Annali del Dipartimento di Metodi e Modelli per l'Economia il Territorio e la Finanza
2026, DOI: 10.13133/2611-6634/1875

E-ISSN 2611-6634
ISSN 2385-0825

Research paper

First published online: April 2, 2026

Alessandra Carleo* **Maria Alessandra Congedo*** **Andrea Gheno*** **Carlo Domenico Mottura*** **Jacopo Maria Ricci*°**

An Empirical Evaluation of Hierarchical Clustering-Based Portfolio Selection Models

Abstract

This paper investigates the empirical performance of hierarchical clustering-based asset allocation models. These models use clustering techniques to inform portfolio construction without relying on traditional optimization.
We evaluate several variants of these methods and compare their out-of-sample performance with standard portfolio selection strategies. Our results suggest that hierarchical clustering models offer competitive and often superior risk-adjusted returns, demonstrating robustness across different market conditions and asset universes.

Keywords: Hierarchical Clustering, Portfolio Selection, Return Correlation Matrix, Portfolio Optimization

*Department of Business Studies, Roma Tre University, Rome, Italy.
°(Corresponding author)

1 Introduction

Portfolio selection is a fundamental task in quantitative finance, aiming at constructing portfolios that balance risk and return. The seminal work of Markowitz (1952) introduced the mean-variance optimization framework, which laid the foundation for modern portfolio theory. However, several works documented the high sensitivity of the mean-variance model to input estimates of expected returns and covariances, often resulting in portfolios that are unstable and poorly diversified out-of-sample (Jobson and Korkie, 1981; Michaud, 1989; Best and Grauer, 1991; Kondor et al., 2007; Cesarone et al., 2020). To address these limitations, researchers have explored a number of alternative strategies, including shrinkage estimators (Ledoit and Wolf, 2004), robust optimization (Fabozzi et al., 2007), and non-parametric methods such as equal-weighting (DeMiguel et al., 2009). Among these, a significative portion of literature has focused on the use of hierarchical structures derived from asset return correlations to improve portfolio robustness. More recently, asset allocation methods based on hierarchical clustering have gained attention for their ability to incorporate the underlying correlation structure of assets into the portfolio construction process (see e.g., Ioannidis et al. (2023), Wang et al. (2024)). In particular, Raffinot (2018) introduced a family of *hierarchical clustering-based asset allocation* (HCAA) models that apply hierarchical clustering algorithms (e.g., single linkage, average linkage, Ward's method) to the asset correlation matrix to uncover a multilevel structure. Unlike Hierarchical Risk Parity, proposed by Lopez de Prado (2016), which focuses on balancing risk contributions, Raffinot's HCAA methods emphasize clustering structure alone without imposing a specific risk budgeting framework. These models are appealing because of their intuitive structure, low reliance on precise parameter estimation, and relative robustness across different market regimes. In this paper, we present a comprehensive empirical study of hierarchical clustering-based portfolio selection models. We evaluate their out-of-sample performance using a range of financial performance metrics, across four real-world datasets encompassing various asset universes and market regimes. The performance is benchmarked against classical portfolio optimization strategies. This work is structured as follows. In Section 2, we introduce the main concept concerning hierarchical clustering. In Section 3, we report the models analyzed in this work. In Section 4, we provide the setting and the discussion of the results of the empirical analysis. Finally, in Section 5, we draw some conclusions.

2 Hierarchical clustering

Hierarchical clustering is a popular unsupervised learning technique that reveals the nested grouping structure among a set of objects without requiring a predefined number of clusters. In financial applications, this method is particularly useful for uncovering latent structures in the correlation matrix of asset returns, aiding portfolio diversification, systemic risk analysis, and as a foundational step in hierarchical risk-based allocation strategies (Lopez de Prado, 2016; Raffinot, 2018). Let $\mathcal{A} = \{1, \ldots, n\}$ denote the universe of n assets. The dissimilarity between pairs of assets is indicated by $D_{i,j}$, which is typically defined as a transformation of their Pearson correlation $\rho_{i,j}$, such that:

$$D_{i,j} = \sqrt{2(1-\rho_{i,j})} \; . \tag{1}$$

This metric satisfies the properties of a Euclidean distance and induces a dissimilarity matrix $\boldsymbol{D} = (D_{i,j})_{=1}^{n}$ which forms the basis for hierarchical clustering algorithms.
The most commonly used hierarchical method in finance is the agglomerative approach, where each asset initially forms its own singleton cluster. At each iteration, the two most similar clusters are merged according to a linkage criterion, and this process is repeated until a single cluster encompasses all assets. The sequence of merges defines a dendrogram, a binary tree that captures the multiresolution structure of the data. Let $\mathcal{C}_1, \mathcal{C}_1 \in \mathcal{A}$ be two disjoint clusters. The distance between the clusters $\mathcal{C}_1$ and $\mathcal{C}_2$, denoted $D(\mathcal{C}_1, \mathcal{C}_1)$, depends on the linkage method used. In the following, we present the main variants relevant in financial contexts.

The *single linkage* method merges clusters based on the closest pair of elements. Thus, given two clusters $\mathcal{C}_i$ and $\mathcal{C}_j$, we have

$$d^{SL}_{\mathcal{C}_i,\mathcal{C}_j} = \min_{x,y}\{D(x,y) | x \in \mathcal{C}_i, y \in \mathcal{C}_j\} \; .$$

The *complete linkage* (CL) method merges based on the most distant pair, producing compact clusters, i.e., for clusters $\mathcal{C}_i$ and $\mathcal{C}_j$

$$d^{CL}_{\mathcal{C}_i,\mathcal{C}_j} = \max_{x,y}\{D(x,y) | x \in \mathcal{C}_i, y \in \mathcal{C}_j\} \; .$$

The *average linkage* (AL) method averages all distances among the clusters, providing a compromise between single and complete linkage. It is particularly effective in financial settings where cluster robustness is important. When applied to clusters $\mathcal{C}_i$ and $\mathcal{C}_j$ with sizes m_i and m_j, respectively, we obtain:

$$d^{AL}_{\mathcal{C}_i,\mathcal{C}_j} = \frac{1}{m_i \cdot m_j} \sum_{i=1}^{m_i} \sum_{i=j}^{m_j} D_{i,j} \; .$$

The *centroid linkage* (CeL) computes the distance between cluster centroids directly. It may result in non-monotonic dendrograms and is less commonly used in financial applications. Thus, for two clusters $\mathcal{C}_i$ and $\mathcal{C}_j$:

$$d^{CeL}_{\mathcal{C}_i,\mathcal{C}_j} = \left\| \mu_{\mathcal{C}_i} - \mu_{\mathcal{C}_j} \right\| \; ,$$

where $\mu_{\mathcal{C}_i}$ and $\mu_{\mathcal{C}_j}$ are the centroids of clusters $\mathcal{C}_i$ and $\mathcal{C}_j$, respectively.

Finally, *Ward's method* (WM, Ward Jr (1963)) minimizes the total within-cluster variance after merging. Though typically defined for Euclidean distances, it can be adapted to correlation-based distances via suitable transformations. Therefore, this method applied to clusters $\mathcal{C}_i$ and $\mathcal{C}_j$ is:

$$d^{WM}_{\mathcal{C}_i,\mathcal{C}_j} = \frac{m_i \cdot m_j}{m_i + m_j} \left\| \mu_{\mathcal{C}_i} - \mu_{\mathcal{C}_j} \right\| \; .$$

After establishing the hierarchical relationships among assets using one of the clustering methods described above, a key step is determining the optimal number of clusters into which the asset universe should be partitioned. Here we adopt an approach inspired by Raffinot (2018), and use the Gap Statistic (Tibshirani et al., 2001) to objectively estimate the most appropriate number of clusters. We choose the Gap Statistic, since it is a robust and general-purpose method, that quantitatively compares the clustering structure of the observed data to that of reference null models. Specifically, the Gap Statistic estimates the number of clusters K by measuring the difference between the within-cluster dispersion of the observed data and that expected under a reference (null) distribution with no apparent cluster structure. For a more detailed explanation of the methodology, we refer the reader to Tibshirani et al. (2001).

3 Portfolio selection models

In this section, we introduce the portfolio selection models analyzed in this study. We begin by establishing the notation used. Let p_{it} denote the price of asset i at time t, with $i = 1, \dots, n$ and $t = 1, \dots, T$, and define the corresponding linear return as $r_{it} = \frac{p_{it} - p_{i(t-1)}}{p_{i(t-1)}}$. Denote the vector of portfolio weights by $x = \{x_1, \dots, x_n\}$, so that the portfolio return at time t is given by $R_t(x) = \sum_{i=1}^{n} r_{it} x_i$. Furthermore, let $\mu = \{\mu_1, \dots, \mu_n\}$ represent the vector of expected returns, and let Σ denote the covariance matrix of asset returns, with entries σ_{ij} representing the covariance between the returns of assets i and j. The portfolio selection models evaluated in this work are presented below.

3.1 Hierarchical asset allocation

To determine portfolio weights, we employ the Hierarchical Clustering Weighting (HCW) method proposed by Raffinot (2018). This approach consists in building a hierarchical clustering tree (dendrogram) based on a pairwise distance matrix of asset returns, typically using correlation-based distances. The assets are organized into a binary tree structure using one of the linkage criteria described in Section 2. Then, Portfolio weights are assigned through a recursive top-down procedure: starting with total capital allocated at the root node, the weight is equally split between each pair of child clusters at every internal node. This process continues until the leaf nodes, corresponding to individual assets, are reached. The final weight of each asset is the product of the splits along the path from the root to the respective leaf. The final allocation reflects the hierarchical structure of asset similarities, without the use of optimization. More in detail, once the clusters have been determined with one of the method explained in Section 2, the weight of the ith asset is computed as follows:

$$x_i = w_{\mathcal{C}_i} \cdot \frac{1}{m_i} ,$$

where $w_{\mathcal{C}_i}$ is the weight assigned to cluster $\mathcal{C}_i$. Therefore, within each cluster, an equal weight is assigned to each asset.
For ease of interpretation, the clustering-based strategies analyzed in this work are also discussed in terms of investor risk profiles: conservative (high risk-aversion, HR), balanced

(low but positive risk-aversion, LR), and return-seeking (no risk-aversion, NR).

3.2 Minimum Variance portfolio

The minimum variance (minV) model, introduced by Markowitz (1952), aims at finding the portfolio that minimizes the portfolio variance, i.e.:

$$\sigma_P^2(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j. \tag{2}$$

Therefore, the corresponding portfolio selection model is the following

$$\begin{cases} \min\limits_{x} & \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} \sigma_{ij} x_i x_j \\ \text{s.t.} & \\ & \sum\limits_{i=1}^{n} x_i = 1 \\ & x_i \geq 0 \qquad\qquad i = 1, \ldots, n \end{cases}$$

3.3 Equal Risk Contribution portfolio

The Equal Risk Contribution (ERC) model, introduced by Maillard et al. (2010), seeks a portfolio in which all assets contribute equally to the overall portfolio risk. The risk is measured by the standard deviation, i.e. $\sigma_P(x) = \sqrt{\sigma_P^2(x)}$, where $\sigma_P^2(x)$ is defined in formula (2). Given the definition of risk contribution, i.e.,

$$RC_i(x) = x_i \frac{\partial \sigma_P x}{\partial x_i} = x_i (\Sigma x)_i \ , \tag{3}$$

the ERC portfolio is characterized by the following condition:

$$RC_i = RC_j \quad \forall i, j \ .$$

3.4 Most Diversified Portfolio

The Most Diversified Portfolio (MDP), introduced by Choueifaty and Coignard (2008), maximizes the diversification ratio, which is is given by

$$DR(x) = \frac{x^T \sigma}{\sqrt{x^T \Sigma x}} \ ,$$

where $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, i.e., the vector of the standard deviations of the assets. Here, we use the reformulation of Choueifaty et al. (2013), i.e.:

$$\begin{cases} \min\limits_{x} & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}y_iy_j \\ \text{s.t.} & \\ & \sum_{i=1}^{n} y_i\sigma_i = 1 \\ & y_i \geq 0 \qquad\qquad i = 1,\ldots,n, \end{cases}$$

and, then, $x_i = \frac{y_i}{\sum_{i=1}^{n} y_i}$.

3.5 Equally Weighted Portfolio

The Equally Weighted (EW) portfolio assigns identical weights to all assets in the portfolio, that is,

$$x_i = \frac{1}{n} \quad \forall i \ .$$

This approach does not require any statistical input and is entirely free of estimation error.

4 Empirical Analysis

In this section, we test the performance of the models described in Section 3 on three real-world datasets, described in Section 4.1, where the experimental setup of the analysis is defined, as well. Then, Section 4.2 deals with the description of the performance measures used to compare the models. Finally, in Section 4.4, we report the results.
All the models have been implemented in Matlab 23.2 on a workstation with Intel(R) Xeon(R) CPU E5-2623 v4 (2.6 GHz, 64 Gb RAM) under MS Windows 10 Pro.

4.1 Datasets and Experimental Setup

Here, we describe the data sets employed in our analysis. Each data set comprises daily return series derived from daily closing prices, which have been adjusted for dividends and stock splits. All data were sourced from LSEG Data & Analytics (formerly known as Refinitiv). The specific data sets utilized are as follows:

Table 1. List of the daily datasets used.

Index	Abbreviation	Country	# Assets	From - To
Dow Jones Industrial Average	DJIA	USA	27	January 2016–May 2024
Euro Stoxx 50	ESTX	EU	47	January 2016–May 2024
FTSE 100	FTSE	UK	78	January 2016–May 2024
NASDAQ-100	NASDAQ	USA	67	January 2016–May 2024

Source: authors' elaboration.

We adopt a rolling window evaluation framework that allows for periodic rebalancing of the portfolio during the investment horizon. We carry out the analysis with two in-sample window lengths (learning periods ΔL), consisting of 250 and 500 daily observations. At each step, the in-sample window is advanced by one month, and the optimal portfolio is re-estimated based on the updated data. Then, the resulting portfolio is evaluated over the subsequent one-month out-of-sample period. This process is repeated until the end of the return time series.

4.2 Performance Measures

In this section, we describe the performance measures used in this work to evaluate and compare the out-of-sample performances of the models analyzed. First of all, let R^{out} be the out-of-sample portfolio return.

- μ^{out} is the average of daily out-of-sample returns, i.e. $\mu^{out} = E[R^{out}]$. Higher values reflect superior performance.
- $\sigma^{out} := \sqrt{E[(R^{out} - \mu^{out})^2]}$ measures the total volatility of the portfolio returns. As a risk indicator, lower values are preferred.
- Maximum Drawdown (MaxDD): represents the greatest observed loss from a peak to a trough in cumulative returns. Given the portfolio wealth at time t, i.e., $W_t = W_{t-1}(1 + R^{out})$ the drawdown at time t is defined as:
$$dd_t = \frac{W_t - \max_{1\leq\tau\leq T} W_\tau}{\max_{1\leq\tau\leq T} W_\tau}, \quad t = 1, \dots, T \ .$$
Then, the Maximum Drawdown is MaxDD $= \min_{1\leq t\leq T} dd_t$. This metric highlights the portfolio's worst-case performance over time.
- Ulcer index (UI): quantifies the depth and duration of drawdowns, providing a smoother view of downside volatility, i.e.:
$$UI = \sqrt{\frac{\sum_{t=1}^{T} dd_t^2}{T}} \ .$$
A lower Ulcer index implies more stable performance.
- Sharpe ratio (SR): assesses the excess return per unit of total risk, and it is defined as
$$SR = \frac{\mu^{out} - r_f}{\sigma^{out}} \ .$$
In our analysis, the risk-free rate r_f is set to zero.
- Sortino ratio (SoR): a downside risk-adjusted version of the Sharpe ratio, where only negative deviations are considered, namely the target downside deviation, where $TDD^{out} = \sqrt{E[(\min\{R^{out} - r_f, 0\})^2]}$. Thus, we have:
$$SoR = \frac{\mu^{out} - r_f}{TDD^{out}} \ ,$$
$r_f = 0$. It is particularly suitable for investors more concerned with losses than with overall volatility.

- Rachev ratio: compares expected returns in the right tail (gains) with those in the left tail (losses), providing insight into asymmetry in risk-reward profiles. It is computed via CVaR at fixed confidence levels, i.e.:

$$Rachev = \frac{CVaR_\alpha(r_f - R^{out})}{CVaR_\beta(R^{out} - r_f)} ,$$

where we set $\alpha = \beta = 5\%$, and $r_f = 0$.

- Normalized Herfindahl index (NHI): measures portfolio risk concentration. It is the normalized version of the Herfindahl index (HI), which is derived from the squared risk contributions of individual assets, mathematically:

$$HI = \sum_{i=1}^{n} RC_i ,$$

where RC_i is defined in formula (3). Therefore, we have:

$$NHI = \frac{1 - HI}{1 - \frac{1}{n}} .$$

It ranges from 0 (risk evenly spread across all the assets) to 1 (risk concentrated in one asset). Finally, we compute the average NHI over the number of rebalances Q, i.e.:

$$E[NHI] = \frac{1}{Q}\sum_{q=1}^{Q} NHI_q ,$$

- Turnover: defined as the average change in asset weights between rebalancing periods. It acts as a proxy for trading activity and transaction costs. Mathematically, it is computed as:

$$Turnover = \frac{1}{Q}\sum_{q=1}^{Q}\sum_{i=1}^{n}|x_{q,i} - x_{q-1,i}|$$

4.3 Statistical Inference

For each strategy, using the results based on the 250-day in-sample window, we assess the statistical uncertainty of the performance measures by computing 5000 circular block bootstrap replications of the out-of-sample series, using a block length $B = \sqrt{T}$ to account for serial dependence. For every metric (except for the Turnover), we report the bootstrap point estimate (Estimate) together with the corresponding 95% percentile confidence interval (CI High and CI Low). This procedure provides robust inference for both moment-based and path-dependent statistics, without relying on analytical standard errors.

4.4 Results

In this section, we report the results for the portfolio selection models described in Section 3. Tables from 2 to 9, show the empirical results of the metrics employed in this work, while Tables from 11 to 14 display the results concerning the robustness analysis. On the other hand, Figs. from 1 to 4 display the time evolution of portfolio wealth of the models used.

4.4.1 DJIA

Descriptive Results Here, we discuss the results of the DJIA data set, which are reported in Tables 2 ($\Delta L = 250$) and 3 ($\Delta L = 500$). Concerning $\Delta L = 250$, the EW portfolio achieves the highest mean return (0.00048), outperforming all other models, however, at the cost of the highest standard deviation (0.01117). In contrast, the MinV model delivers the lowest volatility (0.00921), as expected, but also yields the lowest mean return (0.00026). In terms of risk-adjusted performance, CeL stands out, achieving the highest Sharpe Ratio (0.04528) and Sortino Ratio (0.06364). Both MDP and EW portfolios perform well on these metrics, although with slightly lower Sharpe Ratios. When examining tail risk through the Maximum Drawdown and Ulcer Index, MinV again emerges favorably, with the smallest Maximum Drawdown (-0.27862) and a relatively low Ulcer Index. In contrast, AL and WM experience the largest Maximum Drawdowns (-0.35913 and -0.36256, respectively). The Rachev Ratio is highest for CeL (0.93088), closely followed by CL (0.91257) and EW (0.91314). From a diversification point of view, the Normalized Herfindahl index is highest for EW (0.99562). Among the optimized strategies, HCL and CeL show high values (0.97306 and 0.93729), whereas MinV exhibits the least diversification (0.86312). Finally, considering portfolio turnover, which affects transaction costs, CeL and SL achieve relatively low values (0.17020 and 0.21927), while CL yields the highest turnover (0.27180), suggesting more frequent rebalancing requirements.

With a longer in-sample window ($\Delta L = 500$), the main performance patterns observed for $\Delta L = 250$ remain stable. EW continues to deliver the highest mean return (0.00042) together with high volatility, while MinV again shows the lowest risk (0.00977) but also one of the lowest mean returns. CeL preserves its strong risk-adjusted profile, achieving the highest Sharpe (0.03261) and Sortino (0.04582) ratios, followed by ERC and EW, in line with the previous results. Tail-risk behaviour is also consistent: MinV maintains relatively small drawdowns, whereas AL, CL, and WM exhibit deeper ones. CeL and CL continue to show high Rachev ratios, while diversification ranks remain similar, with EW nearly fully diversified and MinV the most concentrated. Turnover patterns are likewise unchanged, with CeL and SL remaining among the most stable and CL requiring more frequent rebalancing. Overall, the $\Delta L = 500$ window confirms the robustness of the earlier findings.

Statistical Robustness Analysis The bootstrap analysis confirms the robustness of most patterns observed in the descriptive results. Mean returns for the top-performing strategies (HCL, HCeL, ERC, MDP and EW) remain significantly above zero, with confidence intervals that do not include negative values, whereas MinV shows a wide interval extending below zero, reinforcing its weaker return profile. Differences in volatility across models are also statistically meaningful: MinV retains the lowest standard deviation with a tight confidence band, while EW and WM consistently display the highest risk levels. Risk-adjusted metrics broadly confirm the superiority of HCeL, whose Sharpe and Sortino ratios exhibit the highest point estimates and confidence intervals that lie above those of most competing strategies. Regarding downside risk, confidence intervals for Maximum Drawdown consistently identify MinV as the most resilient strategy and WM as one of the weakest, in line with the descriptive evidence. Tail-risk behaviour measured through the Rachev Ratio does not reveal statistically significant differences across clustering-based methods: most intervals overlap considerably, indicating similar downside–upside asymmetry. Finally, diversification levels (as reflected by the Herfindahl index) exhibit negligible sampling variability for all strategies, confirming that the relative ranking of models in terms of concentration is stable and not driven by noise.

Table 2. Out-of-sample performance results for the DJIA dataset, with $\Delta L = 250$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0,00040	0,00045	0,00041	0.00045	0.00043	0.00026	0.00044	0.00045	0.00048
σ^{out}	0.01051	0.01077	0.01057	0.01005	0.01089	0.00921	0.01046	0.01054	0.01117
MaxDD	-0.33255	-0.34762	-0.35913	-0.33332	-0.36256	-0.27862	-0.32027	-0.33882	-0.33381
UI	0.06525	0.06151	0.05925	0.05642	0.07158	0.07070	0.05989	0.05968	0.06538
SR	0.03831	0.04218	0.03845	0.04528	0.03952	0.02818	0.04238	0.04308	0.04299
SoR	0.05353	0.05901	0.05357	0.06364	0.05490	0.03933	0.05923	0.05983	0.06037
Rachev	0.92051	0.91257	0.90990	0.93088	0.90598	0.92173	0.90305	0.89860	0.91314
NHI	0.94039	0.97306	0.95581	0.93729	0.97772	0.86312	-	0.92749	0.99562
Turnover	0.21927	0.27180	0.23613	0.17020	0.21900	0.23890	0.03480	0.23100	-

Source: authors' elaboration.

4.4.2 ESTX

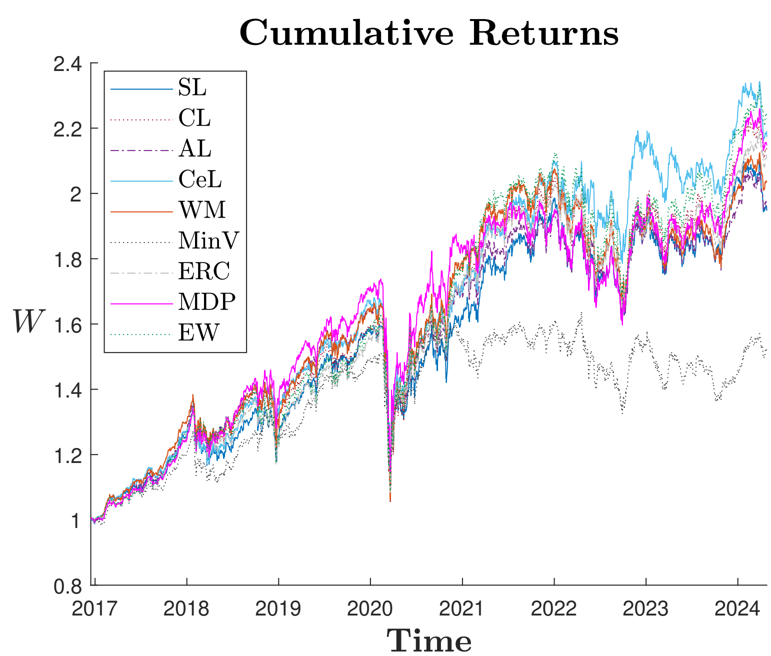
Descriptive Results Here, we deal with the computational results concerning the ESTX dataset, shown in Tables 4 ($\Delta L = 250$) and 5 ($\Delta L = 500$). Regarding $\Delta L = 250$, MDP exhibits the highest mean return (0.00047), slightly outperforming CeL and EW. Meanwhile, SL records the lowest average return (0.00030). In terms of risk, measured by standard deviation, MinV leads with the lowest volatility (0.00839), consistent with its design. On the other end, EW displays the highest volatility (0.01176). The CeL model stands out with the best balance of high return and low volatility, achieving strong Sharpe and Sortino Ratios. These metrics are slightly exceeded by MDP (Sharpe: 0.04993, Sortino: 0.06798), placing both models as top performers in risk-adjusted terms. From a downside risk per-

Table 3. Out-of-sample performance results for the DJIA dataset, with $\Delta L = 500$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0.00022	0.00036	0.00032	0.00033	0.00034	0.00011	0.00038	0.00032	0.00042
σ^{out}	0.01052	0.01117	0.01083	0.01013	0.01114	0.00977	0.01117	0.01091	0.01191
MaxDD	-0.28752	-0.35792	-0.29659	-0.25176	-0.30080	-0.30524	-0.32264	-0.33632	-0.33381
UI	0.08539	0.07298	0.07645	0.06870	0.07804	0.11101	0.06579	0.08324	0.07013
SR	0.02130	0.03192	0.02962	0.03261	0.03086	0.01109	0.03365	0.02915	0.03548
SoR	0.02961	0.04356	0.04134	0.04582	0.04301	0.01503	0.04686	0.03972	0.04972
Rachev	0.91651	0.87174	0.91217	0.92060	0.90788	0.88443	0.89868	0.86362	0.90914
NHI	0.94454	0.97602	0.96142	0.94228	0.98026	0.86076	-	0.93114	0.99627
Turnover	0.15445	0.20305	0.14050	0.08756	0.17674	0.13830	0.01756	0.13687	-

Source: authors' elaboration.

Figure 1. Cumulative portfolio wealth for the DJIA dataset, with $\Delta L = 250$ days.



Source: authors' elaboration.

spective, CeL and MinV, once again, show a strong performance, with the lowest Ulcer index values (0.06069 and 0.07391, respectively), and Maximum Drawdowns of -0.30704 and -0.31894, respectively. EW has the worst Maximum Drawdown (-0.40587) and the highest Ulcer Index (0.09120). Looking at tail risk via the Rachev Ratio, EW leads with a value of 0.91540, followed closely by CL, CeL, and ERC. In terms of portfolio diversification, the NHI is the highest for EW (0.99744), followed by WM (0.98837) and CL (0.97295). The MinV model records the lowest NHI (0.89179). Finally, Turnover is lowest for ERC (0.03688), making it potentially appealing for cost-sensitive portfolios. In contrast, CL and WM have higher turnover values (0.29787 and 0.25432, respectively). Using a longer in-sample window ($\Delta L = 500$) yields results broadly consistent with the previous findings. Mean returns remain highest for MDP (0.00044) and EW (0.00039), although differences across models slightly shrink; SL again delivers the weakest performance. As before, MinV achieves the lowest volatility, while EW remains the most volatile. CeL and MDP continue to rank among the strongest performers in risk-adjusted terms, maintaining competitive Sharpe and Sortino ratios. Regarding downside risk, CeL and MinV again show favorable behavior, with relatively low Ulcer Index values and comparatively mild drawdowns, while EW still records the deepest Maximum Drawdown (-0.40587) and the highest Ulcer Index (0.09676). Tail-risk patterns are also similar: EW achieves the highest Rachev Ratio (0.90090), with CL, CeL, and WM close behind. Finally, in terms of diversification, EW and WM remain the most diversified according to the NHI, whereas MinV stays the least diversified; turnover rankings remain stable, with EW showing minimal rebalancing intensity.

Statistical Robustness Analysis The bootstrap results broadly confirm the descriptive patterns observed in the ESTX dataset, while also highlighting meaningful uncertainty across performance metrics. The superior risk-adjusted behaviour of CeL and MDP remains visible in the resampled distributions, yet the wide confidence intervals for Sharpe and Sortino ratios indicate that their advantage over the remaining models is not statistically tight. MinV shows the most stable behaviour, with consistently low volatility and drawdown across resamples, reinforcing its defensive profile described earlier. By contrast, EW, WM, and CL exhibit larger dispersion in tail-risk measures and volatility, in line with their higher drawdowns and turnover reported in the descriptive section. Overall, the robustness analysis suggests that although point estimates identify CeL and MDP as leading strategies in the ESTX dataset, the differences across models largely overlap within confidence bands, implying that performance rankings should be interpreted with caution.

4.4.3 FTSE

Descriptive Results In Tables 6 ($\Delta L = 250$) and 7 ($\Delta L = 500$), we report the empirical results regarding the FTSE 100 dataset. As for the shorter in-sample window ($\Delta L = 250$), in terms of mean return, the CL strategy achieves the highest value (0.00033), closely followed by SL and EW (both at 0.00030). In contrast, MinV generates the lowest return (0.00015). As for volatility, MinV offers the most stable performance with the lowest standard deviation (0.00751), while EW is, again, the most volatile (0.01073). When consid-

Table 4. Out-of-sample performance results for the ESTX dataset, with $\Delta L = 250$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0.00030	0.00037	0.00032	0.00043	0.00040	0.00037	0.00039	0.00047	0.00042
σ^{out}	0.00934	0.00971	0.00938	0.00890	0.01050	0.00839	0.01071	0.00935	0.01176
MaxDD	-0.33872	-0.35709	-0.34874	-0.30704	-0.37091	-0.31894	-0.38622	-0.32980	-0.40587
UI	0.07718	0.07313	0.08118	0.06069	0.07388	0.07391	0.08366	0.06545	0.09120
SR	0.03194	0.03766	0.03432	0.04864	0.03831	0.04423	0.03655	0.04993	0.03546
SoR	0.04352	0.05160	0.04634	0.06685	0.05246	0.05984	0.05009	0.06798	0.04903
Rachev	0.87703	0.90210	0.86822	0.90178	0.89609	0.87768	0.89929	0.87908	0.91540
NHI	0.92316	0.97295	0.94338	0.92417	0.98837	0.89179	-	0.92073	0.99744
Turnover	0.17275	0.29787	0.24306	0.16298	0.25432	0.25737	0.03688	0.25677	-

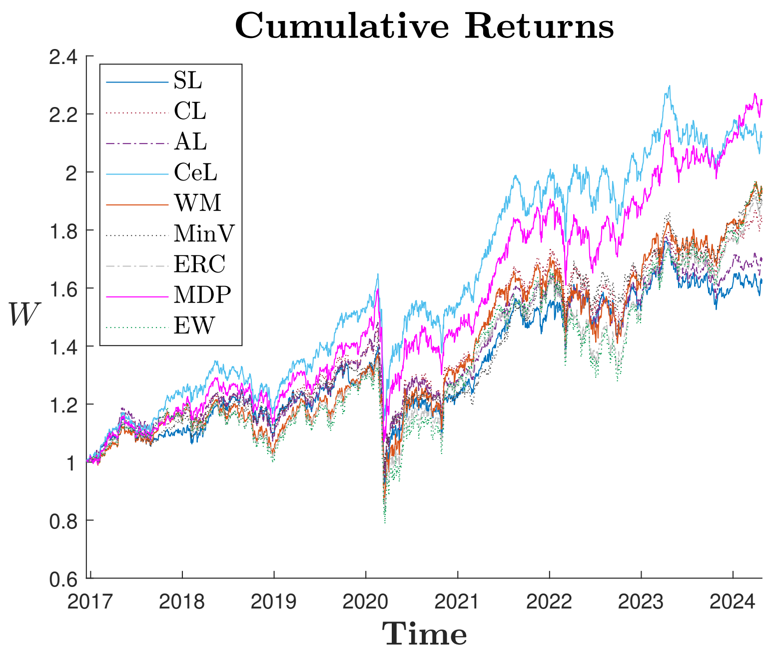
Source: authors' elaboration.

Table 5. Out-of-sample performance results for the ESTX dataset, with $\Delta L = 500$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0.00020	0.00028	0.00026	0.00026	0.00034	0.00037	0.00036	0.00044	0.00039
σ^{out}	0.00963	0.01024	0.00982	0.00943	0.01119	0.00879	0.01129	0.00965	0.01241
MaxDD	-0.34226	-0.37045	-0.34608	-0.31199	-0.37878	-0.31567	-0.38365	-0.32725	-0.40587
UI	0.08598	0.08245	0.07586	0.07806	0.08270	0.07418	0.08773	0.06624	0.09767
SR	0.02128	0.02740	0.02616	0.02734	0.03025	0.04183	0.03222	0.04589	0.03124
SoR	0.02856	0.03708	0.03569	0.03703	0.04124	0.05616	0.04401	0.06232	0.04299
Rachev	0.86136	0.87897	0.88130	0.88119	0.88199	0.85437	0.88580	0.86336	0.90090
NHI	0.92687	0.96862	0.94234	0.92707	0.98838	0.88960	-	0.92186	0.99784
Turnover	0.11486	0.17723	0.11109	0.07791	0.18419	0.13354	0.01861	0.13732	-

Source: authors' elaboration.

Figure 2. Cumulative portfolio wealth for the ESTX dataset, with $\Delta L = 250$ days.



Source: authors' elaboration.

ering risk-adjusted returns, CL and AL stand out, yielding the highest Sharpe Ratios and strong Sortino Ratios. Conversely, MinV shows the lowest risk-adjusted performance with Sharpe and Sortino ratios of 0.01935 and 0.02608, respectively.

As for drawdown risk, assessed by Maximum Drawdown and Ulcer index, MinV achieves the highest Maximum Drawdown (-0.27535), significantly better than EW and SL. However, the Ulcer index for MinV (0.10949) is relatively high. CL and AL balance both metrics effectively, with moderate Maximum Drawdowns and the lowest Ulcer indices (0.07136 and 0.07620, respectively). The Rachev ratio is highest for EW (0.94674), followed by WM (0.91464) and AL (0.90820). Regarding diversification, the Normalized Herfindahl Index is highest for EW (0.99793), reflecting a well-balanced asset distribution. Among the optimized portfolios, WM and CL also demonstrate high diversification levels (0.98995 and 0.97335, respectively), while other strategies such as MinV and CeL show lower NHI values, indicating more concentrated allocations. Finally, in terms of portfolio Turnover, ERC has the lowest value (0.04930), highlighting its practical advantage in minimizing transaction costs. In contrast, CL and AL have the highest turnover rates (0.43757 and 0.37006), which affect net performance depending on trading costs.

Using a longer in-sample window ($\Delta L = 500$) leads to patterns broadly consistent with those observed for $\Delta L = 250$. Mean returns remain highest for CL and EW, with SL close behind (0.00026), while MinV again yields the lowest average return (0.00015). Volatility rankings are stable: MinV continues to be the least volatile (0.00751) and EW the most volatile (0.01073). In terms of risk-adjusted performance, CL again leads with the highest

Sharpe and Sortino ratio (0.05133), followed closely by AL, whereas MinV remains the weakest performer on both metrics. Regarding drawdown risk, MinV continues to exhibit comparatively mild Maximum Drawdowns, although its Ulcer Index remains elevated; CL and AL again provide a favorable balance, with moderate drawdowns and the lowest Ulcer indices. Tail-risk patterns remain similar as well, with EW achieving the highest Rachev Ratio and WM and AL following closely. Diversification levels also align with previous evidence: EW and WM remain the most diversified, while MinV and CeL stay the most concentrated. Turnover results preserve the earlier structure, with ERC again achieving the lowest turnover and CL and AL remaining among the most transaction-intensive strategies.

Table 6. Out-of-sample performance results for the FTSE 100 dataset, with $\Delta L = 250$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0.00030	0.00033	0.00032	0.00024	0.00027	0.00015	0.00026	0.00022	0.00030
σ^{out}	0.00934	0.00885	0.00871	0.00890	0.00969	0.00751	0.00949	0.00832	0.01073
MaxDD	-0.36893	-0.32682	-0.32720	-0.35076	-0.34989	-0.27535	-0.35203	-0.32757	-0.37683
UI	0.10432	0.07136	0.07620	0.10768	0.08302	0.10949	0.07983	0.10228	0.08754
SR	0.03263	0.03734	0.03669	0.02702	0.02769	0.01935	0.02768	0.02632	0.02764
SoR	0.04439	0.05133	0.05045	0.03651	0.03818	0.02608	0.03805	0.03532	0.03863
Rachev	0.89000	0.90108	0.90820	0.86124	0.91464	0.87422	0.90810	0.85755	0.94674
NHI	0.90944	0.97335	0.93263	0.90623	0.98995	0.91007	-	0.94123	0.99793
Turnover	0.30746	0.43757	0.37006	0.29696	0.26073	0.30739	0.04930	0.29307	-

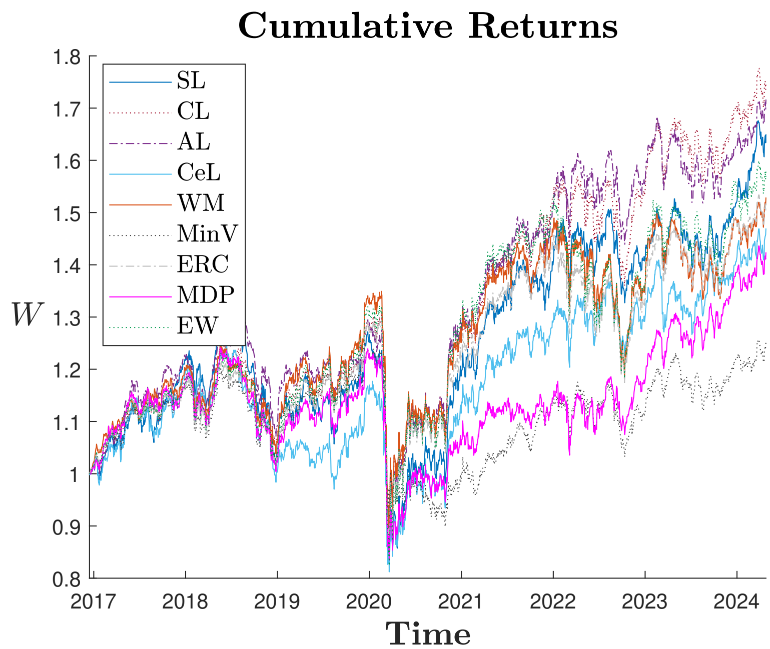
Source: authors' elaboration.

Table 7. Out-of-sample performance results for the FTSE 100 dataset, with $\Delta L = 500$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0.00026	0.00026	0.00020	0.00017	0.00019	0.00002	0.00021	0.00017	0.00026
σ^{out}	0.00943	0.00934	0.00906	0.00907	0.01014	0.00805	0.01012	0.00875	0.01137
MaxDD	-0.31412	-0.31294	-0.29806	-0.42459	-0.36541	-0.30942	-0.35481	-0.34033	-0.37683
UI	0.08663	0.07509	0.08064	0.16786	0.10252	0.14155	0.08767	0.09608	0.09394
SR	0.02714	0.02781	0.02248	0.01822	0.01826	0.00299	0.02123	0.01929	0.02314
SoR	0.03747	0.03825	0.03070	0.02471	0.02497	0.00396	0.02921	0.02611	0.03229
Rachev	0.90074	0.91103	0.89861	0.88256	0.89908	0.85578	0.91470	0.88785	0.94063
NHI	0.91680	0.97261	0.94180	0.90994	0.99303	0.92090	1.00000	0.94410	0.99819
Turnover	0.19053	0.34036	0.21562	0.24884	0.19037	0.17784	0.02611	0.16101	-

Source: authors' elaboration.

Figure 3. Cumulative portfolio wealth for the FTSE dataset, with $\Delta L = 250$ days.



Source: authors' elaboration.

Statistical Robustness Analysis The statistical robustness assessment confirms that the main patterns observed in the point estimates are stable. Across all strategies, the bootstrapped confidence intervals for mean returns, volatilities, and drawdown measures remain relatively tight, indicating that the performance differences highlighted in the descriptive analysis are not driven by sampling noise. In particular, CL and AL preserve their superior risk-adjusted performance, with Sharpe and Sortino ratio intervals clearly dominating those of MinV and EW. Similarly, the drawdown metrics remain consistent: MinV maintains the lowest downside risk but with wide Ulcer-index uncertainty, whereas CL and AL continue to offer a balanced and statistically robust profile. Diversification levels, reflected in fixed Herfindahl indices, show no sampling-induced variation, confirming the structural nature of these allocations. Overall, the robustness checks support the reliability of the empirical findings.

4.4.4 NASDAQ

Descriptive Results In Tables 8 ($\Delta L = 250$) and 9 ($\Delta L = 500$), we provide the results of the NASDAQ dataset. Taking $\Delta L = 250$, concerning the mean return, the EW strategy leads with 0.00077, followed by ERC and WM. In contrast, MinV model yields the lowest return (0.00036). On the volatility front, MinV again demonstrates its intended function by achieving the lowest standard deviation (0.00993), while EW exhibits the highest risk (0.01302). Despite this, EW ranks first in Sharpe ratio (0.05896) and Sortino ratio (0.08414), suggesting that its high returns more than compensate for its increased risk.

Among optimized strategies, ERC yields favorable performances, with a Sharpe ratio of 0.05785 and the second-best Sortino Ratio (0.08233), outperforming most other models in risk-adjusted terms. When examining the Maximum Drawdown behavior, the AL model achieves the largest Maximum Drawdown (-0.22876), while MinV and EW suffer the most (-0.31336 and -0.30163, respectively). The Ulcer Index is lowest for CeL (0.04920), followed closely by AL and MDP. EW, despite strong average returns, shows the highest Ulcer Index (0.07719), highlighting its drawdown persistence. Regarding tail risk, measured by the Rachev Ratio, AL (1.02342) and CeL (1.01176) show the most favorable asymmetry toward upside gains, while ERC and EW post the lowest values (0.93956 and 0.93610, respectively), suggesting relatively less attractive tail profiles. In terms of diversification, EW, again, maximizes the NHI at a level 0.99722, indicating an even allocation across assets. Among optimized portfolios, WM (0.98827) and CL (0.95965) also maintain strong diversification. MinV continues to lag in this aspect with the lowest NHI (0.88411), suggesting concentration in low-risk assets. Finally, ERC stands out in terms of portfolio efficiency, exhibiting the lowest turnover (0.04220). In contrast, CL has the highest turnover (0.41690), reflecting more frequent rebalancing that may erode net returns in practice. Using a longer estimation window ($\Delta L = 500$) preserves most of the patterns observed under $\Delta L = 250$. EW continues to deliver the highest mean return (0.00072), followed by ERC and WM, while MinV again records the lowest (0.00027). Volatility rankings remain stable, with MinV as the least risky model (0.01045) and EW as the most volatile (0.01380). In terms of risk-adjusted performance, EW again achieves the highest Sharpe and Sortino ratios, with ERC maintaining strong second-best results among optimized strategies. Maximum Drawdown behavior remains comparable, with MinV showing one of the mildest drawdowns and AL and CeL experiencing relatively larger ones. CeL and AL also retain favorable Ulcer Index values, while EW again exhibits the highest. Tail-risk patterns shift only slightly: AL and CeL continue to show strong upside asymmetry, whereas ERC and EW remain at the lower end of Rachev ratios. Diversification levels follow the same hierarchy as before, with EW and WM being the most diversified and MinV the most concentrated. Finally, turnover results remain consistent: ERC continues to achieve the lowest turnover, whereas CL and WM remain among the more active strategies.

Statistical Robustness Analysis The robustness checks confirm that the main empirical patterns observed for the NASDAQ dataset remain stable under resampling. The confidence intervals for mean returns and volatilities are sufficiently narrow to preserve the relative ranking across strategies, with EW, ERC, and WM consistently achieving higher average returns, while MinV maintains the lowest risk profile. Risk-adjusted performance remains robust as well: EW and ERC continue to display the strongest Sharpe and Sortino ratios, with their confidence intervals largely dominating those of MinV and other lower-return strategies. The behavior of drawdown risk is similarly preserved, as the intervals for Maximum Drawdown and Ulcer Index do not overturn the descriptive conclusions—AL, CeL, and MDP retain favorable Ulcer dynamics, whereas EW continues to show more persistent drawdowns. Tail-risk assessments also remain consistent, with AL and CeL sustaining their advantage in Rachev ratios. Finally, diversification results are structurally robust, as NHI values do not vary across bootstrap samples, confirming that EW, WM, and CL systemati-

Table 8. Out-of-sample performance results for the NASDAQ dataset, with $\Delta L = 250$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0.00049	0.00054	0.00053	0.00053	0.00061	0.00036	0.00067	0.00053	0.00077
σ^{out}	0.01127	0.01110	0.01101	0.01100	0.01130	0.00993	0.01156	0.01088	0.01302
MaxDD	-0.27770	-0.26410	-0.22876	-0.25829	-0.28342	-0.31336	-0.29537	-0.26227	-0.30163
UI	0.05626	0.05680	0.05028	0.04920	0.06155	0.05889	0.05995	0.04939	0.07719
SR	0.04367	0.04847	0.04797	0.04773	0.05434	0.03602	0.05785	0.04834	0.05896
SoR	0.06280	0.06932	0.06922	0.06897	0.07733	0.05046	0.08233	0.06956	0.08414
Rachev	1,00802	0.96966	1,02342	1,01176	0.95059	0.94972	0.93956	1,00931	0.93610
NHI	0.90289	0.95965	0.91680	0.90147	0.98827	0.88411	-	0.92468	0.99722
Turnover	0.27037	0.41690	0.27370	0.23365	0.30302	0.29086	0.04220	0.27759	-

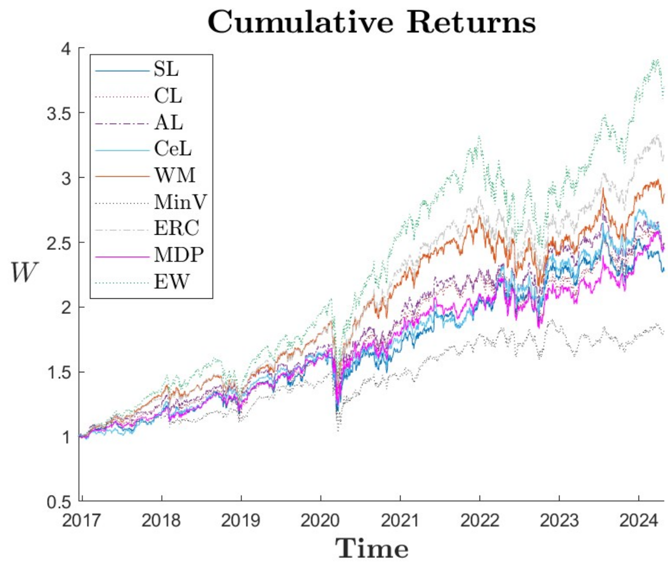
Source: authors' elaboration.

Table 9. Out-of-sample performance results for the NASDAQ dataset, with $\Delta L = 500$ days.

	SL	CL	AL	CeL	WM	MinV	ERC	MDP	EW
μ^{out}	0.00037	0.00046	0.00038	0.00041	0.00048	0.00027	0.00061	0.00039	0.00072
σ^{out}	0.01195	0.01180	0.01184	0.01178	0.01189	0.01045	0.01244	0.01155	0.01380
MaxDD	-0.29811	-0.27355	-0.29156	-0.30187	-0.27513	-0.34573	-0.30394	-0.31130	-0.30163
UI	0.07119	0.05601	0.06332	0.06747	0.06612	0.08635	0.07170	0.07527	0.08275
SR	0.03076	0.03920	0.03193	0.03467	0.04049	0.02567	0.04869	0.03365	0.05184
SoR	0.04352	0.05581	0.04544	0.04857	0.05698	0.03540	0.06912	0.04766	0.07392
Rachev	0.98683	0.97221	0.99246	0.96756	0.92004	0.91858	0.93532	0.98030	0.93215
NHI	0.91660	0.96177	0.93205	0.91075	0.98449	0.88228	-	0.92539	0.99775
Turnover	0.18007	0.25070	0.17964	0.13543	0.19481	0.15338	0.02174	0.15096	-

Source: authors' elaboration.

Figure 4. Cumulative portfolio wealth for the NASDAQ dataset, with $\Delta L = 250$ days.



Source: authors' elaboration.

cally deliver broader asset exposure while MinV remains more concentrated. Overall, the robustness analysis supports the qualitative findings and indicates that the conclusions are not driven by sampling variability.

Table 10. Market features. $\bar{\sigma}$ denotes the average standard deviation, $\bar{\rho}$ denotes the average correlation, and $\lambda_{\%}^{max}$ is the leading eigenvalue share, computed as the ratio between the largest eigenvalue, and the sum of all the eigenvalues of the correlation matrix.

	DJIA	ESTX	FTSE 100	NASDAQ
$\bar{\sigma}$	0.0164	0.0175	0.0186	0.0198
$\bar{\rho}$	0.4097	0.4482	0.3112	0.3890
$\lambda_{\%}^{\max}$	0.4400	0.4710	0.3456	0.4208

Source: authors' elaboration.

4.4.5 Synthesis of Empirical Results

Across all four datasets and both rolling-window configurations, a consistent trade-off emerges between return-seeking and defensive portfolio constructions. Equal-weighted portfolios frequently deliver the highest average returns, but at the cost of elevated volatility, drawdown persistence, and turnover sensitivity, particularly in equity-heavy universes. Conversely, the minimum-variance strategy systematically achieves the lowest risk and drawdown exposure, although with weaker return profiles and more concentrated allocations. Hierarchical clustering–based approaches generally occupy an intermediate and more balanced region of the risk–return spectrum, often improving diversification and stabilizing risk-adjusted performance relative to classical optimized portfolios. These patterns are robust across market regimes and confirmed by bootstrap analyses, indicating that the observed performance differences are structural rather than driven by sampling variability.

5 Conclusions

This study evaluates and compares multiple portfolio selection models in four major equity markets: DJIA, ESTX, FTSE 100, and NASDAQ. The analysis consists of classical risk-return metrics, tail risk measures, diversification indices, and turnover, providing a general view of the strengths and weaknesses of each strategy in varied market environments. Across all datasets, the Equally Weight portfolio consistently achieves high average returns and robust tail performance (via Rachev ratios). However, it does so at the cost of higher volatility and greater drawdown exposure, particularly visible in the FTSE 100 and NASDAQ datasets. The Minimum Variance strategy yields the lowest volatility and maximum drawdown across datasets. Nonetheless, this conservative approach frequently underperforms in terms of average return and risk-adjusted ratios, and its low diversification raises concentration concerns. Risk-adjusted leaders differ slightly across datasets, but strategies based on Hierarchical Clustering, in particular Average linkage, Centroid linkage, and Complete linkage, consistently outperform the other strategies in terms of Sharpe and Sortino ratios, with relatively balanced Maximum Drawdowns and moderate turnover. These models excel particularly in the FTSE 100 and DJIA datasets, combining solid returns with acceptable levels of risk. Among the optimization methods, the Equal Risk Contribution model distinguishes itself as a strong all-rounder, especially in the NASDAQ and DJIA datasets. It combines competitive returns, strong risk-adjusted metrics, low turnover, and superior tail behavior, making it a practical and implementable solution. The Maximum Diversification Portfolio also performs robustly in most scenarios, often showing strong returns and high values of Maximum Drawdown, though its turnover is moderate-to-high depending on the dataset. The Ward's method is more variable, excelling in some instances like NASDAQ but underperforming in others like FTSE 100.

Overall, the robustness analysis confirms that the qualitative rankings of the strategies remain stable across all datasets: the relative advantages of EW, MinV, ERC, and hierarchical clustering-based models persist within the confidence intervals, indicating that the conclusions are not driven by sampling variability. The variability in performance across hierarchical clustering methods can be partly explained by the underlying market characteristics. Markets such as ESTX and DJIA exhibit higher average correlations and larger leading eigenvalue shares, favoring linkage rules (e.g., AL, CL) that emphasize the extraction of dominant comovement structures. In contrast, markets like FTSE 100, characterized by lower average correlation and higher average dispersion, tend to benefit from linkage schemes that better capture heterogeneous cluster sizes and weaker systemic structure. From an investor-profile perspective, the results also suggest distinct suitability across models: risk-averse investors may prefer MinV or CeL (HR) given their lower volatility and controlled drawdown behavior; moderate-risk investors may find ERC and clustering-based strategies attractive (LR) due to balanced risk-adjusted performance and moderate turnover; and return-seeking investors may gravitate toward EW, WM, or MDP (NR), accepting higher volatility in exchange for stronger average returns. Finally, some limitations should be acknowledged. The analysis relies on a single correlation-based distance and may inherit the sensitivity of clustering methods to noisy correlation estimates. Future research could explore alternative distance metrics, statistically validated clustering, network-based methods, or extensions to multi-asset universes to assess the robustness of clustering structures in broader settings.

References

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The Review of Financial Studies*, 4(2):315–342.

Cesarone, F., Mango, F., Mottura, C. D., Ricci, J. M., and Tardella, F. (2020). On the stability of portfolio selection models. *Journal of Empirical Finance*, 59:210–234.

Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management*, 35(1):40.

Choueifaty, Y., Froidure, T., and Reynier, J. (2013). Properties of the most diversified portfolio. *Journal of Investment Strategies*, 2(2):49–70.

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.

Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., and Focardi, S. M. (2007). *Robust Portfolio Optimization and Management*. John Wiley and Sons.

Ioannidis, E., Sarikeisoglou, I., and Angelidis, G. (2023). Portfolio construction: a network approach. *Mathematics*, 11(22):4670.

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance*, pages 889–908.

Kondor, I., Pafka, S., and Nagy, G. (2007). Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking and Finance*, 31(5):1545–1573.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Lopez de Prado, M. (2016). Building diversified portfolios that outperform out-of-sample. *Journal of Portfolio Management*.

Maillard, S., Roncalli, T., and Teiletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management*, 36(4):60.

Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance*, 7(1):71–91.

Michaud, R. O. (1989). The markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45(1):31–42.

Raffinot, T. (2018). Hierarchical clustering-based asset allocation. *Journal of Portfolio Management*, 44(2):89–99.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63(2):411–423.

Wang, G. J., Huai, H., Zhu, Y., Xie, C., and Uddin, G. S. (2024). Portfolio optimization based on network centralities: Which centrality is better for asset selection during global crises? *Journal of Management Science and Engineering*, 9(3):348–375.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Table 11. Robustness analysis for the DJIA dataset.

	SL			CL			AL		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00040	0.00002	0.00074	0.00045	0.00005	0.00082	0.00041	0.00002	0.00076
σ^{out}	0.01051	0.00749	0.01427	0.01077	0.00758	0.01475	0.01057	0.00742	0.01451
MaxDD	0.33255	0.12760	0.51444	0.34762	0.13063	0.52501	0.35913	0.12409	0.53553
UI	0.06526	0.03720	0.25758	0.06153	0.03399	0.25350	0.05927	0.03519	0.26723
SR	0.60812	0.02319	1.37541	0.66958	0.06455	1.46181	0.61044	0.02343	1.39229
SoR	0.84973	0.03217	2.02167	0.93680	0.08704	2.16746	0.85033	0.03165	2.04775
Rachev	0.92435	0.80121	1.07019	0.91647	0.79835	1.06558	0.91388	0.79057	1.06607
NHI	0.04941	0.04941	0.04941	0.02284	0.02284	0.02284	0.03511	0.03511	0.03511
	CeL			WM			MinV		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00045	0.00007	0.00081	0.00043	0.00002	0.00080	0.00026	-0.00008	0.00058
σ^{out}	0.01005	0.00728	0.01362	0.01089	0.00760	0.01499	0.00921	0.00682	0.01231
MaxDD	0.33332	0.11937	0.50057	0.36256	0.13513	0.54990	0.27862	0.13109	0.49466
UI	0.05644	0.03362	0.23573	0.07160	0.03606	0.27644	0.07072	0.04142	0.27009
SR	0.71887	0.09993	1.55741	0.62738	0.02232	1.44687	0.44739	-0.12916	1.17320
SoR	1.01030	0.13383	2.30795	0.87156	0.03044	2.13405	0.62428	-0.17352	1.69603
Rachev	0.93479	0.81594	1.07733	0.90972	0.79242	1.05673	0.92569	0.80089	1.06912
NHI	0.05466	0.05466	0.05466	0.01845	0.01845	0.01845	0.13689	0.13689	0.13689
	ERC			MDP			EW		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00044	0.00006	0.00078	0.00045	0.00005	0.00082	0.00048	0.00008	0.00085
σ^{out}	0.01046	0.00748	0.01420	0.01054	0.00765	0.01418	0.01117	0.00812	0.01499
MaxDD	0.32027	0.13098	0.49029	0.33882	0.12863	0.51589	0.33381	0.13991	0.51355
UI	0.05991	0.03374	0.23211	0.05970	0.03618	0.25215	0.06540	0.03648	0.24675
SR	0.67270	0.08432	1.45654	0.68386	0.06142	1.47423	0.68246	0.09192	1.44627
SoR	0.94026	0.11331	2.14515	0.94984	0.08335	2.17211	0.95842	0.12412	2.13903
Rachev	0.90686	0.78873	1.04945	0.90208	0.78879	1.04020	0.91674	0.80172	1.05794
NHI	0.00338	0.00338	0.00338	0.07704	0.07704	0.07704	0	0	0

Source: authors' elaboration.

Table 12. Robustness analysis for the ESTX dataset.

	SL			CL			AL		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00040	0.00002	0.00074	0.00045	0.00005	0.00082	0.00041	0.00002	0.00076
sigma_out	0.01051	0.00749	0.01427	0.01077	0.00758	0.01475	0.01057	0.00742	0.01451
MaxDD	0.33255	0.12760	0.51444	0.34762	0.13063	0.52501	0.35913	0.12409	0.53553
UI	0.06526	0.03720	0.25758	0.06153	0.03399	0.25350	0.05927	0.03519	0.26723
SR	0.60812	0.02319	1.37541	0.66958	0.06455	1.46181	0.61044	0.02343	1.39229
SoR	0.84973	0.03217	2.02167	0.93680	0.08704	2.16746	0.85033	0.03165	2.04775
Rachev	0.92435	0.80121	1.07019	0.91647	0.79835	1.06558	0.91388	0.79057	1.06607
NHI	0.04941	0.04941	0.04941	0.02284	0.02284	0.02284	0.03511	0.03511	0.03511
	CeL			**WM**			**MinV**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00045	0.00007	0.00081	0.00043	0.00002	0.00080	0.00026	-0.00008	0.00058
σ^{out}	0.01005	0.00728	0.01362	0.01089	0.00760	0.01499	0.00921	0.00682	0.01231
MaxDD	0.33332	0.11937	0.50057	0.36256	0.13513	0.54990	0.27862	0.13109	0.49466
UI	0.05644	0.03362	0.23573	0.07160	0.03606	0.27644	0.07072	0.04142	0.27009
SR	0.71887	0.09993	1.55741	0.62738	0.02232	1.44687	0.44739	-0.12916	1.17320
SoR	1.01030	0.13383	2.30795	0.87156	0.03044	2.13405	0.62428	-0.17352	1.69603
Rachev	0.93479	0.81594	1.07733	0.90972	0.79242	1.05673	0.92569	0.80089	1.06912
NHI	0.05466	0.05466	0.05466	0.01845	0.01845	0.01845	0.13689	0.13689	0.13689
	ERC			**MDP**			**EW**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00044	0.00006	0.00078	0.00045	0.00005	0.00082	0.00048	0.00008	0.00085
σ^{out}	0.01046	0.00748	0.01420	0.01054	0.00765	0.01418	0.01117	0.00812	0.01499
MaxDD	0.32027	0.13098	0.49029	0.33882	0.12863	0.51589	0.33381	0.13991	0.51355
UI	0.05991	0.03374	0.23211	0.05970	0.03618	0.25215	0.06540	0.03648	0.24675
SR	0.67270	0.08432	1.45654	0.68386	0.06142	1.47423	0.68246	0.09192	1.44627
SoR	0.94026	0.11331	2.14515	0.94984	0.08335	2.17211	0.95842	0.12412	2.13903
Rachev	0.90686	0.78873	1.04945	0.90208	0.78879	1.04020	0.91674	0.80172	1.05794
NHI	0.00338	0.00338	0.00338	0.07704	0.07704	0.07704	0	0	0

Source: authors' elaboration.

Table 13. Robustness analysis for the FTSE 100 dataset.

	SL			**CL**			**AL**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00030	0.00019	0.00072	0.00033	0.00009	0.00068	0.00032	0.00010	0.00069
σ^{out}	0.00934	0.00760	0.01160	0.00885	0.00704	0.01110	0.00871	0.00709	0.01077
MaxDD	0.36893	0.13283	0.60856	0.32682	0.11205	0.54748	0.32720	0.11961	0.54666
UI	0.10434	0.04022	0.37359	0.07137	0.03278	0.30514	0.07622	0.03624	0.31213
SR	0.51792	0.27529	1.35175	0.59275	0.14563	1.35996	0.58251	0.16912	1.36241
SoR	0.70459	0.35080	1.99634	0.81477	0.18798	2.02056	0.80090	0.21993	2.02122
Rachev	0.89292	0.73626	1.06843	0.90426	0.74957	1.07069	0.91114	0.75357	1.07904
NHI	0.07255	0.07255	0.07255	0.02069	0.02069	0.02069	0.04765	0.04765	0.04765
	CeL			**WM**			**MinV**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00024	0.00020	0.00062	0.00027	0.00020	0.00067	0.00015	0.00009	0.00047
σ^{out}	0.00890	0.00730	0.01093	0.00969	0.00781	0.01201	0.00751	0.00619	0.00915
MaxDD	0.35076	0.12932	0.57752	0.34989	0.13323	0.62054	0.27535	0.11411	0.54946
UI	0.10771	0.04087	0.35518	0.08304	0.04104	0.37688	0.10951	0.03557	0.34898
SR	0.42886	0.31505	1.20734	0.43961	0.29597	1.22113	0.30711	0.42518	1.10324
SoR	0.57958	0.39966	1.77546	0.60615	0.38612	1.82346	0.41402	0.54311	1.59675
Rachev	0.86404	0.72220	1.02668	0.91787	0.76112	1.09284	0.87721	0.72254	1.05090
NHI	0.07753	0.07753	0.07753	0.00741	0.00741	0.00741	0.08993	0.08993	0.08993
	ERC			**MDP**			**EW**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00026	0.00020	0.00065	0.00022	-0.00022	0.00058	0.00030	-0.00022	0.00073
σ^{out}	0.00949	0.00744	0.01199	0.00832	0.00671	0.01039	0.01073	0.00848	0.01339
MaxDD	0.35203	0.12944	0.61525	0.32757	0.11711	0.59888	0.37683	0.14602	0.65210
UI	0.07985	0.03807	0.36743	0.10231	0.03491	0.37186	0.08756	0.04419	0.39670
SR	0.43944	0.29669	1.21893	0.41786	-0.35159	1.26108	0.43870	-0.29128	1.18687
SoR	0.60400	0.37913	1.81564	0.56064	-0.43845	1.84605	0.61317	-0.37719	1.81021
Rachev	0.91151	0.74738	1.09229	0.86043	0.69759	1.04000	0.95054	0.78095	1.14067
NHI	0.00249	0.00249	0.00249	0.06508	0.06508	0.06508	0	0	0

Source: authors' elaboration.

Table 14. Robustness analysis for the NASDAQ dataset.

	SL			**CL**			**AL**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00049	0.00009	0.00086	0.00054	0.00015	0.00090	0.00053	0.00016	0.00088
σ^{out}	0.01127	0.00884	0.01440	0.01110	0.00848	0.01441	0.01101	0.00858	0.01417
MaxDD	0.27770	0.14065	0.46605	0.26410	0.13721	0.42907	0.22876	0.13578	0.40226
UI	0.05627	0.04233	0.22205	0.05682	0.03913	0.18999	0.05029	0.03903	0.18030
SR	0.69326	0.11547	1.36170	0.76936	0.20561	1.47403	0.76146	0.21292	1.41745
SoR	0.99686	0.15847	2.07131	1.10042	0.28253	2.22382	1.09884	0.29831	2.15110
Rachev	1.01213	0.88686	1.15715	0.97363	0.84157	1.11954	1.02749	0.90580	1.16686
NHI	0.06965	0.06965	0.06965	0.02580	0.02580	0.02580	0.05066	0.05066	0.05066
	CeL			**WM**			**MinV**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00053	0.00015	0.00088	0.00061	0.00023	0.00096	0.00036	0.00000	0.00069
σ^{out}	0.01100	0.00877	0.01392	0.01130	0.00860	0.01471	0.00993	0.00707	0.01353
MaxDD	0.25829	0.12620	0.42733	0.28342	0.13237	0.43004	0.31336	0.12687	0.49911
UI	0.04921	0.03860	0.19673	0.06157	0.03663	0.17991	0.05890	0.03709	0.25827
SR	0.75777	0.19412	1.38060	0.86264	0.28692	1.59046	0.57186	-0.00564	1.31737
SoR	1.09491	0.27088	2.11668	1.22751	0.39613	2.37903	0.80099	-0.00753	1.95455
Rachev	1.01578	0.88305	1.16315	0.95424	0.84193	1.07968	0.95377	0.83494	1.09913
NHI	0.07476	0.07476	0.07476	0.01191	0.01191	0.01191	0.11590	0.11590	0.11590
	ERC			**MDP**			**EW**		
	Estimate	CI Low	CI High	Estimate	CI Low	CI High	Estimate	CI Low	CI High
μ^{out}	0.00067	0.00027	0.00104	0.00053	0.00018	0.00085	0.00077	0.00031	0.00119
μ^{out}	0.01156	0.00877	0.01501	0.01088	0.00817	0.01428	0.01302	0.01022	0.01641
MaxDD	0.29537	0.13381	0.43640	0.26227	0.12311	0.40562	0.30163	0.15546	0.46578
UI	0.05997	0.03549	0.17517	0.04940	0.03539	0.17209	0.07721	0.04152	0.19531
SR	0.91842	0.32753	1.66065	0.76737	0.23100	1.39976	0.93600	0.34114	1.66258
SoR	1.30696	0.44894	2.48832	1.10425	0.32359	2.12884	1.33560	0.47143	2.47963
Rachev	0.94322	0.83371	1.07274	1.01359	0.88838	1.16304	0.93947	0.83829	1.05693
NHI	0.00319	0.00319	0.00319	0.08418	0.08418	0.08418	0	0	0

Source: authors' elaboration.