algorithms

MDPI

Article

Efficient Selection of Investment Portfolios in Real-World Markets: A Multi-Objective Optimization Approach

Antonio J. Hidalgo-Marín [1], Antonio J. Nebro [1,2] and José García-Nieto [1,2,*]

[1] Department de Lenguajes y Ciencias de la Computación, University of Málaga, 29071 Málaga, Spain; antonio.hidalgo@uma.es (A.J.H.-M.); ajnebro@uma.es (A.J.N.)
[2] ITIS Software, University of Málaga, Ada Byron Research Building, 29071 Málaga, Spain
* Correspondence: jnieto@uma.es

Abstract

As financial markets become increasingly complex, optimizing investment portfolios under multiple conflicting objectives has become a central challenge for decision-makers. This paper presents a comprehensive benchmarking framework for multi-objective portfolio optimization based on metaheuristics, designed to operate on real-world financial data. This framework integrates preprocessing, and optimization using four state-of-the-art algorithms: NSGA-II, MOEA/D, SMS-EMOA, and SMPSO. Using historical data from over 11,000 assets listed on U.S. exchanges, including ARCA, NYSE, NASDAQ, OTC, AMEX, and BATS, we define a suite of benchmark scenarios with increasing dimensionality and constraint complexity. Our results highlight algorithmic strengths and limitations, reveal significant trade-offs between return and risk, and demonstrate the effectiveness of multi-objective metaheuristics in constructing diversified, high-performance investment portfolios. Each portfolio is encoded as a real-valued vector combining asset selection and allocation, enabling fine-grained diversification control. All datasets and source code are publicly available to ensure reproducibility.

Keywords: multi-objective optimization; portfolio optimization; financial asset; evolutionary algorithms; swarm intelligence; algorithm comparison

1. Introduction

Portfolio selection is a foundational problem in finance [1], involving the allocation of a limited budget across a set of financial assets to achieve desirable returns. However, return maximization alone is insufficient for sound investment decisions, as risk is an intrinsic factor that must also be considered. Markowitz's seminal work in 1952 [2] introduced Modern Portfolio Theory (MPT), emphasizing that risk and return should be evaluated jointly. According to MPT, the overall behavior of a portfolio depends not only on the individual assets but also on their interactions, as captured by covariances.

Efficient portfolio selection is naturally formulated as a multi-objective optimization problem [3,4], given the conflicting nature of its objectives: maximizing expected return while minimizing risk. The standard deviation of returns is commonly used as a proxy for risk, while the expected return captures the average projected gain. Solving this problem yields a set of trade-off solutions forming the Pareto front, which represents the most efficient risk–return combinations. Investors can then choose the solution that best aligns with their risk tolerance and performance expectations.

Check for updates

Academic Editor: Mohammad Majid al-Rifaie

Received: 18 November 2025
Revised: 14 December 2025
Accepted: 19 December 2025
Published: 24 December 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Portfolio optimization is an NP-hard problem [3], and the challenge becomes more pronounced in large-scale scenarios involving thousands of assets. In such cases, traditional exact methods become computationally infeasible. Metaheuristics [5], a class of non-exact stochastic optimization algorithms, offer a compelling alternative due to their ability to efficiently explore complex and high-dimensional search spaces. Among them, Evolutionary Algorithms (EAs) and Particle Swarm Optimization (PSO) have been widely used in multi-objective settings for their robustness and flexibility.

Working with real-world financial data introduces additional complexities. Data must be collected, cleaned, and transformed into a format suitable for optimization. This involves acquiring raw stock information, filtering out irrelevant or erroneous entries, aggregating and normalizing the data, and finally structuring it as input for the optimization process. The quality of the input data directly affects the reliability and validity of the optimization results.

In this work, we propose a multi-objective portfolio optimization framework that integrates these essential data-handling steps and applies state-of-the-art metaheuristics to derive efficient investment solutions. Our approach leverages multi-objective evolutionary algorithms (MOEAs) [6,7] implemented within the jMetal framework [8,9], a Java-based platform for metaheuristic optimization. To provide a broad and balanced comparison, we select three well-established MOEAs, each representing a key category in the field: NSGA-II [10] (Pareto dominance-based), MOEA/D [11] (decomposition-based), and SMS-EMOA [12] (indicator-based). We also include SMPSO [13], a PSO variant tailored for multi-objective optimization, to provide an alternative metaheuristic paradigm.

Building on Markowitz's theory, our framework adopts a novel encoding scheme in which each portfolio is represented by a real-valued vector of size $2 \times n$, where n is the number of available assets. For each asset, the first variable determines whether the asset is included in the portfolio, while the second specifies the investment proportion. Weights are normalized to ensure that the total allocation sums to one, and short selling is not allowed. To promote diversification and reflect practical investment constraints, lower and upper bounds are imposed on individual asset allocations.

To guide the selection of portfolios from the obtained Pareto fronts, we incorporate the Sharpe ratio [14] as a reference metric for ranking. The Sharpe ratio, defined as the difference between the expected portfolio return (R_p) and the risk-free rate (R_f), divided by the standard deviation of returns (σ_p), quantifies the excess return per unit of risk. This measure provides an effective way to assess risk-adjusted performance and compare competing solutions.

A distinctive feature of our study is its application to real financial data obtained from major U.S. stock exchanges, including ARCA, NYSE, NASDAQ, OTC, AMEX, and BATS. We construct realistic problem instances based on over 11,000 assets, enabling an extensive evaluation of the proposed framework in practical settings.

The main contributions of this paper are summarized as follows:

- We design and implement a comprehensive framework for the systematic benchmarking and empirical comparison of portfolio optimization strategies based on real-world financial data. The framework includes modules for data acquisition, preprocessing, and multi-objective optimization using metaheuristics.
- An approach for portfolio optimization is proposed, in which the selected assets and their respective investment percentages are represented using real-valued vectors. This encoding differs from previous works such as [5,15], as it uses a first part for company selection (from the entire stock market index) and a second part for the investment percentage, although it adds a low complexity to the model. It enables a structured and flexible representation of investment portfolios, facilitating the application of

optimization algorithms by allowing precise control over asset allocation. It ensures that the proportion of capital allocated to each asset aligns with the investor's risk tolerance and return expectations.

- We develop a benchmark suite of portfolio optimization instances using real financial data retrieved through a custom integration pipeline. The datasets and source code are publicly available (datasets and source code: https://github.com/AntHidMar/ParetoInvest, https://github.com/AntHidMar/Portfolio_Study_DataSources (accessed on 17 November 2025)), promoting reproducibility and transparency.
- We conduct a comparative study of four leading multi-objective metaheuristics, analyzing their performance across different evaluation criteria, including convergence, solution quality, and computational efficiency. The results provide insights into the strengths and limitations of each algorithm and offer guidance for practical deployment in investment scenarios.

The remainder of the paper is organized as follows. Section 2 reviews related literature in portfolio optimization using multi-objective metaheuristics. Section 3 defines the optimization problem and describes our solution approach. Section 4 outlines the data acquisition and preprocessing workflow. Sections 5 and 6 detail the experimental setup and present the results. Section 7 discusses the findings, and Section 8 concludes the paper and outlines future research directions.

2. Related Works

The dilemma of choosing an efficient investment portfolio has been widely studied by researchers and practitioners [16], and indeed, commercial solutions exist that promise to address it effectively, such as: Quicken (https://www.quicken.com/ (accessed on 17 November 2025)), Sharesight (https://www.sharesight.com/ (accessed on 17 November 2025)), or Empower (https://www.empower.com/personal-investors/ (accessed on 17 November 2025)). Most of these solutions rely on inference engines based on artificial intelligence processes, where traditional Machine Learning (ML) methods [17] and modern Deep Learning (DL) approaches [18] have been proposed in the last years. Specifically, metaheuristic techniques have been showing a great performance in the last two decades [19–21], since they are able to obtain efficient or near-optimal solutions in scenarios where traditional methods may be ineffective due to the complexity and dimensionality of the problem.

In this context, a widely recognized methodology by the research community is based on the Markowitz's mean–variance (MV) model [15,22–24], which uses the premise that the expected returns of the portfolio are calculated using the means of the individual returns, while the portfolio risk is quantified as the variance of the returns of these assets. A significant extension of the Markowitz's model is diversification, which considers not only the variance of individual returns but also the covariances between the returns of different portfolio assets. Effective diversification can significantly reduce the total portfolio risk without compromising expected returns. This characteristic is illustrated in the efficient frontier, which shows all possible combinations of assets that offer the maximum return for a given level of risk or the minimum risk for a given level of expected return [25,26].

Following this model, multi-objective metaheuristics [27] are currently the most used techniques for portfolio optimization [4], since they offer a direct optimization learning model to address the two fundamental objectives of reducing risk and obtaining maximum profitability. Among the different techniques used in this sense, there exists comparative studies, such as those proposed in [28,29], which evaluate the performance of various MOEAs and multi-objective versions of PSO applied to the portfolio asset selection. These approaches optimizes conflicting objectives, such as risk and return, under certain con-

straints. Specifically, the Non-dominated Sorting Multi-objective PSO (NS-MOPSO) is introduced and compared with four non-dominated sorting-based MOEAs, in addition to one decomposition-based approach (MOEA/D). However, these studies were conducted with synthetic benchmarking data, which can distort results in a real-world application.

Regarding constraints, there are also similar articles where constraints are introduced to limit the dimension of the problem to be addressed. For instance in [23], a study is conducted with three objectives: risk, return, and the number of values in the portfolio, where authors also introduced quantity and class constraints to limit the proportion of the portfolio invested in assets with common characteristics and avoid very small holdings. This study highlighted the computational complexities introduced by the nonlinear objective function and discrete constraints, using public benchmark data sources, which contrasts with our proposal that employs real data.

In this regard, while traditional portfolio selection focuses on optimizing based on the return–risk relationship under specific and often unrealistic assumptions, robust portfolio selection seeks to create viable solutions under a much wider range of market conditions [30] and less restrictive assumptions [31]. This can be especially valuable in volatile and uncertain market environments, where parameters such as returns and covariances can be difficult to estimate accurately.

Recent developments in portfolio optimization tend to seek hybrid systems that help genetic algorithms become more efficient. In recent years, several articles have addressed these trends; for example, Aithal et al. (2023) [32] developed a real-time portfolio management system that uses machine learning techniques to select the most promising stocks based on various financial indicators, with the aim of optimizing portfolio selection and management. The methodology employed integrates K-Means clustering algorithms for portfolio selection and genetic algorithms, considering various types of portfolio, such as the global minimum variance portfolio, and introducing the maximum Sharpe ratio. This indicator, due to its simplicity and representational capacity, has been widely used in recent times [15].

From a different perspective, PMP-WOA (Periodic Multi-Population Whale Optimization Approach) [33] has shown high effectiveness in multi-objective optimization, both in ZDT benchmark functions and portfolio optimization problems. It leverages dynamic use of multiple sub-populations and information exchange between them to improve accuracy and convergence toward the true Pareto front. In addition, learning automata are used to adapt algorithm parameters, enabling faster and more efficient convergence. PMP-WOA also maintains an archive of non-dominated solutions, enhancing solution diversity. The results indicate a lower generational distance, a higher number of Pareto-optimal solutions, greater spread, and lower spacing, which all reflect improved accuracy and diversity compared to existing methods.

It is also worth mentioning EvoFolio [34], which based on NSGA-II, tries to maximize the yield and minimize the risk simultaneously. EvoFolio has been experimented on stock datasets in a three-years time-frame and varying the configurations/specifics of algorithmic operators. The complete solution allows an interactive method, where users can provide their own insights and suggestions to the algorithm such that it takes into account users' preferences for some stocks.

Finally, from a large-scale perspective, a novel evolutionary algorithm called LSWOEA [35] is developed to effectively solve a transformed model, which incorporates uncertainty and random variables within a mean–variance–skewness framework In this model, the future returns of long-established and newly listed actives are modeled as random and uncertain variables, respectively. To better reflect real-world investment conditions, the model includes constraints such as cardinality, minimum transaction lots,

boundary limits, and a prohibition on short selling. An encoder–decoder approach is then introduced to transform the constrained model into an unconstrained one, enabling the application of a multi-objective evolutionary algorithm (MOEA).

The proposed approach in this article is based on the design of random real-number vectors that iterate stochastically, by analyzing the covariances between all potential companies. Its main difference with respect to the previous works is that it aims to reduce linearity and correlation among the companies invested, thus significantly improving the results obtained. It performs a multi-objective iterative process to evaluate the historical performance of a group of stocks, weighing them using "chromosomes" composed of real-number vectors. In this regard, an exhaustive and realistic study of portfolio creation for specific markets and companies is included, as well as offering a selection mechanism by cardinality within each potential set or portfolio in which to invest.

This allows the practitioner to perform executions not only for different numbers of cycles, but also for different markets and various sets of companies within each market, thus providing a more robust and comprehensive analysis.

3. Multi-Objective Portfolio Optimization Approach

In this section, we define formally the bi-objective formulation of the portfolio optimization problem and describe the solution encoding and the multi-objective metaheuristics selected to solve it.

3.1. Problem Formulation

In the field of optimization and investment management, Markowitz's Efficient Portfolio Management theory [2] has been a fundamental tool for asset selection, balancing expected returns with associated risk. Due to its complexity, this problem has been studied from the perspective of algorithm optimization, especially in the context of evolutionary algorithms [36].

Markowitz's theory is based on the following fundamental principles:

Diversification: Investors can reduce portfolio risk by distributing investments across a variety of assets or different types of investments. Diversification helps mitigate the specific risk of each asset.

Return and Risk: The theory aims to maximize return and minimize risk associated with the portfolio. Markowitz introduced the concept of the "efficient frontier", which represents all possible combinations of portfolios offering the best-expected return for a given level of risk or the lowest risk for a given level of expected return.

Formally, the portfolio selection is treated as a bi-objective minimization problem. The objective vector $F(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x})]^T$ is defined as follows:

- **Objective 1 (Risk):** Minimize the Portfolio Variance.

$$\text{Minimize } f_1(\vec{x}) = \sigma_p^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} x_{N+i} x_{N+j} \sigma_{ij} \tag{1}$$

where σ_{ij} is the covariance between asset i and asset j (derived from the annualized covariance matrix).

Variance and Covariance Analysis: Markowitz's theory uses variance and covariance analysis to measure risk and the relationship between different assets in a portfolio. The portfolio variance measures the dispersion of returns. It is calculated (Equation (1)) as the weighted sum of the variances of individual assets and the covariances between pairs of assets in the portfolio.

- **Objective 2 (Return):** Maximize the Expected Return. In the minimization framework of standard MOEAs, this is formulated as minimizing the negative expected return:

$$\text{Minimize } f_2(\vec{x}) = -E(R_p) = -\sum_{i=1}^{N} x_{N+i} R_i \tag{2}$$

where R_i represents the simple annual return of asset i.

The expected return of the portfolio is calculated by means of Equation (2) as the weighted sum of the expected returns of the individual assets in the portfolio, where w_i is the weight (proportion of invested capital) of asset i and R_i is its expected return.

All objectives and parameters (including the risk-free rate R_f) are expressed on an **annualized basis** to ensure dimensional consistency.

Investor Indifference Curve: The idea of the efficient frontier was introduced in [37]. Basically, investors have different levels of risk aversion and therefore are willing to accept different levels of expected return for a given amount of risk. This is represented by investor indifference curves, which show combinations of return and risk that are equally preferable for a given investor.

Regarding computational complexity, it is important to distinguish between the search space size and the evaluation cost. While the search space dimension is determined by the total universe of assets (N), the cost of evaluating a solution depends on the number of selected assets (k), where typically $k \ll N$. The calculation of the Expected Return (Equation (1)) has a linear complexity of $O(k)$, while the Portfolio Variance (Equation (2)) involves a quadratic complexity of $O(k^2)$ due to the covariance terms. This decoupling allows the optimization core to handle large-scale scenarios ($N > 10,000$) efficiently, as the computational effort for fitness evaluation remains bounded by the square of the portfolio size rather than the total market dimension.

3.2. Solution Encoding and Decoding Procedure

The solution encoding is a critical component of the framework. We adopt a continuous solution representation where each individual is encoded as a real-valued vector $\vec{x} \in \mathbb{R}^{2N}$, where N represents the total number of assets in the market universe. The vector is structured as follows:

- **Selection Part ($x_1, \ldots, x_N$):** The first N variables act as positional pointers to select specific assets from the universe.
- **Allocation Part ($x_{N+1}, \ldots, x_{2N}$):** The remaining N variables represent the raw weight associated with the selected assets.

3.2.1. Decoding Mechanism (Genotype to Phenotype)

To construct a portfolio with exactly k assets (Cardinality Constraint) from the universe N, we apply a **Direct Index Mapping with Linear Probing** procedure. The algorithm iterates through the selection variables until k unique assets are identified:

1. **Index Mapping:** For each variable x_i (where $0 \leq x_i \leq 1$), a candidate asset index (C_{idx}) is computed by scaling the value to the universe size:

$$C_{idx} = \lfloor x_i \times N \rfloor \tag{3}$$

The value is clamped to ensure it lies within $[0, N-1]$.

2. **Collision Resolution (Linear Probing):** If the asset at C_{idx} has already been selected (a collision), the algorithm searches for the next available asset using a circular linear probe:

$$C'_{idx} = (C_{idx} + 1) \pmod{N}$$

This step repeats until a non-selected asset is found.

3. **Selection & Weighting:** Once a unique asset index j is found, it is added to the portfolio. Its raw weight is retrieved from the corresponding allocation variable ($w'_j = x_{N+j}$).
4. **Termination:** The process repeats for subsequent variables (x_{i+1}) until exactly k unique assets are selected.
5. **Normalization:** Finally, the weights of the k selected assets are normalized to sum to unity ($w_j = w'_j / \sum w'$).

3.2.2. Illustrative Example

Consider a small universe of $N = 4$ assets $\{A, B, C, D\}$ (indices 0, 1, 2, 3) and a target cardinality $k = 2$. Suppose the solver generates:

$$\vec{x} = [\underbrace{0.20, 0.22, \ldots}_{\text{Selection}}, \underbrace{0.5, 0.4, 0.9, 0.1}_{\text{Weights}}]$$

1. **First Selection ($x_1 = 0.20$):**

$$Index = \lfloor 0.20 \times 4 \rfloor = 0 \rightarrow \text{Asset A is selected.}$$

Current Set: $\{A\}$.

2. **Second Selection ($x_2 = 0.22$):**

$$Index = \lfloor 0.22 \times 4 \rfloor = 0 \rightarrow \text{Collision! (A is already in set).}$$

Collision Resolution: Try next index: $(0 + 1) \pmod 4 = 1$. Asset at index 1 is B. B is free.

$$\rightarrow \text{Asset B is selected.}$$

Current Set: $\{A, B\}$. Target size $k = 2$ reached.

3. **Weight Extraction:** Retrieve weights for indices 0 (A) and 1 (B).

$$w'_A = 0.5 \quad (\text{from } x_{N+0}), \quad w'_B = 0.4 \quad (\text{from } x_{N+1})$$

4. **Normalization:** Sum $= 0.5 + 0.4 = 0.9$.

$$w_A = 0.5/0.9 \approx 0.56, \quad w_B = 0.4/0.9 \approx 0.44$$

3.3. Investment Constraints and Bounds

The solution encoding scheme is a key factor when solving optimization problems with metaheuristics. We adopt a continuous solution representation. Each solution is encoded as a vector $\vec{x} \in \mathbb{R}^{2N}$, where N is the total number of assets or financial companies under consideration (the universe).

The first N variables of $\vec{x}$ (i.e., $x_1, \ldots, x_N$) represent the selected companies to invest in, chosen from among all available options in the market. The remaining N variables (i.e., $x_{N+1}, \ldots, x_{2N}$) specify the corresponding investment percentages.

To promote diversification and reflect practical investment constraints, lower and upper bounds are imposed on individual asset allocations. These bounds are determined by the target portfolio cardinality, denoted as k (the number of assets to be selected):

$$MIN_{inv} = \frac{1.0}{2k} \quad ; \quad MAX_{inv} = \frac{2.0}{k} \tag{4}$$

where

- MIN_{inv} is the minimum allowable investment percentage in any individual asset.
- MAX_{inv} is the maximum allowable investment percentage in any individual asset.
- k is the target number of assets in the portfolio (e.g., $k = 5$ or $k = 20$).

Thus, the investment for any selected asset must lie within the interval $[MIN_{inv}, MAX_{inv}]$. For instance, in a scenario where $k = 20$, the minimum position size is $1/(2 \times 20) = 0.025$ (2.5%), ensuring that the portfolio is not diluted into negligible positions even if the search space N is very large (e.g., 6000 assets).

This strategy ensures a balanced portfolio distribution, avoids concentration risk, and reflects realistic investment constraints.

3.4. Evaluating and Manipulating Solutions

The use of a continuous solution encoding enables the application of not only multi-objective evolutionary algorithms but also algorithms specifically designed for continuous domains, such as particle swarm optimization (PSO) techniques. In the case of evolutionary algorithms, the standard crossover and mutation operators for continuous representations available in the `jMetal` framework can be directly applied. Consequently, algorithms such as NSGA-II, MOEA/D, and SMS-EMOA are suitable for solving instances of the bi-objective portfolio optimization problem.

As noted in the previous section, the total investment in any solution must sum to 1. However, applying crossover and mutation operators may generate solutions that violate this constraint. To address this, we adopt a strategy in which normalization is performed immediately before evaluating a new solution. This approach avoids the need to alter the standard variation operators to ensure feasibility, maintaining compatibility with existing algorithm implementations while preserving the validity of the solution space.

4. Data Acquisition and Processing

A key contribution of this work is the use of real financial market data instead of synthetic datasets. To this end, we have developed a comprehensive strategy for acquiring and preprocessing market data to generate valid instances of a multi-objective optimization problem suitable for metaheuristic techniques. This strategy required the development of a supporting software infrastructure that ensures continuous access to up-to-date information via specialized Application Programming Interface (APIs).

In addition, a strategic filtering mechanism has been implemented to extract relevant subsets of assets based on criteria such as market capitalization, trading volume, historical volatility, and sector classification. These filtered subsets are used to define distinct portfolio profiles (e.g., high-liquidity technology stocks, low-risk diversified portfolios), which serve as the foundation for the optimization process.

Our complete dataset encompasses more than 11,000 financial assets listed on the major U.S. stock exchanges, briefly characterized as follows:

- **ARCA (NYSE Arca)**: Specializes in the trading of Exchange-Traded Funds (ETFs) and derivatives; recognized as the global leader in ETF trading volume.
- **NYSE (New York Stock Exchange)**: The largest stock exchange in the world by market capitalization and monetary volume, featuring highly liquid and reputable companies.
- **NASDAQ**: The second-largest exchange, known for listing technology-driven and growth-oriented companies.
- **OTC (Over-The-Counter)**: A decentralized market for trading financial instruments directly between parties, including unlisted stocks, corporate bonds, and derivatives.
- **AMEX (American Stock Exchange)**: Historically focused on small- and mid-cap companies, offering specialized securities and investment funds.

- **BATS Global Markets**: One of the largest electronic trading platforms globally, now merged with CBOE, known for high-performance execution and low-latency infrastructure.

4.1. Temporal Selection and Justification

Although the full dataset spans more than 20 years, analyzing the entire time horizon simultaneously is neither computationally efficient nor methodologically appropriate. Instead, we conduct experiments within delimited temporal windows (e.g., days, months, or years), which allows for improved computational resource management, enhanced experimental replicability, and more meaningful comparative analysis [38].

We selected the year 2023 as the primary analysis period due to its recency and relevance. This year encompasses significant financial events, including heightened market volatility, post-pandemic recovery dynamics, fluctuations in interest rates, and macroeconomic developments that directly impact portfolio performance.

Extending the analysis to multi-year periods is highly recommended to evaluate the robustness and stability of the proposed solutions under diverse market conditions, including financial crises, speculative bubbles, and periods of sustained economic growth. Such longitudinal analysis also enables a more comprehensive investigation into the influence of exogenous factors (e.g., global pandemics, geopolitical conflicts, or regulatory changes) on optimal portfolio configurations.

The inclusion of multi-annual data in future research will be crucial for validating the robustness and adaptability of portfolio optimization models. It is important to note that, as this study is limited to the 2023 market regime, the robustness of the proposed portfolios during varying economic cycles (e.g., recessions or speculative bubbles) remains to be tested. While an analysis limited to a single year can offer valuable insights into performance under specific market conditions, we acknowledge that extending the evaluation across multiple time periods, encompassing complete economic cycles, growth phases, recessions, and disruptive events, would allow for a deeper understanding of how investment strategies behave under diverse pressures. Adopting this longitudinal approach in future research will not only strengthen the external validity of the findings but also facilitate the identification of persistent patterns and the resilience of proposed solutions against the inherent volatility of financial markets.

4.2. Strategic Grouping and Segmentation

Within each market, a series of grouping and segmentation mechanisms have been applied to categorize assets according to criteria relevant to the optimization process:

- Average daily trading volume (as an indicator of liquidity).
- Market capitalization (distinguishing between large-cap, mid-cap, and small-cap companies).
- Economic sector or industry (based on GICS or NAICS classification).
- Historical volatility (as an estimator of systemic risk).
- Availability of fundamental and technical financial data.

These criteria enable the construction of customized datasets tailored to specific portfolio profiles, for example, high-liquidity technology portfolios or low-risk, sector-diversified portfolios.

This study focuses on company sets organized by market and segmented by average trading volume. Such segmentation facilitates the identification of behavioral patterns driven by liquidity and market-specific dynamics. Furthermore, this structure supports comparative analyses across groups with similar characteristics, offering a more controlled and interpretable framework for evaluating the effectiveness of multi-objective optimization strategies.

4.3. Data Acquisition

To enable rigorous analysis in the construction and optimization of investment portfolios, the initial stage focuses on the systematic acquisition and subsequent processing of financial data. This process is divided into two interconnected phases:

- **Retrieval of Historical Financial Data from the Broker**: The first phase involves the implementation of a script that establishes a programmatic connection with a financial broker via an API. This connection allows for the automated retrieval of historical data for the financial assets selected for the study. Then, based on the historical price data, an additional file is generated for each asset, containing the total return for the analysis period, which spans one year, as previously mentioned. The total return R for the entire period is explicitly defined as the **Simple Annual Return**. It is calculated as the percentage change between the closing price on the last trading day of the year (P_f) and the closing price on the first trading day (P_i):

$$R = \frac{P_f - P_i}{P_i} \tag{5}$$

  Consequently, the variance–covariance matrix is computed using the **daily returns** of the assets throughout the 2023 trading year (approximately 252 trading days) and is subsequently annualized to ensure consistency with the return vector R.
  This return file for each asset summarizes the overall performance of the asset throughout the study period, providing an aggregated measure that serves as a reference for performance evaluation. To maintain an organized data structure and facilitate later access, the historical data for each asset is stored in a separate Comma-Separated Values (CSV) file. All files are consolidated in a dedicated directory, which serves as the initial repository of raw financial information. This process is illustrated in Figure 1(1).
- **Construction of the Analytical Dataset**: After retrieving the raw data, a second phase of processing is performed to extract structured and relevant information for portfolio analysis. This stage results in the generation of two key datasets: the return table and the covariance table, which are essential inputs for the portfolio optimization process. This transformation pipeline is shown in Figure 1(2).

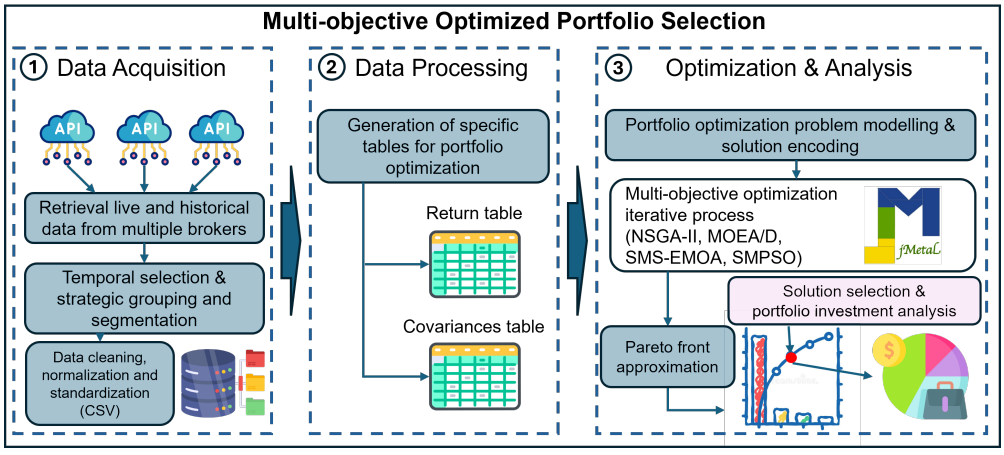

Figure 1. General overview of the multi-objective proposed approach for portfolio optimization. The workflow consists in three main phases of: (**1**) data acquisition and standardization, (**2**) specific data tables processing and (**3**) algorithmic optimization.

The **returns table** contains the identifier (ticker or name) of each analyzed asset along with its corresponding return for the defined evaluation period. This structured format

enables direct comparison of asset performance and serves as a primary input for the portfolio optimization process. An illustrative example is shown in Table 1.

Table 1. Example of a returns table showing the individual returns of selected assets over the evaluation period.

Asset	Return
IT	0.7490606452503557
CACC	0.7162124534909408
EPAM	0.6525867207231031
ROG	0.5849993443006836
AZO	0.574363453765806
HUBS	0.530616114292918
...	...
TEAM	0.4987212212400905

The **variance–covariance matrix** is a fundamental component for quantifying both the individual risk of each asset and the relationships between asset pairs. In this matrix, the *diagonal elements* represent the variances of individual assets, while the *off-diagonal elements* correspond to the covariances between different asset pairs. Formally, the variance of an asset x_i is defined as the covariance of the asset with itself:

$$\mathrm{Var}(x_i) = \mathrm{Cov}(x_i, x_i)$$

For example, the variance of the asset labeled IT is given by:

$$\mathrm{Var}(\mathrm{IT}) = \mathrm{Cov}(\mathrm{IT}, \mathrm{IT})$$

This identity emphasizes that variance is a special case of covariance, where both arguments refer to the same asset.

The covariance between two assets i and j is calculated as:

$$\mathrm{Cov}(i, j) = \mathbb{E}[(r_i - \mu_i)(r_j - \mu_j)]$$

where μ_i and μ_j are the expected returns of assets i and j, respectively. The resulting covariance matrix, which assesses the shared risk relationship between the assets, is stored in a file that will be used by the optimization system to assess the total risk of the portfolio. Table 2 illustrates a sample covariance matrix.

Table 2. Example of a covariance matrix. Diagonal elements correspond to asset variances; off-diagonal elements denote covariances between asset pairs.

	IT	CACC	EPAM	ROG	AZO
IT	0.0000	0.0002	0.0001	0.0001	0.0001
CACC	0.0002	0.0003	0.0001	0.0002	0.0001
EPAM	0.0001	0.0001	0.0001	0.0001	0.0001
ROG	0.0001	0.0002	0.0001	0.0002	0.0001
AZO	0.0001	0.0001	0.0001	0.0001	0.0001

The availability of both the *returns vector* and the *covariance matrix* is essential for formulating instances of the multi-objective portfolio optimization problem solvable by metaheuristics within the `jMetal` framework. These datasets define the input parameters for the objective functions and guide the search process of algorithms such as NSGA-II or MOEA/D. For each experimental run, a consistent pair of return and covariance

files are supplied, ensuring that evaluated solutions are based on valid and synchronized market data.

5. Experimentation

This section is devoted to detail the experimental study we have conducted and to analyze the obtained results. Previously, the multi-objective metaheuristics used are described.

5.1. Multi-Objective Algorithms

We briefly describe the multi-objective metaheuristics selected for comparison in solving the portfolio selection problem. These algorithms take as an initial source of information the modeling procedure of the problem and the encoding of solutions, so that after an iterative process of evolutionary refinement, they produce the results in the form of a Pareto front. This process is illustrated in Figure 1 step (3). These algorithms have been selected and adapted in this approach, since they constitute a representative set of well-grounded techniques that evolve heterogeneous learning models.

NSGA-II [10] is one of the most widely used multi-objective metaheuristics. It ranks the population using a non-dominated sorting strategy that organizes individuals into fronts of non-dominated solutions, promoting convergence toward the Pareto front. To maintain diversity, it employs a density estimator known as *crowding distance*, which encourages the spread of solutions across the objective space.

MOEA/D [11] decomposes a multi-objective optimization problem into multiple single-objective sub-problems, each associated with a different weight vector using an aggregation approach. The optimization of each sub-problem is guided by information shared among neighboring sub-problems, fostering both convergence and diversity.

SMS-EMOA [12] shares certain components with NSGA-II, such as non-dominated sorting, but differs in its selection mechanism. Instead of using crowding distance, SMS-EMOA relies on *hypervolume contribution* as a density estimator, favoring solutions that contribute most to the overall hypervolume. It also differs in selection dynamics: SMS-EMOA follows a *steady-state* approach, whereas NSGA-II employs a *generational* scheme.

SMPSO [13] is a multi-objective extension of the particle swarm optimization (PSO) algorithm. Each particle encodes a potential solution and moves within the search space influenced by its personal best and a global best, selected from an external bounded archive. This archive is managed using crowding distance to maintain diversity. A constrained velocity update mechanism is introduced to prevent erratic particle movements and maintain stable convergence and a perturbation step (using a mutation operator) is included to promote the diversification of the search.

The parameter settings of the compared algorithms are included in Table 3. In all the algorithms share as common settings a population or swarm size of 100 solutions or particles, and the use of a polynomial mutation operator.

Table 3. Parameter settings of the compared multi-objective metaheuristics (n is the number of decision variables).

Common Parameters	**Value**
Population/Swarm size	100 individuals/particles
Mutation operator	Polynomial
Mutation probability	$1.0/n$
Mutation distribution index	20.0

Table 3. *Cont.*

NSGA-II & SMS-EMOA & MOEA/D	
Crossover operator	Simulated Binary (SBX)
Crossover probability	0.9
Crossover distribution index	20.0
MOEA/D	
Normalize objectives	False
Aggregation function	Penalty Boundary Intersection (PBI)
PBI θ value	5.0
Neighborhood size (T)	20
Neighborhood selection prob. (δ)	0.9
Max. replaced solutions (n_r)	2
SMPSO	
Archive Size	100
C_1, C_2	rand [1.5, 2.5]
Inertia Weight (W)	0.1
Lower limit velocity change	-1.0
Upper limit velocity change	**1.0**

5.2. Datasets

As discussed in Section 4, we have selected six of the largest stock market data sources: NYSE, NASDAQ, OTC, AMEX, ARCA, and BATS. From these, we have defined two types of datasets. In the first group, the problem instances define a search space consisting of the top 100 companies with the highest trading volume (liquidity) in each market. The optimization goal is to identify the optimal subset of 5 assets (cardinality constraint) that maximizes return and minimizes risk within this pool. The resulting portfolio instances are named `NYSE_5_100`, `NASDAQ_5_100`, and so forth. In addition, in order to create generalized scenarios, we have also included an aggregated instance combining all markets, named `ALL_5_100`.

In the second group, we have extended the study to consider all markets together, but in this case by selecting 20 companies among the top 100, 500, 1000, 3000, and 6000 ones. These problem instances are referred to as `ALL_20_100`, `ALL_20_500`, `ALL_20_1000`, `ALL_20_3000`, and `ALL_20_6000`.

All these datasets constitute a varied and realistic benchmark for algorithmic evaluation in portfolio optimization scenarios, so they are made available to research and academic communities through the project's GitHub repository (https://github.com/AntHidMar/Portfolio_Study_DataSources (accessed on 17 November 2025)).

5.3. Experimental Methodology

To ensure statistically significant results and account for the stochastic nature of the metaheuristics, each problem instance was solved independently 30 times per algorithm. This approach provides a sufficiently large sample size to compute reliable average performance metrics and analyze the variability of the algorithms under study.

As the complexity of the problem instance varies, the required computational effort must be adjusted accordingly. Therefore, rather than using a fixed number of function evaluations per algorithm execution, we use a base of 50,000 evaluations for all instances of type `5_100`, and multiply this number by 2, 3, 4, and 5 for instances of type `20_500`, `20_1000`, `20_3000`, and `20_6000`, respectively.

To assess algorithm performance, we employed two widely used quality indicators: the generational distance plus (I_{IGD+}), which measures the average distance from each point

in a reference set to the nearest point in the approximation set, providing a comprehensive assessment of both convergence and diversity while being weakly Pareto compliant [39], and the hypervolume indicator (I_{HV}), which assesses both convergence and diversity by calculating the volume of the objective space dominated by the approximation set bounded by a reference point [40].

Both indicators require suitable reference elements: I_{IGD+} requires a reference Pareto front to evaluate approximation quality, while I_{HV} needs a reference point, typically derived from the extreme objective values of the reference Pareto front. To ensure reproducibility, the reference point z_{ref} is computed using the Nadir point of the aggregated reference front (PF_{ref}), applying a slight offset to preserve the contribution of boundary solutions. Specifically, the coordinates are defined as $z_i = \max_{s \in PF_{ref}} f_i(s) \times 1.1$ for each objective i. This scaling factor ensures that all extreme non-dominated solutions are strictly dominated by the reference point and thus contribute to the hypervolume value. Since we are dealing with a real-world problem and the true Pareto front is unknown, we construct an approximate reference front by merging all the non-dominated solutions obtained from all runs of all algorithms.

To enable statistical analysis, we include tables of indicator values reporting the median and interquartile range as measures of central tendency and dispersion. In these tables, we highlight the best and second-best indicator values for each problem with dark and light gray backgrounds, respectively. To assess whether differences between algorithms are significant, we provide tables with the results of applying the Wilcoxon rank-sum test at a 5% significance level.

We complement these results with Critical Difference (CD) plots [41], which summarize the average ranks of the algorithms across multiple problems. Algorithms connected by horizontal lines are not significantly different according to the CD criterion at the chosen confidence level. The CD value indicates the minimum rank difference required to establish statistical significance.

To ensure full reproducibility of the experimental results, the optimization framework was implemented using jMetal version 6.0 running on Java JDK 18.0.1. Regarding the stochastic nature of the algorithms, the pseudo-random number generation was handled by the `JMetalRandom` class, initialized with a fixed reference seed (12345). This configuration ensures that the sequence of random numbers used across the 30 independent runs is deterministic and identical across replications. Finally, the financial data snapshots used for these experiments correspond to the market state as of 17 November 2025, consistent with the datasets deposited in the public repository.

The experiments were conducted on a standalone machine deployed in a local system, equipped with an Intel i7 processor running at 2.5 GHz. The system has 8 cores, 32 GB of RAM, and SSD storage, providing a robust yet accessible environment for the evaluation. The operating system is Microsoft Windows 11 Pro, and we have used Java JDK 18.0.1.

Although the hardware supports parallel processing, the execution was carried out using a single core due to limitations imposed by the Windows operating system. Specifically, the current implementation does not support concurrent access to multiple files from different processes, which prevents efficient parallel execution. For future improvements, we plan to adapt the system to overcome this limitation and take full advantage of parallelism.

An execution consisting of one instance of the problem ALL_5_100 requires an average runtime of 14 s, whereas the corresponding run for the ALL_20_6000 problem takes approximately 1 min. This difference highlights the substantial impact of problem size on computational time, which is influenced not only by the increased number of assets and the greater complexity of the efficient frontier to be estimated, but also by the higher number of evaluations needed to obtain high-quality solutions in large-scale instances.

6. Results

In this section, we present the results obtained in the experimentation phase. First, we perform a quantitative analysis following the experimental methodology described in the preceding section. Second, we conduct a qualitative analysis to provide guidance to decision makers on which algorithms appear most promising, depending on their particular investment preferences. In this regard, an analysis of representative solutions in form of actual portfolios suggested by the proposal is also conducted, from the perspective of the domain expert in portfolio investment generation.

6.1. Quantitative Analysis

The first observation in this analysis concerns the results of the indicator I_{IGD+}, which are presented in Table 4, which contains the median and interquartile range of the algorithm-independent runs conducted in the experiments. In this table, we can observe that SMS-EMOA achieves the best (lowest) indicator values across the twelve problem instances used. Consequently, according to this indicator, SMS-EMOA is able to find fronts with the best degree of convergence and diversity towards the Pareto front. NSGA-II yields the second-best results in ten out of the twelve problems. Notably, there are no significant differences between the two groups of problem instances, i.e., the performance of the algorithms does not appear to depend on whether the optimization focuses on individual markets or considers all markets jointly.

Table 4. Median and Interquartile Range of the I_{IGD+} quality indicator values.

	SMPSO	**MOEA/D**	**SMS-EMOA**	**NSGA-II**
NYSE_5_100	$1.07e-02_{1.4e-03}$	$6.99e-03_{2.2e-03}$	$2.70e-03_{5.8e-04}$	$4.79e-03_{4.8e-04}$
NASDAQ_5_100	$1.29e-02_{2.8e-03}$	$8.13e-03_{3.3e-03}$	$3.65e-03_{8.7e-04}$	$5.14e-03_{7.0e-04}$
OTC_5_100	$7.86e-03_{2.2e-03}$	$4.84e-03_{1.2e-03}$	$2.06e-03_{9.6e-05}$	$3.65e-03_{1.5e-04}$
AMEX_5_100	$1.70e-02_{2.5e-03}$	$4.70e-03_{1.1e-03}$	$2.72e-03_{1.5e-04}$	$4.87e-03_{2.9e-04}$
ARCA_5_100	$9.17e-03_{1.1e-03}$	$5.80e-03_{2.6e-03}$	$3.47e-03_{1.1e-03}$	$5.71e-03_{1.0e-03}$
BATS_5_100	$8.62e-03_{1.1e-03}$	$4.60e-03_{6.1e-04}$	$2.82e-03_{1.8e-04}$	$4.89e-03_{2.8e-04}$
ALL_5_100	$9.85e-03_{2.3e-03}$	$8.69e-03_{4.5e-03}$	$2.94e-03_{3.5e-04}$	$4.81e-03_{2.9e-04}$
ALL_20_100	$5.40e-02_{3.8e-03}$	$3.16e-02_{5.6e-03}$	$9.61e-03_{1.7e-03}$	$1.01e-02_{1.9e-03}$
ALL_20_500	$8.53e-02_{5.4e-03}$	$4.44e-02_{6.8e-03}$	$1.65e-02_{3.1e-03}$	$1.71e-02_{3.5e-03}$
ALL_20_1000	$6.21e-02_{2.2e-03}$	$4.03e-02_{6.0e-03}$	$1.62e-02_{5.2e-03}$	$2.21e-02_{5.9e-03}$
ALL_20_3000	$6.90e-02_{6.4e-03}$	$2.79e-02_{7.6e-03}$	$1.40e-02_{5.8e-03}$	$2.49e-02_{7.1e-03}$
ALL_20_6000	$9.06e-02_{8.4e-03}$	$3.49e-02_{8.5e-03}$	$1.80e-02_{8.4e-03}$	$2.90e-02_{7.7e-03}$

Statistically, these results are also checked with a Wilcoxon rank-sum test, as shown in Table 5. Each cell in this table contains a symbol representing one of the twelve problems considered. There are three symbols: '–" indicates no statistically significant difference between the algorithms in the corresponding row and column; ▲ indicates that the algorithm in the row is significantly better; and ▽ indicates that the algorithm in the column is significantly better than the corresponding one in the row. The results show statistical significance in most comparisons. In particular, when comparing SMS-EMOA and NSGA-II, the differences are statistically significant in favor of SMS-EMOA in nine problems, and there is no significant difference in two.

Table 5. Wilcoxon rank-sum test of the I_{IGD+} indicator. The symbols ▲, ▽, and − indicate that the algorithm significantly outperforms, is significantly outperformed by, or has no statistically significant difference compared to the reference algorithm, respectively, according to the Wilcoxon rank-sum test with a 5% significance level.

	MOEA/D	SMS-EMOA	NSGA-II
SMPSO	▽ ▽ ▽

Table 6. Median and Interquartile Range of the I_{HV} quality indicator values.

	SMPSO	MOEAD	SMS-EMOA	NSGAII
NYSE_5_100	$7.24e-01_{2.2e-03}$	$7.26e-01_{9.8e-03}$	$7.38e-01_{2.2e-03}$	$7.34e-01_{8.8e-04}$
NASDAQ_5_100	$5.72e-01_{5.0e-03}$	$5.76e-01_{1.5e-02}$	$5.91e-01_{2.8e-03}$	$5.87e-01_{2.6e-03}$
OTC_5_100	$7.36e-01_{5.0e-03}$	$7.33e-01_{1.1e-02}$	$7.49e-01_{1.1e-03}$	$7.47e-01_{8.4e-04}$
AMEX_5_100	$7.08e-01_{3.3e-03}$	$7.19e-01_{3.3e-03}$	$7.28e-01_{8.0e-04}$	$7.25e-01_{6.6e-04}$
ARCA_5_100	$6.11e-01_{2.0e-03}$	$6.10e-01_{4.5e-03}$	$6.18e-01_{1.7e-03}$	$6.14e-01_{1.1e-03}$
BATS_5_100	$5.97e-01_{2.3e-03}$	$6.00e-01_{4.6e-03}$	$6.07e-01_{8.1e-04}$	$6.04e-01_{9.1e-04}$
ALL_5_100	$5.85e-01_{3.6e-03}$	$5.84e-01_{1.8e-02}$	$5.96e-01_{1.8e-03}$	$5.93e-01_{1.9e-03}$
ALL_20_100	$5.72e-01_{7.3e-03}$	$5.91e-01_{1.6e-02}$	$6.43e-01_{7.4e-03}$	$6.46e-01_{7.1e-03}$
ALL_20_500	$5.36e-01_{7.2e-03}$	$5.83e-01_{1.6e-02}$	$6.39e-01_{8.4e-03}$	$6.40e-01_{7.9e-03}$
ALL_20_1000	$5.58e-01_{4.0e-03}$	$5.82e-01_{1.5e-02}$	$6.46e-01_{7.4e-03}$	$6.44e-01_{1.3e-02}$
ALL_20_3000	$6.00e-01_{1.7e-02}$	$6.64e-01_{1.5e-02}$	$7.01e-01_{1.0e-02}$	$6.89e-01_{1.3e-02}$
ALL_20_6000	$5.80e-01_{1.9e-02}$	$6.65e-01_{1.3e-02}$	$7.12e-01_{1.6e-02}$	$6.99e-01_{1.6e-02}$

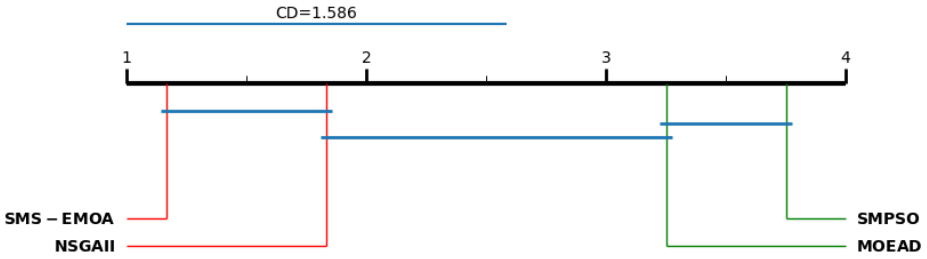

Figure 3. Critical difference (CD) plot. Quality indicator: I_{HV}.

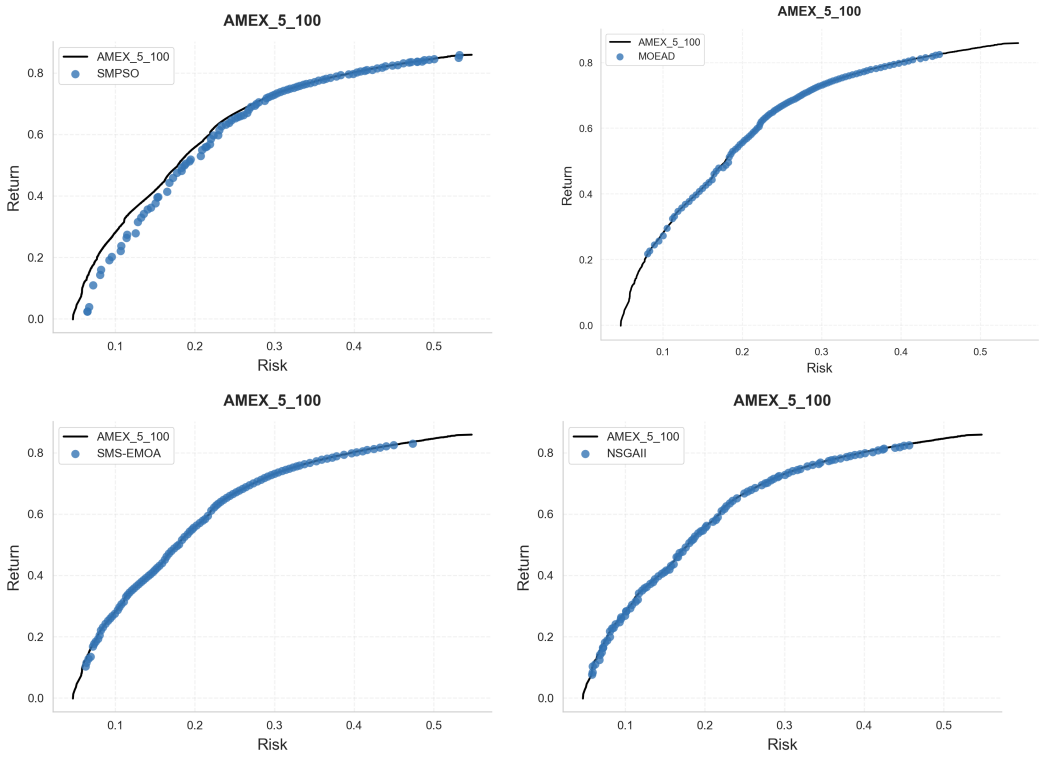

Figure 4. Examples of fronts obtained by SMPSO, MOEA/D, SMS-EMOA, and NSGA-II for the `AMEX_5_100` problem. Each front corresponds to the one having the median I_{HV} indicator value. The reference front is shown as a continuous line.

Table 7. Wilcoxon rank-sum test of the I_{HV} indicator. The symbols ▲, ▽, and − indicate that the algorithm significantly outperforms, is significantly outperformed by, or has no statistically significant difference compared to the reference algorithm, respectively, according to the Wilcoxon rank-sum test with a 5% significance level.

	MOEA/D	SMS-EMOA	NSGAII
SMPSO	− − ▲ ▽ − ▽ − ▽		
MOEAD		▽ ▽	
SMS-EMOA			▲ ▲ ▲ ▲ ▲ ▲ ▲ − − − ▲ ▲

6.2. Qualitative Analysis

While the quantitative analysis enables us to assess algorithm performance in terms of convergence and diversity, it is important to consider that the approximation fronts produced by the optimizers will ultimately be evaluated by a decision maker, i.e., an expert in the problem domain, who selects solutions based on their preferences. In this section, we perform a qualitative analysis by examining the extent to which each algorithm contributes to the reference front generated for each problem instance. This allows us to identify algorithms that, due to the nature of their learning models, tend to favor the optimization of one objective over another, as well as those that display more balanced behavior, with better diversity and coverage of the central regions of the reference front. Such analysis supports domain experts in selecting the most appropriate algorithm based on their interest in a particular objective, such as minimizing risk or maximizing return.

We have selected two problem instances: `BATS_5_100`, which represents a problem focused on a specific market, and `ALL_20_6000`, a more complex instance that combines all markets and, consequently, involves a significantly larger search space.

Figure 5 shows the reference front computed from all the executions (depicted as a continuous line), along with the contributions of each algorithm to this front for the specific market instance `BATS_5_100`. It can be observed that SMS-EMOA contributes the most to the reference front. However, like NSGA-II and MOEA/D, it fails to find solutions in the extreme region corresponding to the highest risk versus highest return (the uppermost part of the reference front). The solutions in this extreme region are found by SMPSO, as was also observed for the `AMEX_5_100` instance discussed in the previous section. Nevertheless, SMPSO performs poorly in the rest of the objective space. NSGA-II and MOEA/D cover a similar portion of the reference front, but their gaps are filled by solutions contributed by SMS-EMOA.

We analyze now the contribution of the four algorithms to the problem instance `ALL_20_6000`. In this complex scenario, we can observe that, except for SMS-EMOA, the algorithms concentrate their contributions in distinct, non-overlapping regions of the reference front. This suggests that each solver may be more suitable depending on the preferred region of the decision maker. Specifically, SMPSO and NSGA-II are preferable when solutions in the extreme regions are desired: the top-right part of the front (high risk and high return) is covered by SMPSO, while NSGA-II finds solutions in the bottom-left region (low risk and low return). SMS-EMOA and MOEA/D, on the other hand, contribute to the central portion of the reference front, offering a balanced trade-off between the two objectives.

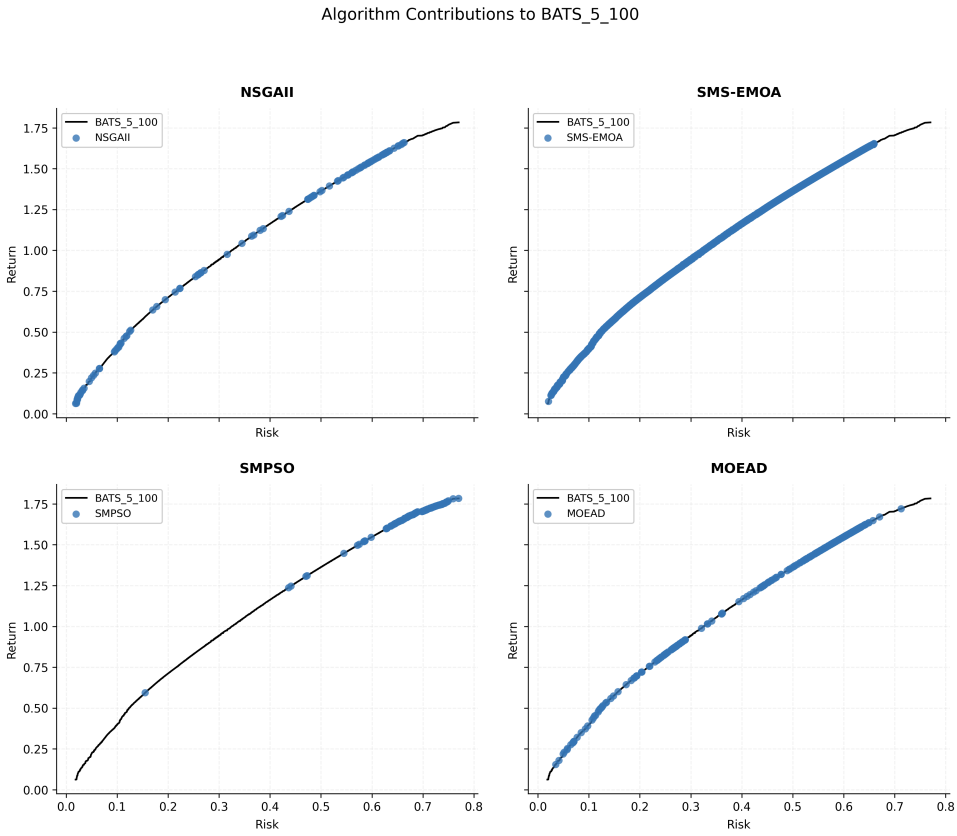

Figure 5. Contribution of the algorithms to the reference front for instance `BATS_5_100`.

6.3. Analysis of Representative Portfolios

In addition to the quantitative and qualitative analysis of the algorithms, it is essential to explore concrete examples of portfolios obtained along the efficient frontiers, in order to illustrate how the risk/return trade-offs manifest in practice and to provide guidance for end investors.

To do so, the Sharpe ratio indicator [14] is used in this section as a fitness value, to evaluate the profitability of an investment relative to the risk assumed. Sharpe's ratio is commonly used to evaluate mutual funds, stock portfolios, and investment strategies. It is formulated as shown in Equation (6):

$$SR = \frac{(R_p - R_f)}{\sigma_p} \tag{6}$$

where

- R_p is the expected return of the investment or portfolio.
- R_f is the risk-free rate of return, such as the interest rate on a government bond. For this study, we defined a strict risk-free rate of $R_f = 0.0697$ (6.97%). This value serves as a high performance threshold, exceeding the average 2023 Treasury yields, to enforce a conservative evaluation and ensure that the selected portfolios demonstrate robust returns even under demanding opportunity cost scenarios.
- σ_p is the volatility or standard deviation of the investment or portfolio.

This metric evaluates the excess return obtained for each additional unit of risk assumed relative to a risk-free investment. A high Sharpe ratio denotes an optimal risk–reward relationship, indicating that the asset or portfolio has generated higher returns relative to the risk assumed.

Among the multiple frontiers generated, the following configurations stand out due to their practical relevance, namely: *High-return extreme portfolio*, *Low-Risk Portfolio*, *Balanced Portfolios*, and *Portfolios with Sectoral or Liquidity Coverage*; which are detailed next.

6.3.1. High-Return Extreme Portfolio

In several instances such as `BATS_5_100` and `ALL_20_6000` (see Figure 6), SMPSO is characterized by generating portfolios located in the upper-right extreme of the frontier. As expected, these portfolios exhibit the highest expected returns and the highest associated risk, making them suitable only for investors with high risk tolerance, such as hedge funds or speculative traders. Their value lies in showcasing the upper limits of return achievable with the available set of assets [6].

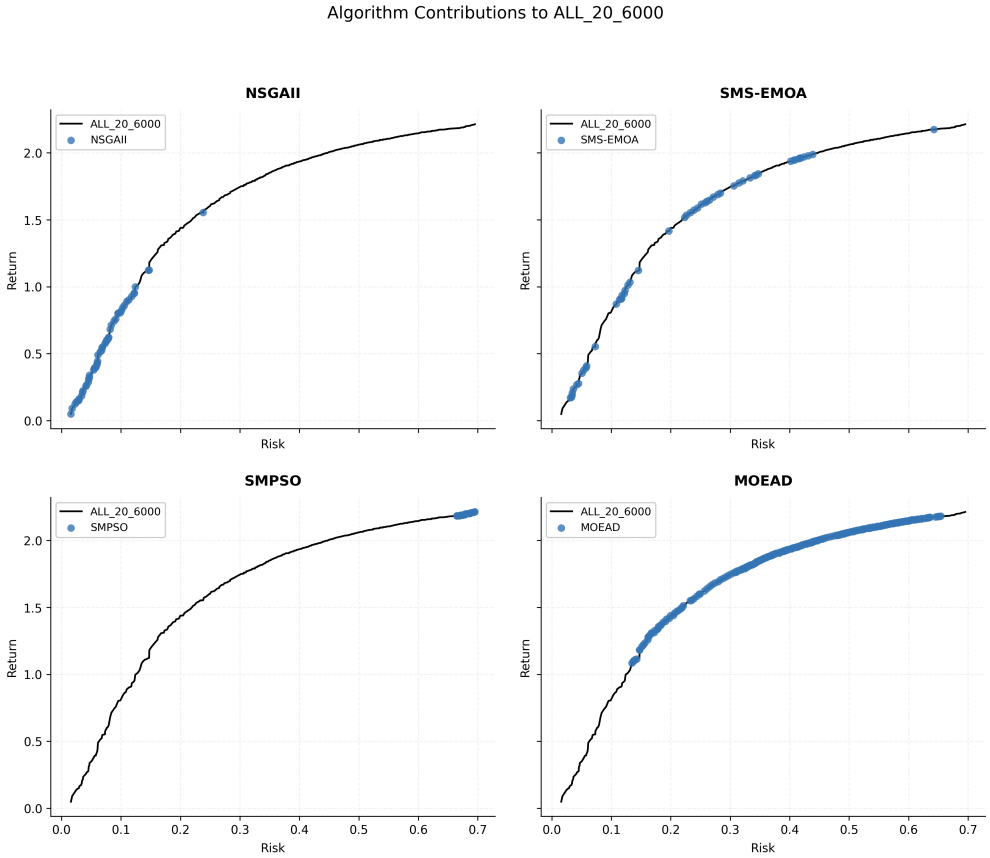

Figure 6. Contribution of the algorithms to the reference front for instance `ALL_20_6000`.

Table 8 presents a **high-return extreme portfolio** located on the efficient frontier generated by the SMPSO algorithm. This portfolio stands out for having one of the highest expected returns (218.85%), along with considerable risk (67.89%), and a calculated fitness value of 2.5716 (based on the Sharpe metric described previously).

This portfolio reflects an aggressive strategy, with a high concentration of capital in six main assets, each receiving around 10% of the total investment. The remaining assets, each around 2.5%, provide marginal diversification and are often linked to volatile sectors such as cryptocurrencies or biotechnology.

The fact that SMPSO was able to find this specific portfolio indicates that, within the search space defined by the model, there exist configurations that allow for exceptionally high returns, demonstrating the algorithm's ability to identify regions of the efficient frontier that strongly prioritize return.

Table 8. SMPSO—ALL_20_6000—Composition of a high-return portfolio generated (Return: 218.85%; Risk: 67.89%; Fitness: 2.5716).

Asset	Weight
CVNY	10.00%
CVNX	10.00%
CVNA	10.00%
AAOI	9.99%
ACIC	9.98%
GREEL	9.98%
CONL	7.53%
CONNQ	2.50%
MARA	2.50%
BITF	2.50%
BITI	2.50%
CIFRW	2.50%
CIFR	2.52%
BTCC	2.50%
BTCL	2.50%
BTCI	2.50%
BTC	2.50%
BTBT	2.51%
EYPT	2.50%
EYUBY	2.50%

However, such high returns do not come without cost, since they are accompanied by equally high levels of risk, implying significant volatility in expected portfolio performance. This relationship points to a very specific investor profile, characterized by a high risk tolerance. These types of portfolios may be attractive to speculative traders, hedge fund managers, or institutional investors aiming to maximize gains over short time horizons and willing to assume significant potential losses.

From a technical standpoint, SMPSO's ability to reach these extremes can be attributed to its particle-based exploratory dynamics, which enable it to escape local optima and explore less visited regions of the solution space [6]. While other algorithms tend to concentrate in intermediate areas of the frontier, SMPSO shows a greater tendency to locate solutions in the high return–high risk region. (It is important to note that the assets in this portfolio are standard Stocks and ETFs traded on major U.S. exchanges, sharing the same commission structure as the rest of the 11,000 assets analyzed in this study. Consequently, our model assumes uniform transaction costs across the board. However, investors should consider that, depending on the specific broker and market liquidity at the time of execution, operational costs (such as bid-ask spreads) for high-volatility tickers could vary compared to blue-chip assets.)

Such solutions are valuable not only for their practical utility for certain investor profiles, but also because they help delineate the theoretical upper bounds of return achievable given the constraints and asset universe. Therefore, the presence of portfolios like this one highlights the importance of including algorithms with strong exploratory capabilities within the toolbox for portfolio optimization.

A second solution, also obtained from the set of portfolios generated by SMPSO is shown in Table 9 for the `AMEX_5_100` dataset. This portfolio shares a similar profile to the first one, maintaining a high return (81.09%) and risk (42.64%), which further confirms SMPSO's ability to locate the highest return and risk portfolios in any analyzed set.

Table 9. SMPSO —AMEX_5_100—Alternative portfolio with high return and moderate risk profile (Return: 81.09%; Risk: 42.64%; Fitness: 1.1351).

Asset	Weight
ARMP	37.69%
ACU	22.27%
BRIA	20.04%
DSS	10.00%
GAU	10.00%

This second portfolio solution shows similar behavior to the previous example, with the investment divided into two different subgroups. However, since this instance involves a smaller set of assets, the allocation appears less diffused. A significant allocation is made to the first subgroup, consisting of three assets (ARMP, ACU, and BRIA), which together absorb more than 80% of the capital, while the remaining investment is distributed between two additional assets. Although this reduces global diversification, it can be interpreted as a focused bet on high-potential stocks, generally associated with a growth strategy or aggressive stock picking, yet still under some degree of risk control.

The presence of such solutions in SMPSO's output confirms that the algorithm does not exclusively explore highly aggressive extremes, but can also identify more balanced configurations within a dynamic investment profile. Thus, this second example complements the previous one and suggests that SMPSO can be especially useful for designing personalized portfolios according to different risk tolerances, provided that its exploration parameters are properly tuned and that solutions are selected from the appropriate regions of the generated frontier.

6.3.2. Low-Risk Portfolio

NSGA-II produces solutions primarily located at the lower-left end of the efficient frontiers, particularly noticeable in problems involving large sets of assets, such as the `ALL_20_6000` instance. These portfolios, characterized by low volatility and moderate return, are ideal candidates for conservative or institutional investors seeking capital preservation, such as pension funds or insurance companies.

Table 10 shows a representative example of this type of portfolio. With a risk of just 4.10% and a moderate return of 26.58%, this optimized solution displays broad diversification without significant concentration in any single asset. Most asset allocations range between 2.5% and 9.5%, reflecting a design focused on stability and volatility control. Assets such as *GTIP*, *DJUN*, or *FLTB*, typically associated with inflation-protected bonds or stable funds, reinforce the hypothesis of a conservative investment approach. The fitness value of this portfolio (0.9680) confirms its position within the safest region of the efficient frontier.

However, NSGA-II also demonstrates a surprising ability to identify configurations with an excellent risk–return ratio beyond the traditional conservative profile. Table 11 presents one of the solutions with the best *fitness* value in the entire experiment (3.1280), combining a high return (156.72%) with a relatively low risk (22.92%). Despite maintaining significant diversification, this portfolio includes growth-oriented and technological assets (such as *MATH*, *AAOI*, *GEOS*, *ELTK*), enabling it to capture profit opportunities without incurring extreme volatility.

Table 10. NSGA —II-ALL_20_6000. Conservative portfolio generated (Return: 26.58%; Risk: 4.10%; Fitness: 0.9680).

Asset	Weight (%)
LIAW	2.50%
FNDE	2.50%
FUTU	4.59%
CRM	2.50%
GVI	5.48%
COLA	2.50%
KTTA	2.50%
BRBR	2.50%
APAM	2.50%
DRV	5.27%
GEOS	2.50%
FMAR	9.15%
GTIP	9.43%
LMBO	2.50%
GLLIW	7.47%
FLTB	9.40%
DJUN	9.31%
FXP	9.03%
AAUC	2.50%
LRGE	5.88%

Table 11. NSGA —II-ALL_20_6000. Portfolio generated with best return–risk ratio (Return: 156.72%; Risk: 22.92%; Fitness: 3.1280).

Asset	Weight (%)
MATH	6.99%
BBIO	3.39%
AAOI	6.47%
EYUBY	2.54%
GREEL	7.53%
ANG.PRB	4.18%
DHC	2.69%
GYRE	3.19%
LMB	8.42%
ACIC	3.25%
ANG.PRD	6.62%
CVNA	2.55%
ELTK	5.47%
CASI	5.59%
ATLX	2.55%
APLU	2.58%
IRON	7.72%
GEOS	10.00%
DRCT	5.03%
KOLD	3.26%

This contrast between both solutions highlights NSGA-II's flexibility, as it not only dominates the conservative end of the frontier, but also identifies highly efficient balance points, adapting to various risk profiles. This makes the algorithm a versatile tool for constructing portfolios tailored both to cautious investors and to managers seeking to maximize efficiency without assuming disproportionate risk.

6.3.3. Balanced Portfolios

In most instances, SMS-EMOA generates a uniform distribution of solutions across the central region of the efficient frontier. This enables the identification of well-diversified portfolios that offer a solid compromise between risk and return, suitable for moderate investor profiles. These solutions are especially valuable in decision-making contexts, as they allow for visual and analytical comparison of the marginal benefits of taking on additional risk [42].

Table 12 presents a representative balanced portfolio for dataset `AMEX_5_100`. This solution lies in an intermediate region of the Pareto front, offering a reasonable return-to-risk ratio and an allocation structure that demonstrates a balance between diversification and concentration. The portfolio consists of five assets with weights ranging from 10% to 35%. Significant allocations to *ACU*, *BRIA*, and *ARMP*, typically stable or moderately growing assets, reinforce the balanced character of this portfolio.

Table 12. SMS —EMOA-AMEX_5_100. Balanced portfolio generated (Return: 72.37%; Risk: 29.29%; Fitness: 1.2084).

Asset	Weight (%)
ARMP	17.36%
BRIA	22.26%
DSS	10.08%
GAU	15.64%
ACU	34.65%

On the other hand, Table 13 shows one of the most efficient portfolios generated by this algorithm for dataset `ALL_20_6000`, achieving a return of 162.60% with a relatively low risk of 24.70%, resulting in a high fitness score of 3.1315. This portfolio, composed of 20 assets with weights ranging from 2.5% to 9.7%, shows excellent diversification without excessive concentration. It includes a combination of technological, energy-related, and speculative assets, such as *GEOS*, *GREEL*, *CASK*, *MATH*, and *DRCT*, allowing for gains across multiple sectors while maintaining effective risk control.

Table 13. SMS—EMOA-ALL_20_6000. Best fitness portfolio generated (Return: 162.60%; Risk: 24.70%; Fitness: 3.1315).

Asset	Weight (%)
IRON	6.79%
GYRE	4.44%
ELTK	4.89%
KOLD	2.80%
ATLX	2.92%
GEOS	8.63%
ANG.PRD	4.07%
AAOI	6.65%
LMB	5.39%
CVNA	2.51%
BTC	2.50%
DRCT	4.33%
MATH	7.53%
CASK	9.58%
ACIC	5.42%
DHC	3.15%
GREEL	9.66%
BBIO	2.56%
APLT	3.68%
EYUBY	2.50%

These results confirm the ability of SMS-EMOA to generate portfolios suitable for moderate profiles with high risk-adjusted performance, making it especially useful for investors seeking efficiency without sacrificing portfolio stability. Its approach, based on the hypervolume indicator [42], enables it to preserve solution diversity and accurately reach central regions of the efficient frontier.

6.3.4. Portfolios with Sectoral or Liquidity Coverage

In certain cases, MOEA/D contributes with portfolios that explore specific regions of the Pareto front associated with subsets of assets. This allows the generation of solutions with a strong thematic component, whether based on economic sector, liquidity, or market behavior. Although its coverage of the frontier may be more localized than that of other algorithms, MOEA/D proves useful for designing strategies aligned with structural preferences or specific sectoral approaches [11].

Table 14 shows one of the most efficient portfolios generated by MOEA/D for dataset `ALL_20_6000`, with a return of 160.16%, a controlled risk of 23.85%, and a fitness score of 3.1368. The portfolio composition reveals a clear orientation toward niche or high-growth assets, many of which belong to specific sectors such as energy, emerging technologies, biotechnology, or mining. For example, significant allocations are made to *GEOS*, *GREEL*, *ANG.PRB*, *CASK*, and *ELTK*, all of which are linked to volatile yet high-potential markets.

Another relevant feature is the selection of assets with good liquidity or trading in secondary markets, such as *BTCC* (crypto), *KOLD* (inverse energy), or *DRCT* (applied technology). This selection suggests a strategy driven not only by mathematical efficiency, but also by the design of portfolios that reflect structural hypotheses or investor-imposed constraints (such as sector preferences, leverage, or hedging).

MOEA/D, by decomposing the objective functions into multiple subproblems, tends to specialize in specific regions of the front. This can be advantageous for generating portfolios with a "strategic bias". Such behavior makes it a particularly useful tool in environments where maintaining sectoral exposure or exploiting specific investment themes is desired, without compromising efficiency in the risk–return trade-off.

Table 14. MOEA/D—ALL_20_6000. Sector-oriented portfolio generated (Return: 160.16%; Risk: 23.85%; Fitness: 3.1368).

Asset	Weight (%)
LMBO	3.26%
ELTK	8.10%
GEOS	9.80%
GREEL	8.81%
ANG.PRB	9.93%
DRCT	4.57%
GYRE	3.67%
CASK	6.80%
AAOI	5.85%
DHC	2.72%
CVNX	2.51%
ATLX	3.02%
EYPT	2.53%
KOLD	2.50%
ACIC	5.66%
APLU	2.51%
BTCC	2.70%
MATH	6.74%
BBIO	2.57%
IRON	5.74%

6.3.5. Comparison with Financial Baselines

To contextualize the computational effort of the metaheuristics, it is essential to compare their results against standard financial benchmarks. For the analyzed period (2023),

the broad market showed strong performance: the S&P 500 index delivered a return of approximately 24%, while the tech-heavy NASDAQ Composite rose by about 43%. A simple "1/N" (Equal-Weight) strategy across the entire asset universe would typically yield a return correlated with these broad market trends. In stark contrast, the optimized portfolios identified by our framework (e.g., the SMS-EMOA balanced portfolio in Table 13) achieved returns exceeding 160% with controlled risk levels (≈24%). This demonstrates that the metaheuristics successfully generated significant "Alpha," outperforming the passive market benchmarks by a factor of three to four, thereby justifying the computational cost of the optimization process.

7. Discussion

The detailed analysis of representative portfolios provides a more comprehensive view of the behavior and specific strengths of each algorithm within the multi-objective optimization process applied to the portfolio selection problem. Each approach exhibits an operational bias that, far from being a limitation, can be leveraged according to the investor's risk profile and the strategic constraints of the fund or manager.

First, **SMPSO** stands out for its exploratory capacity and its ability to reach the extremes of the efficient frontier. Its solutions are typically located in the *high-return and high-risk* region, making them especially suitable for aggressive investors aiming to maximize returns even at the cost of bearing significant volatility. Additionally, SMPSO has demonstrated the ability to generate intermediate solutions with a good trade-off between risk and return, reinforcing its versatility when properly tuned.

In contrast, **NSGA-II** offers efficient coverage of the opposite end: the *low-risk and moderate-return* region. This behavior is ideal for building stable, diversified, and robust portfolios aimed at conservative profiles, such as pension funds, insurers, or institutional investors with capital preservation policies. Nevertheless, it has also shown the ability to generate efficient solutions in mid-range regions of the frontier, with excellent *fitness* values, indicating a good balance between exploration and exploitation.

SMS-EMOA, on the other hand, excels at generating *balanced* portfolios well distributed along the efficient frontier. Its hypervolume-based approach promotes solution diversity and allows for the identification of multiple configurations suitable for moderate-risk investors. Its strength lies in offering a broad range of portfolios with a reasonable compromise between return and risk, facilitating marginal analysis of the benefits associated with assuming more risk or accepting lower returns.

Finally, **MOEA/D** adds further value by identifying portfolios with a structural or thematic orientation. Its tendency to specialize in specific regions of the frontier makes it particularly useful in scenarios where the goal is to implement *sectoral strategies*, liquidity-driven allocations, or exposure to specific market segments. Although its global coverage may be more limited, it can be highly effective for designing strategically biased portfolios while maintaining good risk-adjusted performance metrics.

Taken together, these results demonstrate that there is no single "optimal" algorithm. Instead, each contributes a distinct perspective to the decision-making process. This algorithmic diversity enhances the system's ability to adapt to different investor profiles and market conditions, reaffirming the relevance of multi-objective evolutionary approaches in the construction of efficient portfolios.

8. Conclusions

In this paper, we have proposed and evaluated a novel framework for the efficient selection of investment portfolios, formulated as a multi-objective optimization problem. The framework integrates data acquisition and preprocessing steps with advanced meta-

heuristic optimization techniques, enabling the generation of realistic investment solutions from a broad universe of financial assets.

Our proposal was validated using real-world data from over 11,000 actively traded assets listed on major U.S. stock exchanges, categorized by market and liquidity. To assess the feasibility of the proposed approach, four representative multi-objective metaheuristics have been used: NSGA-II, SMPSO, MOEA/D, and SMS-EMOA. As the main evaluation metrics, we use Inverted Generational Distance Plus and Hypervolume, with the aim of measuring convergence and comparing the performance and efficiency of each algorithm.

We have highlighted the inherently multi-objective nature of the portfolio selection problem, which requires balancing conflicting goals such as maximizing expected return, minimizing risk, and considering liquidity or diversification constraints. Moreover, the complexity of the problem arises from the vast number of possible asset and investment combinations, as well as the incorporation of constraints such as investment limits, sectoral rules, or investor preferences. Additionally, the implementation of constraints has enabled us to obtain results that are more coherent, applicable, and meaningful.

Beyond the quantitative performance analysis, a qualitative evaluation of representative portfolios offered further insight into the behavioral profiles of each algorithm. This analysis showed that SMPSO excels at generating portfolios located at the extreme of the Pareto front, making it ideal for speculative investor profiles. In contrast, NSGA-II demonstrated a notable ability to generate conservative portfolios with low risk and high diversification, making it a solid option for stable environments and capital preservation policies. SMS-EMOA showed strength in generating well-balanced intermediate solutions, useful for moderate-risk investors and marginal analysis-based decision-making. Lastly, MOEA/D proved valuable in constructing thematic or sector-based portfolios, aligned with structural biases or specific investment strategies. This diversity in behavior highlights that each algorithm contributes a complementary perspective, and their integration into a recommendation system could significantly enrich financial decision-making processes.

Overall, the study reinforces the relevance of using advanced optimization techniques to manage the inherent uncertainty of financial markets. The value of multi-objective metaheuristics lies not only in their numerical competitiveness but also in their ability to generate diverse portfolios aligned with different investment strategies and risk profiles. This dual perspective, quantitative and qualitative, is crucial for supporting effective and informed financial decisions.

However, it is important to acknowledge that this study focuses on the structural validation of the optimization framework using historical data. Given the extensive scope of the proposed methodology—which integrates data acquisition, preprocessing, and a comprehensive analysis of multiple metaheuristics—longitudinal testing and Out-of-Sample (OOS) validation were considered beyond the scope of the current work. Consequently, the assessment of the model's robustness in unseen market conditions remains a limitation to be addressed in future research.

From a technical perspective, although the fitness evaluation complexity is effectively decoupled from the total market size (N), the data acquisition and management phase remains dependent on the physical processing of the entire asset universe. The simultaneous access to the file system for a large number of assets currently induces input/output (I/O) bottlenecks that restrict the system's scalability. For future improvements, we plan to overcome this platform limitation by implementing a data partitioning strategy, allowing concurrent threads to process exclusive subsets of asset files without I/O contention, thus taking full advantage of parallel processing capabilities.

Building on the contributions of this work, several promising research directions emerge. One avenue involves enriching the problem formulation by incorporating ad-

ditional investment factors such as trading volume, compound interest, or regulatory constraints. Another opportunity lies in integrating macroeconomic indicators and employing machine learning models to improve market forecasting and enhance the adaptability of investment strategies. Moving beyond static historical analysis, future implementations will explore adaptive MOO frameworks where objectives are dynamically updated based on predictive models. Finally, hybrid or ensemble approaches, particularly those inspired by evolutionary machine learning, could offer new synergies by combining optimization and predictive analytics in a unified framework.

Author Contributions: Conceptualization, A.J.H.-M., J.G.-N. and A.J.N.; methodology, A.J.H.-M. and A.J.N.; software, A.J.H.-M. and A.J.N.; validation, A.J.H.-M., J.G.-N. and A.J.N.; formal analysis, A.J.H.-M. and J.G.-N.; investigation, A.J.H.-M.; resources, J.G.-N. and A.J.N.; data curation, A.J.H.-M.; writing—original draft preparation, A.J.H.-M. and A.J.N.; writing—review and editing, A.J.H.-M., J.G.-N. and A.J.N.; visualization, A.J.N.; supervision, J.G.-N. and A.J.N.; project administration, J.G.-N. and A.J.N.; funding acquisition, J.G.-N. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially funded by grant PID2024-155363OB-C41 funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, KOSMOS (Boosting Federated Data-Sharing Ecosystems with a Multidimensional Approach) and by grant DGP_PIDI_2024_01174, Junta de Andalucia.

Data Availability Statement: All data, obtained results and source code of the experiments are available in https://github.com/AntHidMar/ParetoInvest and https://github.com/AntHidMar/Portfolio_Study_DataSources (accessed on 17 November 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mandal, P.K.; Thakur, M. Higher-order moments in portfolio selection problems: A comprehensive literature review. *Expert Syst. Appl.* **2024**, *238*, 121625. [CrossRef]
2. Markowitz, H. Portfolio Selection. *J. Financ.* **1952**, *7*, 77–91.
3. Tatsumura, K.; Hidaka, R.; Nakayama, J.; Kashimata, T.; Yamasaki, M. Real-Time Trading System Based on Selections of Potentially Profitable, Uncorrelated, and Balanced Stocks by NP-Hard Combinatorial Optimization. *IEEE Access* **2023**, *11*, 120023–120033. [CrossRef]
4. Ponsich, A.; Jaimes, A.L.; Coello, C.A.C. A Survey on Multiobjective Evolutionary Algorithms for the Solution of the Portfolio Optimization Problem and Other Finance and Economics Applications. *IEEE Trans. Evol. Comput.* **2013**, *17*, 321–344. [CrossRef]
5. Erwin, K.; Engelbrecht, A. Meta-heuristics for portfolio optimization. *Soft Comput.* **2023**, *27*, 19045–19073. [CrossRef]
6. Coello, C.A.C.; Lamont, G.B.; Van Veldhuizen, D.A. *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2007.
7. Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*; John Wiley & Sons: Hoboken, NJ, USA, 2002.
8. Durillo, J.J.; Nebro, A.J. jMetal: A Java framework for multi-objective optimization. *Adv. Eng. Softw.* **2011**, *42*, 760–771. [CrossRef]
9. Nebro, A.; Durillo, J.J.; Vergne, M. Redesigning the jMetal Multi-Objective Optimization Framework. In Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation , New York, NY, USA, 11–15 July 2015; pp. 1093–1100.
10. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
11. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE T. Evolut. Comput.* **2007**, *11*, 712–731. [CrossRef]
12. Beume, N.; Naujoks, B.; Emmerich, M. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *Eur. J. Oper. Res.* **2007**, *181*, 1653–1669. [CrossRef]
13. Nebro, A.J.; Durillo, J.J.; García-Nieto, J.; Coello, C.A.C.; Luna, F.; Alba, E. SMPSO: A new PSO-based metaheuristic for multi-objective optimization. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM), Nashville, TN, USA, 30 March–2 April 2009; pp. 66–73.
14. Sharpe, W.F. The Sharpe Ratio. *J. Portf. Manag.* **1994**, *21*, 49–58. [CrossRef]
15. Zhu, H.; Wang, Y.; Wang, K.; Chen, Y. Particle Swarm Optimization (PSO) for the constrained portfolio optimization problem. *Expert Syst. Appl.* **2011**, *38*, 10161–10169. [CrossRef]

16. Elton, E.J.; Gruber, M.J. Modern portfolio theory, 1950 to date. *J. Bank. Financ.* **1997**, *21*, 1743–1759. [CrossRef]
17. Pinelis, M.; Ruppert, D. Machine learning portfolio allocation. *J. Financ. Data Sci.* **2022**, *8*, 35–54. [CrossRef]
18. Ma, Y.; Han, R.; Wang, W. Portfolio optimization with return prediction using deep learning and machine learning. *Expert Syst. Appl.* **2021**, *165*, 113973. [CrossRef]
19. Iba, H.; Sasaki, T. Using genetic programming to predict financial data. In Proceedings of the First International Conference on Computational Intelligence for Financial Engineering (CIFEr), Washington, DC, USA, 6–9 July 1999; pp. 244–253.
20. Lai, K.K.; Yu, L.; Wang, S.; Zhou, C. A double-stage genetic optimization algorithm for portfolio selection. In Proceedings of the International Conference on Neural Information Processing (ICONIP), Hong Kong, China, 3–6 October 2006; Volume 3, pp. 928–937.
21. Shoaf, J.; Foster, J.A. Efficient set GA for stock portfolios. In Proceedings of the Proceedings of the Congress on Evolutionary Computation (CEC), Anchorage, AK, USA, 4–9 May 1998; IEEE: New York, NY, USA, 2000; pp. 1775–1781.
22. Abi Jaber, E.; Miller, E.; Pham, H. Markowitz Portfolio Selection for Multivariate Affine and Quadratic Volterra Models. *SIAM J. Financ. Math.* **2021**, *12*, 369–409. [CrossRef]
23. Anagnostopoulos, K.; Mamanis, G. A portfolio optimization model with three objectives and discrete variables. *Comput. Oper. Res.* **2010**, *37*, 1285–1297. [CrossRef]
24. Davis, M.H.A.; Norman, A.R. Portfolio selection with transaction costs. *Math. Oper. Res.* **1990**, *15*, 676–713. [CrossRef]
25. Gültekin, M.N.; Shohfi, T.D.; Guerard, J.B. The Construction of Efficient Portfolios: A Verification of Risk Models for Investment Making. *Front. Financ.* **2023**, *6*, 456346 [CrossRef]
26. Kwok, Y.K. Markowitz Efficient Portfolio Theory. *Glob. J. Bus. Res.* **2007**, *7*, 59–70.
27. Mahapatra, B.; Mohapatra, S.; Samanta, B.; Guha Deb, S. An Investigation of Portfolio Optimization using Modified NSGA-II Algorithm. *South Asian J. Manag.* **2019**, *26*, 134.
28. Mishra, S.K.; Panda, G.; Majhi, R. A comparative performance assessment of a set of multiobjective algorithms for constrained portfolio assets selection. *Swarm Evol. Comput.* **2014**, *16*, 38–51. [CrossRef]
29. Mishra, S.K.; Panda, G.; Meher, S.; Majhi, R. Multi-objective evolutionary algorithms for financial portfolio design. *Int. J. Comput. Vis. Robot.* **2010**, *1*, 120–138. [CrossRef]
30. Georgantas, A.; Doumpos, M.; Zopounidis, C. Robust optimization approaches for portfolio selection: A comparative analysis. *Ann. Oper. Res.* **2021**, *339*, 1205–1221. [CrossRef] [PubMed]
31. Liu, S.; Xiao, C. Application and comparative study of optimization algorithms in financial investment portfolio problems. *Mob. Inf. Syst.* **2021**, *2021*, 3462715. [CrossRef]
32. Aithal, P.K.; Geetha, M.; Acharya, U.D.; Savitha, B.; Menon, P. Real-Time Portfolio Management System Utilizing Machine Learning Techniques. *IEEE Access* **2023**, *11*, 32595–32608. [CrossRef]
33. Morteza, H.; Jameii, S.M.; Sohrabi, M.K. An improved learning automata based multi-objective whale optimization approach for multi-objective portfolio optimization in financial markets. *Expert Syst. Appl.* **2023**, *224*, 119970. [CrossRef]
34. Guarino, A.; Santoro, D.; Grilli, L.; Zaccagnino, R.; Balbi, M. EvoFolio: A portfolio optimization method based on multi-objective evolutionary algorithms. *Neural Comput. Applic.* **2024**, *36*, 7221–7243. [CrossRef]
35. Liu, W.; Zhang, Y.; Liu, K.; Quinn, B.; Yang, X.; Peng, Q. Evolutionary Multiobjective Optimization for Large-Scale Portfolio Selection With Both Random and Uncertain Returns. *IEEE Trans. Evol. Comput.* **2025**, *29*, 76–90. [CrossRef]
36. Chen, Y.; Mabu, S.; Hirasawa, K. Genetic relation algorithm with guided mutation for the large-scale portfolio optimization. *Expert Syst. Appl.* **2011**, *38*, 3353–3363. [CrossRef]
37. Salo, A.; Doumpos, M.; Liesiö, J.; Zopounidis, C. Fifty years of portfolio optimization. *Eur. J. Oper. Res.* **2023**, *318*, 1–18. [CrossRef]
38. Bernardo, A.E.; Welch, I. Market efficiency and real returns on stocks. *Annu. Rev. Econ.* **2004**, *4*, 597–620.
39. Ishibuchi, H.; Masuda, H.; Nojima, Y. A Study on Performance Evaluation Ability of a Modified Inverted Generational Distance Indicator. In Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2015, Madrid, Spain, 11–15 July 2015; Silva, S., Esparcia-Alcázar, A.I., Eds.; ACM: New York, NY, USA, 2015; pp. 695–702.
40. Zitzler, E.; Thiele, L.; Laumanns, M.; Fonseca, C.; da Fonseca, V. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Trans. Evol. Comput.* **2003**, *7*, 117–132. [CrossRef]
41. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
42. Emmerich, M.; Deutz, A.; Klinkenberg, J.W. Hypervolume-based expected improvement: Monotonicity properties and exact computation. *Evol. Comput.* **2011**, *19*, 381–405.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.