



SCHOOL OF
ECONOMICS AND
MANAGEMENT

Decomposing Hierarchical Portfolio Performance: Tree Structure or Allocation Rule?

Adam Bursic

Department of Economics

Lund University School of Economics and Management

Supervisor: Andreas Johansson

NEKH02 - Bachelor Thesis in Economics 15 ECTS

Seminar date: 2 June 2026

Decomposing Hierarchical Portfolio Performance: Tree Structure or Allocation Rule?*

Adam Bursic[†]

28 May 2026

Abstract

Hierarchical portfolio allocation methods produce diversified portfolios while avoiding the instability of mean-variance optimisation. The framework depends on two separable choices: how assets are grouped and how capital is allocated across those groups. This thesis asks which of these choices matters more for out-of-sample portfolio performance. The empirical design crosses four linkage methods with four allocation kernels, evaluated on a multi-asset ETF universe from 2008 to 2026 across 3,200 resampled walk-forward paths. The results show that allocation kernels account for most differences in risk-adjusted performance, while the linkage method mainly affects turnover. Risk-aware kernels improve risk-adjusted performance and reduce downside risk relative to $1/N$. Practitioners should treat the allocation rule as the primary modelling choice in hierarchical portfolio construction, with linkage choice used mainly to manage implementation stability.

Keywords: Asset allocation, Hierarchical Risk Parity, Diversification, Clustering, Linkage methods, Risk-based allocation, Backtesting

JEL classification: G11, G17, C38, C58

* **Acknowledgement:** I would like to express my sincere gratitude to my supervisor, Andreas Johansson, for his valuable guidance and constructive feedback throughout the writing process. I would also like to thank Dr Ralf Seiz, whose course helped shape the ideas that informed this thesis.

[†]Lund University School of Economics and Management, adambursic@gmail.com

Contents

1	INTRODUCTION	3
2	BACKGROUND	5
2.1	Risk-based allocation	5
2.2	Hierarchical allocation	9
2.2.1	Tree clustering	12
2.2.2	Quasi-diagonalisation	15
2.2.3	Recursive bisection	16
3	LITERATURE	19
4	EMPIRICAL METHOD	22
4.1	Data universe	23
4.2	Factorial design	23
4.3	Backtesting setup	25
4.4	Performance metrics	27
5	RESULTS	28
5.1	Linkage method versus allocation kernel	28
5.2	Full-universe walk-forward illustration	32
6	CONCLUSION	33
	References	36
A	Recursive bisection algorithm	i
B	Data universe	ii
C	Figures	iv
D	Statement on the use of generative AI	vii

1 INTRODUCTION

Portfolio construction concerns the allocation of wealth across assets to achieve an investment objective, typically by balancing expected return against risk. The classical framework is mean-variance optimisation (MVO), introduced by Markowitz (1952, 1959), which systematically optimises the trade-off between expected returns (μ) and risk, measured by variance (σ^2). However, a major empirical fact is that MVO is very sensitive to estimation errors in the inputs. Errors in expected returns have been shown to have the largest impact (Chopra and Ziemba, 1993). This sensitivity leads to the tendency of *error maximisation*, where MVO produces portfolios that are mathematically “optimal” but economically extreme, characterised by concentrated positions and unstable weights (Jobson and Korkie, 1980; Michaud, 1989). To combat this, substantial literature has been devoted to countermeasures such as resampling, shrinkage, and constraints (e.g. Black and Litterman, 1992; Idzorek, 2002; Jagannathan and Ma, 2003; Jorion, 1985; Ledoit and Wolf, 2003, 2004; Michaud, 1998; Michaud and Michaud, 2008; Stein, 1955).

One strand of literature responded to this problem by disregarding expected returns and constructing portfolios from risk estimates alone. *Risk-based allocation*, encompassing strategies such as global minimum-variance and risk parity, emerged as the academic and practitioner response to the error maximisation problem (Lee, 2011; Reigner et al., 2020). However, within this family, a further trade-off arises. Methods that rely only on individual asset volatilities are stable and simple, but ignore how assets co-move, which is central to diversification. Methods that use the full covariance matrix, including the relationships between assets, capture that information but bring back sensitivity to estimation error.

Hierarchical allocation occupies a middle position. By organising assets into a tree structure based on their similarity and then distributing capital through that structure, it considers co-movements while avoiding the estimation sensitivity of full-matrix methods. The motivation also has economic roots: Simon (1962) argued that complex systems naturally exhibit hierarchical organisation, and Mantegna (1999) showed empirically that financial markets do too. Thus, hierarchical methods regularise the allocation without imposing explicit weight constraints and produce portfolios with transparent economic interpretation (Cotton, 2024). López de Prado (2016) formalised this idea in Hierarchical Risk Parity (HRP), which has become one of the most widely adopted regularised portfolio construction methods (Wuebben, 2026).

Hierarchical portfolio construction is defined by several independent modelling choices. Two central choices, and the focus of this thesis, are the *linkage method*, which determines how the tree is constructed from the data, and the *allocation kernel*, which determines how capital is distributed once the tree is fixed. Existing literature has examined these dimensions largely in isolation, typically validating hierarchical methods against traditional allocation methods rather than varying the design choices within the hierarchical framework. This thesis therefore asks:

In hierarchical portfolio construction, is out-of-sample performance driven primarily by the tree structure or by the allocation rule?

The study evaluates a 4×4 factorial design that crosses four linkage methods against four allocation kernels on a multi-asset universe of 38 ETFs spanning global equities, fixed income, commodities and real assets, observed from January 2008 to April 2026. Each combination is evaluated out-of-sample on a set of performance, risk, and implementation metrics. The evaluation uses the Multiple Randomised Backtest of [Palomar \(2025\)](#), which reduces reliance on a single historical path by repeatedly resampling asset subsets from the universe. A single walk-forward backtest on the full universe is included as an illustration of cumulative wealth paths and portfolio behaviour over time.

The results show that allocation kernels account for most observed differences in risk-adjusted performance, while the linkage method matters mainly for turnover. The contribution is one of attribution rather than method development. By holding all other modelling choices fixed across the factorial grid, performance differences can be interpreted as associated with the linkage method or the allocation kernel, rather than other implementation choices that differ across studies. This addresses the question of which dimension of the hierarchical framework warrants greater modelling effort, and responds to the broader observation in the literature that examining linkage choice and alternative risk measures could yield valuable insights for practitioners ([Trucíos, 2026](#)).

The rest of the thesis is structured as follows. [Section 2](#) discusses the theoretical background. [Section 3](#) reviews the empirical literature. [Section 4](#) outlines the empirical study. [Section 5](#) presents the results. [Section 6](#) concludes.

2 BACKGROUND

2.1 Risk-based allocation

A natural response to the instability of mean-variance optimisation is to reduce, or even remove, reliance on expected-return forecasts. The intuition is straightforward: if future returns are difficult to estimate precisely, portfolios that depend on those forecasts can become highly unstable. This concern is central to the “Markowitz optimisation enigma”: mean-variance optimisation can amplify errors in its inputs, especially errors in expected returns (Michaud, 1989; Best and Grauer, 1991). Empirical and theoretical work also shows that errors in estimated returns are typically more damaging for portfolio choice than errors in variances or covariances, which helps explain why risk-only methods remain attractive in practice (Chopra and Ziemba, 1993).

Risk-based allocation methods therefore construct portfolios primarily from risk information rather than from forecasts about which assets will earn the highest returns. These methods typically use the covariance matrix Σ , but downside-risk variants replace variance with tail- or drawdown-based risk measures. Lee (2011, p. 11) characterises risk-based approaches as portfolios for which “only risk forecasts are used as inputs, and no forecasts of returns of any assets are required”. Common examples include inverse-volatility, minimum-variance, equal-risk-contribution, maximum-diversification and related risk-parity approaches (Ardia et al., 2018; Reigner et al., 2020).

Two observations motivate this design. First, expected returns are too imprecisely estimated for portfolio weights to depend heavily on them. Second, capital allocation and risk allocation are not the same. A traditional 60/40 equity-bond portfolio may look diversified in capital terms, but its risk is often dominated by equities. Ardia et al. (2018) note that in a typical 60/40 US equity-bond portfolio, the equity sleeve can account for more than 90 per cent of total volatility. More broadly, Allen (2010) reports that, in typical institutional portfolios, equities explain roughly 70 to 85 per cent of total portfolio volatility. Risk-based allocation responds to this mismatch by allocating according to risk contributions rather than nominal weights. If expected returns are noisy but risks can be estimated with comparatively greater stability, then framing portfolio construction around risk is a natural and defensible alternative.

A first family of risk-based methods assigns portfolio weights through an explicit

formula that uses, at most, the individual asset volatilities (Reigner et al., 2020). The simplest member is the equally weighted, or $1/N$, portfolio where each asset i receives the same weight $w_i = 1/N$, with N denoting the number of assets. The $1/N$ allocation is mean-variance efficient under the assumption that all assets share the same expected return and volatility, and that pairwise correlations are equal (Windcliff and Boyle, 2004). DeMiguel et al. (2009) document that none of the sample-based optimisers they examine reliably outperforms $1/N$ out-of-sample at standard sample sizes. The estimation error introduced by more complex methods tends to erode the theoretical efficiency advantage they offer on paper. The $1/N$ portfolio therefore serves as an empirical floor that any more sophisticated risk-based method should be expected to outperform out-of-sample (Palomar, 2025).

An extension is the inverse-volatility portfolio (IVP), also referred to as naive risk parity, where each asset's weight is set as

$$w_i^{\text{IVP}} = \frac{1/\sigma_i}{\sum_{j=1}^N 1/\sigma_j}, \quad (1)$$

where σ_i denotes the volatility of asset i . This rule assigns lower weights to more volatile assets and higher weights to less volatile assets. As a result, each asset's volatility-weighted position, $w_i\sigma_i$, is equalised. IVP uses only the individual volatility of each asset, derived from the diagonal elements (the variances) of the covariance matrix. It ignores the off-diagonal elements, which contain the pairwise covariances between assets. IVP is therefore less naive than equal weighting because it adjusts for asset volatility, but it ignores correlations between assets. Both the $1/N$ portfolio and IVP thus avoid a central source of estimation instability in mean-variance optimisation. In MVO, the optimal weights are obtained using the inverse of the estimated covariance matrix, because the method needs to account for how all assets co-move when determining the risk-minimising allocation. When the covariance matrix is estimated with error, this inversion can amplify small estimation errors and produce unstable portfolio weights (Jagannathan and Ma, 2003). By not using the full covariance matrix, $1/N$ and IVP avoid this instability. The cost is that they ignore cross-asset risk information.

A second family of risk-based methods defines an objective function whose constrained minimisation yields the portfolio weights. These methods condition on the full covariance

matrix, including the off-diagonal correlation information, at the cost of requiring a numerical solution and reintroducing exposure to estimation error (Reigner et al., 2020). The global minimum-variance portfolio (GMV) is defined as the portfolio that minimises total portfolio variance:

$$\min_w w^\top \Sigma w \quad \text{subject to} \quad \mathbf{1}^\top w = 1, \quad (2)$$

where w is the vector of portfolio weights, Σ is the covariance matrix of asset returns and $\mathbf{1}^\top w = 1$ imposes the requirement that all portfolio weights sum to one. Although it requires no expected return forecast, it remains exposed to the same covariance-related instability as mean-variance optimisation.

The equal risk contribution (ERC) portfolio, also referred to as risk parity, extends the inverse-volatility idea by accounting not only for each asset's own volatility, but also for how assets co-move with the rest of the portfolio. The aim is that no single asset should dominate total portfolio risk. Instead, each asset is assigned a weight such that its contribution to total portfolio risk is the same as that of every other asset. The marginal risk contribution (MRC) of asset j measures how portfolio volatility changes when the weight of asset j changes. Multiplying this by the asset's weight gives its total risk contribution (TRC).

$$\text{Marginal risk contribution} \quad MRC_j = \frac{\partial \sigma_p}{\partial w_j} = \frac{(\Sigma w)_j}{\sigma_p}, \quad (3)$$

$$\text{Total risk contribution} \quad TRC_j = w_j MRC_j, \quad (4)$$

$$\text{Portfolio risk} \quad \sigma_p = \sum_j TRC_j. \quad (5)$$

ERC equalises risk proportions across all assets i and j :

$$w_i(\Sigma w)_i = w_j(\Sigma w)_j \quad \text{for all } i, j. \quad (6)$$

Unlike $1/N$ and IVP, ERC generally has no closed-form solution and is typically obtained through numerical optimisation, for example by minimising the differences in risk contributions across assets (Maillard et al., 2010). Risk parity was implemented in institutional practice by Bridgewater Associates from the mid-1990s and formalised under that name by

Qian (2005). It gained broader traction after the 2008 financial crisis exposed the equity dominance of traditional allocations. It has also drawn criticism, however, particularly for its reliance on leverage to reach equity-like returns and its tendency to underperform during sustained equity bull markets (Chaves et al., 2011; Anderson et al., 2012). If pairwise correlations are assumed to be zero or equal across all assets, the ERC weights collapse to the IVP solution defined above, which is why IVP is also known as naive risk parity (Roncalli, 2013).

The objective-function-based allocation methods introduced so far use variance and volatility as risk measures. These are symmetric measures: deviations above and below the mean enter with equal weight. Most investors, however, are more concerned with losses than with equivalent gains (Kahneman and Tversky, 1979). Risk measures that focus on the left tail of the return distribution are therefore conceptually closer to what most investors are trying to control. The most familiar example is the Value-at-Risk (VaR) at a given confidence level $\alpha \in (0, 1)$, defined as the α -quantile of the loss distribution:

$$\text{VaR}_\alpha(L) = \inf \{l \in \mathbb{R} : \Pr(L \leq l) \geq \alpha\}, \quad (7)$$

where $L = -r_p$ denotes the portfolio loss. VaR identifies a loss threshold but does not say anything about the magnitude of losses beyond it. The Conditional Value-at-Risk (CVaR) (Rockafellar and Uryasev, 2000), also known as expected shortfall, addresses this limitation by averaging losses conditional on VaR being exceeded:

$$\text{CVaR}_\alpha(L) = \mathbb{E}[L \mid L \geq \text{VaR}_\alpha(L)]. \quad (8)$$

The ERC framework introduced above generalises naturally to non-variance risk measures (Roncalli, 2013). Applied to CVaR, this gives an ERC-CVaR portfolio in which weights are chosen so that assets contribute equally to portfolio CVaR. The same can be applied to Conditional Drawdown-at-Risk (CDaR), a drawdown-based analogue of CVaR. Let V_t denote the portfolio value at time t , and let $M_t = \max_{s \leq t} V_s$ denote the running peak. The percentage drawdown at time t is

$$D_t = \frac{M_t - V_t}{M_t}. \quad (9)$$

CDaR at confidence level α is then the CVaR of the drawdown series (Chekhlov et al., 2005).

Among risk-based methods that exploit the full covariance matrix, minimum variance and ERC both depend on the inversion of the covariance matrix or on its use during optimisation. Thus, these methods become unstable when the covariance matrix is difficult to estimate reliably. This is especially dangerous in stressed markets, when correlations rise and many assets start moving together, reducing the number of truly independent sources of risk (Longin and Solnik, 2001; Ang and Chen, 2002). However, this is precisely the environment in which robust diversification is most needed. This echoes what López de Prado (2016) terms the “Markowitz curse” in the context of mean-variance optimisation: the optimiser is most likely to fail in exactly the cases where it should help most. The explicit weighting schemes, $1/N$ and IVP, avoid the problem by construction, at the cost of discarding the off-diagonal correlation information that is, in principle, the most useful input to a diversification problem.

2.2 Hierarchical allocation

Hierarchical allocation can be understood as an attempt to find a middle ground and sidestep the trade-off between estimation risk and information loss. Rather than either ignoring correlations or relying on the full covariance matrix in one large optimisation problem, hierarchical methods first organise assets into a tree structure based on their similarity and then allocate capital through this structure.

Financial assets are rarely viewed as unrelated individual objects. From a practitioner’s perspective, portfolio construction often begins with broad allocation decisions across major asset classes, such as equities and bonds, followed by diversification within each category. Graham (1949), for example, emphasised the importance of periodic rebalancing between stocks and bonds, while also warning against excessive concentration within each group. In this sense, portfolio construction can be understood as a sequence of allocation decisions across and within substructures, rather than as a single allocation decision across all available assets (Cotton, 2024). A related idea appears in the literature on factor premia, where asset returns are often interpreted as being driven by common underlying sources of risk. Although a comprehensive review of the factor literature is beyond the scope of this thesis (see Pukthuanthong et al., 2019, for an overview), the central point is

that assets are not independent in an economic sense. They tend to share exposures to broader risk drivers, meaning that assets within the same group often display stronger relationships with each other than with assets outside the group.

Treating the asset universe as “flat”, where every pair of assets enters the optimisation problem on equal footing, ignores the possibility that assets are naturally organised into groups (López de Prado, 2016). Simon (1962) argues that complex systems are typically hierarchical, with subsystems whose internal interactions are stronger than their interactions with the rest of the system. Financial markets are no exception (Mantegna, 1999). Applied to portfolio construction, this means that diversification should not only be considered across individual assets, but also across groups of related assets. In this view, portfolio construction becomes a problem of allocating both across and within clusters, rather than allocating across all assets simultaneously without reference to their underlying structure.

López de Prado (2016) interprets the correlation matrix through this hierarchical perspective. A standard correlation matrix describes the relationship between every asset pair in the investment universe and can be visualised as a complete graph with N nodes and $N(N - 1)/2$ edges, as illustrated in Figure 1 (left panel). While this representation contains a large amount of information, it does not by itself distinguish between stable economic relationships and weaker relationships that may reflect estimation noise. Nor does it directly reveal whether some assets form closer groups than others. As a result, traditional optimisers may react too strongly to noisy pairwise correlations, producing portfolio weights that are difficult to interpret and sensitive to small changes in the estimated covariance matrix. This is one reason why traditional optimised portfolios can struggle to outperform a naive $1/N$ allocation out of sample (DeMiguel et al., 2009).

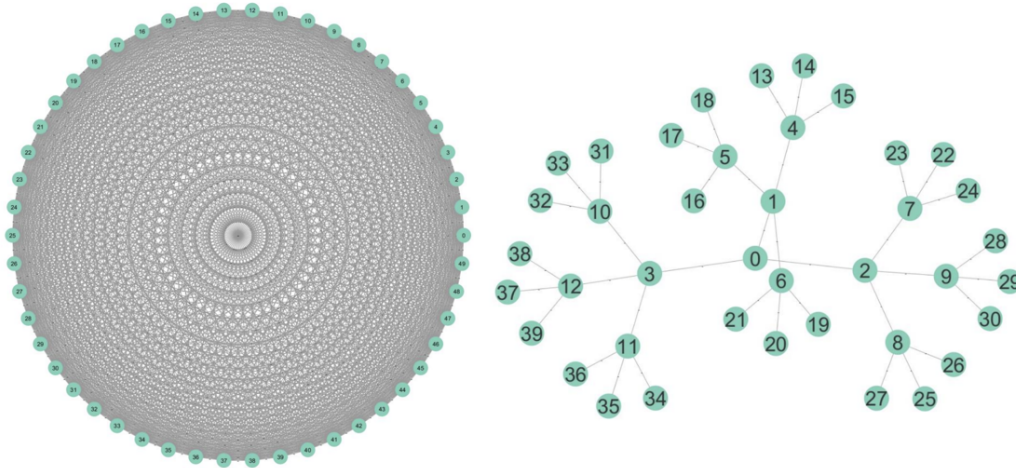


Figure 1: Complete-graph and tree-graph representations of asset dependence. The left panel shows the full set of pairwise relationships, where every asset is connected to every other asset. The right panel shows a simplified tree representation, where only the strongest links needed to connect all assets are kept.

Source: Adapted from López de Prado (2016).

Mantegna (1999) proposes a way to make this dense correlation structure easier to interpret by reducing it to a simpler network. Instead of keeping all possible pairwise relationships between assets, the minimum spanning tree (MST) keeps only the connections needed to link all assets together, while removing redundant links, as illustrated in Figure 1 (right panel). With N assets, this means retaining only $N - 1$ links out of the $N(N - 1)/2$ possible pairwise relationships. The result is a filtered representation of the correlation matrix that highlights the main structure of dependence between assets. Assets that are closely related tend to appear near each other in the tree, while more weakly related assets are connected through longer paths. In this way, the MST helps reveal the hierarchical organisation of the asset universe.

This network representation is closely related to hierarchical clustering. Mantegna (1999) shows that the MST corresponds to single-linkage agglomerative (bottom-up) clustering when both are applied to the same correlation-based distance matrix. Put simply, the two methods describe the same dependence structure in different forms: the MST represents it as a network, while hierarchical clustering represents it as a dendrogram, as shown in Figure 2. This connection is important because it allows clustering methods, originally developed in statistics, to be interpreted as tools for filtering noisy correlation matrices and organising assets before portfolio construction.

The idea of allocating capital by walking through a clustering tree of asset returns was

introduced by [Papenbrock \(2011\)](#). As illustrated in Figure 2, moving top-down, capital is split equally at each branching of a correlation-based dendrogram, yielding what later writers refer to as the hierarchical 1/ N portfolio. Building on this, [López de Prado \(2016\)](#) introduces Hierarchical Risk Parity (HRP), formalising the construction into a three-stage pipeline: tree clustering, quasi-diagonalisation and recursive bisection. In HRP, capital is again allocated top-down through the dendrogram, but at each branching the split is made in inverse proportion to the two halves’ variances, so the lower-variance half receives more weight. [Raffinot \(2017, 2018\)](#) extends the framework by replacing the inverse-variance allocation with ERC allocation and introducing downside risk measures, CVaR and CDaR. The remainder of this section reviews the three stages of the hierarchical allocation process.

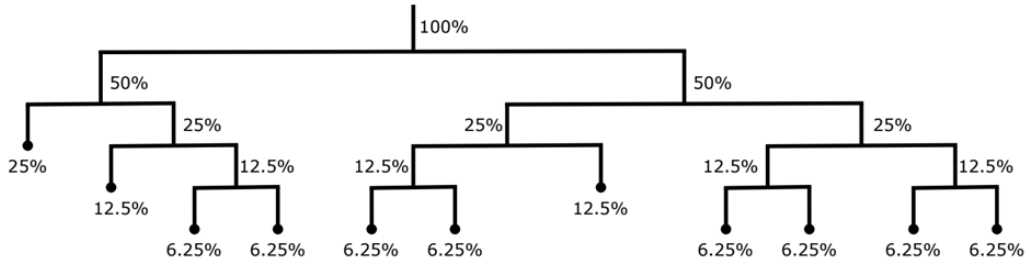


Figure 2: Hierarchical 1/ N allocation on a correlation-based dendrogram. Starting from the full asset universe, capital is split equally at each branching. The percentages show the cumulative portfolio weights implied by the top-down allocation rule.

Source: Adapted from [Palomar \(2025\)](#).

2.2.1 Tree clustering

To construct a hierarchical tree, the correlation matrix must first be translated into a measure of distance between assets. This step is important because clustering algorithms do not group assets directly from correlations, but from a notion of how “close” or “far apart” assets are. In portfolio terms, assets that move very similarly should appear close together in the tree, while assets with weaker or opposite co-movement should be placed further apart. Following [Mantegna \(1999\)](#) and [López de Prado \(2016\)](#), the standard transformation is

$$d_{ij} = \sqrt{\frac{1}{2}(1 - \rho_{ij})}, \quad (10)$$

where ρ_{ij} is the Pearson correlation between assets i and j , and d_{ij} is the corresponding distance. This transformation maps correlations into distances between zero and one. Two perfectly positively correlated assets have distance zero, meaning they will always

be clustered together. Two uncorrelated assets have distance $1/\sqrt{2}$, while two perfectly negatively correlated assets have distance one. The transformation therefore preserves the intuitive ordering of relationships: higher correlations imply shorter distances and lower correlations imply larger distances. It also satisfies the formal properties required of a proper metric, including non-negativity, symmetry and the triangle inequality (Gower, 1966; Mantegna, 1999).

The choice of distance metric matters because it determines the structure of the clustering tree. A Pearson correlation distance captures average linear co-movement, which makes it transparent and easy to implement, but it may not fully capture nonlinear dependence or dependence during market stress. For this reason, alternative co-dependence measures have been proposed, including distance correlation, mutual information and lower-tail dependence coefficients (Barziy and Chlebus, 2020; Lohre et al., 2020). These alternatives may be useful when the objective is to capture broader or more tail-sensitive forms of dependence. However, the Pearson-based distance remains the standard in the hierarchical allocation literature. Barziy and Chlebus (2020) find that Pearson correlation produces higher Sharpe ratios than alternative co-dependence measures on a comparable multi-asset ETF universe, while Salas-Molina et al. (2025a) confirm that correlation-based metrics outperform non-correlation alternatives across bull, sideways and bear market regimes.

Once the asset distances have been computed, the next step is to organise the assets into a hierarchical tree. This is done using an agglomerative hierarchical clustering algorithm. The agglomerative approach begins with each asset in its own cluster. At each step, the two closest clusters are merged, and this process continues until all assets are contained in a single hierarchy. The resulting dendrogram shows the order in which assets and clusters are combined. Clusters that merge at low levels of the tree are more similar, while clusters that merge only at higher levels are more dissimilar (Murtagh, 2004). The main distinction between hierarchical clustering algorithms is the linkage rule, which defines how the distance between two clusters is measured from the pairwise distances of their members, and thereby determines the order in which clusters are merged.

Papenbrock (2011) discusses four common linkage rules and the cluster structures they tend to produce. Under single linkage, the distance between two clusters C_i and C_j is

defined as the smallest pairwise distance between any two members of the clusters:

$$d(C_i, C_j) = \min\{d(x, y) : x \in C_i, y \in C_j\}. \quad (11)$$

This is the linkage rule adopted by [López de Prado \(2016\)](#). Complete linkage instead defines the cluster distance as the largest pairwise distance between their members:

$$d(C_i, C_j) = \max\{d(x, y) : x \in C_i, y \in C_j\}. \quad (12)$$

Average linkage uses the average pairwise distance across all members of the two clusters:

$$d(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{y \in C_j} d(x, y). \quad (13)$$

Finally, Ward’s method differs from the previous rules because it does not define cluster distance directly from the minimum, maximum or average pairwise distance between members. Instead, it merges the two clusters that generate the smallest increase in within-cluster variance ([Ward, 1963](#)). Under Euclidean distances, this can be expressed as:

$$d(C_i, C_j) = \frac{|C_i||C_j|}{|C_i| + |C_j|} \|\mu_i - \mu_j\|^2, \quad (14)$$

where μ_i and μ_j denote the centroids of clusters C_i and C_j , respectively.

The four rules produce dendrograms with different shapes and, consequently, different cluster sizes. Single linkage is vulnerable to a chaining effect. Because only one close connection is needed for two clusters to merge, the method is prone to forming one long dominant branch by adding assets one at a time, as shown in [Figure 3](#). This may leave weakly related assets as small outlier clusters. When capital is later allocated through the tree, these small outlier clusters can receive disproportionately large weights ([Jain and Dubes, 1988](#); [Pfitzinger and Katzke, 2019](#)). Complete linkage avoids chaining by requiring all members of two clusters to be close before merging, producing more compact and similarly sized clusters at the cost of sensitivity to outliers. Average linkage offers an intermediate compromise. Ward’s method produces compact, similarly sized clusters that often correspond cleanly to a priori groupings such as sector or asset class ([Papenbrock, 2011](#)). The linkage rule is not a property of the data but a modelling choice that conditions

the tree on which all subsequent allocation steps depend.

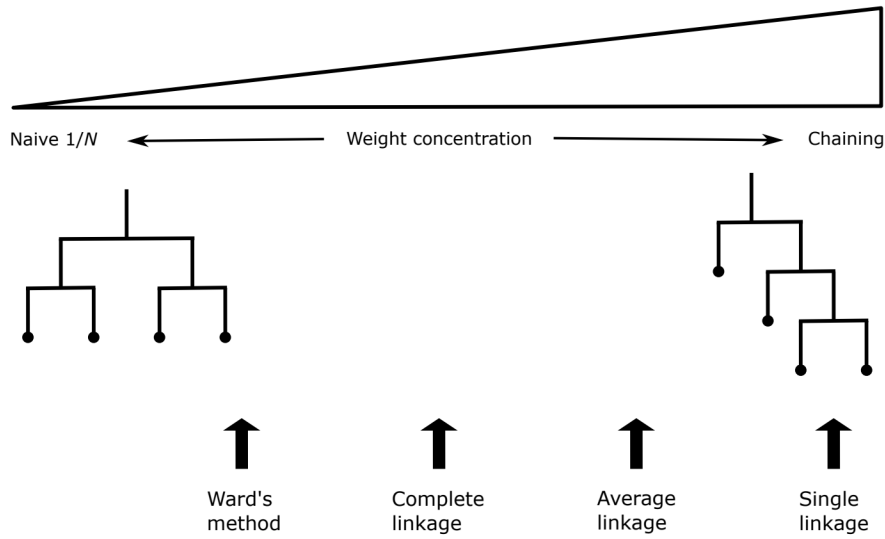


Figure 3: Linkage rules, chaining and hierarchical allocation. The figure illustrates how different hierarchical clustering linkage rules can produce different dendrogram shapes and therefore different allocation outcomes.

Source: Adapted from [Palomar \(2025\)](#).

2.2.2 Quasi-diagonalisation

The second stage, quasi-diagonalisation, reorganises the correlation matrix so that assets belonging to the same cluster are placed next to each other. Importantly, the correlation values themselves are not changed. Only the order of the rows and columns is rearranged according to the leaf order of the dendrogram. This makes the structure identified in the clustering step visible in the matrix: highly related assets are moved closer to the main diagonal, while weaker relationships are placed further away, as shown in Figure 4. The result is a sorted list of assets, $L_0 = \{n_1, n_2, \dots, n_N\}$, that reflects the hierarchical ordering produced by the dendrogram. Quasi-diagonalisation does not determine portfolio weights directly. Instead, it prepares the matrix for the recursive allocation step by ensuring that it follows the cluster structure rather than splitting the asset universe arbitrarily.

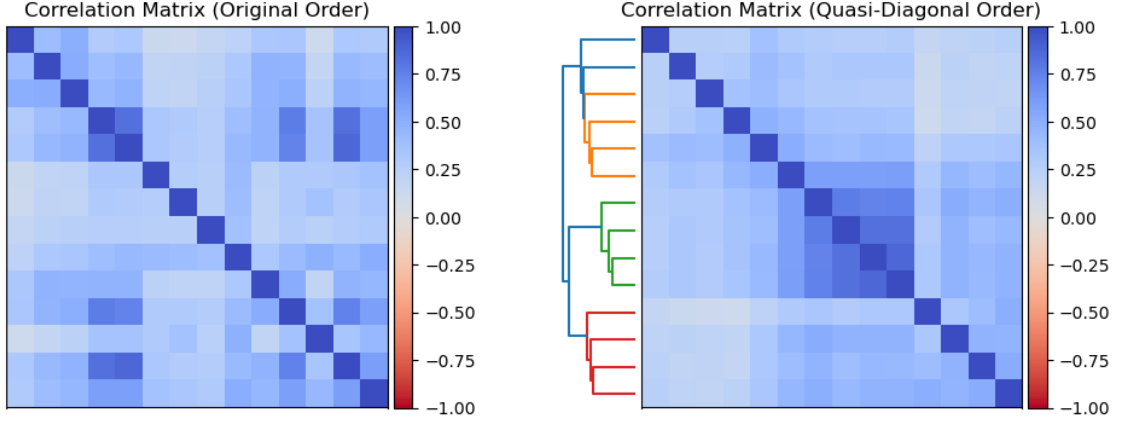


Figure 4: Correlation matrix before and after quasi-diagonalisation. The left panel shows the original correlation matrix with assets ordered arbitrarily. The right panel shows the same correlation matrix after quasi-diagonalisation and the corresponding dendrogram.

Source: Author's calculations.

2.2.3 Recursive bisection

Once the tree has been built, the remaining question is how much capital should be sent down each branch. The third stage assigns portfolio weights using the sorted asset list L_0 . The defining idea of HRP (López de Prado, 2016) is to allocate capital top-down, rather than solve one global optimisation problem across all assets at once. The algorithm starts with the full asset universe and repeatedly splits it into two sub-clusters according to the hierarchical ordering. At each split, the available capital is divided between the two sub-clusters, with more capital allocated to the sub-cluster with lower estimated risk and less capital allocated to the sub-cluster with higher estimated risk. The same procedure is then applied recursively within each sub-cluster until each final branch contains a single asset. The capital assigned to each terminal asset becomes its portfolio weight.

This construction is important because each allocation decision is local. Instead of asking the optimiser to determine all asset weights simultaneously, HRP breaks the problem into a sequence of simpler two-way allocation decisions between clusters. The full covariance matrix is therefore not inverted. Correlation information is still used, first through the clustering tree and then through the estimation of each sub-cluster's risk, but it is used in a more constrained and structured way than in traditional minimum-variance optimisation. This is the main architectural feature that gives HRP its appeal: it seeks to retain useful dependence information while reducing the sensitivity to noisy covariance estimates, particularly when the covariance matrix is ill-conditioned. The formal recursive

bisection algorithm is presented in Appendix A.

The first design choice in the recursive bisection step concerns how each cluster is split before capital is allocated. When a cluster contains more than two assets, the algorithm must decide which assets should form the two sub-groups that compete for capital. In [López de Prado \(2016\)](#)’s original HRP algorithm, this is done by splitting the sorted asset list into two contiguous halves. The ordering produced by quasi-diagonalisation is therefore preserved, but the split is based on position in the list rather than directly on the branching structure of the dendrogram. In practice, this means that the algorithm may divide a cluster in a way that is natural for a sorted list, but not necessarily identical to the split implied by the clustering tree, as illustrated in Figure 5.

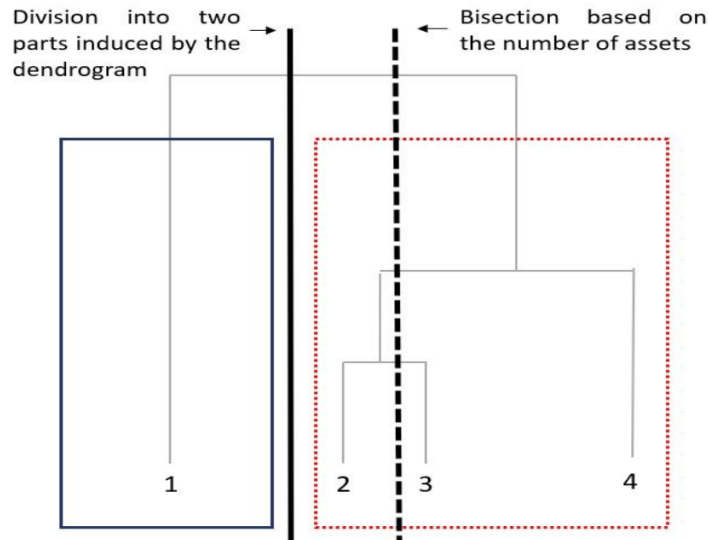


Figure 5: Difference between dendrogram-based bisection and equal-size bisection in hierarchical allocation. The solid line shows a split that follows the structure of the dendrogram, where the two sub-clusters are separated according to the clustering tree. The dashed line shows a split based only on the number of assets, where the ordered asset list is divided into two equally sized groups.

Source: Adapted from [Raffinot \(2018\)](#).

[Raffinot \(2018\)](#)’s Hierarchical Equal Risk Contribution (HERC) uses a different splitting rule. Rather than cutting the sorted list at its midpoint, it follows the internal branches of the dendrogram. Each split therefore corresponds to the two sub-clusters identified by the clustering algorithm, as illustrated in Figure 5. The difference between the two approaches is small when the dendrogram is balanced, because midpoint splits and tree-based splits often coincide. However, when the tree is unbalanced, the distinction becomes more important. [López de Prado \(2016\)](#)’s rule prioritises a simple and stable list-based

allocation, while [Raffinot \(2018\)](#)’s rule stays closer to the hierarchical structure produced by the clustering stage.

The second design choice concerns the allocation kernel, meaning the rule used to divide capital between the two sub-groups at each split, and has two components. The within-cluster rule assigns a risk number to each sub-group, typically by treating the sub-group as a small portfolio and computing its variance, CVaR or other risk measure. The across-cluster rule then determines how much capital should be allocated to each sub-group. If fraction $\alpha \in [0, 1]$ denotes the share of capital allocated to the first sub-group, then $1 - \alpha$ is allocated to the second. Several kernels appear in the literature.

The simplest is hierarchical $1/N$ ([Papenbrock, 2011](#); [Raffinot, 2017](#)), which allocates capital equally between the two sub-groups at every node:

$$\alpha = \frac{1}{2}, \quad (15)$$

regardless of their estimated risk. The resulting portfolio therefore depends on the structure of the dendrogram, because the tree determines the sequence of splits, but not on the covariance matrix at the allocation stage. It is therefore the most naive member of the family.

The inverse-variance kernel, applied in [López de Prado \(2016\)](#)’s original HRP, treats each sub-group as a portfolio in which assets are weighted inversely to their individual variances. It then computes the variance of that portfolio as the sub-group’s cluster variance $\tilde{V}$. Capital is split in inverse proportion to the two cluster variances:

$$\alpha = \frac{\tilde{V}_2}{\tilde{V}_1 + \tilde{V}_2}, \quad (16)$$

so the half with lower variance is allocated more capital. Thus, an inverse-variance allocation is applied both at the within-cluster and across-cluster levels.

[Raffinot \(2018\)](#)’s HERC kernel generalises the inverse-variance rule by allowing the risk of each sub-cluster to be measured by arbitrary risk measures. Within each sub-cluster, assets are weighted by naive risk parity, that is inverse-volatility weighting. The sub-cluster’s risk R_i is then computed as the risk-measure value of this sub-portfolio. Capital

is split between the two sub-clusters in inverse proportion to their risks:

$$\alpha = \frac{R_2}{R_1 + R_2}, \quad (17)$$

so the lower-risk sub-cluster receives more capital. When R is variance, the rule coincides with the inverse-variance kernel above. [Raffinot \(2018\)](#) also examines two downside measures: CVaR and CDaR.

3 LITERATURE

The empirical literature on HRP-style allocation begins with [López de Prado \(2016\)](#), while earlier work on financial market trees and asset clusters provides the structural foundation. The work of [Raffinot \(2017, 2018\)](#) extended the approach with alternative cluster-allocation kernels and tail-aware risk measures. Section 2 covers the theoretical content of those papers. This section traces how subsequent studies have tested, modified and stress-tested the method on real markets.

[Munich Re \(2020\)](#) apply HRP to a multi-asset universe of global equity indices, government bonds and commodities in the construction of the FIVE Robust Multi-Asset Index. The index is based on 20 futures markets, rebalanced monthly and scaled to a realised volatility target of 5 per cent per annum. They compare HRP with common allocation approaches, including equal weighting, naive risk parity, minimum variance and ERC. In their historical analysis, HRP delivers the highest Sharpe ratio, the lowest maximum drawdown and the strongest Sortino and Calmar ratios among the reported strategies. Their interpretation is framed less as an alpha claim than as a robustness claim: HRP is presented as a way to use correlation information without inverting the covariance matrix, producing more stable weights, broader diversification and improved drawdown control.

[Eidenvall \(2021\)](#) runs a comparable exercise on a multi-asset, multi-factor universe combining developed and emerging equities, commodities, rates, credit and factor portfolios. Using bootstrap-simulated returns and a walk-forward backtest, he reports that hierarchically clustered portfolios achieve statistically better risk-adjusted performance than commonly used optimisation techniques, with particular attention to left-tail risk and drawdown control. The thesis combines a regime-shift overlay with the hierarchical

allocation step, but its core empirical claim, that hierarchical methods deliver tail-aware diversification on a broad multi-asset universe, aligns with the findings of [Munich Re \(2020\)](#).

[Deković and Šimović \(2025\)](#) approach the method from a computational angle. They reduce the time complexity of the HRP algorithm and test it on S&P 500 constituents over 2005-2023. Across all their experimental setups, the naive $1/N$ strategy outperforms HRP on risk-adjusted return, although the HRP produces portfolios with approximately one percentage point lower annualised volatility in every configuration. The authors conclude that HRP is preferable “when risk is the primary concern” rather than as a universal allocator. This conclusion is consistent with the broader pattern that hierarchical methods tend to compete on volatility, turnover and concentration rather than on raw return.

A second strand of the literature decomposes HRP into its component design choices. This is the strand to which the present thesis contributes. [Salas-Molina et al. \(2025a\)](#) compare six distance functions, four correlation-based and two non-correlation, across bull, sideways and bear market scenarios on real stock data. Two findings of relevance stand out: correlation-based metrics outperform non-correlation metrics in all three market regimes, and HRP outperforms the equal-weight portfolio in every scenario while losing to quadratic optimisers only in the bear market. Their preferred distance is the standard transformation of Pearson correlation used by [López de Prado \(2016\)](#).

[Barziy and Chlebus \(2020\)](#) substitute Pearson correlation with three alternative co-dependence measures, namely distance correlation, mutual information and variation of information, across thirty-two ETFs spanning equities, bonds and commodities from 2007 to 2019. Transaction costs from a standard retail brokerage schedule are included throughout. They find that alternative co-dependence measures worsen both Sharpe ratio and maximum drawdown relative to Pearson correlation across the allocation algorithms: IVP, the Critical Line Algorithm and HRP. HRP and IVP also show lower turnover than mean-variance and CLA, indicating that hierarchical methods compete on implementation stability as well as on return. Together with [Salas-Molina et al. \(2025a\)](#), these results motivate using Pearson correlation as the basis for the dissimilarity measure in this study.

[Salas-Molina et al. \(2025b\)](#) study the estimation window. Using daily data from the S&P 500 and the Euro Stoxx 50, they test past-data windows ranging from one to seven years and report that five years of daily data, roughly 1,300 observations, gives the

best out-of-sample Sharpe ratio for HRP across both indices. A parallel synthetic-data experiment confirms the result. HRP outperforms both IVP and equal weighting in their tests, while minimum variance wins on the Euro Stoxx 50, but the authors flag this result is partly driven by concentrated weights.

[Arévalo et al. \(2019\)](#) compare linkage methods within a hierarchical portfolio setting. Using two thousand liquid US stocks and one-minute returns, they compare seven linkage methods and select average linkage because it produces the highest cophenetic correlation coefficient, meaning that its dendrogram best preserves the original distance matrix. They then test a portfolio based on this selected tree, combining Markowitz optimisation within clusters with inverse-variance weighting across clusters. The study shows that linkage choice affects the structure of the clustering tree, but it does not test whether different linkage methods lead to different out-of-sample portfolio performance. Nor can the effect of the tree be separated cleanly from the subsequent allocation rule. This thesis departs from their approach by treating tree construction and capital allocation as separate design choices and evaluating all linkage-kernel combinations directly on out-of-sample performance.

[Pfitzinger and Katzke \(2019\)](#) present a methodological extension of HRP. Their constrained HRP algorithm introduces three main modifications. First, it allows for weight constraints at both the asset and group level while preserving the relative structure of the unconstrained allocation as far as possible. Second, it compares [López de Prado \(2016\)](#)'s single-linkage clustering with Ward clustering, top-down divisive clustering and a genetic optimisation-based alternative. Third, it replaces [López de Prado \(2016\)](#)'s simple halving rule with a dissimilarity-based recursive allocation tree, where splits are informed by asset similarity at the edges of the cluster structure. Using daily returns from six equity universes and simulated portfolios of varying size, the authors find that HRP produces diversified portfolios with low volatility, low turnover and competitive risk-adjusted performance. Their results also show that clustering choices matter most when the allocation rule makes direct use of the tree structure: under dissimilarity-based allocation.

[Katrancı and Kundakcı \(2026\)](#) embed HRP inside a multi-criteria decision-making framework that combines an objective-weighting method, MEREC, with a performance-ranking method, WEDBA, on BIST 100 stocks for 2018–2022. They test all seven standard linkage methods within HRP and compare standalone HRP against the hybrid procedure.

The hybrid variant outperforms standard HRP on Sharpe ratio under Ward, single, complete and median linkage. Standard HRP wins under average, centroid and weighted linkage. The result that linkage choice changes the relative ranking of the two allocation procedures is one of the closest precedents for studying how linkage and allocation rule interact, though the comparison is not framed as a factorial design.

The most relevant recent benchmark for the present thesis is [Trucíos \(2026\)](#), who evaluates hierarchical allocation methods out-of-sample and examines their sensitivity to covariance matrix estimation. Using Brazilian and US stock data, [Trucíos \(2026\)](#) compares hierarchical methods such as HRP, HERC and constrained HRP with traditional risk-based portfolios under several covariance estimators. The findings are cautious: hierarchical strategies are sensitive to covariance estimation and show no consistent evidence of outperforming traditional risk-based alternatives out-of-sample. This is directly relevant because it suggests that hierarchical performance depends materially on implementation choices. However, [Trucíos \(2026\)](#) focuses on covariance estimation and broad method comparisons, rather than isolating the relative contribution of tree construction and capital allocation.

[Trucíos \(2026\)](#) closes by suggesting that “exploring alternative risk measures beyond the covariance matrix and experimenting with different linkage functions in hierarchical risk clustering methods may yield valuable insights for practitioners” ([Trucíos, 2026](#), p. 21). The literature reviewed here has separately tested distance metrics ([Barziy and Chlebus, 2020](#); [Salas-Molina et al., 2025a](#)), estimation windows ([Salas-Molina et al., 2025b](#)), allocation rules ([Raffinot, 2018](#); [Pfitzinger and Katzke, 2019](#); [Noguer i Alonso, 2026](#); [Katrancı and Kundakcı, 2026](#)) and covariance estimators ([Trucíos, 2026](#)). The mixed empirical picture, with hierarchical methods outperforming on volatility, turnover and concentration but not consistently on Sharpe ratio or return, suggests that HRP performance may depend materially on implementation choices that are not always separated in the literature.

4 EMPIRICAL METHOD

This section specifies the empirical design used to test whether out-of-sample performance in hierarchical portfolio construction is driven primarily by tree construction or by the allocation rule. The linkage methods and recursive bisection allocation were defined in Section 2. The purpose here is to specify the data, the fixed modelling choices, the

backtesting procedures and the performance metrics that underlie the empirical comparison.

4.1 Data universe

The empirical universe consists of 38 exchange-traded funds (ETFs) covering six asset groups: nine US equity sector funds, five US equity size and style funds, eleven international equity country and regional funds, eight fixed-income funds, three commodity funds and two real-estate funds. The full list of tickers and descriptions is provided in Appendix B. Daily net asset value (NAV) prices are obtained from FactSet for the period January 2008 to April 2026. Returns are total returns computed from NAV rather than from market prices, because NAV reflects the value of the underlying portfolio and avoids the temporary premia and discounts at which ETF shares can trade (Kashner, 2018).

The structure of the universe reflects the kind of multi-asset space that an allocator or risk-parity fund would face in practice, and it provides meaningful dispersion in volatility, correlation and tail behaviour across asset groups. Single-asset-class equity universes tend to display higher and more homogeneous cross-asset correlations, which could give rise to different linkage-kernel interactions. Generalisation to such settings is outside the scope of this study.

Daily returns are used to estimate correlations, covariances and risk measures. Portfolios are rebalanced monthly: at each rebalancing date, weights are estimated from the preceding 1,260 daily observations and then held for the following month before the estimation window advances. After reserving the initial 1,260-day estimation window, the out-of-sample period runs from November 2012 to April 2026, yielding 162 monthly out-of-sample observations.

4.2 Factorial design

The empirical design is a 4×4 factorial experiment. The first dimension is the *linkage method* used to construct the hierarchical tree. The second is the *allocation kernel* used to distribute capital through that tree. Crossing four linkage methods with four allocation kernels produces sixteen hierarchical portfolio strategies. All remaining modelling choices are held fixed: the dissimilarity measure, the splitting rule, the estimation-window length and the rebalancing frequency, as shown in Table 1. Holding these constant is what allows performance differences across the sixteen combinations to be interpreted as differences associated with the linkage method or the allocation kernel.

Table 1: Empirical design. The linkage method and allocation kernel are varied across the 4×4 grid; all other modelling choices are held fixed.

Design dimension	Values	Role in the model
Linkage method	Single, complete, average, Ward	Determines the tree structure
Allocation kernel	$1/N$, inverse-variance, CVaR, CDaR	Determines capital allocation
Distance metric	Pearson-correlation distance	Held fixed
Splitting rule	Equal-size recursive bisection	Held fixed
Rebalancing	Monthly	Held fixed
Estimation window	1,260 trading days	Held fixed

The four linkage methods span a range of tree structures, from the chaining tendency of single linkage to the more balanced clusters typically produced by Ward (Section 2.2.1). The four kernels span a range of allocation principles, from risk-blind allocation through variance-, tail- and drawdown-aware allocation (Section 2.2.3).

The dissimilarity measure is the Pearson correlation transformation of Section 2.2.1, applied across all four linkage methods, so that differences between tree structures reflect the linkage rule alone. The splitting rule is the halving recursive bisection of López de Prado (2016) rather than the dendrogram-following recursion of HERC. Palomar (2025) notes that halving bisection tends to produce more balanced splits than following the tree.

The four allocation kernels differ only in how capital is split at each node. Hierarchical $1/N$ splits capital equally at every node, $\alpha = 1/2$, and weights assets equally within each sub-cluster. The three risk-aware kernels split capital in inverse proportion to the measured risk of the two sub-clusters. Letting R_1 and R_2 denote the risk of the left and right sub-clusters, the share allocated to the left sub-cluster is

$$\alpha = 1 - \frac{R_1}{R_1 + R_2}, \quad (18)$$

so the lower-risk sub-cluster receives the larger share. The inverse-variance kernel measures sub-cluster risk as the variance of an inverse-variance-weighted internal portfolio, the original kernel of López de Prado (2016). The CVaR and CDaR kernels replace variance with conditional value-at-risk and conditional drawdown-at-risk respectively. Both downside-risk kernels are implemented at the 95 per cent confidence level. The purpose of varying these four kernels is not to propose a new allocation rule, but to compare how different risk principles behave when applied within the same hierarchical structure.

4.3 Backtesting setup

The sixteen strategies are evaluated using the Multiple Randomised Backtest (MRB) of [Palomar \(2025\)](#). A single walk-forward backtest on the full universe is included separately for visual illustration of cumulative wealth paths and portfolio behaviour over time. Both are implemented in Python using the `skfolio` library ([Delatte et al., 2024](#)).

The main performance evidence comes from the MRB. A single historical walk-forward backtest can be sensitive to the particular assets that happen to be included in the universe, so that performance differences reflect the specific path rather than the strategy. The MRB reduces this dependence by repeatedly evaluating each strategy on resampled subsets of the full data universe. [Palomar \(2025\)](#) defines the procedure as drawing k resamples, each consisting of a randomly selected subset of assets and a randomly selected contiguous time window, and running a walk-forward backtest on each (see [Figure 6](#)). In each resample, 20 ETFs are drawn from the full 38-ETF universe, and a walk-forward backtest is run on that subset for all sixteen linkage-kernel combinations. The procedure is repeated 200 times over each combination, producing 3,200 resampled walk-forward strategy paths in total. These paths are not independent: every resample shares the same time period, and the drawn asset subsets overlap across resamples. They are therefore treated as resampled realisations rather than independent observations.

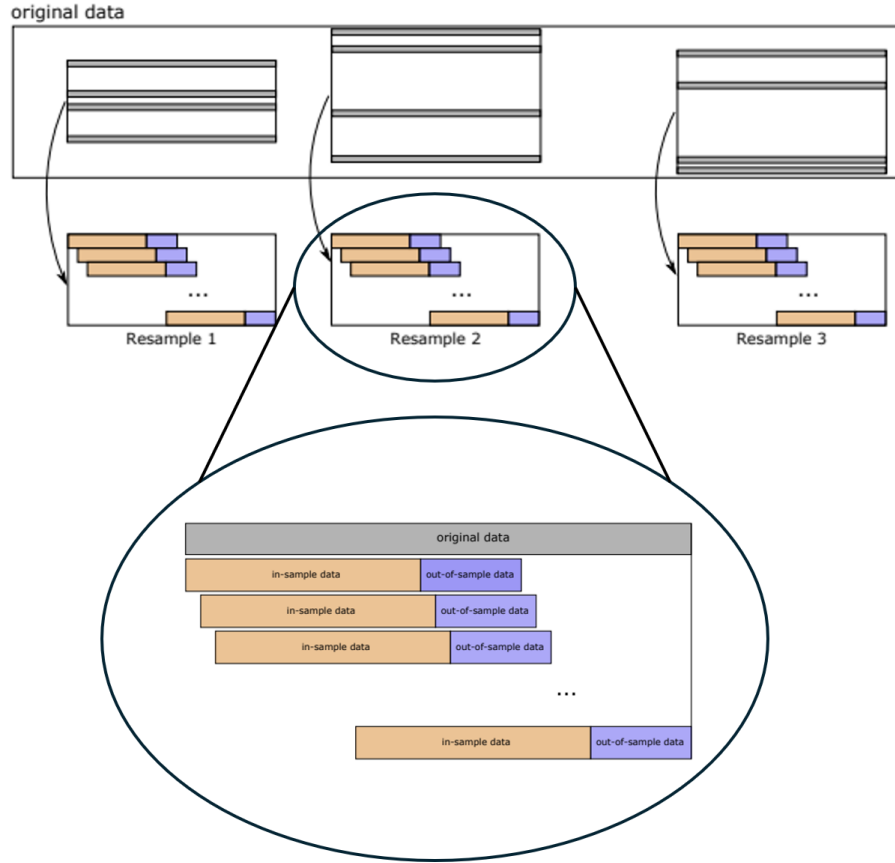


Figure 6: Multiple Randomised Backtests procedure. The original time series is divided into several resampled backtest paths. Within each resample, portfolio methods are estimated on rolling in-sample windows and evaluated on the following out-of-sample windows.

Source: Adapted from [Palomar \(2025\)](#).

Unlike the full MRB design described by [Palomar \(2025\)](#), the contiguous time window is not randomly resampled. The historical sequence from January 2008 to April 2026 is held fixed in every resample, so robustness comes from repeated asset resampling rather than from randomising time windows. Randomising the time window introduces a trade-off between window length and statistical reliability. Short windows reduce overlap across draws but leave few out-of-sample observations for reliable estimation of performance statistics. Longer windows improve reliability but, given a sample of 4,781 trading days, generate overlap across subsamples. Holding the time dimension fixed ensures that performance differences are attributable to asset composition through the hierarchical allocation rather than to which time period a particular draw happened to cover.

The second procedure is a single walk-forward backtest on the full 38-ETF universe. It is used to illustrate cumulative wealth paths and portfolio behaviour over time. It is not

the basis for the performance comparison, because it remains a single historical realisation and is exposed to the backtesting pitfalls that motivate the MRB (Wang et al., 2014; Palomar, 2025). The flat $1/N$ portfolio is also included as a conventional benchmark.

In both procedures the estimation window is fixed at 1,260 trading days, approximately five years, following the empirical recommendation of Salas-Molina et al. (2025b), who report that this lookback maximises out-of-sample Sharpe ratios for HRP portfolios. The window rolls forward through the sample with monthly rebalancing. At each rebalancing date, weights are estimated using only the preceding 1,260 observations, so that no future information enters estimation and look-ahead bias is avoided (see Figure 6). The covariance matrix is estimated every rebalancing date with the Ledoit-Wolf shrinkage estimator (Ledoit and Wolf, 2003, 2004). It enters the inverse-variance kernel and the performance-metric calculations, but not the distance matrix used for tree construction, which is computed from sample Pearson correlations. Because recursive bisection allocates a strictly positive fraction of capital at every node, all weights are non-negative and sum to one, so every portfolio is fully invested and long-only. No further single-name weight constraints are imposed. The universe is defined ex ante as a broad multi-asset set rather than a specific asset class, and the backtest is run once, without ex post adjustment of parameters, to limit overfitting (Palomar, 2025).

4.4 Performance metrics

Each strategy is evaluated out-of-sample on a set of performance, diversification and implementation metrics, computed from the out-of-sample return series of each resample under the MRB.

Performance is summarised by annualised return, annualised volatility, the Sharpe ratio, the maximum drawdown and expected shortfall. Annualised return and volatility scale the mean and standard deviation of the out-of-sample returns to an annual frequency. The risk-free rate is set to zero, so the Sharpe ratio is computed as annualised portfolio return divided by annualised volatility,

$$\text{SR} = \frac{R_p}{\sigma_p}, \quad (19)$$

where R_p is the annualised portfolio return and σ_p the annualised portfolio volatility (Palomar, 2025; Trucíos, 2026). The maximum drawdown is the largest peak-to-trough

decline in cumulative wealth over the evaluation period, and expected shortfall is the conditional value-at-risk of portfolio returns at the 95 per cent level, defined in Section 2.1, measuring the average loss in the worst 5 per cent of return observations.

Diversification is measured by the *diversification ratio* (Choueifaty and Coignard, 2008), defined as the weighted average of standalone asset volatilities divided by portfolio volatility,

$$\text{DR} = \frac{\sum_j w_j \sigma_j}{\sqrt{w^\top \Sigma w}}, \quad (20)$$

where w_j and σ_j are the weight and standalone volatility of asset j and Σ is the covariance matrix. The ratio equals one for a single-asset portfolio and rises as the portfolio combines assets whose risks partly offset, so higher values indicate that portfolio volatility is lower relative to the volatility of its constituents.

Turnover is defined as the mean absolute change in target weights between consecutive rebalancing dates,

$$\Gamma = \frac{1}{N_{\text{reb}} - 1} \sum_{n=2}^{N_{\text{reb}}} \sum_{i=1}^N |w_i(n) - w_i(n-1)|, \quad (21)$$

where N_{reb} is the number of rebalancing events, $w_i(n)$ the target weight of asset i at rebalancing date n , and $w_i(n-1)$ the target weight at the previous rebalancing date (Reigner et al., 2020). This measure does not account for weight drift between rebalancing dates and therefore understates true implementation turnover. It is used as a relative diagnostic for comparing strategies rather than as an estimate of actual trading costs.

5 RESULTS

5.1 Linkage method versus allocation kernel

Table 2 reports the mean out-of-sample performance of the sixteen linkage-kernel combinations across the 200 resampled walk-forward paths. Table 3 then summarises the same results as marginal averages, first by allocation kernel and then by linkage method, making the relative importance of each modelling choice easier to interpret.

The main result is that the allocation kernel drives substantially larger differences in risk-adjusted performance than the linkage method. In Table 3, the average Sharpe ratio ranges from 0.713 for hierarchical $1/N$ to 0.800 for CVaR, a spread of 0.087 across kernels. By contrast, the average Sharpe ratio across linkage methods ranges from 0.758 for Ward

to 0.775 for Single, a spread of 0.017. The spread of marginal average Sharpe ratios across kernels is therefore approximately five times larger than the spread across linkage methods.

Table 2: Mean performance across 200 resampled walk-forward paths.

Kernel	Linkage	Return	Volatility	Sharpe	Max DD	ES 95%	DR	Turnover
1/N	Single	8.99%	12.50%	0.720	29.75%	1.84%	1.358	6.72%
	Complete	8.81%	12.39%	0.712	29.36%	1.82%	1.356	7.35%
	Average	8.83%	12.44%	0.711	29.52%	1.83%	1.354	6.49%
	Ward	8.77%	12.39%	0.708	29.38%	1.82%	1.356	5.94%
IVar	Single	3.33%	4.46%	0.776	14.86%	0.67%	1.693	3.75%
	Complete	3.27%	4.42%	0.770	14.76%	0.66%	1.692	4.12%
	Average	3.28%	4.44%	0.769	14.79%	0.66%	1.694	3.79%
	Ward	3.21%	4.43%	0.758	14.80%	0.66%	1.694	3.66%
CVaR	Single	5.55%	6.93%	0.810	17.51%	1.03%	1.668	5.46%
	Complete	5.41%	6.83%	0.802	17.15%	1.01%	1.669	5.97%
	Average	5.42%	6.87%	0.798	17.30%	1.02%	1.667	5.43%
	Ward	5.33%	6.85%	0.788	17.25%	1.01%	1.668	5.18%
CDaR	Single	6.17%	7.79%	0.792	19.14%	1.17%	1.598	7.01%
	Complete	6.03%	7.66%	0.787	18.73%	1.15%	1.598	7.56%
	Average	6.06%	7.71%	0.786	18.88%	1.16%	1.598	6.92%
	Ward	5.98%	7.68%	0.778	18.87%	1.15%	1.598	6.63%

Note: Return and volatility are annualised. Sharpe denotes the annualised Sharpe ratio, Max DD maximum drawdown, ES 95% expected shortfall at the 95 per cent level, DR diversification ratio and average turnover per rebalancing.

Source: Author's calculations.

The kernel results show a trade-off rather than a single dominant method. The $1/N$ kernel records the highest average annualised return, at 8.85 per cent, but also the highest volatility, maximum drawdown and expected shortfall. Inverse-variance is the most defensive kernel, with the lowest volatility, maximum drawdown, expected shortfall and turnover, but it gives up substantial return, with an average annualised return of 3.27 per cent. This is a consequence of inverse-variance weighting which heavily tilts the portfolio toward the lowest-volatility assets in this multi-asset universe. CVaR and CDaR occupy the middle ground. They retain more return exposure than inverse-variance while reducing downside risk relative to $1/N$. CVaR records the highest average Sharpe ratio, at 0.800, while CDaR produces a higher average return, at 6.06 per cent, but also higher volatility, drawdown and turnover. The risk-aware kernels therefore improve risk-adjusted performance and downside risk relative to $1/N$, but they do not dominate across all performance and implementation metrics.

Table 3: Marginal averages by allocation kernel and linkage method.

Dimension	Category	Return	Volatility	Sharpe	Max DD	ES 95%	DR	Turnover
Kernel	1/N	8.85%	12.43%	0.713	29.50%	1.83%	1.356	6.62%
	IVar	3.27%	4.43%	0.768	14.80%	0.66%	1.693	3.83%
	CVaR	5.43%	6.87%	0.800	17.30%	1.02%	1.668	5.51%
	CDaR	6.06%	7.71%	0.786	18.91%	1.16%	1.598	7.03%
Linkage	Single	6.01%	7.92%	0.775	20.31%	1.18%	1.579	5.74%
	Complete	5.88%	7.83%	0.768	20.00%	1.16%	1.579	6.25%
	Average	5.90%	7.86%	0.766	20.12%	1.17%	1.578	5.66%
	Ward	5.82%	7.84%	0.758	20.07%	1.16%	1.579	5.35%

Note: Metric definitions are as in Table 2.

Source: Author’s calculations.

The full factorial grid in Table 2 confirms that this pattern is not driven by a single linkage choice. Within every linkage column, CVaR produces the highest Sharpe ratio, ranging from 0.788 under Ward to 0.810 under Single, and 1/ N records the lowest Sharpe ratio under all four linkage methods.

The effects of the linkage method are smaller in magnitude. Single linkage records the highest average Sharpe ratio across kernels, while Ward records the lowest. However, the differences are modest compared with the kernel differences. The clearest systematic linkage pattern appears in turnover. Ward has the lowest turnover across all kernels, with an average of 5.35 per cent, while Complete has the highest with an average of 6.25 per cent. This suggests that linkage choice is more relevant for implementation stability than for gross performance. The result is consistent with the intuition that Ward produces more balanced tree structures, while Single linkage can create more unbalanced trees and slightly higher risk-taking.

To assess robustness, the kernel and linkage spreads are recomputed within each of the 200 resamples and compared as paired differences, since both are evaluated on the same resampled universe. Figure 7 illustrates that the within-resample kernel spread has a median of 0.126, with a central 95% range of 0.057 to 0.332, compared with a median linkage spread of 0.038, with a central 95% range of 0.012 to 0.082. The relevant paired difference, defined as the kernel spread minus the linkage spread within each resample, has a median of 0.084 and a central 95% range of 0.010 to 0.279, which remains entirely above zero. The kernel spread exceeds the linkage spread in 196 of the 200 resamples,

corresponding to 98 per cent. This indicates that performance is more sensitive to the allocation kernel than to the linkage rule across resampled universes, rather than only in the marginal averages.

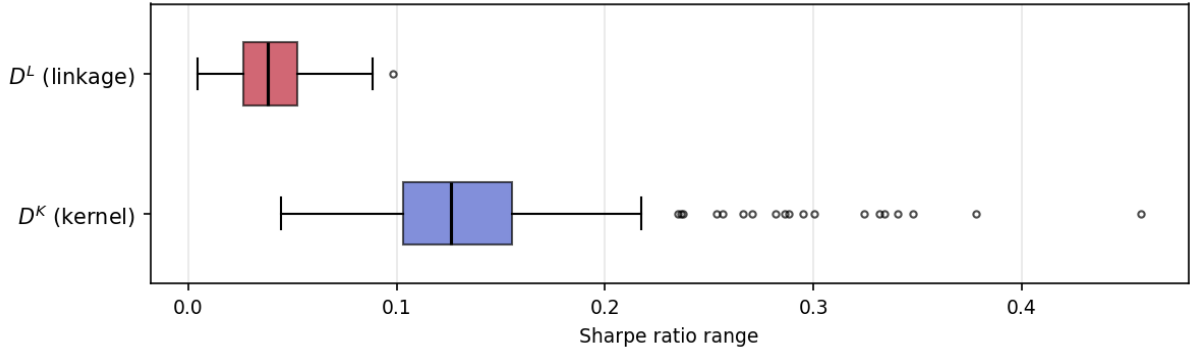


Figure 7: Distribution of within-resample Sharpe-ratio spreads across the 200 resamples. D^K is the range of the four kernel marginal Sharpe ratios within a resample. D^L is the range of the four linkage marginals. Boxes show the interquartile range and median, whiskers extend to $1.5 \times \text{IQR}$ and points beyond are individual resamples.

Source: Author's calculations.

Taken together, the results refine the implementation message in the hierarchical allocation literature. Prior studies show that hierarchical methods are sensitive to modelling choices, but the factorial design shows that not all choices matter equally. In this setting, the allocation kernel is the primary driver of risk-adjusted performance, while the linkage method plays a secondary role and matters most clearly through turnover.

5.2 Full-universe walk-forward illustration

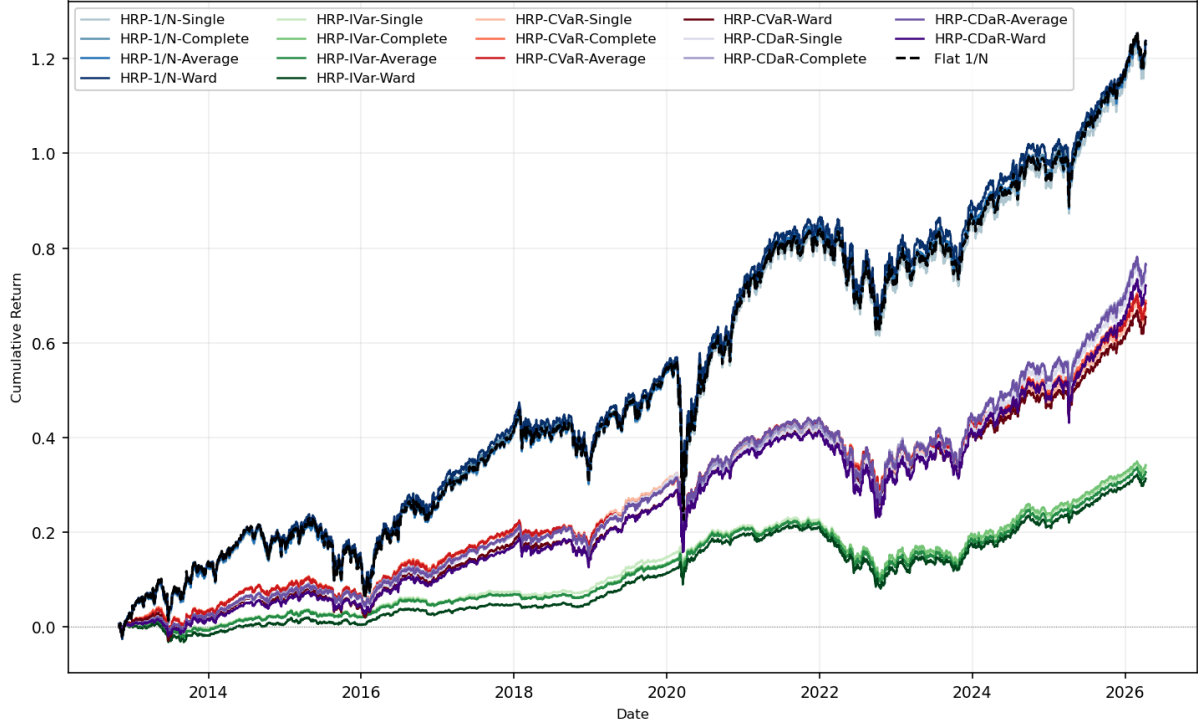


Figure 8: Full-universe walk-forward cumulative returns. The figure shows the cumulative return paths of the sixteen hierarchical allocation strategies over the full 38-ETF universe, together with a flat $1/N$ benchmark.

Source: Author's calculations.

Figure 8 presents the cumulative returns from the full-universe backtest. The figure reinforces the main finding from Section 5.1. The cumulative return paths separate mainly by allocation kernel, while the four linkage variants within each kernel remain tightly clustered. The flat $1/N$ benchmark and the hierarchical $1/N$ strategies produce the highest cumulative returns, but also experience the most visible drawdowns, especially around the COVID-19 shock in early 2020. This is consistent with Table 3, where $1/N$ has the highest average return but also the highest volatility, maximum drawdown and expected shortfall.

The inverse-variance strategies form the lowest cumulative-return group throughout the sample. The CVaR and CDaR strategies sit in the intermediate range. They do not match the cumulative return of $1/N$, but they retain more return exposure than inverse-variance while controlling downside risk. This visual pattern is consistent with the kernel trade-off reported in Tables 2 and 3.

Appendix Figures C2 and C3 show the evolution of capital weights and volatility

contributions over the walk-forward period. The kernel dimension produces visibly larger differences in both series than the linkage dimension, which is consistent with the MRB result.

6 CONCLUSION

Hierarchical portfolio construction requires two separate decisions: how to group assets into a tree, and how to allocate capital through that tree once it has been built. This thesis asked which of these two decisions matters more for out-of-sample performance. In the multi-asset ETF setting studied here, the answer is the allocation kernel. The spread of marginal average Sharpe ratios between the best and worst kernel is 0.087, approximately five times larger than the 0.017 spread between the best and worst linkage method, and this order holds in 196 of 200 resampled universes. A similar pattern appears across the main return and risk measures. In practical terms, the allocation kernel determines the portfolio’s broad risk-return profile, while the linkage method mainly adjusts its implementation characteristics.

The kernel result is a trade-off rather than a contest with a single winner. Each kernel produces a risk-return profile consistent with its construction. The risk-blind $1/N$ kernel takes the most risk and earns the highest return, but records the highest volatility, maximum drawdown and expected shortfall. Inverse-variance is the most defensive kernel: by weighting toward the lowest-volatility assets it achieves the lowest risk and turnover, at the cost of substantial return. In this multi-asset universe that tilt corresponds to a heavy concentration in the fixed-income block. The downside-aware kernels, CVaR and CDaR, sit between these extremes, retaining more return than inverse-variance while reducing downside risk relative to $1/N$. They spread volatility contributions more evenly across asset groups, and post Sharpe ratios that are close to one another and above $1/N$. The defensible conclusion is that the risk-aware kernels improve risk-adjusted performance and downside risk relative to ignoring risk. If the objective is risk-adjusted return, the downside-aware kernels post the highest Sharpe ratios among the four kernels and distribute volatility contributions more evenly across asset groups in this multi-asset universe, albeit at higher turnover than inverse-variance.

The linkage method plays a secondary role. Within a given allocation kernel, the four linkage methods produce comparatively similar performance outcomes. Under the CVaR

kernel, for example, annualised return varies by about two-tenths of a percentage point and the Sharpe ratio by 0.022 across the four linkages. The clearest linkage effect appears in turnover. Ward produces the lowest turnover within every kernel, while Complete produces the highest. This is consistent with Ward’s tendency to produce more balanced trees (Papenbrock, 2011). Linkage choice therefore matters more for implementation stability than for gross risk-adjusted performance.

The contribution of the thesis is one of attribution rather than method development. Earlier studies have varied the modelling choices but often studied them separately. By crossing linkage methods and allocation kernels in a single factorial design, while holding the dissimilarity measure and splitting rule fixed, this thesis compares the two dimensions on the same empirical footing. That the allocation kernel drives out-of-sample performance more consistently than the linkage method, with this ordering holding in 98 per cent of the resampled universes, provides clear evidence that the allocation rule is the more important modelling choice in this setting. It also qualifies the recommendation in Trucíos (2026) that linkage functions deserve systematic examination. Linkage is worth examining, but in this setting its main role is implementation-related rather than performance-driving.

For practitioners choosing among hierarchical allocation methods, the order of modelling decisions should follow these results. The risk principle embedded in the allocation kernel should be chosen first, because it determines whether the portfolio is aggressive, defensive or intermediate. The tree-building rule should be treated as a secondary implementation choice. In this study, Ward’s lower turnover makes it a natural default when implementation stability is the priority. This refines rather than overturns the existing literature. Hierarchical methods remain useful because they impose structure on a noisy dependence problem, but their realised performance depends primarily on the rule used to allocate capital through that structure.

These conclusions are bounded by the study’s design. The covariance estimator, distance metric and estimation window are all held fixed, and Trucíos (2026) shows that hierarchical methods are sensitive to covariance estimation in particular. The results reported here should therefore not be interpreted as invariant to alternative modelling choices. Pfitzinger and Katzke (2019) show that linkage matters most when the allocation rule makes direct use of the tree structure in dissimilarity-based allocation. In this thesis, the tree determines the ordering used by the allocation procedure, rather than serving as

the direct basis for allocation. The weak linkage effects should therefore be interpreted as specific to this implementation. The evidence also comes from a single multi-asset ETF universe over one fixed historical period, and may not generalise to single-asset-class equity universes with more homogeneous correlations. This study also sits entirely within the hierarchical family. Benchmarking against conventional methods lies outside its scope and is left to prior and future work. Finally, transaction costs are not modelled directly; turnover is reported as a proxy, so a natural extension would be to examine the same linkage-kernel effects net of a credible cost model.

In hierarchical portfolio construction, performance is shaped less by finding the “right” tree than by choosing the risk principle used to allocate capital through it. The tree organises the problem; the allocation rule determines the portfolio.

References

- Allen, G. C. (2010). The Risk Parity Approach to Asset Allocation. Technical report, Callan Investments Institute Research, Callan Associates.
- Anderson, R. M., Bianchi, S. W., and Goldberg, L. R. (2012). Will My Risk Parity Strategy Outperform? *Financial Analysts Journal*, 68(6):75–93.
- Ang, A. and Chen, J. (2002). Asymmetric Correlations of Equity Portfolios. *Journal of Financial Economics*, 63(3):443–494. [https://doi.org/10.1016/S0304-405X\(02\)00068-5](https://doi.org/10.1016/S0304-405X(02)00068-5).
- Ardia, D., Boudt, K., and Nguyen, G. (2018). Beyond Risk-Based Portfolios: Balancing Performance and Risk Contributions in Asset Allocation. *Quantitative Finance*, 18(8):1249–1259. <https://doi.org/10.1080/14697688.2018.1424349>.
- Arévalo, A., León, D., and Hernandez, G. (2019). Portfolio Selection Based on Hierarchical Clustering and Inverse-Variance Weighting. In *Computational Science – ICCS 2019*, volume 11538 of *Lecture Notes in Computer Science*, pages 315–328. Springer, Cham. https://doi.org/10.1007/978-3-030-22744-9_25.
- Barziy, I. and Chlebus, M. (2020). HRP Performance Comparison in Portfolio Optimization under Various Codependence and Distance Metrics. Working Paper 2020-21, Faculty of Economic Sciences, University of Warsaw. <https://ideas.repec.org/p/war/wpaper/2020-21.html>.
- Best, M. J. and Grauer, R. R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means. *The Review of Financial Studies*, 4(2):315–342.
- Black, F. and Litterman, R. (1992). Global Portfolio Optimization. *Financial Analysts Journal*, 48(5):28–43. <https://doi.org/10.2469/faj.v48.n5.28>.
- Chaves, D., Hsu, J., Li, F., and Shakernia, O. (2011). Risk Parity Portfolio vs. Other Asset Allocation Heuristic Portfolios. *The Journal of Investing*, 20(1):108–118.
- Chekhlov, A., Uryasev, S., and Zabarankin, M. (2005). Drawdown Measure in Portfolio

- Optimization. *International Journal of Theoretical and Applied Finance*, 8(1):13–58. <https://doi.org/10.1142/S0219024905002767>.
- Chopra, V. K. and Ziemba, W. T. (1993). The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *The Journal of Portfolio Management*, 19(2):6–11.
- Choueifaty, Y. and Coignard, Y. (2008). Toward Maximum Diversification. *The Journal of Portfolio Management*, 35(1):40–51. <https://doi.org/10.3905/JPM.2008.35.1.40>.
- Cotton, P. (2024). Schur Complementary Allocation: A Unification of Hierarchical Risk Parity and Minimum Variance Portfolios. arXiv preprint arXiv:2411.05807. <https://arxiv.org/abs/2411.05807>.
- Deković, D. and Šimović, P. P. (2025). Hierarchical Risk Parity: Efficient Implementation and Real-World Analysis. *Future Generation Computer Systems*, 167:107744. <https://doi.org/10.1016/j.future.2025.107744>.
- Delatte, H., Nicolini, C., and Manzi, M. (2024). skfolio: Portfolio Optimization in Python. Python library. <https://doi.org/10.5281/zenodo.16148630>.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953. <https://doi.org/10.1093/rfs/hhm075>.
- Eidenvall, A. (2021). Hierarchical Clustering to Improve Portfolio Tail Risk Characteristics. Master’s thesis, Centre for Mathematical Sciences, Lund University. LUTFMS-3409-2021.
- Gower, J. C. (1966). Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika*, 53(3–4):325–338. <https://doi.org/10.1093/biomet/53.3-4.325>.
- Graham, B. (1949). *The Intelligent Investor*. Harper & Brothers, New York.
- Idzorek, T. M. (2002). A Step-by-Step Guide to the Black-Litterman Model: Incorporating User-Specified Confidence Levels. Working paper, Ibbotson Associates.

- Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 58(4):1651–1683. <https://doi.org/10.1111/1540-6261.00580>.
- Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs, NJ.
- Jobson, J. D. and Korkie, B. (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association*, 75(371):544–554. <https://doi.org/10.1080/01621459.1980.10477507>.
- Jorion, P. (1985). International Portfolio Diversification with Estimation Risk. *The Journal of Business*, 58(3):259–278.
- Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291. <https://doi.org/10.2307/1914185>.
- Kashner, E. (2018). Shopping Across ETF Legal Structures. FactSet Insight. <https://insight.factset.com/shopping-across-etf-legal-structures> [Accessed 26 May 2026].
- Katrancı, A. and Kundakcı, N. (2026). A Novel Hybrid Portfolio Optimization Framework: Integrating HRP with MEREC and WEDBA Methods. *Informatica*. <https://doi.org/10.15388/26-INFOR618>.
- Ledoit, O. and Wolf, M. (2003). Honey, I Shrunk the Sample Covariance Matrix. Economics and Business Working Paper 691, Universitat Pompeu Fabra. <https://doi.org/10.2139/ssrn.433840>.
- Ledoit, O. and Wolf, M. (2004). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis*, 88(2):365–411. [https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4).
- Lee, W. (2011). Risk-Based Asset Allocation: A New Answer to an Old Question? *The Journal of Portfolio Management*, 37(4):11–28. <https://doi.org/10.3905/jpm.2011.37.4.011>.

- Lohre, H., Rother, C., and Schäfer, K. A. (2020). Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-Asset Multi-Factor Allocations. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3513399>.
- Longin, F. and Solnik, B. (2001). Extreme Correlation of International Equity Markets. *The Journal of Finance*, 56(2):649–676. <https://doi.org/10.1111/0022-1082.00340>.
- López de Prado, M. (2016). Building Diversified Portfolios that Outperform Out of Sample. *The Journal of Portfolio Management*, 42(4):59–69. <https://doi.org/10.3905/jpm.2016.42.4.059>.
- Maillard, S., Roncalli, T., and Teïletche, J. (2010). The Properties of Equally Weighted Risk Contribution Portfolios. *The Journal of Portfolio Management*, 36(4):60–70. <https://doi.org/10.3905/JPM.2010.36.4.060>.
- Mantegna, R. N. (1999). Hierarchical Structure in Financial Markets. *The European Physical Journal B*, 11(1):193–197. <https://doi.org/10.1007/s100510050929>.
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>.
- Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, New York.
- Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal? *Financial Analysts Journal*, 45(1):31–42. <https://doi.org/10.2469/faj.v45.n1.31>.
- Michaud, R. O. (1998). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Harvard Business School Press, Boston, MA.
- Michaud, R. O. and Michaud, R. O. (2008). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press, New York, 2 edition.
- Munich Re (2020). Towards Robust Portfolios: Hierarchical Risk Parity: Persistent Diversification Uncovered by Graph Theory and Machine Learning. FIVE report, Munich Re, Munich.

- Murtagh, F. (2004). On Ultrametricity, Data Coding, and Computation. *Journal of Classification*, 21(2):167–184.
- Noguer i Alonso, M. (2026). Return-Adjusted Hierarchical Risk Parity and Schur Portfolios: Mathematical Foundations and Empirical Evidence. Working paper, Artificial Intelligence Finance Institute.
- Palomar, D. P. (2025). *Portfolio Optimization: Theory and Application*. Cambridge University Press, Cambridge.
- Papenbrock, J. (2011). *Asset Clusters and Asset Networks in Financial Risk Management and Portfolio Optimization*. Doctoral dissertation, Karlsruher Institut für Technologie (KIT).
- Pfützinger, J. and Katzke, N. (2019). A Constrained Hierarchical Risk Parity Algorithm with Cluster-Based Capital Allocation. Stellenbosch Economic Working Papers WP14/2019, Department of Economics, Stellenbosch University.
- Pukthuanthong, K., Roll, R. W., and Subrahmanyam, A. (2019). A Protocol for Factor Identification. *The Review of Financial Studies*, 32(4):1573–1607. <https://doi.org/10.1093/rfs/hhy093>.
- Qian, E. (2005). Risk Parity Portfolios: Efficient Portfolios through True Diversification. Research paper, PanAgora Asset Management.
- Raffinot, T. (2017). Hierarchical Clustering-Based Asset Allocation. *The Journal of Portfolio Management*, 44(2):89–99. <https://doi.org/10.3905/jpm.2018.44.2.089>.
- Raffinot, T. (2018). The Hierarchical Equal Risk Contribution Portfolio. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3237540>.
- Reigner, P.-A., Nguyen, V., Ciliberti, S., Seager, P., and Bouchaud, J.-P. (2020). Agnostic Allocation Portfolios: A Sweet Spot in the Risk-Based Jungle? *The Journal of Portfolio Management*, 46(4):22–38. <https://doi.org/10.3905/jpm.2020.1.129>.
- Rockafellar, R. T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2(3):21–41.

- Roncalli, T. (2013). *Introduction to Risk Parity and Budgeting*. Chapman and Hall/CRC, Boca Raton, FL.
- Salas-Molina, F., Pla-Santamaria, D., Garcia-Bernabeu, A., and Hilario-Caballero, A. (2025a). An Empirical Evaluation of Distance Metrics in Hierarchical Risk Parity Methods for Asset Allocation. *Computational Economics*, 66:5189–5206. <https://doi.org/10.1007/s10614-025-10848-w>.
- Salas-Molina, F., Pla-Santamaria, D., Garcia-Bernabeu, A., and Reig-Mullor, J. (2025b). Estimation Windows in Hierarchical Risk Parity Methods for Portfolio Selection. In *Data Science and Artificial Intelligence – DSA ISC 2024*, volume 14779 of *Lecture Notes in Computer Science*, pages 42–55. Springer, Cham. https://doi.org/10.1007/978-3-031-78241-1_5.
- Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482.
- Stein, C. (1955). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 197–206, Berkeley, CA. University of California Press.
- Trucíos, C. (2026). Hierarchical Risk Clustering versus Traditional Risk-Based Portfolios: An Empirical Out-of-Sample Comparison. *Empirical Economics*, 70(58). <https://doi.org/10.1007/s00181-026-02900-x>.
- Wang, S., Luo, Y., Alvarez, M.-A., Jussa, J., Wang, A., and Rohal, G. (2014). Seven Sins of Quantitative Investing. DB Quant Handbook, Part II, Deutsche Bank Securities Inc., New York.
- Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244. <https://doi.org/10.1080/01621459.1963.10500845>.
- Windcliff, H. and Boyle, P. (2004). The 1/N Pension Investment Puzzle. *North American Actuarial Journal*, 8(3):32–45. <https://doi.org/10.1080/10920277.2004.10596151>.

Wuebben, B. J. (2026). Beyond De Prado and Cotton: Hierarchical and Iterative Methods for General Mean-Variance Portfolios. arXiv preprint arXiv:2604.23833. <https://arxiv.org/abs/2604.23833>.

Appendix A: Recursive bisection algorithm

This appendix states the recursive bisection algorithm of [López de Prado \(2016\)](#) used to allocate capital through the quasi-diagonalised asset ordering. Let $L_0 = \{n_1, \dots, n_N\}$ denote the ordered list of assets obtained from quasi-diagonalisation, and let Σ denote the covariance matrix of asset returns.

1. Initialise the list of clusters as $L = \{L_0\}$.
2. Initialise the portfolio weights as $w_i = 1$, for $i = 1, \dots, N$.
3. For each cluster $L_i \in L$ with $|L_i| > 1$, split L_i into two ordered sub-clusters,

$$L_i^{(1)} \cup L_i^{(2)} = L_i, \quad L_i^{(1)} \cap L_i^{(2)} = \emptyset,$$

where the ordering inherited from L_0 is preserved.

4. For each sub-cluster $L_i^{(j)}$, $j = 1, 2$, compute inverse-variance weights within the sub-cluster:

$$\tilde{w}^{(j)} = \frac{\text{diag}(V^{(j)})^{-1}}{\text{tr}[\text{diag}(V^{(j)})^{-1}]},$$

where $V^{(j)}$ is the covariance matrix of the assets in sub-cluster $L_i^{(j)}$.

5. Compute the variance of each sub-cluster portfolio:

$$\tilde{V}^{(j)} = \tilde{w}^{(j)\top} V^{(j)} \tilde{w}^{(j)}, \quad j = 1, 2.$$

6. Compute the split factor:

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

7. Rescale the weights in the first and second sub-clusters:

$$w_n \leftarrow \alpha_i w_n, \quad n \in L_i^{(1)}, \quad w_n \leftarrow (1 - \alpha_i) w_n, \quad n \in L_i^{(2)}.$$

8. Replace each split cluster in L with its two sub-clusters and repeat steps 3–7 until all clusters contain a single asset.

The final values of w_i are the portfolio weights.

Appendix B: Data universe

Table 4: ETF universe used in the empirical analysis. The table reports the 38 ETFs included in the empirical universe, grouped by asset category

Name	Provider	Ticker	Asset class	Short description
US equity sectors				
Technology Select Sector SPDR	State Street	XLK	US equity sector	S&P 500 technology stocks, market-cap weighted
Financial Select Sector SPDR	State Street	XLF	US equity sector	S&P 500 financial stocks, market-cap weighted
Energy Select Sector SPDR	State Street	XLE	US equity sector	S&P 500 energy stocks, market-cap weighted
Health Care Select Sector SPDR	State Street	XLV	US equity sector	S&P 500 healthcare stocks, market-cap weighted
Industrial Select Sector SPDR	State Street	XLI	US equity sector	S&P 500 industrial stocks, market-cap weighted
Consumer Discretionary Select Sector SPDR	State Street	XLY	US equity sector	S&P 500 consumer discretionary stocks, market-cap weighted
Consumer Staples Select Sector SPDR	State Street	XLP	US equity sector	S&P 500 consumer staples stocks, market-cap weighted
Utilities Select Sector SPDR	State Street	XLU	US equity sector	S&P 500 utilities stocks, market-cap weighted
Materials Select Sector SPDR	State Street	XLB	US equity sector	S&P 500 materials stocks, market-cap weighted
US equity size and style				
SPDR S&P 500 ETF Trust	State Street	SPY	US equity broad	S&P 500 equity index
iShares Russell 2000 ETF	iShares	IWM	US equity small cap	Russell 2000 small-cap index
SPDR S&P MidCap 400 ETF Trust	State Street	MDY	US equity mid cap	S&P MidCap 400 index
iShares Russell 1000 Growth ETF	iShares	IWF	US equity growth	Russell 1000 growth stocks
iShares Russell 1000 Value ETF	iShares	IWD	US equity value	Russell 1000 value stocks
Intl equity				
iShares MSCI EAFE ETF	iShares	EFA	Intl equity developed	MSCI EAFE index, developed markets excluding US and Canada
iShares MSCI Emerging Markets ETF	iShares	EEM	Intl equity emerging	MSCI Emerging Markets index
iShares MSCI Japan ETF	iShares	EWJ	Intl equity country	MSCI Japan index
iShares MSCI Germany ETF	iShares	EWG	Intl equity country	MSCI Germany index
iShares MSCI United Kingdom ETF	iShares	EWU	Intl equity country	MSCI United Kingdom index
iShares MSCI Australia ETF	iShares	EWA	Intl equity country	MSCI Australia index
iShares MSCI Canada ETF	iShares	EWG	Intl equity country	MSCI Canada index
iShares MSCI Brazil ETF	iShares	EWZ	Intl equity country	MSCI Brazil index
iShares MSCI Taiwan ETF	iShares	EWT	Intl equity country	MSCI Taiwan index
iShares MSCI South Korea ETF	iShares	EWY	Intl equity country	MSCI South Korea index

Continued on next page

Table 4: ETF universe used in the empirical analysis, continued

Name	Provider	Ticker	Asset class	Short description
iShares MSCI Sweden ETF	iShares	EWD	Intl equity country	MSCI Sweden index
Fixed income				
iShares 20+ Year Treasury Bond ETF	iShares	TLT	US Treasury long	ICE US Treasury 20+ Year index
iShares 7–10 Year Treasury Bond ETF	iShares	IEF	US Treasury intermediate	ICE US Treasury 7–10 Year index
iShares 1–3 Year Treasury Bond ETF	iShares	SHY	US Treasury short	ICE US Treasury 1–3 Year index
iShares Core US Aggregate Bond ETF	iShares	AGG	US aggregate bond	Bloomberg US Aggregate Bond index
iShares iBoxx \$ Investment Grade Corporate Bond ETF	iShares	LQD	US IG corporate	Investment-grade USD corporate bonds
iShares iBoxx \$ High Yield Corporate Bond ETF	iShares	HYG	US HY corporate	High-yield USD corporate bonds
iShares J.P. Morgan USD Emerging Markets Bond ETF	iShares	EMB	Emerging-market bonds	USD-denominated emerging-market sovereign bonds
iShares TIPS Bond ETF	iShares	TIP	US TIPS	Bloomberg US Treasury Inflation-Protected Securities
Commodities				
SPDR Gold Shares	State Street	GLD	Commodity gold	Physical gold bullion
iShares Silver Trust	iShares	SLV	Commodity silver	Physical silver bullion
Invesco DB Commodity Index Tracking Fund	Invesco	DBC	Commodity broad	Diversified commodity index
Real estate				
Vanguard Real Estate ETF	Vanguard	VNQ	US REITs	MSCI US Investable Market Real Estate 25/50 index
SPDR Dow Jones Intl Real Estate ETF	State Street	RWX	Intl REITs	Dow Jones Global ex-US Select Real Estate Securities

Source: Author’s compilation based on FactSet and ETF provider descriptions.

Appendix C: Figures

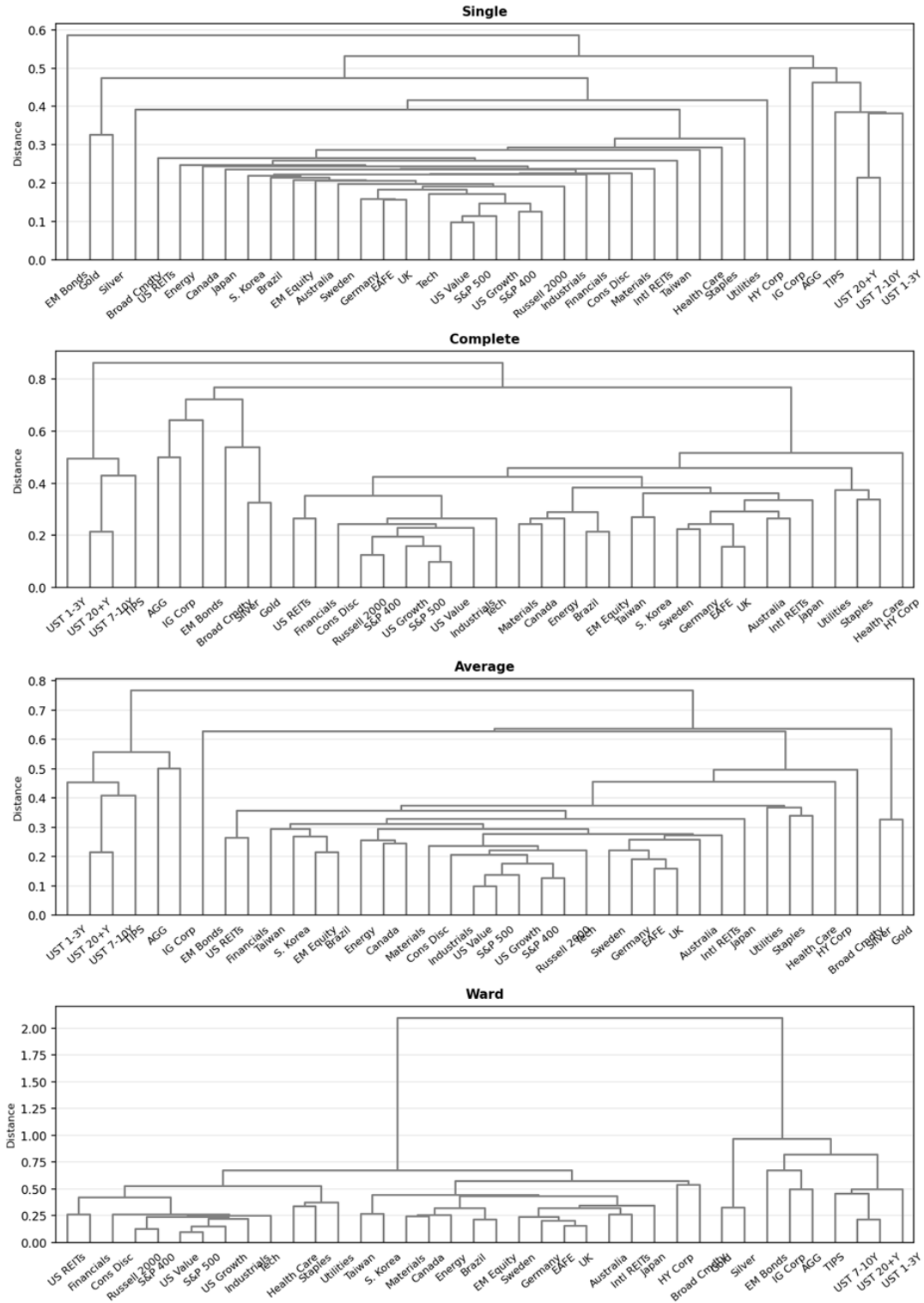


Figure C1: Dendrograms by linkage method over the 38-ETF universe. The vertical axis shows the clustering distance at which assets or clusters are merged.

Source: Author's calculations.

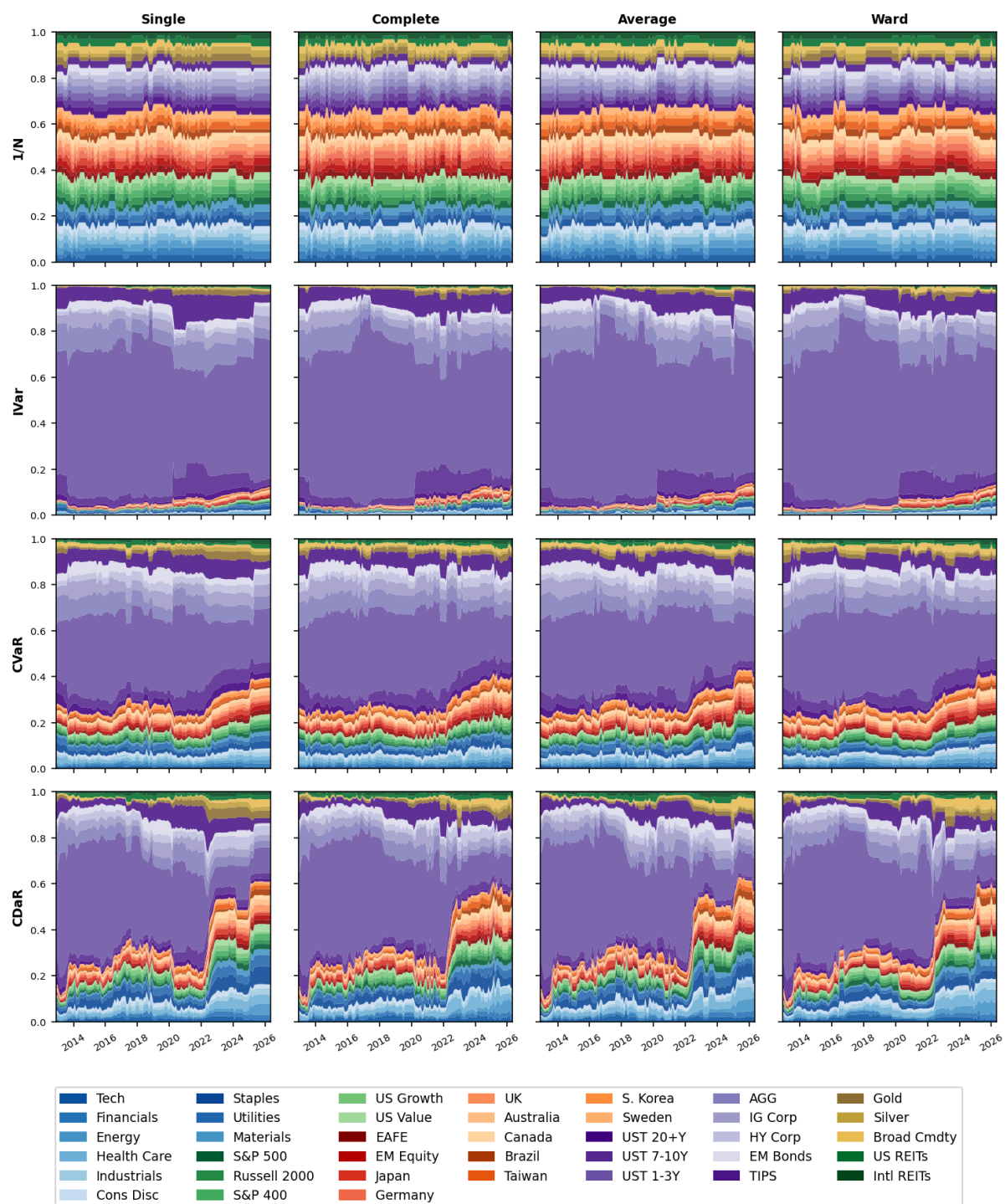


Figure C2: Rolling portfolio weights in the full-universe walk-forward backtest. Each panel shows the portfolio weights through time for one combination of allocation kernel and linkage method, with kernels shown by row and linkage methods by column.

Source: Author's calculations.

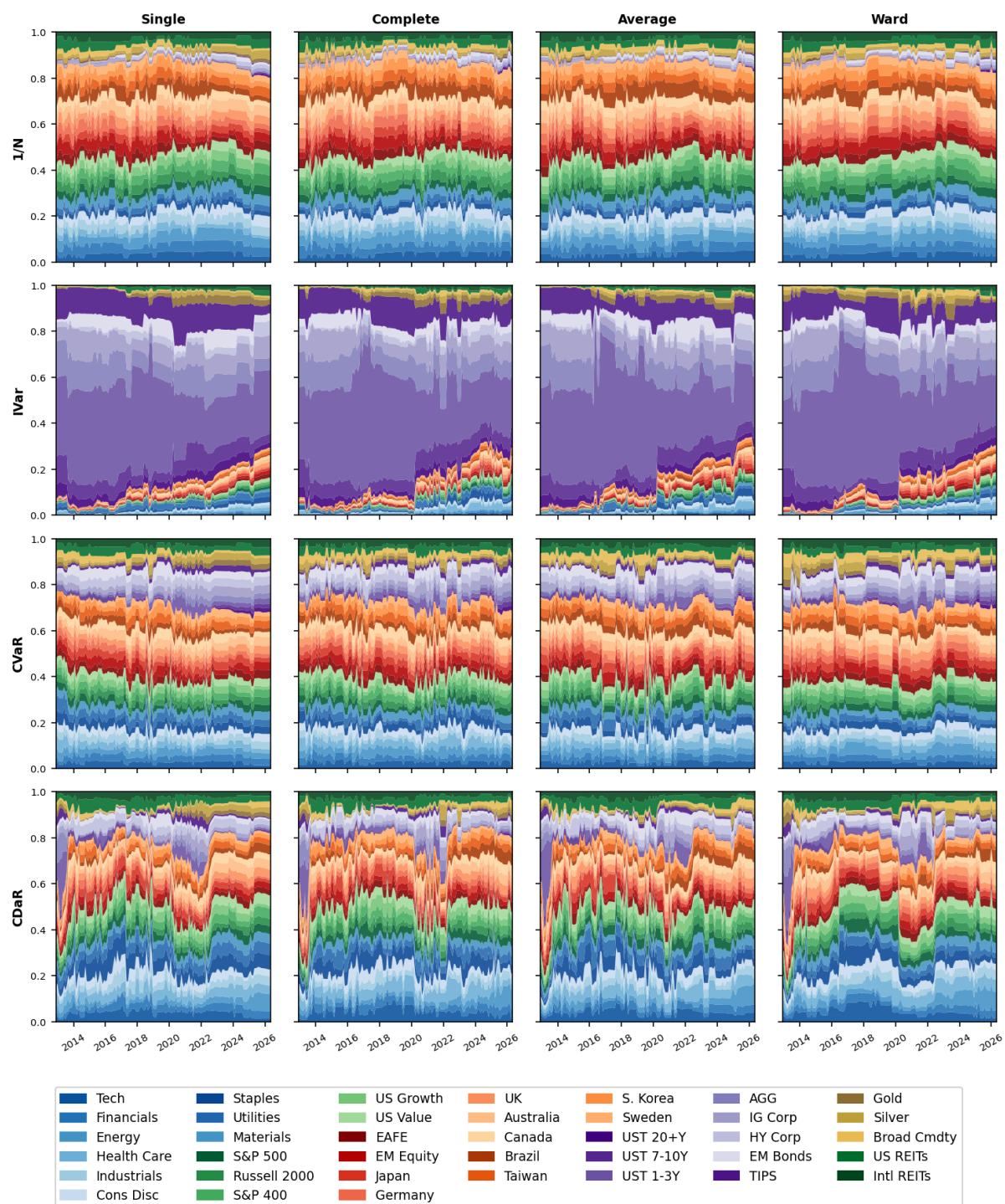


Figure C3: Rolling portfolio volatility contributions in the full-universe walk-forward backtest. Each panel shows the asset volatility contributions through time for one combination of allocation kernel and linkage method, with kernels shown by row and linkage methods by column.

Source: Author's calculations.

Appendix D: Statement on the use of generative AI

ChatGPT and Claude were used as support tools during the writing process. These AI tools were used to improve academic tone, clarity and structure, as well as to assist with wording, synonyms and L^AT_EX implementation. AI-tools have also been used for practical help with implementation and debugging in Python and to brainstorm possible ways of presenting the results visually. Regarding language assistance and academic wording, AI-tools have been used to support across the whole thesis. All key academic decisions, research design, data collection, result interpretation, and conclusions are the work of the author. The author takes full responsibility for the content of the thesis.