Journal of Mathematics and Modeling in Finance (JMMF)
Vol. 4, No. 2, Summer & Autumn 2024
Research paper

Asset Allocation Using Nested Clustered Optimization Algorithm: A Novel Approach to Risk Management in Portfolio

Mahsa Safavi Iranji[1], Majid Zanjirdar[2], Mojgan Safa[3], Hossein Jahangirnia[4]

[1] Department of Finance, Qom Branch, Islamic Azad University, Qom, Iran.
mahsa.safavi@ut.ac.ir

[2] Department of Finance, Arak Branch, Islamic Azad University, Arak, Iran.
zanjirdar08@gmail.com

[3] Department of Accounting, Qom Branch, Islamic Azad University, Qom, Iran.
mojgansafa@gmail.com

[4] Department of Accounting, Qom Branch, Islamic Azad University, Qom, Iran.
hosein_jahangirnia@yahoo.com

Abstract:
Given the widespread increase in classical and emerging models for asset allocation in investment portfolios available in the capital market, investors find it challenging to easily compare classical methods and machine learning techniques to identify the optimal investment combination. The aim of this research is to compare asset allocation based on the Nested Clustering Algorithm (NCO) with classical portfolios. This study has been conducted in a practical and descriptive-analytical manner, with the statistical population consisting of all companies listed on the Tehran Stock Exchange and the Iran Farabourse from 2013 to 2022. After screening, adjusted daily data from 88 companies were selected as the final sample for statistical analysis. In this context, the Kruskal-Wallis test was used to examine the hypotheses, and Python, SPSS, and Excel software were utilized. Based on the overall performance evaluation criteria for portfolios (Sharpe ratio, Sortino ratio, maximum drawdown, value at risk, and expected shortfall), the results of the hypothesis tests in this research indicate that the methods based on the Nested Clustering Optimization Algorithm outperform their classical counterparts significantly. Therefore, it can be concluded that portfolios based on machine learning algorithms perform better than classical portfolios.

Keywords: Asset Allocation, Hierarchical Clustering, Risk Management, Machine Learning.
*M*SC Classification: 91G40 or JEL Classifications: G11, C61, G32

1 Introduction

In financial markets, risk management and portfolio optimization are among the primary goals and key challenges for financial professionals and academics. So far, finding an efficient model for financial asset allocation has been one of the main issues because asset allocation decisions are always made under uncertainty and with

[2]Corresponding author
Received: 15/10/2024 Accepted: 18/01/2025
https://doi.org/10.22054/JMMF.2025.82388.1149

incomplete information [13]. In financial markets, asset allocation refers to the distribution of an investment portfolio across various assets such as stocks, bonds, and cash in order to balance risk and return. The portfolio optimization and asset allocation model is based on the foundational idea introduced by Markowitz (1952), which operates within a framework that optimizes based on mean and variance. Markowitz emphasized that the optimization problem depends on the structure of the covariance and the expected returns of the assets, not just the returns alone. Given the practical challenges associated with predicting returns, many have turned to risk-based portfolio allocation methods that utilize the covariance matrix as a key input. So far, extensive research has been conducted to reduce errors in modeling related to asset allocation and portfolio optimization. Due to the lack of a suitable return index for calculation, which could reduce computational errors in portfolio optimization, in risk-based asset allocation methods, return is not considered a primary input of the model. Therefore, in asset allocation methods focused on risk distribution, the goal is to maximize return [23] [24]. On the other hand, today portfolio construction is recognized as an emerging and promising field in machine learning. For several decades, the asset management industry has relied on modifying and refining the Markowitz efficient frontier to create optimal portfolios. It is clear that many of these models perform optimally in-sample; however, due to computational instabilities in convex optimization, these models may perform poorly out-of-sample. In general, there are three common approaches to mitigating instability in optimal portfolios. First, some researchers have tried to regularize the solutions by injecting additional information about the mean or variance in a prior form. Second, others have suggested reducing the solution's feasible range by incorporating additional constraints. Third, some researchers have proposed methods to improve the numerical stability of the inverse covariance matrix [13]. Several classical approaches have attempted to solve these computational instabilities, each with varying degrees of success. Among them, machine learning algorithms have demonstrated the potential to generate robust portfolios that also perform well out-of-sample. Thanks to their ability to detect sparse hierarchical relationships, traditional methods are gradually being replaced [12]. The primary goal of this research is to compare classical methods in this field with the Nested Clustered Optimization (NCO) algorithm as a representative of unsupervised machine learning algorithms. In this context, the theory and performance of clustering-based methods are examined and compared with minimum variance portfolios, maximum diversification, and equal risk contribution approaches. Additionally, the performance of portfolios based on machine learning algorithms is compared to one another as well as to their classical counterparts. To meaningfully evaluate the differences in the mean values of performance evaluation metrics for portfolios formed by each model in terms of risk and return, the results from the performance evaluation criteria of various portfolio models will be analyzed using the Kruskal-Wallis test (or the H test), which is a non-parametric test.

2 Theory

Machine learning (ML) methods are designed to manage large datasets and have proven to be particularly accurate in fields such as investment and computer science. Today, some research in the field of finance examines whether ML models can be adapted to improve the performance of traditional asset allocation models in finance [32]. In fact, such applications allow us to more easily solve both linear and nonlinear problems, which traditional models cannot address. This has led to the widespread application of deep learning and machine learning techniques, as subsets of artificial intelligence, in the financial domain [1]. Given the importance of this topic, the main goal of this research is to review and introduce recent developments in computational methods. This research highlights findings that enhance our understanding of the application of ML in asset allocation and risk management. Some of the risk-based asset allocation strategies examined in this study include:

2.1 Minimum Variance Portfolio (MV)

Markowitz (1952) introduced the mean-variance framework, presenting a new solution for portfolio optimization to the world. His theory was further developed into a more general model known as Modern Portfolio Theory (MPT) with the publication of his later book [19]. Markowitz (1952) formulated a model in which, for a portfolio composed of diversified risky assets, the investor can achieve the minimum risk for a given level of return. Mathematically, this formulation can be expressed as a quadratic problem, as shown in Equation (1):

$$\begin{aligned} &\min_{\omega} \quad \omega^\top \Sigma \omega \\ &\text{s.t.} \quad \omega^\top \mathbf{1} = 1 \\ &\omega^\top \mu \geq R \end{aligned} \tag{1}$$

Where ω represents the weight vector of the assets in the portfolio and Σ is the covariance-variance matrix of the assets. In general, portfolios constructed using the Markowitz model tend to produce concentrated weights because the primary goal of optimization is to optimize volatility diversification, not the weights. The result of solving the optimization problem from Equation (1) is presented in Equation (2):

$$\omega_{\text{MV}} = \frac{\sum^{-1} 1}{1^\top \sum^{-1} 1} \tag{2}$$

Where Σ is the covariance matrix, $\mathbf{1}$ is a vector of ones, and $\mathbf{1}^\top$ is the transpose of the vector of ones.

2.2 Maximum Diversification Portfolio (MD)

Choueifaty (2006) introduced the maximum diversification framework to the academic literature and provided a ratio for measuring diversification within a portfolio. Choueifaty and Coignard (2008) expanded this framework with the Maximum Diversification Portfolio. This method operates based on the diversification ratio, serving as an efficient alternative to minimum variance portfolios [6]. Choueifaty and his colleagues (2013) defined the diversification ratio as shown in Equation (3), representing the ratio of the weighted average volatility of the portfolio to its overall volatility [7].

$$\mathrm{DR} = \frac{w^\top \sigma}{\sqrt{w^\top \Sigma w}} \tag{3}$$

Where w is the weight vector of the assets and σ is the vector of asset volatilities.

$$\sum_{i=1}^{n} w_i \sigma_i = w^\top \sigma \tag{4}$$

Equation (4) serves as a metric for the diversification index, as in a portfolio of assets, the volatilities of the assets are less than or equal to the weighted sum of the asset volatilities, given that correlations are not perfectly positive. Therefore, the Diversification Ratio (DR) estimates the level of diversification achieved by holding assets that are not perfectly correlated. It is logical that portfolios in which capital is spread across many unrelated assets also have a high DR. Choueifaty and his colleagues (2013) formalized this insight by decomposing the DR of a portfolio into its weighted correlation metrics and its concentration metrics, denoted as Asset Concentration (AC) and Risk Concentration (CR).

$$DR = \frac{1}{\sqrt{AC(1 - CR) + CR}} \tag{5}$$

AC is defined as a weighted average correlation of the volatilities of the components in the portfolio:

$$AC = \frac{\sum_{i \neq j} (w_i \sigma_i w_j \sigma_j \rho_{i,j})}{\sum_{i \neq j} (w_i \sigma_i w_j \sigma_j)} \tag{6}$$

CR is defined as the weighted concentration ratio of the portfolio's volatilities:

$$CR = \frac{\sum_i (w_i \sigma_i)^2}{\left(\sum_i w_i \sigma_i\right)^2} \tag{7}$$

A portfolio consisting of a single long position will have a CR equal to 1, while a portfolio with equal weights across its assets, reflecting their volatilities, will have the lowest possible CR, which is equal to the inverse of the number of assets in it. Therefore, the Diversification Ratio (DR) of a portfolio has an inverse relationship with both AC and CR, increasing as either of these two metrics decreases. Thus, the most diversified portfolio is represented by the case where both AC and CR are minimized. Additionally, the weight vector W_t^{MDP} is based on maximizing the diversification ratio.

$$W_t^{MDP} = \frac{argmaxDR(W_t)}{W_t} \tag{8}$$

2.3 Equal Risk Contribution Portfolio (ERC)

The core idea of this strategy is that the weights of the assets are allocated in proportion to their risk in the portfolio. In this approach, risk is measured using variance. This means that if an asset has lower risk (variance), it will receive a larger weight compared to an asset with higher risk. This method was first introduced to the capital markets by Bridgewater in 1996, and the term was coined by Edward Qian in 2005. Qian et al. (2005) argue that the investment portfolio should be allocated based on risk rather than capital allocation [28]. To understand the ratios of portfolios with equal risk contribution, Milard et al. (2008) first define the minimum and maximum risk contributions of different assets using Equation (9) in the portfolio [18].

$$MRC_i = \frac{\partial \sigma(w)}{\partial w_i} = \frac{(\Sigma w)_i}{2w^\top \Sigma w} \tag{9}$$

On the other hand, $RM(x_1, \ldots, x_n)$ is calculated as a measure of portfolio risk, which is the sum of the products of each asset's weight and the overall risk of that asset.

$$RM(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \frac{\partial RM(x_1, \ldots, x_n)}{\partial x_i} \tag{10}$$

Therefore, the risk contribution of asset iii in the portfolio is defined according to Equation (11) as follows:

$$RC_i(x_1, \ldots, x_n) = x_i \frac{\partial RM(x_1, \ldots, x_n)}{\partial x_i} \tag{11}$$

The risk contribution of asset iii is represented in Equation (12) as follows:

$$RC_i(x_1, \ldots, x_n) = x_i \left(\frac{\partial \sigma(x)}{\partial x_i} \right) = x_i \left(\frac{(\Sigma x)_i}{\sqrt{x^T \Sigma x}} \right) \tag{12}$$

2.4 Hierarchical Clustering Models

Hierarchical clustering is a clustering method that hierarchically divides data into clusters in a layered partitioning manner, creating a hierarchy of these clusters. In this approach, distances between data objects are calculated to be used as a similarity metric. Various distance measures, such as Euclidean, Manhattan, or Minkowski distances, can be used to determine the distance between clusters. The distance matrix also aids the hierarchical clustering algorithm in deciding whether to merge or split clusters. Furthermore, the linkage method significantly impacts

the determination of similar clusters. In this method, each object is initially assigned to its own cluster, resulting in n clusters for n objects. The most similar pairs of clusters are then merged based on the defined distance metric and linkage type, forming a new single super-cluster. Subsequently, the distance between this new cluster and the remaining clusters is recalculated, and this process continues until the number of clusters matches the desired count or all objects merge into a single cluster. This process is crucial, as incorrect selection of a linkage method can lead to improper and inefficient clustering [30] [26].
Various Linkage Models Based on Distance Measurement Between Clusters in Hierarchical Clustering:

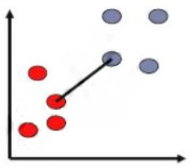
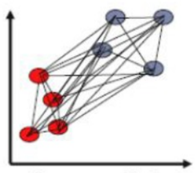
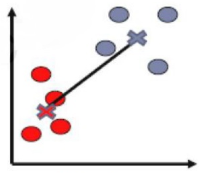
- **Single Linkage**: This method measures the distance between the closest points of two clusters. It proves to be proficient in recognizing non-elliptical geometries, rendering it advantageous for discerning intricate patterns within datasets. However, it can lead to chaining, where clusters are formed based on a single close connection, which may cause elongated or stringy clusters. Despite this, its simplicity and wide usage in various clustering applications make it a valuable choice for understanding the initial structure of the data.

- **Average Linkage**: This method computes the average distance between all points in two clusters, offering a balance between the tendencies of Single Linkage and Complete Linkage. It performs well in the case of noise between clusters and produces clusters with equal variance. However, it is somewhat biased toward globular-shaped clusters. The average linkage method is selected for its robustness in handling noise and ensuring that the resulting clusters are not overly sensitive to outliers or extreme variations.

- **Ward's Method**: Ward's Method minimizes the total within-cluster variance by merging clusters that result in the smallest increase in total variance. This method is effective for creating compact and spherical clusters, making it a preferred choice for situations where cluster compactness is crucial. While it may be sensitive to outliers and less effective in recovering elongated clusters, its ability to minimize variance and produce distinct, well-formed clusters makes it a solid choice for portfolio clustering in financial contexts [14].

These three methods were chosen due to their provision of an all-encompassing strategy for clustering, effectively tackling various issues including outliers, noise, and the compactness of clusters. By combining these methods, the goal is to achieve robust clustering results that enhance the overall portfolio optimization process. In Table 1, supplementary explanations are provided. These explanations examine the criteria based on the relationships between clusters for evaluating clustering performance. This table also illustrates how different clustering methods can be compared using these definitions. With this information, the reader can gain a better understanding of the correlation coefficient metric and its significance in

Table 1: Types of hierarchical clustering models

 Single-link	 Average-link	 Ward's
$\text{dist}_{\text{SL}}(C_i, C_j) =$ $\min_{(x \in C_i, x' \in C_j)} \text{dist}(x, x')$	$\text{dist}_{\text{AL}}(C_i, C_j) =$ $\frac{1}{\lvert C_i\rvert \cdot \lvert C_j\rvert} \sum_{x \in C_i} \sum_{x' \in C_j} D(x, x')$	$\text{dist}_{\text{Ward}}(C_i, C_j) =$ $\frac{\lvert C_i\rvert \cdot \lvert C_j\rvert}{\lvert C_i\rvert + \lvert C_j\rvert} \text{dist}(c_i, c_j)$

Where:
$\text{dist}_{\text{SL}}(C_i, C_j)$: The distance between two clusters using the single linkage method.
$\text{dist}_{\text{AL}}(C_i, C_j)$: The distance between two clusters using the average linkage method.
$\text{dist}_{\text{Ward}}(C_i, C_j)$: The distance between two clusters using Ward's method.
C_i: Center of the first group.
C_j: Center of the second group.
x: Member of the first group.
x': Member of the second group.

hierarchical clustering. Using these hierarchical clustering models, the results of the NCO clustering can be enhanced, leading to a better categorization of the data. Each of these models has its unique characteristics based on how they utilize distances and similarities.

Nested Cluster Optimization Algorithm (NCO)

According to Markowitz's mean-variance model, Markowitz's curse is an issue that arises in portfolio optimization, especially when there are either strong or low correlations between the assets in the portfolio [13]. To address Markowitz's curse in highly correlated portfolio assets, Lopez de Prado presented the Nested Clustered Optimization (NCO) algorithm. The process, which has its roots in Markowitz's mean-variance technique, attempts to distribute the portfolio's volatility among its component blocks. To do this, a clustering technique is used, and a set of stable weights is obtained by computing the best intracluster and intercluster allocations [29]. Ref. [11] describes the NCO algorithm's initial suggestion. The nested cluster assignment is implemented through a **four-phase** approach:

Phase 1: Hierarchical Correlation Clustering
Initially, the correlation matrix is converted into a distance matrix. The distance

matrix is then used to form clusters through hierarchical clustering algorithms. The structure of the clusters formed in this phase is utilized in the next phase to estimate the actual cluster constituents.

Phase 2: Optimal Number of Clusters
In this phase, the optimal number of clusters is determined. The optimal number of clusters ultimately defines the final constituents of the clusters from which the cluster correlations are extracted.

Phase 3: Intra-cluster Weight Assignment
In this stage, the intra-cluster weight assignment is conducted using the clusters formed in the previous phase. This is achieved by creating a correlation matrix for the cluster constituents and optimizing it using a portfolio optimization method. Given that NCO is considered a framework that can be applied to any convex optimization-based portfolio allocation method.

Phase 4: Inter-cluster Weight Assignment
In this phase, the reduced covariance matrix is established, considering only the variances and correlations between clusters. Using this matrix, the weight assignment between clusters is calculated using the same optimization method employed in the intra-cluster weight assignment phase [31] [13].

In this research, after determining the weights of each asset, the research variables are calculated as per the equations presented in Table 2.

The following abbreviations, as outlined in Table 3, will be used to facilitate better understanding of the concepts and to avoid lengthy explanations in the discussion and evaluation of various portfolio assessment models.

3 Literature Review

Many decision-makers, such as fund managers and investors, face serious challenges in their decision-making processes due to a lack of accurate and comprehensive information. A common approach to making these decisions is to provide answers as solutions to a convex optimization problem, where the goal is to maximize a specified objective function subject to a series of inequality constraints. The Critical Line Algorithm (CLA) is one of the most common methods for solving convex optimization problems with inequality constraints [19], [3]. Although this algorithm is mathematically valid, it is known to be a weak estimator for obtaining out-of-sample optimal solutions [20], [10], [16]. Two main reasons that indicate significant estimation errors in CLA are: (a) noisy inputs and (b) signal structure that destabilizes CLA. Various analytical approaches have attempted to reduce the estimation error of CLA by solving this error with three alternative methods: (1) introducing strong constraints [15], (2) adding preliminaries [4], [9], and (3) shrinking the covariance matrix [17]. Although these methods seem practically useful, they do not directly address the two contextual reasons that lead to instability in CLA. Marcus Lopez de Prado first stated in 2016 in his article that the instability of convex

Table 2: Measurement Methods for Research Variables

No.	Variable	Formula
1	r_t: Stock Return	$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$ p_t: Current Price, p_{t-1}: Previous Price
2	σ_t^2: Stock Variance	$\sigma_t^2 = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})^2 \cdot d$ r_t: Daily Geometric Return, d: Number of Trading Days
3	σ_{mn}: Covariance Between Two Stocks	$\sigma_{mn} = E\left[(r_m - E[r_m])(r_n - E[r_n])\right]$ r_m: Return of Stock m, $E[r_m]$: Expected Value of Stock m r_n: Return of Stock n, $E[r_n]$: Expected Value of Stock n
4	Sharpe Ratio	$\text{Sharpe} = \frac{\mu - r_f}{\sigma}$ μ: Average Return, r_f: Risk-Free Rate, σ: Standard Deviation
5	Sortino Ratio	$\text{Sortino} = \frac{\bar{r} - r_{MAR}}{\sigma_{MAR}}$ $\bar{r}$: Expected Portfolio Return, r_{MAR}: Minimum Acceptable Return σ_{MAR}: Downside Risk
6	MDD: Maximum Drawdown	$\text{MDD} = \max_{\tau \in [0,t]} \frac{W_\tau - W_t}{W_\tau}$ W_t: Lowest Portfolio Value in the Interval $[0, T]$
7	VaR: Value at Risk	$\text{VaR} = \mu + \sigma_r \cdot N^{-1}(\alpha)$ $N^{-1}(\alpha)$: Cumulative Distribution Function at confidence level α
8	CVaR: Conditional Value at Risk	$\text{CVaR} = \frac{1}{\alpha}\int_0^\alpha \text{VaR}(\alpha, X)\, dx$

optimization solutions can be traced to two sources: (1) noise in input variables and (2) signal structure that amplifies estimation errors in input variables. The innovation of this paper is the introduction of the Nested Clustering Optimization (NCO) algorithm, a method that addresses both sources of instability. Vito Cisirti and Andrea Biasi (2023) introduced a new model for portfolio optimization using a tree method with minimal nested clusters in their research. This model is capable of overcoming the limitations of asset allocation through classical methods, such as instability and excessive concentration of portfolio weights, and it provides a defensive mechanism against increased systematic risk during high-volatility periods. To achieve this, the authors utilized graph theory and followed a multi-step clustering-based approach to ultimately improve volatility. They also claimed that

Table 3: Abbreviations Used in Machine Learning and Classical Portfolio Models

Row	Abbreviation	Model Type	Name
1	NCO ward MV	Machine Learning	Minimum Variance Model implemented in the Nested Cluster Optimization Algorithm with Ward linkage
2	NCO ward MD	Machine Learning	Maximum Diversification Model Implemented in Nested Cluster Optimization Algorithm with Ward linkage
3	NCO ward ERC	Machine Learning	Equal Risk Contribution Model Implemented in Nested Cluster Optimization Algorithm with Ward linkage
4	NCO single MV	Machine Learning	Minimum Variance Model implemented in the Nested Cluster Optimization Algorithm with Single Linkage
5	NCO single MD	Machine Learning	Maximum Diversification Model implemented in the Nested Cluster Optimization Algorithm with Single Linkage
6	NCO single ERC	Machine Learning	Equal Risk Contribution Model implemented in the Nested Cluster Optimization Algorithm with Single Linkage
7	NCO average MV	Machine Learning	Minimum Variance Model implemented in the Nested Cluster Optimization Algorithm with Average Linkage
8	NCO average MD	Machine Learning	Maximum Diversification Model implemented in the Nested Cluster Optimization Algorithm with Average Linkage
9	NCO average ERC	Machine Learning	Equal Risk Contribution Model implemented in the Nested Cluster Optimization Algorithm with Average Linkage
10	MD	Classical	Maximum Diversification
11	MV	Classical	Minimum Variance
12	ERC	Classical	Equal Risk Contribution

by employing the bootstrap method, they can produce more diverse and stable portfolios that, compared to competing methods, result in improved skewness metrics and lower tail risk [8]. Banouchir and Khadri (2021) sought to find a method to enhance the out-of-sample performance of portfolio weights in their research. Using hierarchical clustering, they proposed an alternative cluster-based portfolio to obtain a sequence of clustered assets. Based on the Gram-Schmidt orthogonalization, the risk estimation of the dataset is converted to the sum of the cluster estimates in the sequence. The performance of this method and its competitors was empirically compared through various simulations in high dimensions [5]. Pouletabo and Spiridonova (2021) proposed a method for reducing the dimensionality of input data based on hierarchical clustering of available securities for investment. Initially, the Pearson pairwise correlation coefficient is used as a measure of proximity among securities for hierarchical clustering. In the next step, the impact of the proposed method on the quality of the optimal solution obtained using several samples, with respect to the Markowitz model, is examined. The effect of hierarchical clustering parameters (distance metrics between clusters and clustering threshold values) on the change in the quality of the optimal solution obtained is also investigated. The dependency between the returns of the target portfolio and dimensionality reduction using the proposed method is analyzed [27]. Noor Ahmadi and Sadeghi (2022) employed hierarchical risk parity machine learning techniques in their research and compared the results with a minimum variance approach. For this research, the adjusted closing prices of 30 listed companies over 760 trading days from 2018 to

2020 were utilized. The Sharpe ratio was used to assess portfolio performance for both in-sample and out-of-sample periods. The results from in-sample and out-of-sample analysis indicated that the hierarchical risk parity approach outperformed the minimum variance approach [25]. Mirlouhi and Toudeshki (2020) aimed to present a suitable method for portfolio optimization using market data and its clustering. The outcome of this comparison illustrates the success rate of optimization based on clustering compared to a benchmark portfolio [21].

4 Hypotheses of the Research

- **Hypothesis 1**: The performance of the Minimum Variance model implemented in the Nested Cluster Optimization algorithm (NCO MV) is better than that of the Minimum Variance model (MV).
- **Hypothesis 2**: The performance of the Equal Risk Contribution model implemented in the Nested Cluster Optimization algorithm (NCO ERC) is better than that of the Equal Risk Contribution model (ERC).
- **Hypothesis 3**: The performance of the Maximum Diversification model implemented in the Nested Cluster Optimization algorithm (NCO MD) is better than that of the Maximum Diversification model (MD).

5 Methodology

This research is classified as a quantitative method due to the use of numerical measurements in the results of the study, and it falls under applied research in terms of outcomes. From the perspective of research philosophy, it is interpretative, and in terms of research approach, it is inductive, aimed at analyzing quantitative and scattered data and transforming them into rich and detailed information. Additionally, the research strategy for identifying and compiling related factors under a common condition is correlational, and the research objectives, in terms of implementation logic, are descriptive. Regarding data collection methods, both field and library studies are utilized. In this research, in the first step, the adjusted price data after the distribution of cash dividends and capital increases of the sample stocks is extracted for a ten-year period. After normalizing and cleaning the data, in the second step, all classic models, including minimum variance portfolio, equal risk contribution portfolio, and maximum diversification portfolio, as well as machine learning models using cumulative hierarchical clustering, are calculated according to the research assumptions in the Python software environment. In the third step, the performance in terms of profitability and risk measurement of different models is assessed using modern and postmodern portfolio indicators. In the fourth step, the Kruskal-Wallis statistical test is employed to measure the performance superiority of each strategy. This test is used to compare the means of two or more populations. In this test, the variables being compared are quantitative, and the sample sizes are small, or their distribution is not normal [22]. The statistical population of this research includes the securities of all active issuers in the Tehran Stock Exchange and Iran Fara bourse during the period from the beginning of 2013 to the end of 2022. The research sample for the period from April 25, 2013, to March 18, 2023, is selected based on 2,408 trading days. The sampling is conducted after applying the following restrictions:

1) The issuers' stocks must be accepted in the Tehran Stock Exchange by the end of 2022.

2) The stock symbol must be listed on the board and have a complete price history on all years considered.

3) In each year, the stocks must have been traded in an eligible state for more than 160 days.

Ultimately, the research sample consists of daily adjusted data for 88 companies among all companies listed on the Tehran Stock Exchange that have performed satisfactorily and met the above conditions.

Data collection was conducted in Excel, while data refinement and processing, algorithm execution, variable calculations relevant to the research, and hypothesis testing to identify the statistical significance of differences in average errors were carried out using Python version 3.9.7. In this software, the libraries NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-learn, FinPy, and CvxPy were utilized. Finally, for hypothesis testing, SPSS software was used, and for data analysis and creating related graphs, Excel was employed.

6 Main results

Table 4 provides a summary of the descriptive statistics for the returns of different portfolio strategies under review over a 10-year period from 2013 to 2022 (Persian calendar years 1392 to 1401).

Based on the information in Table 4, the following conclusions can be drawn:

1. **High Mean Return of NCO ward MD:** Despite having a higher standard deviation than the average among the models (17.56%), the NCO ward MD strategy shows the highest average return of 141.27%. This indicates that this strategy, while potentially riskier, yields greater returns.
2. **Lowest Standard Deviation for MV:** The MV strategy has the lowest standard deviation (11.20%), suggesting that it offers more consistent performance compared to other strategies. This can be particularly appealing for risk-averse investors.

Table 4: Statistical Summary of Machine Learning and Classic Algorithms

Algorithm	Mean	Standard Deviation	Min	25th Percentile	Median	75th Percentile	Max	Skewness	Kurtosis
Machine Learning Algorithms									
NCO ward MV	1.153	0.154	-0.396	0.106	0.654	1.774	5.235	1.906	4.351
NCO ward MD	1.413	0.176	-0.327	0.113	0.741	1.692	7.348	2.461	6.654
NCO ward ERC	0.945	0.168	-0.323	0.049	0.585	1.389	4.439	2.026	4.850
NCO single MV	1.157	0.148	-0.418	0.110	0.661	1.779	5.104	1.788	3.853
NCO single MD	1.363	0.164	-0.343	0.089	0.702	1.896	6.628	2.214	5.473
NCO single ERC	1.019	0.156	-0.340	0.125	0.627	1.563	4.555	1.926	4.470
NCO average MV	0.746	0.115	-0.302	0.077	0.355	1.229	3.208	1.614	3.011
NCO average MD	0.737	0.112	-0.309	0.053	0.351	1.227	3.163	1.556	2.750
NCO average ERC	0.753	0.115	-0.321	0.078	0.443	1.201	3.159	1.544	2.915
Classic Models									
MD	1.028	0.127	-0.384	0.067	0.671	1.697	4.216	1.600	3.097
MV	0.731	0.112	-0.303	0.042	0.359	1.162	3.165	1.625	3.086
ERC	0.926	0.160	-0.341	0.061	0.569	1.325	4.372	2.011	4.751

3. **Skewness and Kurtosis of NCO ward MD:** The NCO ward MD strategy also exhibits the highest skewness (2.461) and kurtosis (6.654), indicating that its return distribution is more asymmetrical and has heavier tails compared to the others. This suggests a higher likelihood of extreme returns (both positive and negative).

4. **Overall Performance of NCO Strategies:** The NCO strategies generally demonstrate high returns; however, they also come with increased standard deviation. This implies that while they may offer attractive returns, they also entail a higher level of risk.

These observations highlight the trade-off between risk and return in different portfolio strategies, which is essential for investors to consider when making decisions. Classic models are suitable for investors seeking simplicity. However, these strategies may not be the best choice for maximizing returns or minimizing risk. Based on these values, it can be said that machine learning algorithms have higher averages compared to classic algorithms. Additionally, the standard deviation of machine learning algorithms is also higher, indicating that their results are more varied. Figure 1 illustrates the return and risk associated with 12 different models.

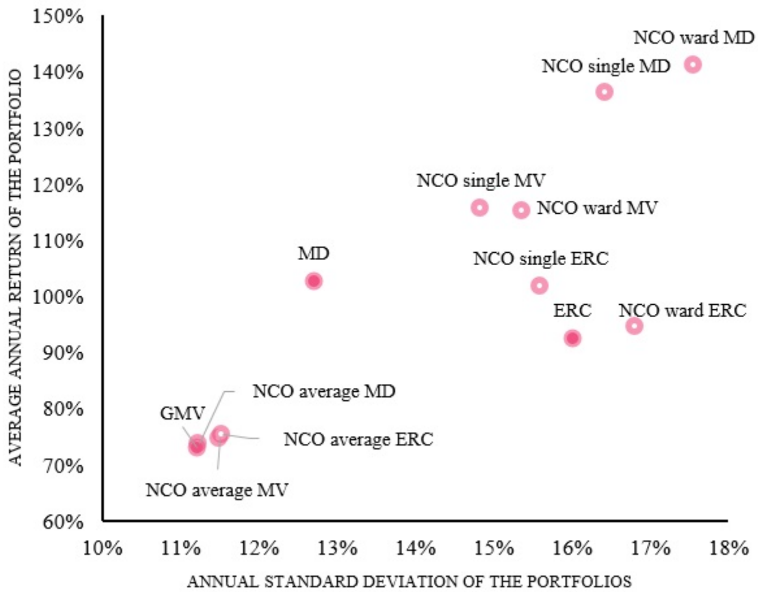


Figure 1: Return and Risk Graph of Different Portfolios

In this Figure, the attention to standard deviation in the NCO ward MD strategy indicates that its average return is the highest. This result suggests that the NCO ward MD method is recognized as a suitable approach for selecting a portfolio while considering an acceptable standard deviation. Regarding the NCO methods, it is observed that their average returns are higher than those of the machine learning methods MD and MV, indicating the superiority of these NCO methods compared to MD and MV. As shown in Figure 2, the NCO ward MD and NCO single MD portfolios have achieved higher cumulative returns compared to other portfolios. The NCO average ERC and MV portfolios have lower cumulative returns than the others. Additionally, among the three classic models, the MD portfolio has outperformed the other two methods in terms of returns.

6.1 Review of Risk-Based Performance Evaluation Variables

In this section, the performance of each of the classic portfolios and machine learning, using the metrics provided in Table 5 including maximum drawdown, annualized risk-adjusted return, annual expected shortfall, Sharpe ratio, and Sortino ratio, has been evaluated. This table shows the risk-based performance metrics of the portfolios over a 10-year period. As observed in the table, based on the average Sortino ratio and the average Sharpe ratio, the machine learning model NCO ward MD has the best overall performance. This algorithm has the highest

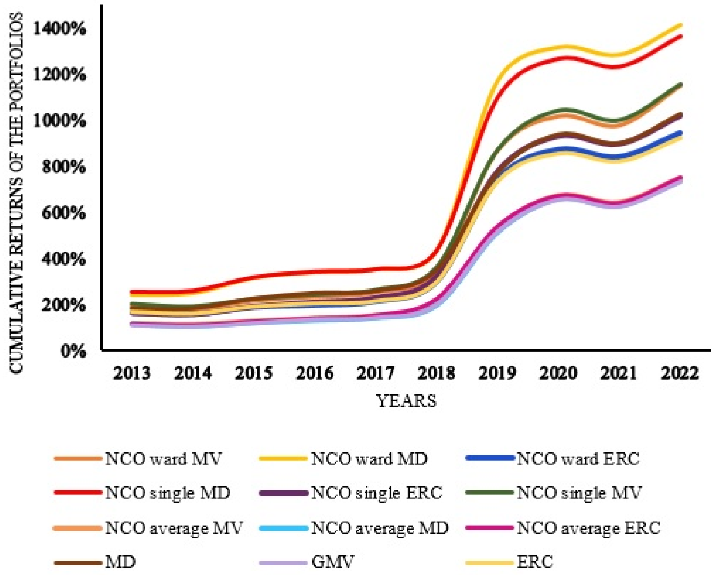


Figure 2: Annual Cumulative Return Trend of Different Portfolios

average Sortino ratio and average Sharpe ratio. Considering the value at risk at the 5% level and the expected shortfall at the 5% level, the classic MV algorithm has lower risk. The NCO average MD algorithm has the lowest value (14.89%) for the maximum drawdown metric. Given the average Sortino ratio, the MD algorithm is more suitable for investors seeking high returns, as it has the highest average Sortino ratio. In contrast, based on the value at risk at the 5% level and the expected shortfall at the 5% level, the classic MV algorithm is more appropriate for investors looking for low risk. However, on the other hand, this algorithm has the lowest return values.

6.2 Hypothesis Testing of the Research

The Kruskal-Wallis test is a non-parametric test used to compare the rankings of different groups, particularly for two or more groups. This test has two outputs: In the first output, groups are categorized based on the rankings of the variables, where higher ranks indicate better performance of the related method in that criterion, while lower ranks signify poorer performance of the method in the same criterion. In the second output, values related to the Kruskal-Wallis test and their significance levels are reported. In this study, if the significance level is less than 10%, the hypothesis under investigation will be confirmed. The use of a 10% significance threshold has been selected to effectively simulate hidden and unforeseen risks. In risk analysis, particularly under critical conditions or economic crises, a higher threshold allows researchers to identify new patterns and the effects of

Table 5: Average Values of Risk-Based Performance Metrics for Machine Learning and Classical Portfolios

Algorithm	**Maximum Drawdown**	**Value at Risk**	**Expected Shortfall**	**Average Sharpe Ratio**	**Average Sortino Ratio**
		Machine Learning Algorithms			
NCO ward MV	19.79%	21.45%	28.55%	2.04	4.73
NCO ward MD	19.66%	21.75%	31.03%	5.68	8.36
NCO ward ERC	22.07%	25.43%	32.74%	2.82	5.41
NCO single MV	18.56%	19.33%	27.38%	4.25	8.18
NCO single MD	18.30%	20.09%	29.05%	5.23	8.14
NCO single ERC	19.21%	22.15%	29.11%	2.82	5.41
NCO average MV	14.93%	16.07%	22.25%	1.88	4.06
NCO average MD	14.89%	15.32%	21.58%	2.12	4.78
NCO average ERC	15.12%	15.90%	21.95%	3.12	6.30
		Classic Models			
MD	16.77%	17.58%	23.59%	2.42	5.41
MV	14.97%	14.93%	21.46%	1.93	4.26
ERC	21.88%	24.80%	31.46%	2.54	4.62

uncertain risks. This choice facilitates the simulation of worst-case scenarios and enhances protection against unexpected risks. Furthermore, when data is subject to noise or significant deviations, a 10% threshold helps identify weaker yet critical relationships. Consequently, this approach is considered logical and beneficial for examining long-term risks and complex economic conditions.

Overall, there is a direct relationship between higher values and ranks of the Sharpe and Sortino ratios and the superior performance of the models. In other words, models that obtain higher values and ranks for these indices demonstrate better performance in terms of risk-adjusted returns and return-to-volatility ratios.

Additionally, there is an inverse relationship between lower values and ranks of maximum drawdown, value at risk, and expected shortfall, and the superior performance of the models. This means that models with lower values and ranks for these criteria show better performance in terms of stability and reduced risk of loss.

The results from the ranking of the Kruskal-Wallis test based on the five relevant criteria, broken down by different clustering linkage methods, are summarized in Table 6.

According to Table 6, in the assessment of the Sharpe ratio, the MV method has an average rank, while the NCO single MV method has performed better with

Table 6: Mean Ranks and Kruskal-Wallis (KW) test results for different models

Variable	Mean Ranks MV	Mean Ranks NCO single MV	KW	Mean Ranks MV	Mean Ranks NCO average MV	KW	Mean Ranks MV	Mean Ranks NCO ward MV	KW
Sharp Ratio	8.20	12.55	3.526**	10.60	10.40	0.006	9.00	10.70	0.023
Sortino Ratio	8.30	12.65	2.929**	10.90	10.10	0.091	9.10	10.90	0.091
Maximum Drawdown	9.30	11.70	0.823	10.20	10.80	0.051	9.00	12.00	1.286
Expected Shortfall	8.90	12.10	1.463	10.10	10.90	0.091	8.50	12.50	2.286
Value at Risk	8.60	12.40	2.063	9.70	10.30	0.366	8.40	12.60	2.520
Variable	ERC	NCO single ERC	KW	ERC	NCO average ERC	KW	ERC	NCO ward ERC	KW
Sharp Ratio	10.00	11.00	0.705	9.90	11.10	0.206	10.60	10.40	0.006
Sortino Ratio	9.80	11.20	0.597	9.70	11.30	0.366	10.70	10.30	0.023
Maximum Drawdown	11.40	9.60	0.496	12.30	8.70	1.851	10.30	10.70	0.023
Expected Shortfall	11.20	9.80	0.597	13.65	8.20	3.023**	10.00	11.00	0.143
Value at Risk	11.40	9.60	0.496	13.45	8.30	2.766**	10.30	10.70	0.023
Variable	MD	NCO single MD	KW	MD	NCO average MD	KW	MD	NCO ward MD	KW
Sharp Ratio	8.40	11.10	3.526**	11.10	9.90	0.206	8.30	11.50	3.738**
Sortino Ratio	8.50	10.60	3.084**	10.90	10.10	0.091	9.90	11.10	3.680**
Maximum Drawdown	10.10	10.90	0.091	11.00	10.00	0.143	9.80	11.20	0.280
Expected Shortfall	9.00	12.00	1.286	11.30	9.70	0.366	8.40	12.60	2.520
Value at Risk	9.40	11.60	0.691	11.90	9.10	1.120	8.90	12.10	1.463

**Statistical significance is indicated at the 10% level.

a higher rank. From the perspective of the Sortino ratio, the NCO single MV method has also shown better performance. In terms of maximum drawdown, and considering its nature, the MV method has performed better with a lower rank. Regarding Value at Risk and expected drawdown, the MV method has outperformed the NCO single MV method despite having a lower rank. With the increase in the significance level to 10%, the significance values for the Sharpe ratio (0.076) and the Sortino ratio (0.086) become less than 0.1. This indicates the rejection of the null hypothesis or, in other words, the acceptance of the alternative hypothesis at the 10% statistical significance level. This reveals evidence of a difference in the ranking of the models for these two criteria. Therefore, it can be concluded that the observed differences in the rankings of the models for the Sharpe ratio and the Sortino ratio are considered statistically significant at the 10% level.

Based on the average values of the performance metrics in Table 5 and the analyses conducted, it can be concluded that the average rank of the Sharpe ratio variable for the minimum variance model within the nested clustering optimization algorithm using the single link method ($\mu_{\text{NCO single MV Sharpe}}$) is better than the average rank of the Sharpe ratio variable for the classical minimum variance model ($\mu_{\text{MV Sharpe}}$). Additionally, the average rank of the Sortino ratio variable for the minimum variance model within the nested clustering optimization algorithm using the single link method - ($\mu_{\text{NCO single MV Sortino}}$) is also better than the average rank of the Sortino ratio variable for the classical minimum variance model ($\mu_{\text{MV Sortino}}$). Therefore, the first hypothesis related to this topic is accepted.

Table 6 also shows that the NCO average ERC model has a better rank compared to the ERC model in terms of the Sharpe ratio, Sortino ratio, expected drawdown, and value at risk variables. The significance level for the expected drawdown variable is 0.082, and for the value at risk variable, it is 0.096, both of which are less than the significance level of 0.1. These results indicate that there is sufficient statistical evidence to reject the null hypothesis for these two criteria. In other words, it can be concluded that the distribution of expected drawdown and value at risk between the NCO average ERC and ERC models likely has a significant difference. Considering Table 5, and noting that the average rank of the NCO average ERC model is better than that of the classical ERC model in terms of risk as shown in Table 6, it can be concluded that the alternative hypothesis is accepted, and no reason is found to reject the claim. Because at an almost equal level of return, this machine learning model performs better than its classical counterpart. Looking at Table 6, it can be observed that the NCO single MD model has higher average ranks compared to the MD model, especially regarding the Sharpe and Sortino ratio variables. However, for the expected drawdown and value at risk variables, the NCO single MD model has lower average ranks compared to the MD model. For the Sharpe and Sortino ratios, the significance levels are below 0.1, indicating a significant difference between the two models. Therefore, based on the results from this table and the results from Table 5, which show a better rank for the NCO single MD model in the Sharpe and Sortino ratios compared to the MD model, it can be concluded that the performance of the NCO single MD model has been better than that of the classical MD model. Additionally, referring to Table 6, it can be observed that the NCO ward MD model has achieved better average ranks than the MD model in the Sharpe and Sortino ratio criteria. This indicates the potential advantage of the NCO ward MD model concerning these performance metrics. However, in the expected drawdown and value at risk criteria, the MD model has demonstrated better performance. According to the table above, for the Sharpe and Sortino ratio variables, the significance levels are less than 0.1, indicating a statistically significant difference between the two models. However, for the other variables, the significance levels are greater than 0.1, meaning there is no statistically significant difference between the two models. Based on the ranking results and the statistical test in Table 6, it can be concluded that the NCO ward MD model performs better than the MD model in the Sharpe and Sortino ratios. Given the lack of a significant difference between the two models in the other variables, the author's decision is to accept the alternative hypothesis.

7 Conclusion and Recommendations

This research examines asset allocation by comparing the Nested Cluster Optimization (NCO) algorithm, a machine learning method, with classical approaches. With the growing use of machine learning in portfolio management, the study rigorously evaluates the performance of this algorithm. NCO, introduced by López de Prado in 2016, was compared against classical models. Results from the Kruskal-Wallis test revealed that the NCO single MV algorithm outperformed the MV model, NCO single MD and NCO ward MD algorithms surpassed the MD model in Sharpe and Sortino ratios, and the NCO average ERC algorithm excelled over the ERC model in value at risk and expected drawdown. Based on the findings, the nested clustering algorithm is recommended for investment banks, portfolio management firms, institutional investors, and stock market managers as an effective machine learning approach for asset allocation. The results align with prior studies. Mirlouhi and Toodeshki (2020) demonstrated the superiority of clustering methods over classical portfolios. Soltani Nejad and Davallou (2016) improved portfolio performance using single and average linkage clustering methods. Ciciretti and Bocci (2023) developed a nested tree-based model to enhance portfolio performance during high volatility using graph theory. García-Medina and Rodríguez-Camejo. (2024) found that the asset allocation problem poses challenges due to instability arising from high correlation among assets and structural changes in financial markets; however, the use of RMT and NCO methods can help improve portfolio performance and reduce investment risk. Raffinot (2017) highlighted the robustness and diversification of hierarchical clustering-based portfolios, showing better risk-adjusted performance and stability than traditional techniques. López de Prado (2016) addressed the inefficiencies of convex optimization in financial calculations and proposed NCO to improve stability and efficiency. The study faced challenges such as data variability, overfitting risks, and market uncertainties, which impacted the accuracy of results. Historical constraints also contributed to variations in outcomes. Future research could explore generalizing these findings to other markets.

Bibliography

[1] S. Ahmed, M. M. Alshater, A. El Ammari, H. Hammami, *Artificial intelligence and machine learning in finance: A bibliometric review*, Research in International Business and Finance, 61 (2022), 101646. https://doi.org/10.1016/j.ribaf.2022.101646

[2] E. Abounori, R. Tehrani, M. Shamani, *Performance of risk-based portfolios under different conditions in the stock market (Evidence from the Iranian stock market)*, Financial Economics, 45(12) (2018), 51-71. [In Persian]

[3] D. Bailey, M. López de Prado, *An open-source implementation of the critical-line algorithm for portfolio optimization*, Algorithms, 6(1) (2013), 169-196. https://doi.org/10.3390/a6010169

[4] F. Black, R. Litterman, *Asset allocation combining investor views with market equilibrium*, Journal of Fixed Income, 1(2) (1991), 7-18. https://doi.org/10.3905/jfi.1991.408013

[5] N. Bnouachir, A. Mkhadri, *Efficient cluster-based portfolio optimization*, Communications in Statistics-Simulation and Computation, 50(11) (2021), 3241-3255. https://doi.org/10.1080/03610918.2019.1621341

[6] Y. Choueifaty, Y. Coignard, *Toward maximum diversification*, The Journal of Portfolio Management, 35(1) (2008), 40-51. https://doi.org/10.3905/jpm.2008.35.1.40

[7] Y. Choueifaty, T. Froidure, J. Reynier, *Properties of the most diversified portfolio*, Journal of Investment Strategies, 2(2) (2013), 49-70. https://doi.org/10.21314/jois.2013.033

[8] V. Ciciretti, A. Bucci, *Building optimal regime-switching portfolios*, The North American Journal of Economics and Finance, 64 (2023), 101837. https://doi.org/10.1016/j.najef.2022.101837

[9] R. Clarke, H. De Silva, S. Thorley, *Portfolio constraints and the fundamental law of active management*, Financial Analysts Journal, 58 (2002), 48-66. https://doi.org/10.2469/faj.v58.n5.2468

[10] V. De Miguel, L. Garlappi, R. Uppal, *Optimal versus naïve diversification: How inefficient is the 1/N portfolio strategy?*, Review of Financial Studies, 22(5) (2009), 1915-1953. https://doi.org/10.1093/rfs/hhm075

[11] M. Lopez de Prado, *A robust estimator of the efficient frontier*, Available at SSRN, (2016). http://dx.doi.org/10.2139/ssrn.3469961

[12] M. M. De Prado, *Advances in financial machine learning*, John Wiley & Sons, 2018.

[13] M. M. De Prado, *Machine learning for asset managers*, Cambridge University Press, 2020.

[14] A. Dogan, D. Birant, *K-centroid link: a novel hierarchical clustering linkage method*, Applied Intelligence, (2022), 1-24. https://doi.org/10.1007/s10489-021-02624-8

[15] F. Fabozzi, P. Kolm, D. Pachamanova, S. Focardi, *Robust portfolio optimization and management*, Wiley Finance, First Edition, (2007).

[16] J. Guerard, *Handbook of portfolio construction*, Springer, First Edition, 2010.

[17] O. Ledoit, M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis, 88(2) (2003), 365-411. https://doi.org/10.1016/s0047-259x%2803%2900096-4

[18] S. Maillard, T. Roncalli, J. Teïletche, *The properties of equally weighted risk contribution portfolios*, The Journal of Portfolio Management, 36(4) (2010), 60-70. https://doi.org/10.3905/jpm.2010.36.4.060

[19] H. M. Markowitz, *Portfolio selection*, The Journal of Finance, 7 (1952), 77-91.

[20] R. Michaud, *Efficient asset allocation: a practical guide to stock portfolio optimization and asset allocation*, MA: Harvard Business School Press, (1998).

[21] S. M. Mirlouhi, N. Mohammadi Toodeshki, *Optimal portfolio construction in Tehran Stock Exchange using hierarchical and divisive clustering methods*, Investment Knowledge, 9(34) (2020), 333-354. [In Persian]

[22] M. Momeni, A. Fa'al Qayoumi, *Statistical analysis using SPSS*, Author, Fifth Edition, 2022. [In Persian]

[23] H. Nikumaram, F. Rahnamay Roodposhti, M. Zanjirdar, *The explanation of risk and expected rate of return by using Conditional Downside Capital Assets Pricing Model*, Financial Knowledge of Securities Analysis, 3(1) (2008), 55-77. [In Persian]

[24] M. Nourahmadi, H. Sadeghi, *The Application of the Main Components in Investment Basket Management: A Case Study of Fifty Stock Exchange Companies*, Budget and Finance Strategic Research, 3(1) (2022), 71-95. [In Persian] https://dor.isc.ac/dor/20.1001.1.27171809.1401.3.1.3.6

[25] M. Nourahamadi, H. Sadeghi, *A Machine Learning-Based Hierarchical Risk Parity Approach: A Case Study of Portfolio Consisting of Stocks of the Top 30 Companies on the Tehran Stock Exchange*, Financial Research Journal, 24(2) (2022), 236-256. [In Persian] https://doi.org/10.22059/frj.2021.319092.1007146

[26] M. Nourahmadi, H. Sadeqi, *Portfolio Diversification Based on Clustering Analysis*, Iranian Journal of Accounting, Auditing and Finance, 7(3) (2023), 1-16. https://doi.org/10.22067/ijaaf.2023.43078.1092

[27] A. Y. Poletaev, E. M. Spiridonova, *Hierarchical clustering as a dimension reduction technique in the Markowitz portfolio optimization problem*, Automatic Control and Computer Sciences, 55(7) (2021), 809-815. https://doi.org/10.3103/s0146411621070270

[28] E. Y. Qian, F. Ying, J. Higgison, *A dynamic decision model for portfolio investment and assets management*, Journal of Zhejiang University-SCIENCE A, 6 (2005), 163-171. https://doi.org/10.1631/jzus.2005.as0163

[29] B. Rodríguez-Camejo, *Random matrix theory and nested clustered optimization on high-dimensional portfolios*, International Journal of Modern Physics C (IJMPC), 35(08) (2024), 1-19. https://doi.org/10.1142/S0129183124500980

[30] M. Soltani-Nejad, M. Davallou, *Portfolio Optimization with Clustering Methods*, Journal of Asset Management and Financing, 4(4) (2016), 1-16. https://doi.org/10.22108/amf.2016.21104

[31] D. Sjöstrand, N. Behnejad, M. Richter, *Exploration of hierarchical clustering in long-only risk-based portfolio optimization*, PhD thesis, CBS, Copenhagen, 2020.

[32] H. O. Zapata, S. Mukhopadhyay, *A bibliometric analysis of machine learning econometrics in asset pricing*, Journal of Risk and Financial Management, 15(11) (2022), 535. https://doi.org/10.3390/jrfm15110535

How to Cite: Mahsa Safavi Iranji[1], Majid Zanjirdar[2], Mojgan Safa[3], Hossein Jahangirnia[4], *Asset Allocation Using Nested Clustered Optimization Algorithm: A Novel Approach to Risk Management in Portfolio*, Journal of Mathematics and Modeling in Finance (JMMF), Vol. 4, No. 2, Pages:137–157, (2024).