	4`765.66 USD	32.20 USD	1`123`421.40 USD
	CLA	13`371.45 USD	90.35 USD	941`352.91 USD
	HRP	6`833.08 USD	46.17 USD	1`101`353.83 USD
Mutual information	MV	-	-	-
	IVP	4`765.66 USD	32.20 USD	1`123`421.40 USD
	CLA	13`375.41 USD	90.37 USD	1`217`025.64 USD
	HRP	6`591.98 USD	44.54 USD	1`095`891.03 USD
Variation of information	MV	-	-	-
	IVP	4`765.66 USD	32.20 USD	1`123`421.40 USD
	CLA	13`585.20 USD	91.79 USD	1`139`257.87 USD
	HRP	6`283.94 USD	42.46 USD	1`097`521.01 USD

Source: Own calculations.

Note: The table above presents the transaction costs incurred by each portfolio optimization algorithm with multiple codependence metrics. The total number of models created was 13.

In each rebalance the transaction costs are relatively low in comparison to the total value of a portfolio. All the transaction costs recorded for each individual ETF are either a minimum transaction cost of 1 USD or a fixed cost. This means that increase in transaction cost per individual order will increase total transaction costs in the same proportion.

For the chosen size of a portfolio the transaction costs are small compared to the returns generated. In order for transaction costs to severely affect the performance either the costs have to be higher or the rebalancing has to occur more frequently. For example, moving to weekly rebalances would increase the total transaction costs by around 4 times.

In terms of robustness, the IVP algorithm shows the best result. The MV and the HRP algorithms have similar robustness of weights and place the second place. The lowest robustness is shown by the CLA algorithm. This goes in line with the results obtained from Figure 7.

6. Conclusion

In this work, the performance of the HRP portfolio optimization algorithm was tested using the modifications of codependence metrics of the instruments in a portfolio, and distance metrics to transform the codependence matrix into the distance matrix. In order to determine whether the alternative metrics possess an advantage over the standard Pearson correlation, the other widely used optimization algorithms, the MV, IVP, and the CLA were applied to the same data with these metrics.

The performance of algorithms was tested on a portfolio of highly liquid ETFs representing such assets as stocks, bonds, and commodities. The stocks ETFs were picked to provide exposure to different world regions and various sectors of the economy. The time period used for testing was 2007 – 2019. Backtesting was made on 30-min data transformed into daily data using the price records 30 minutes prior to the close of trades. Data used was purchased from Kibot data provider. The transaction costs were included in the testing and the numbers were used from the Interactive Brokers platform.

The performances of the algorithms were compared based on the following metrics – annualized Sharpe ratio, average daily return, standard deviation of daily returns, annualized rates of return, maximum drawdowns, and the transaction costs incurred by the algorithm.

Results show that the HRP algorithm performs better on each of the codependence metrics. The Pearson correlation codependence metric has shown the highest Sharpe ratio results for each of the compared algorithms, which raises doubts about the effectiveness of using the distance correlation, mutual information, and the variance of information metrics for the portfolio optimization purposes, at least in the standard form.

The optimal modification of the HRP algorithm in terms of the distance metric was found to be either the absolute or the squared angular distance, as it offsets some risk in comparison to the standardly used angular distance metric. This result may be different for another type of portfolio where the instruments have higher negative correlations. In that situation, the angular distance is expected to give better performance.

The algorithms have shown different robustness in terms of weights allocations and as a result different transaction costs. The IVP and the HRP algorithms are more robust in comparison to the MV and the CLA algorithms.

In the future, this research can be extended by using another dataset with negatively correlated instruments and allowing short positions for the optimization algorithms. Also, deeper modifications to the algorithms can be made to use the data obtained from the alternative codependence metrics in a more efficient way. When using a portfolio with a higher number of instruments also the effectiveness of the de-noising and de-toning techniques can be used to improve the performance of the algorithms, as proposed by Lopez de Prado (2019).

References

Alipour, Elham, Clemens Adolphs, Arman Zaribafiyan, and Maxwell Rounds. 2016. "Quantum-inspired hierarchical risk parity." 1QB Information Technologies: 11.

Arkhangel'skii, A. V.; Pontryagin, L. S. 1990. General Topology I: Basic Concepts and Constructions Dimension Theory, Encyclopaedia of Mathematical Sciences: 60.

Bailey, David H., and Marcos Lopez de Prado, 2013. "An Open-Source Implementation of the Critical-Line Algorithm for Portfolio Optimization." Algorithms 6: 169-196.

Bravais, Auguste, 1844. "Analyse Mathematique. Sur les probabilités des erreurs de situation d'un point" Mem. Acad. Roy. Sci. Inst. France, Sci. Math, et Phys., 9:255-332.

Burke, James, 2020. "Lecture Notes on Finance: Markowitz Mean-Variance Portfolio Theory." University of Washington: 1-10.

Chavalle, Luc, and Luis Chavez-Bedoya, 2019. "The impact of transaction costs in portfolio optimization." Journal of Economics, Finance & Administrative Science 24: 288-311.

Clarke, Roger, Harindra de Silva, and Steven Thorley, 2006. "Minimum-Variance Portfolios in the U.S. Equity Market". Journal of Portfolio Management 33 (September): 10-24.

Cover, T. M., and Joy A. Thomas, 2006. "Elements of Information Theory." John Wiley and sons: 13-54.

Gábor J. Székely, Maria L. Rizzo, Nail K. Bakirov, 2007. "Measuring and testing dependence by correlation of distances." Annals of Statistics 35: 2769-2794.

Hacine-Gharbi, Abdenour, and Philippe Ravier, 2018. "A binning formula of bi-histogram for joint entropy estimation using mean square error minimization." Pattern Recognition Letters 101: 21-28.

Jaconetti, Colleen, Francis M. Kinniry Jr., and Yan Zilbering. 2010. "Best practices for portfolio rebalancing." Vanguard research (July): 17.

Jain, Prayut, and Shashi Jain, 2019. "Can Machine Learning-Based Portfolios Outperform Traditional Risk-Based Portfolios? The Need to Account for Covariance Misspecification." Risks 7: 74.

Ledoit, Olivier, and Michael Wolf. 2008. "Robust Performance Hypothesis Testing With the Sharpe Ratio." Journal of Empirical Finance 15 (March): 850-859.

Ledoit, Olivier, and Michael Wolf. 2011. "Robust Performance Hypothesis Testing With the Variance." Working Paper, University of Zurich 516 (October): 11.

Lopez de Prado, Marcos, 2016. “A Robust Estimator of the Efficient Frontier.” SSRN Electronic Journal: 1-13.

Lopez de Prado, Marcos, 2016. “Building Diversified Portfolios that Outperform Out-of-Sample. ” Journal of Portfolio Management (May): 31.

Lopez de Prado, Marcos, 2019. “Estimation of Theory-Implied Correlation Matrices.” SSRN Electronic Journal: 1-11.

Lopez de Prado, Marcos, 2020. “Machine Learning for Asset managers.” Cambridge University Press: 24-37, 38-51.

Magill, Michael J. P., and George M. Constantinides. 1976. “Portfolio Selection with Transaction Costs.” Journal of Economic Theory 13 (October): 245-263.

Markowitz, Harry. 1952. “Portfolio Selection.” Journal of Finance 7 (March): 77-91.

Meila, Marina, 2007. “Comparing clusterings—an information based distance.” Journal of Multivariate Analysis 98: 873-895.

Merton, Robert C. 1969. “Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case.” The Review of Economics and Statistics 51 (August): 247- 257.

Mitchell, John E., and Stephen Braun. 2002. “Rebalancing an Investment Portfolio in the Presence of Transaction Costs.” Working Paper, Department of Rensselaer Polytechnic Institute.

Pfitzinger, Johann, and Nico Katzke, 2019. “A constrained hierarchical risk parity algorithm with cluster-based capital allocation.” Stellenbosch Economic Working Papers WP14/2019: 1-21.

Schlanger, Todd, Scott J. Donaldson, and Maria Grant, 2018. “Vanguard asset allocation ETFs: A simple yet sophisticated approach to portfolio construction.” Vanguard Research:1-17.

Shannon, C.E., Weaver, W. 1964. “The Mathematical Theory of Communication.” Urbana: University of Illinois Press:379–423.

Sharpe, William, 1994. “The Sharpe Ratio.” Journal of Portfolio Management 21 (October): 49-58.

Szekely, Gabor J., Rizzo, Maria L., Bakirov, Nail K., 2007. “Measuring and testing independence by correlation of distances.” The Annals of Statistics. 35 (6): 2769–2794.

Zakamulin, Valeriy. 2016. “Volatility Weighting Over Time in the Presence of Transaction Costs.” University of Adger (January): 24.

Appendix A

Example of a dendrogram of instruments in a portfolio – a result of the tree clustering stage of the HRP algorithm on Figure 8 and the comparison of the Pearson correlation matrices before and after the quasi-diagonalization step on Figure 9.

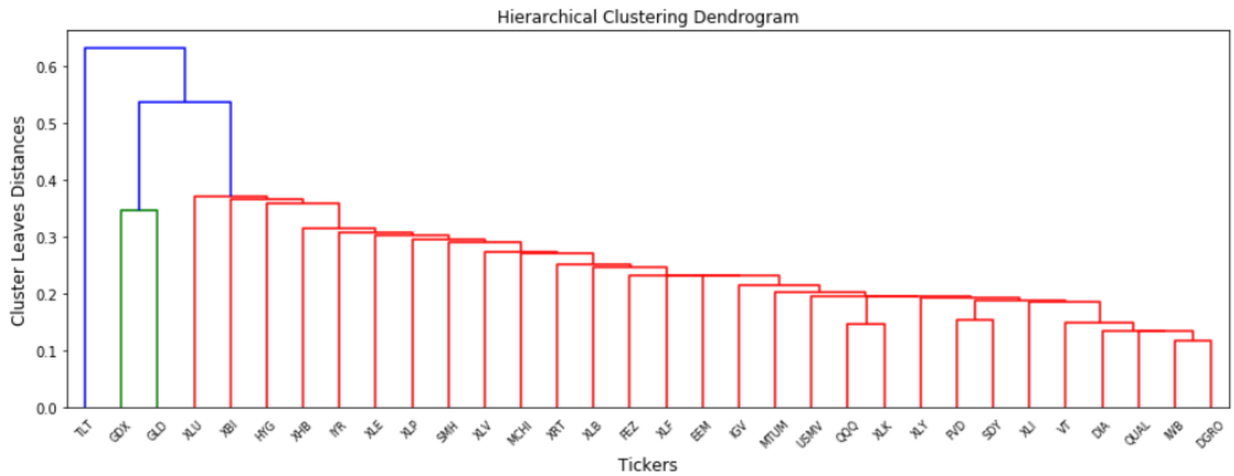

Figure 8: Dendrogram of instruments in a portfolio based on the whole period of data.
Source: Own calculations.

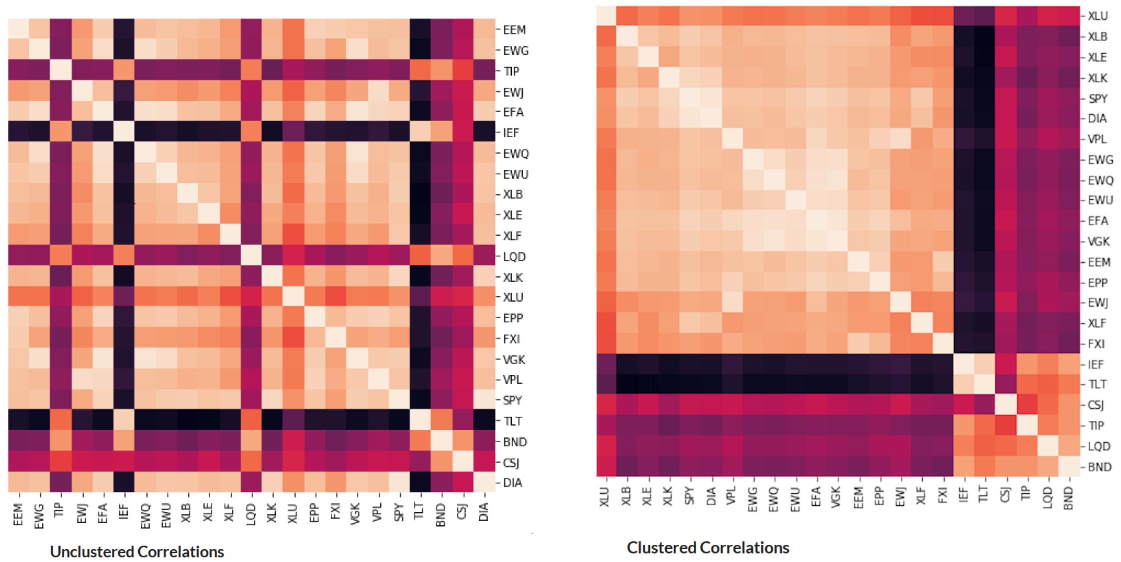

Figure 9: Comparison of correlation matrices before and after the quasi diagonalization step.
Source: Own calculations.

Appendix B

Tickers, full names, description of the representations, periods for which the information is available, and the annualized Sharpe ratios of each ETF available in Table 4.

Table 4: Information about ETFs used in a portfolio.

Ticker	Full name	Represents	Period for which data is available	Annualized Sharpe ratio	Average daily return
DIA	SPDR Dow Jones Industrial Average ETF	US industrial equity	01.01.2007 – 20.12.2019	0.413794	0.000295
EEM	iShares MSCI Emerging Markets ETF	Emerging markets equity	01.01.2007 – 20.12.2019	0.179114	0.000216
FEZ	SPDR EURO STOXX 50 ETF	EU top equity	01.01.2007 – 20.12.2019	0.046078	0.000050
FVD	First Trust Value Line Dividend Index	Dividend equity	01.01.2007 – 20.12.2019	0.420422	0.000275
GDX	VanEck Vectors Gold Miners ETF	Gold mining	01.01.2007 – 20.12.2019	0.138073	0.000225
GLD	SPDR Gold Trust	Physical gold	01.01.2007 – 20.12.2019	0.256152	0.000156
HYG	iShares iBoxx $ High Yield Corporate Bond ETF	US Corporate bonds	11.04.2007 – 20.12.2019	-0.060984	-0.000028
IGV	iShares Expanded Tech-Software Sector ETF	US Technology sector	01.01.2007 – 20.12.2019	0.636762	0.000566
IWB	iShares Russell 1000 ETF	US broad top equity	01.01.2010 – 20.12.2019	0.747228	0.000435
IYR	iShares U.S. Real Estate ETF	US real estate	01.01.2007 – 20.12.2019	0.183211	0.000222
QQQ	Invesco QQQ	US top tech sector	01.01.2007 – 20.12.2019	0.658965	0.000535
SDY	SPDR S&P Dividend ETF	US dividend equity	01.01.2010 – 20.12.2019	0.671095	0.000352

SMH	VanEck Vectors Semiconductor ETF	US semiconductor companies	21.12.2011 – 20.12.2019	0.953392	0.000799
TLT	iShares 20+ Year Treasury Bond ETF	US long-term bonds	01.01.2007 – 20.12.2019	0.311598	0.000173
VT	Vanguard Total World Stock ETF	World equity	01.01.2010 – 20.12.2019	0.436400	0.000275
XBI	SPDR S&P Biotech ETF	US biotech	01.01.2007 – 20.12.2019	0.578194	0.000665
XHB	SPDR S&P Homebuilders ETF	US home building	01.01.2007 – 20.12.2019	0.207929	0.000263
XLB	Materials Select Sector SPDR ETF	US materials	01.01.2007 – 20.12.2019	0.282171	0.000267
XLE	Energy Select Sector SPDR Fund	US energy	01.01.2007 – 20.12.2019	0.143413	0.000158
XLF	Financial Select Sector SPDR Fund	US financial	01.01.2007 – 20.12.2019	0.150885	0.000193
XLI	Industrial Select Sector SPDR Fund	US industrial	01.01.2007 – 20.12.2019	0.396373	0.000329
XLK	Technology Select Sector SPDR Fund	US technology	12.04.2007 – 20.12.2019	0.571830	0.000463
XLP	Consumer Staples Select Sector SPDR Fund	US consumer staples	01.01.2007 – 20.12.2019	0.542954	0.000290
XLU	Utilities Select Sector SPDR Fund	US utilities	01.01.2007 – 20.12.2019	0.328570	0.000227
XLV	Health Care Select Sector SPDR Fund	US health care	01.01.2007 – 20.12.2019	0.543429	0.000357
XLY	Consumer Discretionary Select Sector SPDR Fund	US consumer discretionary	01.01.2007 – 20.12.2019	0.517679	0.000431
XRT	SPDR S&P Retail ETF	US retail	01.01.2007 – 20.12.2019	0.361500	0.000357
MCHI	iShares MSCI China ETF	China equity	31.03.2011 – 20.12.2019	0.174572	0.000160
USMV	iShares Edge MSCI Min Vol USA ETF	US low volatility equity	20.10.2011 – 20.12.2019	1.121525	0.000460

MTUM	iShares Edge MSCI USA Momentum Factor ETF	US high momentum equity	18.04.2013 – 20.12.2019	0.940691	0.000541
QUAL	iShares Edge MSCI USA Quality Factor ETF	US quality factor equity	18.07.2013 – 20.12.2019	0.812114	0.000420
DGRO	iShares Core Dividend Growth ETF	US dividend equity	12.06.2014 – 20.12.2019	0.753542	0.000370

Source: Own calculations.

Note: The table above describes each of the ETFs included in the portfolio. The period for which data was available as well as the annualized Sharpe ratios help to understand how each individual asset has performed.

Appendix C

Transaction costs used in the backtesting of the algorithm are the Interactive Brokers' commissions for ETFs according to the fixed pricing structure.

Table 5: Transaction costs used in a backtest.

Action	Fixed cost	Minimum cost per order	Maximum cost per order
Buy an ETF	USD 0.005	USD 1.00	1.0% of trade value
Sell an ETF	USD 0.005	USD 1.00	1.0% of trade value
Holding a position	No transaction fees		

Source: Interactive Brokers' Commissions https://www.interactivebrokers.ca/en/index.php?f=45251&p=stocks1

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl