Computational Economics (2025) 66:5189–5206
https://doi.org/10.1007/s10614-025-10848-w

An Empirical Evaluation of Distance Metrics in Hierarchical Risk Parity Methods for Asset Allocation

Francisco Salas-Molina[1] · **David Pla-Santamaria**[1] · **Ana Garcia-Bernabeu**[1] · **Adolfo Hilario-Caballero**[2]

Accepted: 5 January 2025 / Published online: 14 February 2025

Abstract
Hierarchical Risk Parity methods address instability, concentration, and underperformance in asset allocation by taking advantage of machine learning techniques to build a diversified portfolio. HRP methods produce a hierarchical structure to the correlation between assets by means of tree clustering that results in a reorganization of the covariance matrix of returns. However, HRP admits multiple variations in terms of clustering algorithms and distance metrics. In this paper, we evaluate the out-of-sample performance of alternative hierarchical distance metrics for clustering purposes using real stock markets in three different market scenarios: bull market, sideways trend, and bear market. We pay special attention to the mean-variance performance of the output portfolios as an estimation of the ability of alternative methods to estimate future return and risk. Our results show that correlation-based metrics provide better performance than non-correlation metrics. In addition, HRP methods outperform quadratic optimizers in two of the three stock market scenarios.

Keywords Portfolio selection · Machine learning · Clustering · Distance metrics · Risk management

✉ Francisco Salas-Molina
frasamo@upv.es

David Pla-Santamaria
dplasan@upv.es

Ana Garcia-Bernabeu
angarber@upv.es

Adolfo Hilario-Caballero
ahilario@upv.es

1 Economics and Social Sciences, Universitat Politècnica de València, Ferràndiz y Carbonell, s/n, 03801 Alcoy, Spain

2 Systems Engineering and Automation, Universitat Politècnica de València, Ferràndiz y Carbonell, s/n, 03801 Alcoy, Spain

1 Introduction

Portfolio selection deals with the problem of allocating a budget to several assets so that the expected return is maximized for a given level of risk. Indeed, investors do not only care about profit but also about the amount of risk accepted to achieve that profit. Modern portfolio theory was initiated by Markowitz (1952) and is an ongoing research topic within computational economics. Despite the historical discussion (Levy, 1974; Kroll et al., 1984; Markowitz, 1991), the mean-variance efficient frontier is assumed to be a suitable technique for obtaining good investment results. Indeed, all the portfolios on the efficient frontier show the highest expected return for each level of risk, and simultaneously, they have the minimum risk for each particular level of expected return.

Next, we will mention some recent examples from the vast literature on the topic to provide an overview. Gulliksson et al. (2023) consider the case when the covariance matrix of the asset returns is rank-deficient to derive a unique analytical solution. The authors also developed an iterative method and showed its convergence. Qi and Steuer (2024) also follow an analytical approach to derive closed-form formulae for the computation of the properly efficient and weakly efficient sets of portfolio selection problems with multiple quadratic objectives. Steuer and Utz (2023) propose a tri-criterion approach that computes efficient surfaces and special non-contour curves in portfolio selection problems with multiple quadratic objectives.

Another line of literature originates from using heuristic approaches inspired by biological processes. These approaches do not provide an exact optimal solution but can generate approximate solutions efficiently. For example, Khan et al. (2023) propose a quadratic interpolated beetle antennae search algorithm to solve the higher dimensional portfolio selection problems as a way to overcome the limitations of traditional beetle antennae search methods through robust approximation. Garcia-Bernabeu et al. (2019) and Garcia-Bernabeu et al. (2024) rely on multiobjective evolutionary genetic algorithms to ensure convergence towards the efficient surface with limited memory resources.

In this paper, we follow a different approach to analytical and heuristic solutions as a continuation of the work initiated by De Prado (2016) to integrate graph theory and machine learning in portfolio selection through Hierarchical Risk Parity (HRP) methods. Machine learning techniques in finance have become very popular among researchers due to their ability to solve complex data-based problems. Some recent examples are the possibility of obtaining cost savings in cash management by improving predicting accuracy (Salas-Molina et al., 2017), or significant economic gains to investors using machine learning forecasts (Gu et al., 2020, 2021). HRP methods rely on machine learning techniques such as tree clustering to produce a hierarchical structure to the correlation between assets that increases stability and yields more intuitive results (De Prado, 2018).

Within the framework of the classical mean-variance model by Markowitz (1952), there is a need to estimate expected returns and the covariance matrix. First, the implications of estimation uncertainty in modern portfolio theory have been extensively researched in the literature by Ledoit and Wolf (2003, 2017, 2022); Bodnar et al. (2018) and Javed et al. (2024) among many others. Second, Markowitz's mean-variance approach requires the use of quadratic optimizers to find the best solutions for portfolio weights. Quadratic optimizers require the invertibility of the estimated covariance matrix, which imposes some limitations: instability, concentration, and underperformance. Instability implies that minor estimation errors may lead to significant deviations in optimal allocations (Kolm et al., 2014). Concentration refers to reducing portfolio weights to a small number of assets, hence becoming exposed to particular shocks (De Prado, 2018). Finally, underperformance indicates that some portfolio optimization methods are optimal in-sample but often provide relatively poor out-of-sample forecast performance (De Miguel et al., 2009).

To overcome these limitations, Black and Litterman (1992) proposed introducing the tactical views of portfolio managers in asset allocation, and Ledoit and Wolf (2003) introduced regularization techniques in the covariance matrix estimators. Other authors followed risk-based asset allocation approaches as described by Roncalli (2013) and Jurczenko (2015). More precisely, risk parity methods do not rely on expected returns but only on estimating risk measures. The term risk parity was coined by Qian (2005). Risk parity methods are based on building a balanced portfolio so that the risk contribution is the same for different assets. Furthermore, the HRP method was introduced by De Prado (2016) and relies on graph theory and machine learning to build a diversified portfolio. Returns covariance and correlation matrices lack the notion of hierarchy leading to portfolio weights to vary freely in unintended ways. HRP provides a new asset allocation method based on three key steps: hierarchical tree clustering, quasi-diagonalization of the covariance matrix, and recursive bisection to obtain portfolio weights through an inverse-variance procedure. Indeed, HRP can compute portfolio weights without requiring the invertibility of the covariance matrix, relying on an ill-degenerated or even a singular covariance matrix (De Prado, 2018; Raffinot, 2017). Recent applications of the HRP method are described in Molyboga (2020); Burggraf (2021); Sen and Dutta (2022), and Cho and Song (2023).

However, the HRP approach admits multiple variations with respect to the use of distance functions, clustering algorithms, calculation of weights, or the use of additional allocation constraints. As an extension of the works by De Prado (2016, 2020), the main goal of this work is to compare alternative HRP strategies to find out which techniques perform best out-of-sample. In this paper, we follow a computational approach to compare the ability of alternative HRP methods to forecast risk and profits under different economic scenarios. We hypothesize that there are neither distance metrics nor clustering algorithms that perform best in all possible scenarios. To this end, we work with portfolios derived from assets within the IBEX35 Spanish stock index in three different trends: bull market (from 2005 to 2008), sideways trend (from 2014 to 2019), and bear market (from 2008 to 2012). We compare different methods by computing a Normalized Mean Square Error (NMSE) of the average returns as a measure of expected profits and the variance of returns as a measure of expected risk. By using Cross-Validation (CV) to determine the generalization error of alternative methods, we depart from the works by De Prado (2016,

2020) to estimate future mean-variance performance. We explore different distance metrics and clustering algorithms. We argue that the Spanish market trends in the three scenarios are an example of the general trends in world markets, so the results presented can be considered sufficiently representative.

In our empirical evaluation, we focus on the combined return and risk generalization error using the out-of-sample mean return as a measure of returns and the out-of-sample variance of returns as a measure of risk. Our results show that correlation-based metrics perform better than non-correlation metrics in the three scenarios. In addition, we also find that the classical quadratic optimizer is the best option in the bear-market scenario and presents a slightly worse performance compared to the best HRP method in the sideways-trend scenario. HRP methods outperform the quadratic optimizer in the bull-market scenario. In addition, HRP methods show better combined return-risk performance than following the market by investing in the market index in the three scenarios considered.

Summarizing, this work distinguishes itself from recent works on portfolio selection through HRP methods, as proposed by De Prado (2016, 2020), in the evaluation of alternative distances through the generalization error about return and risk performance in different market scenarios. More precisely, the main contributions of our paper are as follows:

1. An empirical evaluation of alternative distance metrics used by tree clustering algorithms in HRP methods.
2. An empirical evaluation of alternative models through the generalization error about expected returns and risk.
3. An empirical evaluation of different investing strategies in different market scenarios.

In addition to this introduction, we provide useful background on hierarchical risk parity methods for asset allocation in Sect. 2. We describe our empirical evaluation results in Sect. 3, and we conclude this paper in Sect. 4, analyzing the implications of the results and pointing out future lines of work.

2 Useful Background on Hierarchical Risk Parity Methods

In this section, we describe the HRP method proposed by De Prado (2016, 2018), which will be later used in the empirical section. The underlying idea of this method is the notion of hierarchy using trees. A tree is an undirected graph in which any two nodes are connected by exactly one path. In our context, nodes represent assets and edges represent the correlation between assets. Starting from a correlation matrix among the returns of any pair of assets, the first goal is to derive a hierarchical structure, namely, a tree structure. This structure is used to reorganize the covariance matrix so that the large values lie along the diagonal. Finally, the reorganized list of assets and the covariance reorganization are used as inputs to a recursive inverse variance allocation algorithm to elicit the weights of the assets. As a result, this method is based on three stages: 1) tree clustering, 2) quasi-diagonalization, and 3) recursive bisection.

2.1 Tree Clustering

Tree clustering aims to combine a matrix of observations into a hierarchical structure of clusters. Hierarchical clustering seeks to build a hierarchy of clusters following an agglomerative strategy in which each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. The input of this step is a matrix containing N column vectors with returns series of N assets over T periods. The goal is to construct a tree so that allocations can flow downstream. The output of this step is a tree represented by a linkage matrix obtained through the following steps:

1. Compute an $N \times N$ correlation matrix with entries ρ_{ij} set to the correlation coefficient between variables X_i and X_j. These correlations are used to establish the strength of the relationship among assets.
2. Define a distance metric d_{ij} and compute $N \times N$ distance matrix $D = \{d_{ij}\}_{i,j=1,\dots,N}$. This distance function allows the selection of pairs of assets that are close in terms of distances. Note that there are multiple distance metrics that can be defined for evaluation purposes as we next further elaborate in Sect. 3.
3. Compute the Euclidean distance $\tilde{d}_{ij}$ between any two column-vectors from D so that we can observe the closeness of any pair of assets.
4. Cluster together columns (i^*, j^*) such that $(i^*, j^*) = \text{argmin}(i,j)_{i \neq j}\{\tilde{d}_{ij}\}$, and denote this cluster as $u[1]$. The minimum Euclidean distance of correlation metrics implies the formation of a cluster of two assets.
5. Define distance $\dot{d}_{i,u[1]}$ between cluster $u[1]$ and any item i of the rest of unclustered items as the linkage criterion typical in hierarchical clustering. This new distance function establishes the closeness of the new cluster of two assets as a whole and the rest of the assets.
6. Update matrix D by appending $\dot{d}_{i,u[1]}$ and dropping the clustered columns and rows $j \in u[1]$. The assets belonging to the new cluster are replaced by the cluster itself.
7. Apply recursively steps 4, 5, and 6 to append $N-1$ clusters such that the final cluster contains all of the original items. The final structure is a tree denoting a hierarchy of assets defined in terms of distances between any pair of merging leaves.

The main result from this algorithm is a $(N-1) \times 4$ linkage matrix with the following structure:

$$Y = \{(y_{m,1}, y_{m,2}, y_{m,3}, y_{m,4})\}_{m=1,\dots,N-1} \tag{1}$$

where $y_{m,1}$ and $y_{m,2}$ are the constituents of the cluster, $y_{m,3} = \tilde{d}_{y_{m,1},y_{m,2}}$ is the distance between $y_{m,1}$ and $y_{m,2}$, and $y_{m,4}$ is the number of original items included in cluster m. Note that constituents $y_{m,1}$ and $y_{m,2}$ and the number of original items from the last row of Y will be used in the next step. Distance between constituents $y_{m,3}$ is useful to plot a tree structure such as a dendrogram in which the vertical axes measure the distance between merging leaves (De Prado, 2018).

2.2 Quasi-Diagonalization of the Covariance Matrix

This step reorganizes the rows and columns of the covariance matrix V so that the largest values lie along the diagonal. After the quasi-diagonalization of the covariance matrix, similar investments are placed together, and dissimilar investments are placed far apart. The quasi-diagonalization method replaces clusters in $(y_{N-1,1}, y_{N-1,2})$ with their constituents recursively until no clusters remain. The replacements preserve the order of the clustering and the output is a sorted list of original (unclustered) items in a quasi-diagonal matrix. This function traverses through a hierarchical clustering linkage matrix in reverse order, decomposing each cluster until it reaches the original items, through the following steps:

1. Initialize list L_0 as a series with the last two cluster constituents in Y.
2. Set *nItems* to the total number of original items from the last row of Y.
3. While the maximum value in L_0 is greater than or equal to *nItems*:
    a. Reindex L_0 to allow insertion between elements.
    b. Find indices in L_0 corresponding to clusters.
    c. Get the row in Y corresponding to this cluster.
    d. Replace the cluster index in L_0 with the first item in the cluster.
    e. Insert the second item of the cluster after the first one in L_0.
    f. Sort L_0 by index to maintain the quasi-diagonal order.
    g. Reset the index of L_0 to be consecutive.
4. Return L_0 as a list.

The ordered list L_0 of assets and the covariance matrix V are then used as inputs in the next step. Figures 1 and 2 show the covariance matrix before and after quasi-diagonalization of the covariance matrix of assets in the IBEX 35 Spanish stock index using a color scale.

2.3 Recursive Bisection

To derive the portfolio weights, the last step of the HRP method takes advantage of the fact that the inverse-variance allocation procedure is optimal for a diagonal covariance matrix. The inverse-variance allocation procedure computes weights inversely to their variances so that assets with lower variance receive higher weights, while those with higher variance receive lower weights. From a bottom-up perspective, this recursive bisection step defines the variance of a contiguous subset as the variance of an inverse-variance allocation. From a top-down perspective, the recursive bisection split allocations between adjacent subsets in inverse proportion to

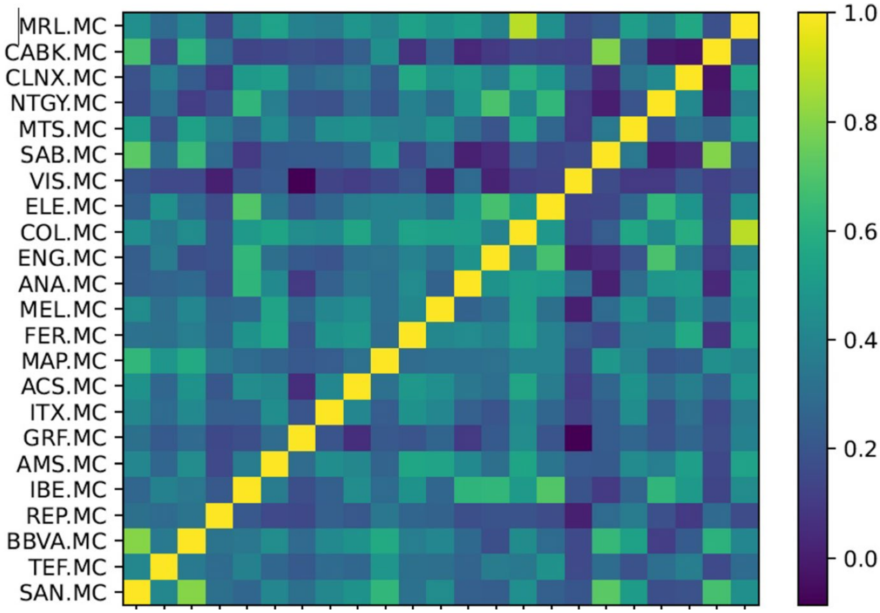

Fig. 1 Covariance matrix before quasi-diagonalization

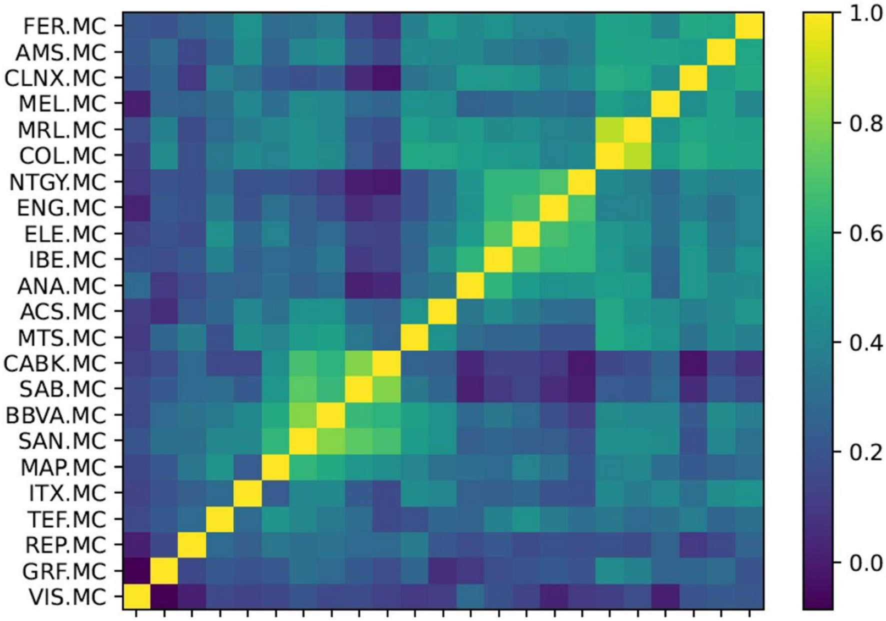

Fig. 2 Covariance matrix after quasi-diagonalization

their aggregated variances. As a result, the recursive bisection method outputs a vector of weights by means of the following algorithm:

1. Initialization:
   a. Set list of items $L = \{L_0\}$, with $L_0 = \{n\}_{n=1,\dots,N}$.
   b. Assign unit weights to all items: $w_n = 1, \forall n = 1,\dots,N$.
2. If $|L_i| = 1, \forall L_i \in L$, then stop.
3. For each $L_i \in L$ such that $|L_i| > 1$:
   a. Bisect L_i into two subsets, $L_i^{(1)} \cup L_i^{(2)}$, where $|L_i^{(1)}| = \text{int}[\frac{1}{2}|L_i|]$, and the order is preserved.
   b. Define the variance of $L_i^{(j)}$, $j = 1, 2$, as the quadratic form:

$$\tilde{V}_i^j \equiv \tilde{w}_i^{(j)\prime} V_i^{(j)} \tilde{w}_i^{(j)} \tag{2}$$

   where $V_i^{(j)}$ is the covariance matrix between the constituents of the $L_i^{(j)}$ bisection, and weights are set to:

$$\tilde{w}_i^{(j)} = \text{diag}\left[V_i^{(j)}\right]^{-1} \frac{1}{\text{tr}[\text{diag}[V_i^{(j)}]^{-1}]} \tag{3}$$

   where diag[.] and tr[.] are, respectively, the diagonal and trace operators.
   c. Compute split factor α_i so that $0 \le \alpha_i \le 1$:

$$\alpha_i = 1 - \frac{\tilde{V}_i^{(1)}}{\tilde{V}_i^{(1)} + \tilde{V}_i^{(2)}}. \tag{4}$$

   d. Re-scale allocations w_n by a factor of α_i, $\forall n \in L_i^{(1)}$.
   e. Re-scale allocations w_n by a factor of $(1 - \alpha_i)$, $\forall n \in L_i^{(2)}$.
4. Loop to step 2.

The recursive bisection algorithm guarantees that all weights are between zero and one ($0 \le w_i \le 1,\ \forall i = 1,\dots,N$), and that all weights add up to one ($\sum_{i=1}^{N} w_i = 1$), because at each iteration we are splitting the weights from higher hierarchical levels. Additional constraints can be introduced in this stage by replacing equations in steps 3c, 3d, and 3e according to the user's preferences.

3 Empirical Evaluation of Metrics in Hierarchical Risk Parity Methods

In this section, we compare alternative distance metrics d_{ij} in the second step of the tree clustering HRP method described in Sect. 2.1. We examine the ability of alternative metrics to estimate the risk and profits of portfolios derived from assets within the IBEX 35 Spanish stock index in three different time trends: bull market, sideways trend, and bear market. More precisely, we compare the Mean Square Error (MSE) of the average returns as a measure of profits and the variance of return as a measure of risk. To this end, we follow a Cross-Validation (CV) approach.

3.1 Data and Benchmarks

To assess the impact of market trends in the performance of alternative estimation methods, we use three different data sets for weekly returns in the Spanish Stock Exchange (IBEX 35) used in Bravo et al. (2022) as summarized in Table 1. More precisely, we identify three different trends: 1) bull market (from 2005 to 2008) with an average weekly return of 0.351%; 2) sideways trend (from 2014 to 2019) with an average weekly return of −0.001%; and 3) bear market (from 2008 to 2012) with an average weekly return of −0.189%. Interested readers can obtain the data themselves at *http://es.finance.yahoo.com*.

The evolution of the market index IBEX of the three different scenarios is shown in Fig. 3. Note that we removed the short bull trend in 2013 to focus on the sideways trend from 2014 onwards.

We use the return-risk performance of the market index IBEX35 as a benchmark. In addition, we use a quadratic optimizer to solve the classical mean-variance model by Markowitz (1952) for a null target return for comparative purposes. To solve the quadratic optimization models, we use the open-source library SciPy in Python (Version 3.10) and Jupyter Notebooks (Kluyver et al., 2016) as described in Hilpisch (2014) that relies on the SLSQP (Sequential Least Squares Programming) method initially implemented by Kraft (1988). We also compare the results obtained to more recent approaches in portfolio selection, such as a risk budgeting approach and a Sharpe Ratio maximization using evolutionary algorithms. More precisely, we follow a risk budgeting allocation in which weights are split in inverse proportion to the assets' variance (Roncalli, 2013; De Prado, 2018). We denote this risk budgeting approach as Inverse Variance Portfolio (IVP). In the case of the Sharpe Ratio (SR) maximization, we follow the approach described by Garcia-Bernabeu et al. (2019) to deal with the non-linearity of the Sharpe Ratio using evolutionary algorithms. We denote this approach as Evolutionary Sharpe Ratio (ESR). Finally, we use a naive equal-weight strategy.

Table 1 Data sets summary (Avg Ret: average weekly return for IBEX 35)

Data set	Initial date	Final date	Weeks	Assets	Avg Ret (%)
Bull market	14/01/2005	28/12/2007	155	38	0.351
Sideways trend	27/04/2014	07/04/2019	259	43	−0.001
Bear market	06/01/2008	12/08/2012	241	41	−0.189

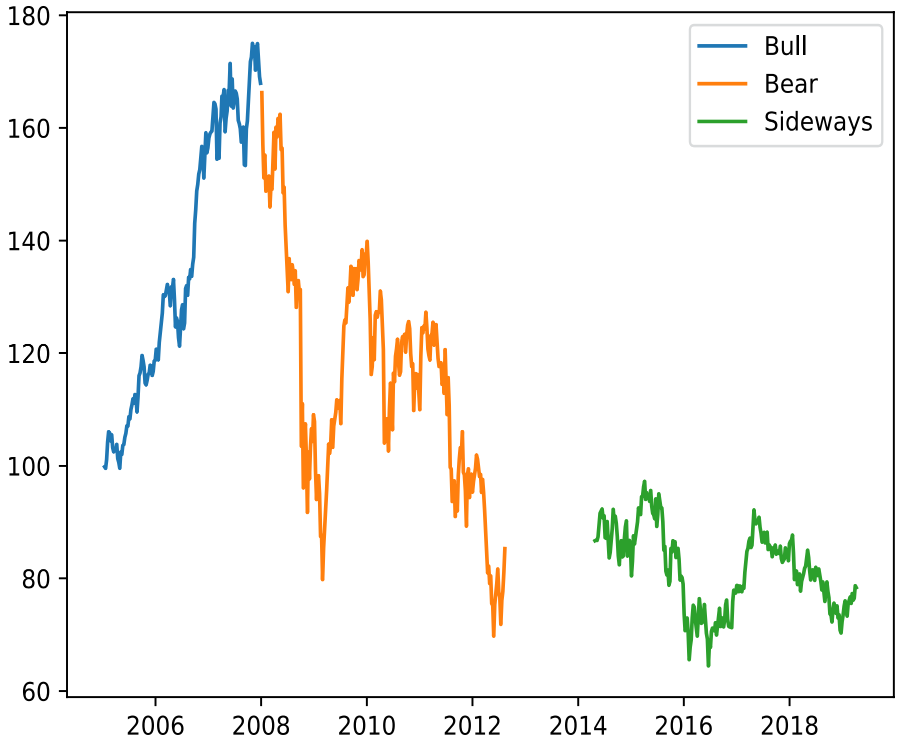

Fig. 3 The three scenarios from Table 1

3.2 Cross-Validation Methodology

The purpose of a CV methodology is to determine the generalization error of a predictive algorithm. In our case, we test the ability of alternative methods to forecast mean-variance performance. By computing the mean-variance performance of portfolios out-of-sample, we can evaluate the predictive ability of any method. Then, we proceed as follows:

1. Split data in k subsets without shuffling.
2. For $i = 1, \dots k$
   a. Obtain fitted portfolios based on all subsets excluding i. Store mean and variance as expected performance measures for in-sample portfolios.
   b. Test fitted portfolios on subset i by comparing out-of-sample mean and variance as real performance measures.
3. Compute average estimation errors for the model with respect to a benchmark.

To compute the average estimation errors for alternative models, we consider a set of distances $A = \{A_0, A_1, \dots, A_n\}$. Then, we compute the Mean Squared Error (MSE) for $A_j \in A$ as follows:

$$MSE(A_j) = \frac{1}{k}\sum_{i=1}^{k}(\hat{y}_i - y_i)^2 \tag{5}$$

where $\hat{y}_i$ is a predicted performance measure, such as return and volatility, using the portfolio derived from the training set, and y_i is the performance measure of the same portfolio but evaluated on the test set. For comparative purposes, we use the MSE for $A_0 = IBEX$ (the market index performance) as a normalization factor:

$$NMSE(A_j) = \frac{MSE(A_j)}{MSE(A_0)}. \tag{6}$$

In addition to the benchmark algorithm A_0, we consider six different HRP methods in which we change distance metric d_{ij} in the second step of the tree clustering algorithm described in Sect. 2.1. Next, we describe the alternative distance metrics used in our experiments.

3.3 Alternative Distance Metrics

We first use the correlation distance metric described in our experiments in De Prado (2016, 2018).

$$d_{ij}^{(1)} = \sqrt{0.5(1 - \rho_{ij})}. \tag{7}$$

An interesting property of $d_{ij}^{(1)}$ is that two random variables with negative correlation are more distant than two random variables with positive correlation, regardless of their absolute value. This property is helpful in a long-only portfolio, where negative-correlated securities can offset risk and should be treated as different for diversification purposes.

Distance $d_{ij}^{(1)}$ is a non-linear function of ρ_{ij} restricted to the domain [0, 1]. In order to examine the influence of non-linearity in the use of distance metrics, we also consider the following linear transformation of ρ_{ij} as an additional distance function:

$$d_{ij}^{(2)} = \frac{1 - \rho_{ij}}{2}. \tag{8}$$

However, correlation-based distances present some limitations (De Prado, 2020). First, correlation is a measure of the linear codependency between two random variables, hence neglecting nonlinear relationships. Second, correlation is highly influenced by outliers. Third, even though the correlation function can be applied to any pair of random variables, its application assumes a multivariate normal distribution of random variables. To overcome these limitations, we consider an additional distance metric based on the concept of normalized variation of information $VI[X, Y]$ for random variables X and Y as described in De Prado (2020):

$$d_{ij}^{(3)} = VI[X, Y] = H[X, Y] - I[X, Y] \tag{9}$$

where $H[X, Y]$ is the joint entropy of random variables X and Y, and $I[X, Y]$ is the informational gain in X that results from knowing the value of Y. For computations, we use the Python implementation of the normalized variation of information provided in De Prado (2020).

Finally, we also consider normalized Minkowski distance $M(X, Y, p)$ between vectors of returns for different values of parameter p ranging in $[1, \infty]$ with normalization factor $1/n$:

$$M(X, Y, p) = \left[\frac{1}{n} \sum_{i=1}^{n} | x_i - y_i |^p \right]^{1/p}. \tag{10}$$

More precisely, we define three new distances by setting p to the common values 1, 2, and ∞, respectively, to obtain the Manhattan, Euclidean, and Chebyshev distances:

$$d_{ij}^{(4)} = \frac{1}{n} \sum_{i=1}^{n} | x_i - y_i | \tag{11}$$

$$d_{ij}^{(5)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2} \tag{12}$$

$$d_{ij}^{(6)} = \frac{1}{n} \max | x_i - y_i | . \tag{13}$$

3.4 Empirical Results

The results of our empirical analysis are summarized in Table 2. This table shows the NMSE for returns and risk estimations provided by the HRP methods using the alternative distance metrics described in Sect. 3.3 and alternative methods as benchmarks. The lower the average NMSE, the better the out-of-sample estimation for returns and risk provided by each method according to the CV methodology. For benchmarking purposes, we include the performance of portfolios obtained using a classical quadratic optimizer, an Inverse Variance Portfolio (IVP), an Evolutionary Sharpe Ratio (ESR) maximization algorithm, and an equal-weight portfolio. Using a quadratic optimizer within the classical Markowitz mean-variance framework is motivated by the fact that this approach is a cornerstone of finance theory and is widely taught and understood, representing a clear standard for portfolio optimization methods. On the other hand, an

Table 2 Combined return-risk NMSE evaluation for different portfolio selection methods (HRP: Hierarchical risk parity; IVP: inverse variance portfolio; ESR: evolutionary Sharpe ratio)

Scenario	Method	Return	Risk	Euclidean
Bull	HRP-$d_{ij}^{(1)}$	1.01	0.58	1.16
	HRP-$d_{ij}^{(2)}$	1.00	0.56	**1.15**
	HRP-$d_{ij}^{(3)}$	1.04	0.58	1.19
	HRP-$d_{ij}^{(4)}$	1.09	0.77	1.33
	HRP-$d_{ij}^{(5)}$	1.07	0.84	1.36
	HRP-$d_{ij}^{(6)}$	1.06	0.72	1.28
	HRP average	1.05	0.68	1.25
	Quadratic	1.03	0.83	1.32
	IVP	1.09	0.84	1.38
	ESR	1.12	0.85	1.41
	Equal-weight	1.36	1.27	1.86
Sideways	HRP-$d_{ij}^{(1)}$	0.74	0.67	**1.00**
	HRP-$d_{ij}^{(2)}$	0.83	0.67	1.07
	HRP-$d_{ij}^{(3)}$	0.76	0.77	1.08
	HRP-$d_{ij}^{(4)}$	0.83	0.79	1.15
	HRP-$d_{ij}^{(5)}$	0.80	0.77	1.11
	HRP-$d_{ij}^{(6)}$	0.79	0.76	1.10
	HRP average	0.79	0.74	1.08
	Quadratic	0.76	0.66	1.01
	IVP	0.77	0.78	1.10
	ESR	1.46	0.72	1.63
	Equal-weight	1.31	1.20	1.78
Bear	HRP-$d_{ij}^{(1)}$	0.64	0.34	0.72
	HRP-$d_{ij}^{(2)}$	0.62	0.36	0.72
	HRP-$d_{ij}^{(3)}$	0.65	0.35	0.74
	HRP-$d_{ij}^{(4)}$	0.74	0.43	0.86
	HRP-$d_{ij}^{(5)}$	0.73	0.44	0.85
	HRP-$d_{ij}^{(6)}$	0.65	0.41	0.77
	HRP average	0.67	0.39	0.78
	Quadratic	0.52	0.30	**0.60**
	IVP	0.75	0.49	0.90
	ESR	0.66	0.25	0.71
	Equal-weight	1.16	0.64	1.32

Bold values represent best values

equal-weight portfolio serves as a natural baseline due to its inherent characteristics. This approach is simple and transparent and avoids the need for parameter estimation, unlike the Markowitz model. Any advanced method should at least outperform this naive approach to justify its added complexity as highlighted by De Miguel et al. (2009). A similar reasoning validates the use of the IVP strategy.

Finally, using evolutionary algorithms allows us to consider non-linear objective functions such as the Sharpe Ratio and to compare to recent approaches described in the portfolio selection literature. Any new method can credibly claim to advance portfolio selection techniques by outperforming these benchmarks.

To allow a combined bicriteria comparison in terms of both return and risk NMSE, we compute the Euclidean distance from every pair of return and risk values to the point (0, 0). With minimum return and risk deviation measured in terms of NMSE, this point is clearly the ideal point and acts as a reference point for global comparison purposes. Bolded figures identify the best values for the alternative portfolio selection methods in the bull market, sideways trend, and bear market scenarios.

In the case of the bull-market scenario, we find that all the methods perform slightly worse than the market index for returns (NMSE values slightly above one) and significantly better for risk (NMSE values below one). This situation is unsurprising because high returns and high volatility characterize bull markets. As a result, the market index benefits from high returns in exchange for high volatility. On the contrary, HRP methods focus on building diversified portfolios with less risk. Indeed, we find that the average return performance of alternative methods is similar to one, but the risk performance is significantly below the risk provided by the market index. Concerning the ability of alternative metrics, we find that the best distance metric is the linear correlation-base metric $d_{ij}^{(2)}$. Minkowski distance metrics are below the average of both the correlation-based metrics and the entropy metric performances. In addition, the HRP methods are significantly better than the naive equal-weight portfolios in terms of return and risk. Finally, the best HRP method is globally better than the quadratic optimizer due to the better risk performance.

In the case of the sideways-trend scenario, all the methods perform better than the market index for both returns and risk, with the exception of the ESR and equal-weight strategy. In this scenario, HRP methods are able not only to diversify risk but also to clearly outperform the masrket index in terms of returns. As a result, both the average return and risk performance of alternative methods are significantly below one. With respect to the ability of alternative metrics, we find that the best distance metric is the correlation-based metric $d_{ij}^{(1)}$. We also observe that the best HRP method performs slightly better than the quadratic optimizer. Focusing on the particular NMSE values for HRP-$d_{ij}^{(1)}$ and the quadratic optimizer, we find the HRP method is better in returns but worse in risk. Finally, all the HRP methods considered in this scenario globally outperform the ESR and equal-weight strategy, with a similar average performance to IVP. Again, the average performance in risk is better than the average performance in returns.

In the case of the bear-market scenario, the average performance of all the methods considered is even better than in the case of the sideways trend, both in terms of returns and risk. We find that the average return performance of alternative HRP methods is remarkably below one, and the average risk performance is below half of the risk provided by the market index. With respect to the ability of alternative metrics, we find that the best distance metrics are the correlation-based metrics $d_{ij}^{(1)}$ and

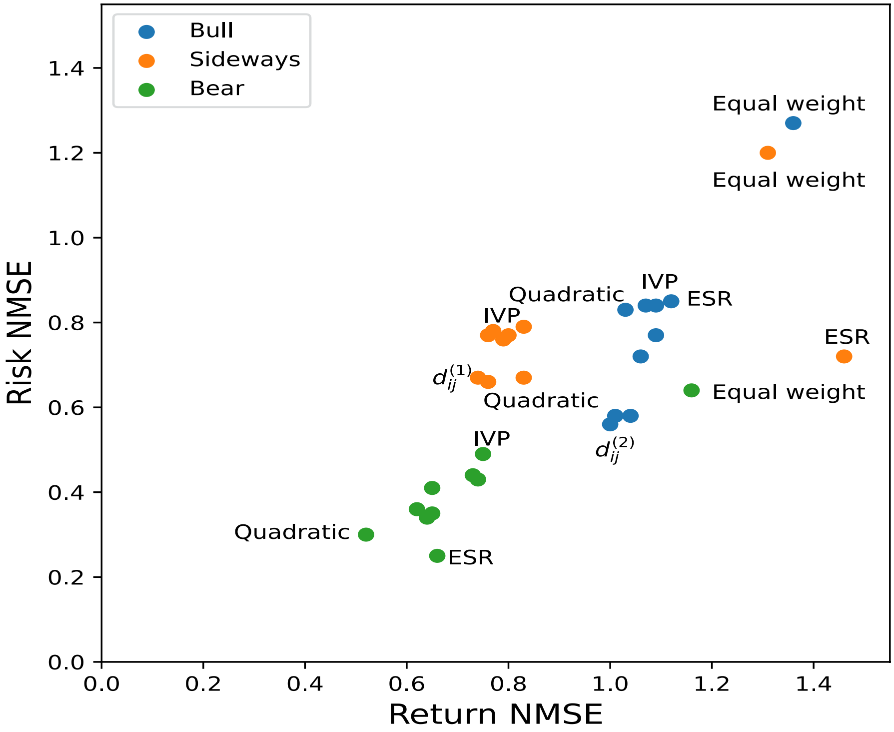

Fig. 4 NMSE values in the return-risk space

the $d_{ij}^{(2)}$. In addition, the HRP methods are significantly better than the equal-weight portfolios but similar to IVP and ESR. However, the quadratic optimizer produced the best results in this scenario.

As a result of our empirical analysis, we observe important deviations between the performance of alternative methods and distance metrics for different market scenarios. These deviations can be visualized in terms of NMSE for returns and risk in Fig. 4, in which some interesting points are labeled for ease of understanding. Note that the ideal point for combined performances is the origin point (0, 0). Therefore, the closer any point to this reference point, the better the combined return-risk performance. On the one hand, we find that the best performance for both HRP and quadratic optimizers is achieved in the bear-market scenario. The ability to outperform the market index is also found in the sideways-trend scenario but is limited to better performance only in the risk perspective when considering the bull-market scenario. On the other hand, when comparing HRP methods with a naive equal-weight investment strategy, we find that the HRP methods provide better combined results in the three market scenarios. Finally, the quadratic optimizer is the best option in the bear-market scenarios, but the best HRP method presents a slightly better performance in the sideways-trend scenario and a significantly better performance in the bull-market scenario.

With respect to the alternative distance metrics, we find that the best alternative in the three market scenarios considered are correlation-based metrics $d_{ij}^{(1)}$ and $d_{ij}^{(2)}$. There is a slight difference in performance in the bull-market scenario in favor of $d_{ij}^{(2)}$, and there is also a difference in performance in the sideways-trend scenario in favor of $d_{ij}^{(1)}$. In general, non-correlation metrics such as the entropy-based metric and the metrics derived from the Minkowski distance function performed worse in the three scenarios. As a result, we conclude that correlation-based distance metrics are globally more appropriate than non-correlation metrics for tree clustering in HRP disregarding the market scenario.

4 Concluding Remarks

Risk parity methods have been recently proposed as an alternative to the classical mean-variance portfolio optimization model by Markowitz to solve some of its limitations, such as the sensitivity to the estimation of input parameters, namely, expected returns and covariance matrix. Risk parity methods rely on the idea of building an equally weighted portfolio in terms of risk, not in terms of weight. HRP methods produce a hierarchical structure to the correlation between assets by means of tree clustering and reorganize the covariance matrix of returns. Clustering algorithms are heuristics that group assets together in terms of the distance among them. As a consequence, there is a need to define a distance function to assess how far the two assets are. Even more important is to evaluate the performance of alternative distance metrics to facilitate the selection task of investors and practitioners.

In this paper, we evaluate the out-of-sample performance of alternative hierarchical distance metrics for tree clustering. In order to examine the ability of alternative distance metrics under different circumstances, we use three real stock-market scenarios: bull market, sideways trend, and bear market. We focus on mean-variance estimation error as a measure of the performance of the output to find that correlation-based metrics provide better performance than non-correlation metrics in different market scenarios. Even though the quadratic optimizer is the best option in the bear market scenario, the HRP methods show better performance in the bull market and the sideways trend scenario.

Summarizing, our paper contributes to the selection of the best investing method for different market scenarios. This selection is not limited to the alternative distance metrics to be used in tree clustering for HRP methods, but also to the classical mean-variance portfolio selection and the use of the market index as a benchmark. Even though our empirical evaluation is based on the Spanish stock market, we focus on general trends that were also present in other financial markets in the three periods considered in our experiments. Finally, interesting future lines of work include but are not limited to, new distance metrics, alternative clustering algorithms, and different weight allocation methods.

Author contributions All authors contributed to the study's conception and design. Material preparation and data collection were performed by Francisco Salas-Molina. The first draft of the manuscript was

written by Francisco Salas-Molina. All authors analyzed the results and commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal, 48*(5), 28–43.

Bodnar, T., Mazur, S., & Parolya, N. (2018). Central limit theorems for functionals of large dimensional sample covariance matrix and mean vector in matrix-variate skewed model. *Scandinavian Journal of Statistics, 46*(2), 636–660.

Bravo, M., Jones, D., Pla-Santamaria, D., & Salas-Molina, F. (2022). Encompassing statistically unquantifiable randomness in goal programming: An application to portfolio selection. *Operational Research, 22*, 5685–5706.

Burggraf, T. (2021). Beyond risk parity-a machine learning-based hierarchical risk parity approach on cryptocurrencies. *Finance Research Letters, 38*, 101523.

Cho, Y., & Song, J. W. (2023). Hierarchical risk parity using security selection based on peripheral assets of correlation-based minimum spanning trees. *Finance Research Letters, 53*, 103608.

De Miguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies, 22*(5), 1915–1953.

De Prado, M. L. (2016). Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management, 42*(4), 59–69.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

De Prado, M. L. (2020). *Machine learning for asset managers*. Cambridge: Cambridge University Press.

Garcia-Bernabeu, A., Hilario-Caballero, A., Tardella, F., & Pla-Santamaria, D. (2024). Esg integration in portfolio selection: A robust preference-based multicriteria approach. *Operations Research Perspectives, 12*, 100305.

Garcia-Bernabeu, A., Salcedo, J. V., Hilario, A., Pla-Santamaria, D., & Herrero, J. M. (2019). Computing the mean-variance-sustainability nondominated surface by ev-moga. *Complexity, 2019*(1), 6095712.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223–2273.

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics, 222*(1), 429–450.

Gulliksson, M., Oleynik, A., & Mazur, S. (2023). Portfolio selection with a rank-deficient covariance matrix. *Computational Economics, 63*, 2247–2269.

Hilpisch, Y. (2014). *Python for finance: Analyze big financial data*. d: O'Reilly Media Inc.

Javed, F., Mazur, S., & Thorsén, E. (2024). Tangency portfolio weights under a skew-normal model in small and large dimensions. *Journal of the Operational Research Society, 75*(7), 1395–1406.

Jurczenko, E. (2015). *Risk-based and factor investing*. Elsevier.

Khan, A. T., Cao, X., & Li, S. (2023). Using quadratic interpolated beetle antennae search for higher dimensional portfolio selection under cardinality constraints. *Computational Economics, 62*(4), 1413–1435.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., et al. (2016). Jupyter notebooks-a publishing format for reproducible computational workflows. In *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87-90).

Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research, 234*(2), 356–371.

Kraft, D. (1988). A software package for sequential quadratic programming. *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*, *28*.

Kroll, Y., Levy, H., & Markowitz, H. (1984). Mean-variance versus direct utility maximization. *The Journal of Finance, 39*(1), 47–61.

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10*(5), 603–621.

Ledoit, O., & Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies, 30*(12), 4349–4388.

Ledoit, O., & Wolf, M. (2022). The power of (non-) linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics, 20*(1), 187–218.

Levy, H. (1974). The rationale of the mean-standard deviation analysis: Comment. *The American Economic Review, 64*(3), 434–441.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77–91.

Markowitz, H. (1991). Foundations of portfolio theory. *The Journal of Finance, 46*(2), 469–477.

Molyboga, M. (2020). A modified hierarchical risk parity framework for portfolio management. *The Journal of Financial Data Science, 2*(3), 128–139.

Qi, Y., & Steuer, R. E. (2024). An analytical derivation of properly efficient sets in multi-objective portfolio selection. *Annals of Operations Research*. https://doi.org/10.1007/s10479-024-05848-2

Qian, E. (2005). *Risk parity portfolios: Efficient portfolios through true diversification*. Panagora Asset Management.

Raffinot, T. (2017). Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management, 44*(2), 89–99.

Roncalli, T. (2013). *Introduction to risk parity and budgeting*. Boca Raton: CRC Press.

Salas-Molina, F., Martin, F. J., Rodriguez-Aguilar, J. A., Serra, J., & Arcos, J. L. (2017). Empowering cash managers to achieve cost savings by improving predictive accuracy. *International Journal of Forecasting, 33*(2), 403–415.

Sen, J., & Dutta, A. (2022). A comparative study of hierarchical risk parity portfolio and eigen portfolio on the NIFTY 50 stocks. *Computational Intelligence and Data Analytics: Proceedings of ICCIDA 2022* (pp. 443–460). Springer.

Steuer, R. E., & Utz, S. (2023). Non-contour efficient fronts for identifying most preferred portfolios in sustainability investing. *European Journal of Operational Research, 306*(2), 742–753.