Investment Portfolio Through Evolutionary Algorithm

Parthi Vishnawat

Supervisor: Dr. Joseph Bonello

September 2025

Submitted in partial fulfilment of the requirements for the degree of MS in Information System, ICT

L-Università ta' Malta
Faculty of Information & Communication Technology

Abstract

This dissertation investigates the application of evolutionary algorithms—specifically the Genetic Algorithm (GA) and Simulated Annealing (SA)—for portfolio optimisation within the S&P 500, addressing the limitations of traditional models. The methodology uses Gower's distance to handle mixed numerical and categorical data, allowing for the construction of factor-aligned portfolios based on Growth, Value, and Quality dimensions. The primary optimisation objective is to maximise the Sortino Ratio, focusing on downside-risk-adjusted returns.

The methodology employs composite feature engineering to rank stocks across Growth, Value, and Quality dimensions, followed by distance-based clustering and anomaly detection to reveal market structures. GA and AGA are then applied with objective functions designed to maximize the Sortino Ratio, which emphasizes downside-risk-adjusted returns. Hyperparameters such as population size, mutation rate, crossover probability, and annealing temperature schedules are tuned to balance exploration and exploitation.

Empirical evaluation demonstrates that both GA and AGA generate highly diversified portfolios with competitive performance. The GA-optimised maximum Sortino portfolio achieved an annualized return of 18.95%, a Sortino Ratio of 1.0999, and a beta of 0.98, indicating strong returns with reduced downside risk relative to the market. Comparative analysis reveals that AGA converges faster and achieves marginally superior downside protection, validating its advantage in complex search landscapes. Complementary tools such as similarity maps, hierarchical clustering, and a diversification recommender further enhance interpretability and practical applicability.

The results underscore the potential of evolutionary algorithms to construct robust, risk-aware investment portfolios that go beyond linear optimisation frameworks. By combining factor-based insights with evolutionary optimization, this work contributes to the growing literature on computational finance and demonstrates actionable applications for institutional and retail portfolio managers.

Acknowledgements

I would first and foremost like to express my sincere gratitude to my supervisor, Dr. Joseph Bonello. His invaluable guidance, unwavering support, and insightful feedback have been instrumental throughout every stage of this research. His expertise and encouragement were essential in navigating the complexities of this topic and shaping this dissertation into its final form.

I would also like to extend my thanks to the Faculty of Information & Communication Technology at the L-Università ta' Malta. The faculty provided an excellent academic environment and the necessary resources that made this work possible.

Finally, I am deeply grateful to my family and friends for their constant encouragement, patience, and support. Their belief in my efforts has been a source of motivation during this challenging and rewarding journey.

"For the past century, statisticians have been using the bell curve to model the world around them... But the financial world is different. It is not the tame randomness of a roulette wheel; it is the ‘wild’ randomness of an earthquake."

— Benoit Mandelbrot, *The Misbehaviour of Markets*

Contents

List of Figures

List of Tables

List of Abbreviations

ICT	Information and Communication Technology
CAMP	Capital Asset Pricing Model
VaR	Value at Risk
CvaR	Conditional Value at Risk
ES	Expected Shortfall
SD	Standard Deviation
Var	Vector Autoregression
AR	Autoregression
OLS	Ordinary Least Squares
SR	Sharpe Ratio
MDD	Maximum Drawdown
RRR	Reward-to-Risk Ratio
GA	Genetic Algorithm
EA	Evolutionary Algorithm
MOEA	Multi-Objective Evolutionary Algorithm
NSGA-II	Non-dominated Sorting Genetic Algorithm II
PSO	Particle Swarm Optimization
SA	Simulated Annealing
MOP	Multi-Objective Problem
RSI	Relative Strength Index
MACD	Moving Average Convergence Divergence
ADF	Augmented Dickey-Fuller (test)
ADF	Autoregressive Distributed Lag
EUR	Euro

1 Introduction

Modern portfolio theory, introduced by Harry Markowitz in 1952, fundamentally transformed investment management from an art of heuristic stock-picking into a quantitative science. Before MPT, asset allocation often relied on simple rules of thumb, such as diversifying across a handful of "blue-chip" stocks. Markowitz's core insight was that a portfolio's risk is not merely the average of its components' risks but is instead a function of how its assets' returns move in relation to one another, their covariance.

By framing portfolio selection as a quadratic optimisation problem, Mean-Variance (MV) theory provided a powerful analytical framework for building diversified portfolios. The set of optimal portfolios generated by this process forms the efficient frontier, a foundational concept representing the best possible return for any given level of risk. This elegant trade-off, where volatile assts are balanced against stable ones to minimise total portfolio variance, remains a cornerstone of financial theory and laid the groundwork for subsequent equilibrium models like Capital Asset Pricing Model (CAPM).

However, the financial markets of today are vastly more complex than they were seven decades ago. The universe of investable assets has expanded dramatically to include small-cap and international equities, sector specific ETFs, commodity futures and cryptocurrencies. This explosion in choice creates a high-dimensional covariance matrices that are difficult to estimate accurately. Empirical research has also confirmed that financial returns are rarely Gaussian; they often exhibit heavy tails and skewness, especially during market crises, which violates the core assumptions of the MV framework. Compounding these issues, frictions such as transaction costs and short-sale constraints have become more significant with the rise of electronic trading. Consequently, a naively computed MV-optimal portfolio is often highly sensitive to estimation errors, leading to unstable weights, severe concentration risk, and poor out-of-sample performance.

1.1 Motivation

This research is motivated by the critical and persistent gap between classical portfolio theory and the practical demands of modern investment management. While numerous advanced methodologies from tail-risk optimisation to machine learning have been proposed to address the shortcomings of MV, there is a lack of systematic, out-of-sample comparison under a unified framework. Investors and asset managers face a complex and often confusing landscape of competing models, each with its own theoretical advantages and practical trade-offs. This thesis seeks to evaluate a diverse set of portfolio construction techniques in realistic market settings.

1.2 Hypothesis

This thesis builds on these innovations by systematically comparing classical, modern, and evolutionary approaches under a unified experimental framework. The central hypothesis is that portfolio optimisation strategies explicitly designed to handle non-normal returns and noisy covariance estimates, such as Conditional Value-at-Risk (CvaR) and Hierarchical Risk Parity (HRP), can produce portfolios with improved risk-adjusted performance and lower drawdowns compared to the classical Mean-Variance model. Furthermore, it is hypothesised that evolutionary algorithms, by virtue of their flexibility and robust search capabilities, can achieve a more attractive balance between risk, return, and portfolio stability.

1.3 Research Question

The review of existing literature reveals that while traditional portfolio optimization models are well-understood, they possess significant limitations in practice. Their sensitivity to input parameters and inability to accommodate complex, multi-faceted investor goals highlight a clear need for more robust and flexible methods. Modern heuristic algorithms offer a promising alternative, yet their practical efficacy in achieving specific, style-based investment objectives remains an area ripe for empirical investigation.

To address this gap, this thesis is guided by the following central research question:

To what extent can modern heuristic optimization algorithms, such as Simulated Annealing and Genetic Algorithms, construct superior investment portfolios compared to traditional benchmarks, when evaluated on out-of-sample performance, risk management, and the achievement of specific factor-based objectives?

1.4 Aims and Objectives

To investigate this, the thesis addresses the problem of identifying which portfolio optimisation approach offers the most robust risk-adjusted returns within a consistent out-of-sample framework, while also considering practical factors like tail risk and turnover. The primary aims are to:

1. Implement five distinct portfolio paradigms, from the benchmark Equal Weight (EW) and classical Mean Variance (MV) to modern approaches like Hierarchical Risk Parity (HRP), Conditional Value-at-Risk (CVaR) optimisation, and Regime-Switching models.
2. Develop a consistent out-of-sample testing framework using a rolling window approach to ensure no look-ahead bias is present in the evaluation.
3. Investigate whether evolutionary algorithms, specifically a Genetic Algorithm (GA) and an Annealing Genetic Algorithm (AGA), can generate competitive portfolios when benchmarked against the other established methods.

1.5 Dissertation Structure

The remainder of this thesis is structured to logically address these objectives. Chapter 2 establishes the theoretical background, consolidating the principles of Modern Portfolio Theory and the alternative frameworks developed to overcome its limitations. Chapter 3 provides a critical literature review, evaluating prior empirical findings on the relative strengths and weaknesses of these methods.

Chapter 4 details the data sources, preprocessing steps, and the mathematical and conceptual design of the portfolio models and evaluation metrics. Chapter 5 describes the practical implementation of the back testing framework, including software, libraries, and specific parameter settings for the optimisation algorithms. Chapter 6 presents the empirical results from the out-of-sample analysis and provides a detailed discussion of the findings and their practical implications. Finally, Chapter 7 concludes the thesis by summarising the key outcomes, acknowledging the study's limitations, and proposing directions for future research.

2 Background

This chapter provides the theoretical and methodological background underpinning the thesis. It begins by introducing Modern Portfolio Theory (MPT), which established the foundation for quantitative portfolio construction. It then discusses the practical limitations of mean–variance optimization in real financial markets. In response to these shortcomings, the chapter reviews a range of alternative approaches, including downside-risk measures, clustering-based diversification, regime-switching models, machine learning methods, and evolutionary optimization techniques. This progression motivates the adoption of evolutionary, multi-objective methods, with particular emphasis on Genetic Algorithms (GA) and an Annealing Genetic Algorithm (AGA).

2.1 Modern Portfolio Theory: The Foundation of Diversification

Quantitative portfolio management was proposed by Harry Markowitz's in 1952 in his paper "Portfolio Selection," a work that transformed the field from a practice based on heuristics to a rigorous analytical discipline. Before Markowitz, diversification was an intuitive but ill-defined concept, often amounting to simply holding a collection of well-regarded "blue-chip" stocks. Markowitz's groundbreaking contribution was to provide a precise mathematical framework to formalise and optimise the principle of diversification. The core insight of MPT is that an asset's risk should not be evaluated in isolation but by its marginal contribution to the overall volatility of a portfolio. This contribution is measured by its covariance with all other assets in the portfolio.

MPT frames portfolio selection as a quadratic optimisation problem with a fundamental trade-off between risk and the return. In this model, risk is quantified by the portfolio's variance (σ_p^2), and return is measured by the weighted average of the expected returns of the constituent assets (μ_p). The goal is to find the vector of asset weights that optimally balances these two objectives. This typically takes one of two forms: either minimising portfolio variance for a target level of expected return or, conversely, maximising expected return for a given level of variance.

The collection of all such optimal portfolios forms a hyperbola in the risk-return space known as the efficient frontier. Any portfolio that lies below the efficient frontier is considered sub-optimal because another portfolio can exist that offers either a higher return for the same level of risk or lower risk for the same level of return. This elegant framework mathematically demonstrates that combining assets with less-than-perfect correlation (i.e., a correlation coefficient of less than +1) can reduce total portfolio risk without sacrificing expected return, thus providing a quantitative justification for diversification.

2.2 Limitations of Mean–Variance Optimization

Despite its foundational importance and theoretical elegance, the practical application of Mean-Variance Optimisation (MVO) is hindered by several critical limitations. These issues stem from both its underlying assumptions about market behaviour and its structural sensitivity to statistical inputs, which often leads to results that are impractical and perform poorly out-of-sample (Michaud, 1989).

2.2.1 The unrealistic assumption of normal returns

A core, implicit assumption of MVO is that asset returns are jointly normally (or elliptically) distributed. This assumption is convenient because it allows the entire return distribution to be described by just two parameters: mean and variance. However, a vast body of empirical research has conclusively shown that financial returns are not normal (Mandelbrot, 1963; Fama, 1965). They consistently exhibit:

Fat tails: Extreme market events, such as crashes and sharp rallies, occur far more frequently than the normal distribution would predict. The bell curve of a normal distribution thins out rapidly at its tails, underestimating the probability of these high-impact events (Mandelbrot, 1963; Fama, 1965).

Skewness: Returns are often asymmetrically distributed. Negative skewness is common in equity markets, meaning large negative returns are more probable than large positive ones (Mandelbrot, 1963; Fama, 1965).

Because variance, as a risk measure, penalises upside volatility and downside volatility equally, it fails to capture the nuances of these non-normal distributions and an investor's asymmetric preference for avoiding large losses over capturing large gains (Mandelbrot, 1963; Fama, 1965).

2.2.2 The impact of estimation error

The MVO framework requires two key sets of inputs: the vector of expected returns for each asset and the covariance matrix describing their interrelationships. In practice, these parameters are unknown and must be estimated from historical data. This estimation process is fraught with uncertainty, and the optimiser is notoriously sensitive to these inputs. As noted by Michaud (1989), the MVO process can act as an "error maximiser": small, statistically insignificant errors in the estimation of expected returns can be amplified into large, economically significant changes in the resulting optimal portfolio weights. This often leads to highly concentrated portfolios, where the model allocates heavily to a few assets that had marginally higher (and likely spurious) historical returns, while ignoring others. The result is an unstable allocation that fluctuates wildly with each re-estimation period and lacks genuine diversification (Best & Grauer, 1991**)**.

2.2.3 Instability in high-dimensional environments

The challenges of estimation error are severely exacerbated in high-dimensional settings, where the number of assets is large relative to the number of historical observations. As the asset universe expands, the size of the covariance matrix grows quadratically. Estimating the vast number of unique covariances with limited data leads to a noisy and unstable matrix that may be ill-conditioned. This means the matrix is nearly singular and its inversion—a critical step in the optimisation process—becomes numerically unstable. This instability can lead to nonsensical results and portfolios that are "optimal" only for the specific noise in the historical data, a classic case of overfitting that guarantees poor out-of-sample performance (Ledoit & Wolf, 2004).

2.3 Alternative paradigms in portfolio construction

In response to these practical failings of MVO, researchers and practitioners have developed a host of alternative methodologies designed for greater robustness and real-world applicability.

2.3.1 Downside-risk measures: Conditional Value-at-Risk (CvaR)

To better capture investors' aversion to large losses, downside-risk measures have been developed. While Value-at-Risk (VaR) is widely adopted, it is often criticized.

VaR is a quantile-based measure that answers the question: "What is the maximum I can lose over a given period with a certain level of confidence (e.g., 95%)?" The flaw in VaR is that it provides no information about the magnitude of the loss if that threshold is breached.

CvaR also known as expected shortfall, was developed to address this shortcoming (Rockafellar & Uryasev, 2000). CVaR answers the more pertinent question: "If things go badly, what is my expected loss?" It measures the expected loss in the tail of the distribution, conditional on the loss exceeding the VaR threshold. By focusing directly on the magnitude of extreme losses, CVaR aligns better with the goals of risk management. Furthermore, CVaR is a coherent risk measure, satisfying mathematical properties (like subadditivity) that VaR lacks, and optimisation problems formulated with CVaR can be cast as efficient linear programs, making it computationally tractable (Artzner, 1999).

2.4 Diversification via Clustering

To circumvent the inherent instability of an inverted covariance matrix, Marcos López de Prado (2016) developed Hierarchical Risk Parity (HRP), a novel approach rooted in graph theory and machine learning. HRP constructs a portfolio without relying on the traditional optimisation process, thereby avoiding its associated pitfalls. The process involves three steps:

1. Hierarchical Tree-Clustering: First, the assets are grouped based on their correlation structure. The algorithm creates a hierarchical dendrogram (a tree diagram) that organizes assets into clusters and sub-clusters, where highly correlated assets are grouped together.
2. Quasi-Diagonalisation: The rows and columns of the covariance matrix are then reordered according to the hierarchy from the clustering step. This places similar assets next to each other, pushing the largest variance values along the diagonal and effectively creating structured blocks of correlated assets.
3. Recursive Bisection: Finally, the algorithm allocates weights in a top-down manner. It starts with the full portfolio and recursively splits it into smaller sub-portfolios based on the hierarchical tree structure. At each split, it allocates capital between the two resulting subsets based on an inverse-

variance weighting, ensuring that risk is distributed evenly across the different clusters. This hierarchical approach provides true diversification, spreading risk not just across assets but across different sources of correlation, resulting in more stable and robust portfolios.

2.5 Regime-Switching Models

The classical MVO framework assumes that the statistical properties of asset returns (means, volatilities, correlations) are constant, or stationary, over time. This assumption is patently false. Financial markets exhibit clear structural breaks and transition between distinct states, or **regimes**. For instance, a market might alternate between a "calm" regime of low volatility and positive returns and a "turbulent" regime of high volatility and negative correlations.

Regime-switching models are designed to explicitly capture this dynamic nature. Using techniques like Gaussian Mixture Models (GMMs) or Hidden Markov Models (HMMs), these models analyse historical data to infer the probability of the market being in a particular latent (unobservable) state at any given time. Once the current regime is identified, the portfolio construction can be adapted accordingly. For example, the expected returns and covariance matrix used as inputs for an optimiser can be conditioned on the current regime. This allows the strategy to be more defensive during turbulent periods (e.g., by increasing allocations to cash or low-volatility assets) and more aggressive during calm periods, leading to improved long-term, risk-adjusted performance.

2.6 Evolutionary and Heuristic Optimization

When portfolio optimisation problems involve complex, real-world constraints such as cardinality constraints (a limit on the number of assets), sector allocation limits, or non-linear transaction costs, traditional convex optimisation methods are often inadequate. Evolutionary and heuristic algorithms, inspired by natural processes, provide a powerful and flexible alternative for navigating these complex search spaces.

2.6.1 Genetic Algorithms (GA)

Genetic Algorithms are a class of search heuristics inspired by Charles Darwin's theory of natural selection. A GA starts by creating an initial population of random candidate solutions—in this case, many different random portfolios. This population then evolved over many generations using three bio-inspired operators:

- Selection: Each portfolio in the population is evaluated using a fitness function (e.g., its Sharpe ratio). Portfolios with higher fitness are given a higher probability of being selected to "reproduce." This is the "survival of the fittest" principle.

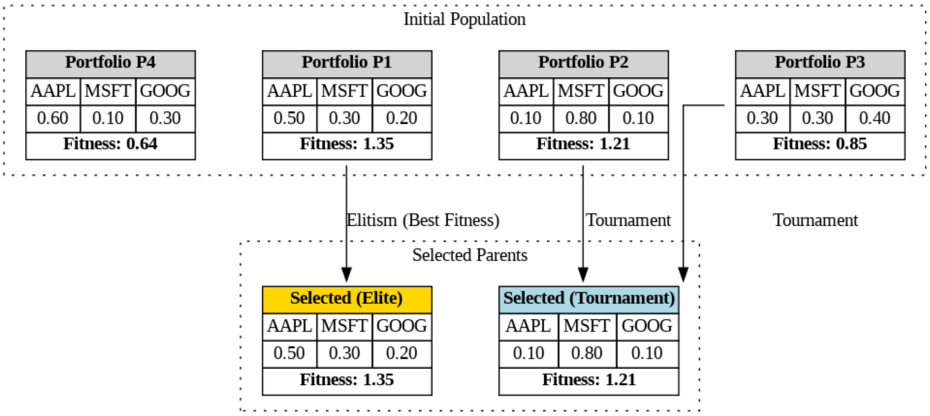

Figure 2.1 The selection phase of a Genetic Algorithm, where parent portfolios are chosen based on fitness scores.

- Crossover: Two high-fitness "parent" portfolios are selected, and their genetic material (their asset weights) is combined to create one or more new "offspring" portfolios. This allows the algorithm to mix and match the best characteristics of successful solutions.

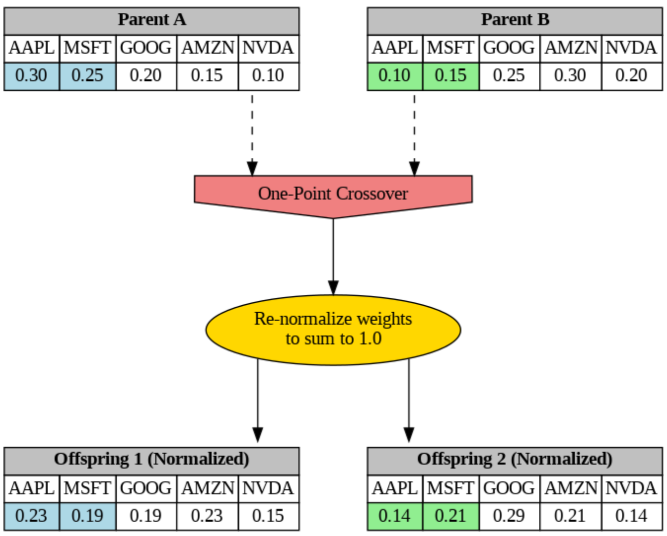

Figure 2.2 The crossover operation in a Genetic Algorithm, where two parent portfolios combine to create new offspring.

- Mutation: Small, random changes are introduced into the weights of an individual portfolio. This maintains genetic diversity in the population, preventing it from converging prematurely to a local optimum and ensuring the entire solution space is explored.

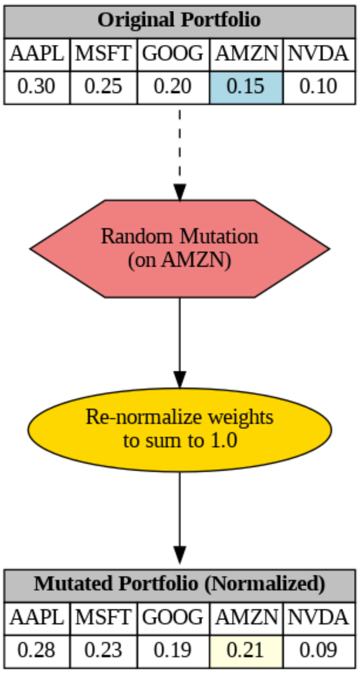

Figure 2.3 The mutation operation in a Genetic Algorithm, where a random change is introduced to a portfolio to maintain genetic diversity.

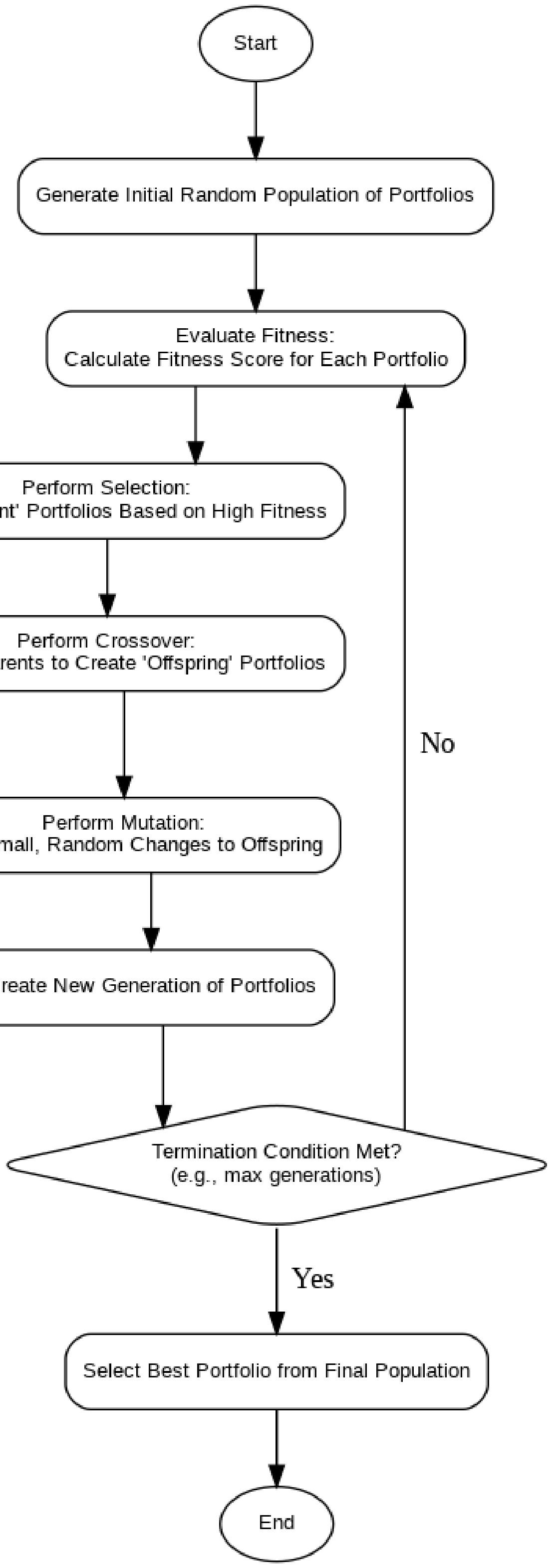

Figure 2.4 Flowchart illustrating the iterative process of a standard Genetic Algorithm.

Through this iterative process, the overall fitness of the population improves from one generation to the next, eventually converging on a highly optimised solution that satisfies all specified constraints.

2.6.2 Annealing Genetic Algorithm (AGA)

The Annealing Genetic Algorithm (AGA) is a sophisticated hybrid that enhances the search capability of a standard GA by incorporating principles from Simulated Annealing (SA). The SA heuristic is itself an analogy for the metallurgical process of annealing, where a metal is heated to a high temperature and then slowly cooled to strengthen its structure.

AGA introduces a "temperature" parameter that governs the algorithm's behaviour over time. This temperature controls the probability of accepting solutions that are worse than the current best.

- In the early stages, the temperature is high. This allows the algorithm to frequently accept inferior solutions, encouraging it to **explore** the solution space broadly and making it capable of jumping out of local optima.
- As the process continues, the temperature is gradually "cooled" according to a pre-defined schedule. As the temperature decreases, the algorithm becomes more selective and is less likely to accept worse solutions. This shifts the focus from exploration to exploitation, allowing the algorithm to fine-tune the high-quality solutions it has discovered.

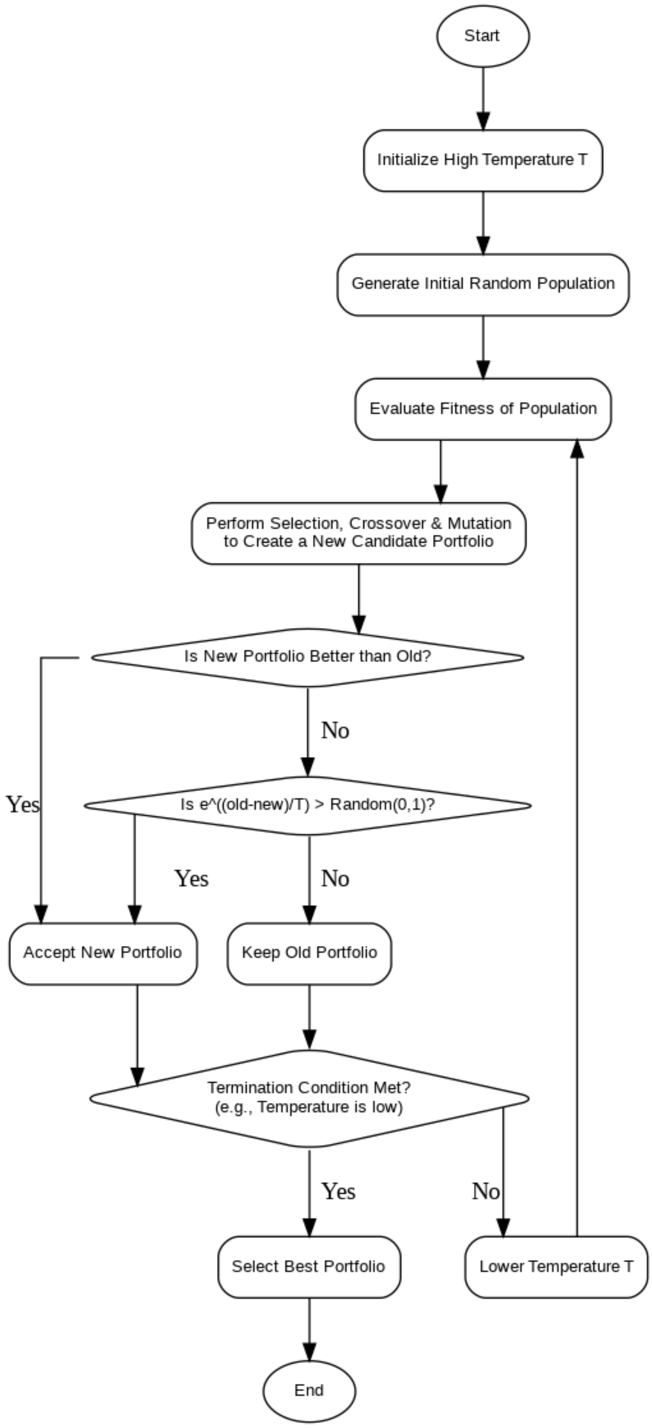

Figure 2.5 Flowchart of the Annealing Genetic Algorithm (AGA) process, incorporating a temperature parameter to balance exploration and exploitation.

By intelligently balancing this trade-off between broad exploration and focused exploitation, AGA provides a more robust and powerful search mechanism than a standard GA, significantly increasing its ability to find the true global optimum in a complex and rugged fitness landscape.

3 Literature Review

The challenge of allocating capital under uncertainty has driven decades of financial research, evolving from elegant foundational theories to complex, data-intensive modern techniques. This chapter provides a critical review of this evolution, tracing the path from the classical mean-variance paradigm to the sophisticated heuristic and machine learning models used today. The review is structured to first examine the foundational models and the critical shortcomings that necessitated subsequent innovation. It then explores the major avenues of research that emerged in response, including robust statistical methods, dynamic market models, and the integration of machine learning. Finally, it focuses on the class of metaheuristic optimisation algorithms, which have become indispensable for solving complex, real-world portfolio problems. The objective is not merely to summarise these works but to critically evaluate their methodologies, assess their performance against relevant benchmarks, and identify their limitations, thereby establishing the context and justification for the methodologies employed in this thesis.

3.1 The Classical Paradigm and Its Practical Shortcomings

The discipline of quantitative portfolio management was formally inaugurated by Markowitz (1952), whose mean-variance (MV) framework was a revolutionary step in formalising diversification. This normative work was followed by positive models describing equilibrium asset pricing, most notably the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965). However, the practical application of MV optimisation is notoriously problematic. Michaud (1989) famously argued that the process acts as an "error maximiser," amplifying estimation errors in input parameters to produce unstable and often nonsensical portfolios that have little to no investment value.

Addressing this instability has been a central theme of subsequent research. The Black-Litterman model (1992) represented a significant step forward. It elegantly addresses the instability of MVO by using a Bayesian approach to blend an investor's subjective views with market equilibrium returns, leading to more stable and diversified allocations. Despite its strengths, its effectiveness is highly sensitive to the specification of its hyperparameters, particularly the matrix representing the confidence in the investor's views, a challenge that remains a topic of practical research.

Another significant challenge is the classical models' assumption of normally distributed returns. Recognizing that financial returns often exhibit "fat tails" and skewness, Konno and Yamazaki (1991) proposed the Mean-Absolute Deviation (MAD) model. This model replaces variance with absolute deviation as the measure of risk, transforming the quadratic optimisation problem into a more computationally efficient linear one. While an improvement in robustness for non-normal returns, the MAD model still relies on a single risk metric and does not fully capture the complexities of tail risk that are of primary concern to risk-averse investors.

3.2 Robust and Dynamic Approaches to Portfolio Construction

Recognizing that market parameters are not static and estimation is fraught with error, modern research has shifted towards more robust and adaptive methods.

3.2.1 Robust Statistical Methods

To mitigate the impact of estimation error, researchers have turned to alternative statistical paradigms. Bodnar, Parolya, and Schmid (2017) developed a Bayesian framework for the Global Minimum Variance (GMV) portfolio, which minimises risk without relying on volatile expected return estimates. Using data from major international indices, their study demonstrated that Bayesian models with informative priors yield more stable portfolio weights and superior out-of-sample Sharpe ratios compared to traditional frequentist methods. The primary limitation of this approach, however, is the potential for subjectivity in the specification of the prior distributions, which can significantly influence the final allocation.

3.2.2 Dynamic Market Regimes

Another avenue of research explicitly models the dynamic, non-stationary nature of financial markets. Alexander, Kapraun, and Koroviy (2023) explored a Markov-switching approach, using historical data clustering to identify distinct market states, or regimes. Their "Markov Markowitz" model dynamically adapts its allocations based on the prevailing regime, demonstrating lower volatility and greater resilience during major market crises, such as the 2008 financial crisis and the 2020 COVID-19 crash, when benchmarked against a static MV portfolio. The main challenge with such models lies in the risk of overfitting the regime detection process and the core assumption that historical regimes will be representative of future market states.

3.2.3 Alternative Risk-Based Strategies

Beyond MVO, a significant body of research has focused on portfolio construction based purely on risk, without requiring problematic expected return estimates. While the GMV portfolio is foundational, more advanced methods have been developed to address different aspects of risk. Racic (2022) proposed a Gini-based **portfolio** model, leveraging the Gini Mean Difference (GMD) as a measure of dispersion. Unlike variance, GMD is less sensitive to outliers, making it more robust in the presence of fat-tailed returns. In an empirical study on the Croatian stock market from 2013 to 2021, the Gini-optimised portfolio consistently demonstrated superior risk-adjusted performance compared to the Equally Weighted (EW) and GMV portfolios. Specifically, the Gini portfolio achieved a Sharpe ratio of 0.49, compared to 0.31 for the EW portfolio and 0.28 for the GMV portfolio. The primary limitation is that, while robust, this approach is purely risk-based and may not be suitable for investors with specific return targets, as it completely ignores expected return signals.

3.2.4 Risk Parity and Factor-Based Strategies

A major departure from the mean-variance paradigm is the development of Risk Parity (RP) portfolios. Unlike MVO, which focuses on the allocation of capital, RP focuses on the allocation of risk, with the goal of creating a more balanced portfolio. Maillard, Roncalli, and Teiletche (2010) provide a foundational analysis of the "Equal Risk Contribution" (ERC) portfolio, a key implementation of the RP concept. Their research demonstrates that the ERC portfolio, where each asset contributes equally to the total portfolio risk, is unique and lies between the GMV and the Equally Weighted portfolios. In their empirical tests on a US equity and bond universe from 1995 to 2008, the ERC portfolio exhibited a Sharpe ratio of 0.84, substantially outperforming the GMV portfolio (Sharpe of 0.61) with only slightly higher volatility. This highlights the strategy's robustness, as it does not rely on expected return estimates. The primary limitation of the traditional RP approach is its tendency to over allocate low-volatility assets like government bonds, which requires the use of leverage to meet return targets—a strategy that introduces its own set of risks (such as funding liquidity risk).

3.3 Advanced Dependency Modelling and Network Theory

The classical reliance on the linear correlation coefficient as the sole measure of inter-asset dependency is another significant weakness, as it fails to capture complex, non-linear relationships, particularly during market stress.

3.3.1 Copula-Based Models for Tail-Risk

To address this, researchers have employed copula functions, a sophisticated statistical tool that separates the marginal distributions of individual asset returns from their joint dependency structure. This allows for the modelling of complex, non-linear dependencies, including asymmetry and tail dependence (the tendency for assets to crash together). Low et al. (2013) conducted a study applying various copula models (including Gaussian, Student's t, and Gumbel copulas) to construct portfolios of international equities. Their out-of-sample analysis from 2005 to 2009 showed that portfolios optimised using a Gumbel copula—which explicitly models asymmetric tail dependence—achieved a 23% reduction in portfolio Value-at-Risk (VaR) compared to a portfolio based on a standard correlation matrix. This demonstrates the superior ability of copulas to mitigate extreme downside risk. The main limitation of copula-based models is their mathematical complexity and the significant computational cost required to estimate them, especially in high-dimensional portfolios, which can make them challenging to implement in practice.

3.4 Machine Learning and Heuristic Optimisation

The rise of computational power has enabled the application of Machine Learning (ML) and heuristic algorithms to navigate the complexities of modern portfolio selection.

3.4.1 Machine Learning for Forecasting and Allocation

Supervised learning models are now widely used for forecasting returns. Pinelis and Ruppert (2022) employed a Random Forest model for dynamic asset allocation, which used macroeconomic indicators to make monthly allocation decisions. Their back test over the 2000-2020 period showed a 28% increase in the Sharpe ratio and reduced drawdowns during crises compared to a 60/40 benchmark. However, the model's complexity reduces its interpretability (a "black box" problem), and crucially, the study did not account for transaction costs, which could significantly erode performance in a real-world dynamic strategy. Unsupervised learning has also proven effective.

3.4.2 Metaheuristic Optimisation for Complex Problems

When real-world constraints like cardinality limits or transaction costs are introduced, the portfolio problem often becomes non-convex and intractable for traditional solvers. Metaheuristic algorithms, such as Genetic Algorithms (GA), have become a standard tool for these problems. For instance, Tai, Wang, and Chen (2023) used a vector-evaluated GA to find the Pareto-optimal frontier for portfolios of Nikkei 225 stocks, successfully balancing risk and return under cardinality constraints where classical solvers would fail.

Hybrid approaches that combine ML with heuristics represent the current state of the art. Al-Muharraqi and Munir (2024) developed a hybrid GA-PSO (Particle Swarm Optimisation) model that used an LSTM neural network for return forecasting. Their model, which optimised for downside risk using higher-order moments like skewness and kurtosis, generated a reported annualized return of 56.4% on Boursa Kuwait data, significantly outperforming traditional benchmarks. However, a key limitation is that such high returns are often specific to the volatile, regional market tested and **may not generalise** to more efficient global markets. Furthermore, the study did not account for the impact of transaction costs on its high-turnover strategy.

3.4.3 Ensemble and Hybrid Strategies

Recognizing that no single model is superior in all market conditions, researchers have explored ensemble strategies that combine the strengths of different approaches. Lim, Kim, and Ahn (2020) developed an integrated strategy combining a Genetic Algorithm with the Capital Asset Pricing Model (CAPM) and a momentum strategy. The GA was used to select an optimal subset of assets, which were then weighted using insights from CAPM and momentum indicators. When tested on both the S&P 500 and KOSPI 200 indices, their ensemble model consistently outperformed a passive buy-and-hold strategy, delivering higher risk-adjusted returns. The key limitation, however, is the model's increased complexity and number of tunable parameters, which elevates the risk of overfitting if not rigorously validated with out-of-sample data.

3.4.4 Fuzzy Logic for Handling Uncertainty

Traditional models require precise inputs, yet financial data is inherently uncertain and imprecise. Fuzzy logic offers a powerful framework for handling this ambiguity. Khan, Kriaa, and Guerpillon (2024) developed a fuzzy rule-based system for asset selection. Their model used technical indicators (like RSI and MACD) as inputs into a fuzzy inference system to classify stocks as "buy," "hold," or "sell." A Genetic Algorithm was then used to optimize the portfolio from the subset of "buy" rated stocks. Benchmarked against a traditional MV portfolio on the CAC 40 index, their fuzzy-GA approach yielded significantly higher Sharpe ratios. The primary challenge of such models is the construction of the fuzzy rule base, which can be subjective and requires significant domain expertise to define effectively. Similarly, Vercher and Bermudez (2013) proposed a Credibility Mean-Absolute Semi-Deviation (CMASD) model that uses fuzzy variables to represent uncertain returns, optimizing the portfolio with a GA. This approach effectively constructed robust and efficient frontiers but is computationally intensive and relies on the investor's ability to accurately define return possibilities as fuzzy distributions.

3.4.5 Advanced Machine Learning Models

A significant advancement is the use of Deep Reinforcement Learning (DRL), which can learn a complete, end-to-end trading policy. Xiong et al. (2021) developed a DRL agent tested on Dow Jones 30 constituents from 2009 to 2020. The agent achieved an annualized return of 21.3% and a Sharpe ratio of 1.88, significantly outperforming the DJIA index (Sharpe ratio of 1.25) and a traditional Min-Variance portfolio (Sharpe ratio of 0.89). A key strength is its ability to implicitly learn transaction costs. However, the primary limitation is the "black box" nature of deep neural networks, making decisions difficult to interpret, and the models are notoriously data hungry.

Another powerful trend is the hybridization of classical econometrics with modern machine learning. Gu, Kelly, and Xiu (2020) conducted a comprehensive study comparing models for return prediction. A portfolio constructed based on their hybrid neural network's predictions generated an out-of-sample Sharpe ratio of 1.73, compared to just 0.51 for a portfolio based on a traditional linear regression model. This highlights the power of ML in capturing non-linear patterns, but its immense computational complexity and the risk of data mining remain significant concerns.

3.4.6 Explainable AI (XAI) in Portfolio Management

A major barrier to the adoption of advanced machine learning models in finance is their "black box" nature, which makes it difficult for regulators and asset managers to trust their decisions. The field of Explainable AI (XAI) aims to address this. Golosnoy and Grieble (2023) developed a portfolio strategy using a Gradient Boosting Trees model and then applied SHAP (SHapley Additive exPlanations) analysis to interpret the model's predictions. Their model, tested on S&P 500 stocks, produced an annualized return of 14.2% with a Sharpe ratio of 1.05, outperforming a value-weighted S&P 500 benchmark which had a Sharpe ratio of 0.85. More importantly, the SHAP analysis revealed which economic features (such as momentum, P/E ratio) were driving the model's decisions at different points in time, providing valuable transparency. The primary limitation of current XAI techniques is that they provide explanations for individual predictions but do not fully illuminate the complex, holistic logic of the entire model. Furthermore, generating these explanations is computationally intensive, making real-time application challenging.

4 Methodology

4.1 Introduction and Research Workflow

This chapter details the systematic, multi-phase methodology designed to construct, optimise, and evaluate the portfolio strategies investigated in this thesis. The workflow is engineered to ensure a transparent, reproducible, and methodologically sound research process, progressing from data acquisition and preparation to algorithmic optimisation and robust out-of-sample performance evaluation.

The methodology is structured into three primary phases:

1. Data Acquisition and Preparation: This phase covers the sourcing of historical market data, the construction of a stratified investment portfolio, the engineering of relevant financial features, and the chronological splitting of data for model training and evaluation.
2. Quantitative Analysis Framework: This phase details the analytical techniques used to understand the dataset, including peer-group comparative analysis, factor-based ranking, and event studies to measure market impact.
3. Algorithmic Portfolio Optimization: This final phase presents the core of the research—the design and implementation of a Genetic Algorithm (GA) for portfolio optimization. It includes a detailed discussion of the innovative distance-based fitness functions, the hybridization of the GA with simulated annealing principles (Annealing GA), and the framework for back testing and performance evaluation.

4.2 Phase 1: Data Acquisition and Preparation

4.2.1 Data Sourcing and Universe Selection

Historical daily data for a broad universe of U.S.-listed equities were programmatically sourced from the Yahoo Finance API via the yfinance Python library. The sample spans January 1, 2016, through December 31, 2024, a window that intentionally captures multiple market regimes, including the COVID-19 crash, the subsequent bull market, and the 2022 bear market.

To enable meaningful comparative analysis, the investment universe was stratified into Peer Groups based on Sector, Industry, and Market Capitalization. This stratification is standard in empirical asset pricing, ensuring that comparisons are conducted within fundamentally comparable cohorts and allowing for the testing of hypotheses related to segment-specific risk-return characteristics.

4.2.2 Feature Engineering and Description

To create a comprehensive analytical dataset, a wide array of features was sourced and engineered for each stock. These features transform raw market data into a structured format suitable for statistical and Machine Learning models and are categorized into three groups:

- Core Market Data: This includes standard daily Open, High, Low, and Close (OHLC) prices, Volume, and the calculated Daily Return ($R_t=(P_t-P_t-1)/P_t-1$), which serves as the primary measure of daily performance.
- Technical Indicators: Features were engineered to capture momentum, trend, and volatility dynamics. These include the 20-day Simple Moving Average (SMA 20), the 14-day Relative Strength Index (RSI 14), the Moving Average Convergence Divergence (MACD diff), and 10-day rolling historical volatility.
- Fundamental Data: To provide economic context, descriptive data such as Company Name, Sector, and Industry were included, along with key financial metrics like Market Cap, Price-to-Earnings (P/E) Ratio, Forward P/E, and Dividend Yield.
- Growth & Financial Health: Revenue Growth and Debt to Equity to assess growth potential and financial leverage.
- Shareholder Metrics: Dividend Yield to measure returns to shareholders.
- Descriptive and Market Data:

1. Ticker, Company Name, Sector, Industry: Categorical data for identification and peer group analysis.
2. Market Cap, Cap Type: To classify companies by size.
3. Beta: To measure systematic risk relative to the market.
4. Analyst Recommendation: A forward-looking qualitative measure of market sentiment.

4.2.3 Data Splitting for Training and Testing

To ensure a robust evaluation of the optimization models and prevent look-ahead bias, the dataset was chronologically divided into two distinct, non-overlapping periods:

- Training Set (2016–2023): Data from January 1, 2016, to December 31, 2023. This dataset was used to discover relationships, construct factors, and train the Genetic Algorithm to identify optimal portfolio allocations.
- Testing Set (2024): Data from January 1, 2024, to December 31, 2024. This dataset was held out and used exclusively for the out-of-sample evaluation of the optimised portfolios' performance.

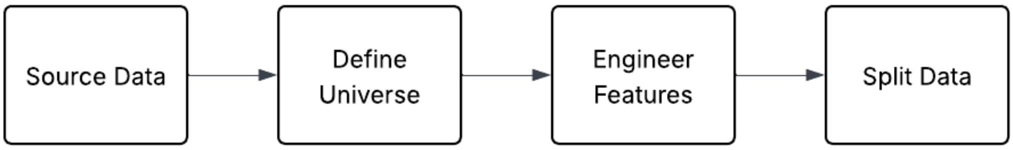

Figure 4.1: The workflow for Phase 1: Data Acquisition and Preparation.

4.2.4 Data Splitting for Rolling-Window Analysis

To rigorously test the strategy's adaptability and to prevent look-ahead bias, a rolling-window back testing approach was implemented. This methodology simulates how a real-world quantitative manager would operate, by periodically updating the model with new data.

- Lookback Window: A rolling 36-month (3-year) window of historical data was used to train the optimization model.
- Test Period: The overall out-of-sample testing period for the strategy spans from January 1, 2022, to December 31, 2024. This single, consistent test period is used for all analyses.

  For example, to determine the portfolio for Q1 2022, the model is trained on data from January 1, 2019, to December 31, 2021. This process is repeated for each quarter throughout the entire testing period, ensuring that all investment decisions are made strictly on data that would have been available at the time.

4.3 Phase 2: Portfolio Construction Paradigms

The evaluation is structured as a direct comparison between an intelligent, active selection strategy and a naive, passive benchmark.

4.3.1 Simulated Annealing Asset Selection Portfolio

This is the core strategy of this research. It is an active approach where the Simulated Annealing algorithm is used to solve an discrete optimisation problem: from the full universe of 30 assets, it must select the optimal subset of just 10 tickers that are expected to deliver the highest risk-adjusted return.

- Justification for Universe Size (30 equities): While the S&P 500 contains over 500 equities, a smaller, curated universe of 30 was selected to create a focused and computationally tractable optimization problem. This subset consists of large-cap, highly liquid stocks from diverse sectors, ensuring the universe is representative while reducing the noise and dimensionality that can negatively impact optimisation performance.
- Justification for Portfolio Size (10 equities): The portfolio size was fixed at 10 equities to strike a balance between diversification and high conviction investing. A smaller portfolio might incur excessive idiosyncratic risk, while a larger one could dilute the impact of the algorithm's top selections and increase transaction costs.

4.3.2 Equal Weight (All Assets) Benchmark

This serves as the passive benchmark. The portfolio assigns an equal weight to every single asset in the full 30-ticker dataset. It represents a strategy of maximum diversification without any active decision-making.

- Justification for Equal Weighting**:** Once the 10 optimal tickers are selected, they are held in an equal-weighted (10% each**)** portfolio. This scheme was chosen deliberately to isolate the performance contribution of the asset selection algorithm itself, removing the secondary effect of a separate weight optimization step. This allows for a clear evaluation of the model's ability to *select* superior assets.

4.3.3 Inverse Volatility

Inverse Volatility is a risk-based portfolio construction technique that assigns weights to assets in inverse proportion to their historical volatility. The core principle is to allocate more capital to less volatile (historically more stable) equities and less capital to more volatile (historically riskier) equities. This approach is a form of risk parity that aims to equalise the risk contribution of each asset in the portfolio, rather than its capital contribution.

The primary motivation for using Inverse Volatility weighting is to enhance portfolio diversification and improve its risk-adjusted return profile. Unlike market-capitalisation-weighted indices, which can become highly concentrated in a few large-cap stocks, an Inverse Volatility portfolio systematically tilts towards lower-risk assets. This often results in a portfolio with lower overall volatility and improved resilience during market downturns. In the context of this thesis, it is used as the final step after the SA and GA models have selected a high-momentum universe, creating a two-stage "smart beta" strategy that combines a momentum factor (for selection) with a low-volatility factor (for weighting).

4.4 Phase 3: Quantitative Analysis Framework

4.4.1 Peer Group Comparative Analysis

To identify best-in-class stocks within key sectors, each company was evaluated along three critical dimensions:

1. Risk, measured by annualised historical volatility.
2. Valuation, assessed via the Price-to-Earnings (P/E) ratio.
3. Performance, defined as total return over the analysis period.

This multi-dimensional framework highlights the trade-offs across different investment objectives. For example, analysis revealed that while Mega-Cap Tech stocks (such as AAPL, MSFT) offered stability and value, high-growth stocks (such as TSLA) delivered superior total returns at the cost of higher valuation premiums and risk.

4.5 Phase 4: Simulated Annealing (SA) for Asset Selection

The "swapping" mechanism is the heart of the Simulated Annealing optimiser. Instead of adjusting weights, the algorithm's primary operation is to literally swap one ticker for another, iteratively improving the quality of the selected 10-stock portfolio.

4.5.1 The "Swapping" Mechanism

4.5.1.1 Model Objective Functions

To ensure clarity and reproducibility, this section explicitly defines the construction rules and objective function for each of the primary optimisation models evaluated in this thesis.

The optimisation process at each rebalancing date is as follows:

1. Initialisation: The algorithm begins with a randomly selected portfolio of 10 equities from the full universe of 30 tickers.
2. Iteration and Swapping: For a total of 1,000 iterations, the algorithm performs a "swap":
   - It randomly selects one ticker to remove from the current 10-stock portfolio.
   - It then randomly selects one equity to add from the remaining 20 assets in the universe.
   - This creates a new, "neighbour" portfolio with one asset swapped.
3. Fitness Evaluation: The quality, or "fitness," of this new portfolio is calculated by computing its historical Sharpe Ratio on the 36-month training data, assuming an equal weight among the 10 selected assets.
4. Acceptance Criteria: A core feature of Simulated Annealing is its intelligent acceptance rule:
   - If the swapped portfolio has a better Sharpe Ratio, the swap is always accepted.
   - If the swapped portfolio has a worse Sharpe Ratio, it may still be accepted based on a probability that decreases over time. This allows the algorithm to escape sub-optimal solutions and explore the search space more effectively.

5. Convergence and hyperparameter justification: The process is repeated for a set number of iterations. To determine this hyperparameter, a convergence analysis was conducted. It was observed that the algorithm's objective function (the Sharpe Ratio) consistently stabilized, or "plateaued," after approximately 5,000 iterations. To ensure thorough exploration of the solution space, a total of 1000 iterations were used for each optimization run, allowing the algorithm to reliably converge on a high-quality solution

4.5.2 Dynamic Back testing with Quarterly Rebalancing

To simulate a realistic and adaptive strategy, the entire SA asset selection process is re-run at the beginning of every quarter.

The training data for the model expands (or rolls forward), ensuring that the algorithm's decisions are always based on the most recent market information. For example:

- For Q1 2024: The SA algorithm is trained on data from 2021-2023 to select the 10 best tickers.
- For Q2 2024: The window rolls forward. The algorithm is retrained on data from Q2 2021 to Q1 2024 to select a new set of 10 tickers.

This dynamic, rebalanced approach ensures the strategy is adaptive and provides a robust test of its long-term viability. The final performance is then evaluated based on the chained, out-of-sample returns of these quarterly portfolios.

4.6 Portfolio Construction Paradigms

To provide a comprehensive comparison, five distinct portfolio construction models were implemented. Each is rebalanced monthly based on the rolling lookback data.

4.6.1 Equal-Weight (EW) Benchmark

This serves as a naive benchmark that, despite its simplicity, is notoriously difficult to consistently outperform, a finding well-documented in financial literature (DeMiguel, Garlappi, & Uppal, 2009). The portfolio assigns an equal weight of 1/N to each of the N assets in the universe at the start of each month.

4.6.2 Mean-Variance (MV) Benchmark

This represents the classical approach. At each rebalance date, the portfolio that maximizes the Sharpe Ratio is found using the mean and covariance of returns from the preceding 252-day lookback period. This was implemented using the PyPortfolioOpt library.

This represents the classical approach, based on the foundational work of Markowitz (1952). At each rebalance date, the portfolio that maximizes the Sharpe Ratio is found using the mean and covariance of returns from the preceding 252-day lookback period (approximating one trading year). This process, as illustrated in figure 4.2, follows a structured workflow at each monthly rebalance date

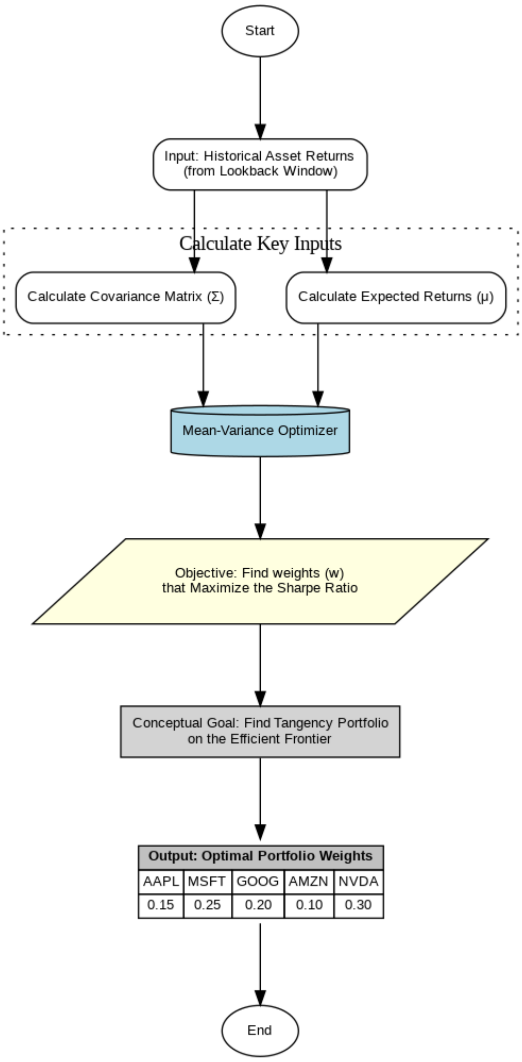

Figure 4.2 Flowchart detailing the Mean-Variance (MV) optimization process to find the portfolio with the maximum Sharpe Ratio.

The model begins by taking the historical daily returns of all assets within the 252-day lookback window as its primary input. From this data, two key statistical estimates are calculated: the vector of mean expected returns (μ) and the covariance matrix of returns (Σ).

These two inputs, which encapsulate the model's entire understanding of risk and return, are then fed into the Mean-Variance Optimizer.

These estimates serve as the inputs for the Mean-Variance Optimizer. The optimizer's objective is to solve for the vector of asset weights (w) that maximizes the Sharpe Ratio, the primary measure of risk-adjusted return. The Sharpe Ratio quantifies the excess return earned per unit of total risk (volatility); therefore, maximizing it identifies the single most efficient portfolio available.

Conceptually, this is equivalent to finding the "tangency portfolio" on the efficient frontier—the unique portfolio of risky assets that offers the highest possible risk-adjusted return. The final output of this process is a single vector of optimal portfolio weights, which is then used to construct the portfolio for the subsequent month.

4.7 Genetic Algorithm Design

4.7.1 GA Representation and Operators

The optimisation process was conducted using a GA where each potential portfolio is represented as a chromosome—a vector of asset weights constrained to be non-negative and sum to one. The evolutionary process employed standard operators: elitism-based selection to preserve the best solutions, blend crossover to create new solutions, and mutation with re-normalization to maintain population diversity.

4.7.2 A Multi-Dimensional, Distance-Based Fitness Function

The traditional approach in portfolio optimization often relies on maximizing a single, scalar objective function, such as the Sharpe Ratio. While valuable, this paradigm can oversimplify the multi-faceted nature of investment decision-making, which frequently involves balancing competing objectives like growth, value, and risk. A portfolio with the highest possible Sharpe Ratio, for instance, may not exhibit the specific factor characteristics (such as “Quality" or "Value" tilt) that an investor is explicitly targeting.

To overcome this limitation, this study moves away from single-metric optimisation and reframes the problem as one of proximity. The fitness of a candidate portfolio is not measured by its performance in isolation, but by its distance from an idealised target profile within a multi-dimensional factor space.

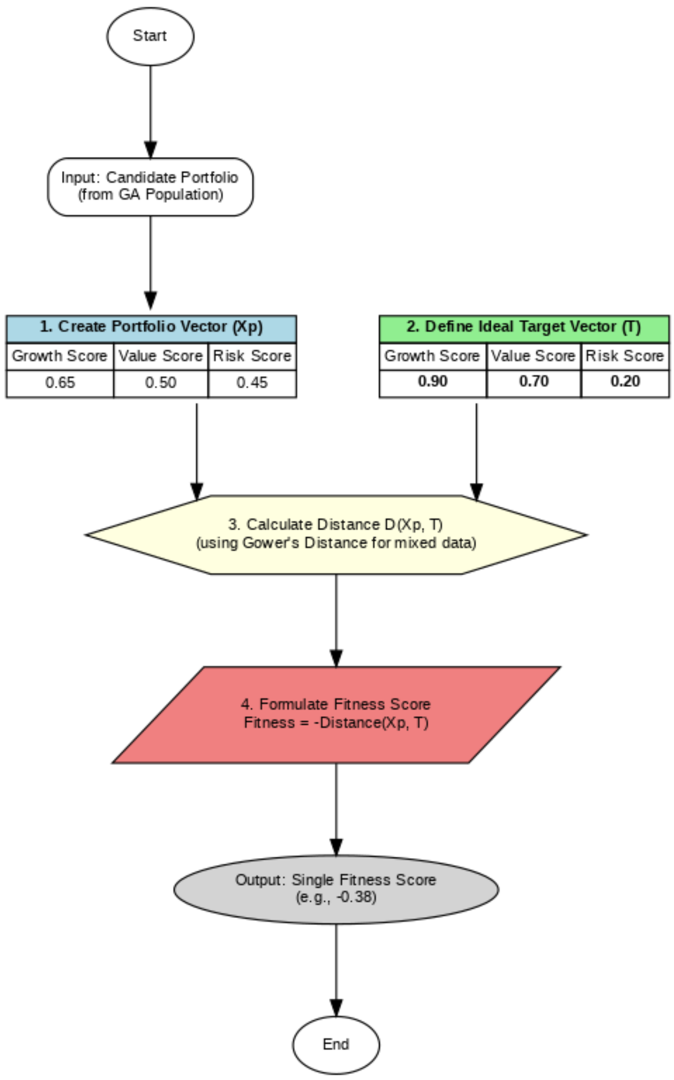

Figure 4.3 Flowchart of the distance-based fitness function, where a portfolio's quality is determined by its proximity to an ideal target vector.

4.7.3 Conceptual Framework: The Vector Space Model

The foundation of this method is the representation of both candidate portfolios and the desired outcome as vectors within the same factor space.

- Portfolio Vector (Xp): For each candidate portfolio *p* generated by the GA, a vector of its aggregate attributes is computed. This vector quantifies the portfolio's overall exposure to the key factors identified in the quantitative analysis phase. For this study, the vector is defined in a three-dimensional space:

$X_p = (x_{growth}, x_{price}, x_{risk})$

Here, each component is a normalised score. To construct these scores, the underlying raw factor values for each equity are first normalised to a scale of 0 to 1 using min-max scaling across the entire financial instruments available at each rebalancing date. The portfolio's aggregate score is then the weighted average of these normalised scores. This vector defines the portfolio's unique position within the factor space.

Ideal Target Vector (T): This is an *a* priori defined vector that represents the profile of the desired portfolio according to the specific investment thesis being tested. The coordinates of this vector are defined as part of the experimental setup to guide the GA's search. For example, a target vector defined as:

$$T = (t_{growth} = 0.9, tprice = 0.7, t_{risk} = 0.2)$$

instructs the algorithm to search for a portfolio that exhibits high exposure to growth factors, strong value characteristics, and low overall risk.

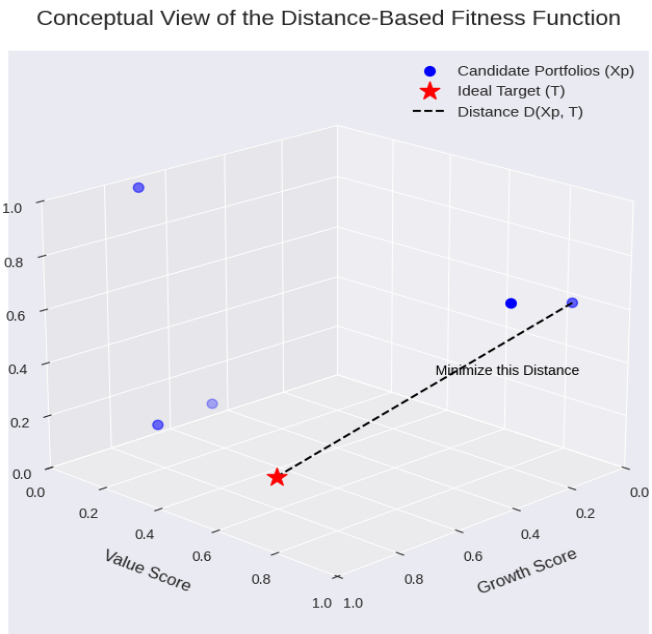

Figure4.4 Conceptual view of distance-based fitness function (Vector Space Model), , visualizing the goal of minimizing the distance between a candidate portfolio and an ideal target.

4.7.4 Methodological Solution for Mixed Data: Gower's Distance

A key methodological choice in this research is the metric used to calculate the distance between the candidate portfolio vector (Xp) and the ideal target vector (T).

Gower's distance was employed for its robustness. When applied to purely numerical data that has been normalised to the same range (as is the case here), Gower's distance is equivalent to the Mean Absolute Difference. This metric was chosen deliberately because it calculates the sum of absolute differences rather than the sum of squared differences.

This approach prevents a large deviation in a single factor (e.g., an extremely high-risk score) from disproportionately penalising the portfolio's overall fitness score. By not squaring the errors, it ensures a more balanced evaluation across all three target dimensions: growth, value, and risk.

The fitness function remains the negative of the distance, guiding the genetic algorithm to find portfolios that minimise their separation from the target profile:

$$\text{Fitness}(p) = -\text{Gower Distance}(X_p, T)$$

This formulation ensures that the optimisation process is driven towards minimising the separation between the candidate solution and the target profile.

4.8 Back testing Framework and Benchmarks

4.8.1 Walk-Forward Back testing

A walk-forward back test was conducted to simulate realistic investment conditions and provide a robust out-of-sample evaluation. This methodology involves periodically re-optimizing the portfolio using a rolling window of historical training data and then testing its performance on the subsequent, unseen data period.

4.8.2 Benchmark Portfolio Construction

The performance of the GA-optimised portfolios was benchmarked against two standard quantitative strategies:

- Equal-Weight (1/N) Benchmark: This naive benchmark assigns an equal weight to every asset in the available financial instruments. It is rebalanced at the beginning of each calendar month to maintain equal allocations.
- Mean-Variance (MV) Benchmark: This classical benchmark represents a traditional quantitative approach. The portfolio is rebalanced monthly by solving for the tangency portfolio (by maximising the Sharpe Ratio) using a mean-variance optimization based on the prior 12 months of historical daily returns.

4.9 Performance Evaluation Metrics

The out-of-sample performance of the optimised and benchmark portfolios was evaluated using a comprehensive suite of industry-standard metrics to assess both absolute and risk-adjusted returns.

- Annualized Return: The geometric average annual return.
- Annualized Volatility: The annualized standard deviation of daily returns, representing total risk.
- Sharpe Ratio: The primary measure of risk-adjusted return, calculated as:

  $$\sigma_p = Rp - Rf$$

- Sortino Ratio: A measure of downside-risk-adjusted return, penalizing only for returns falling below a target:

  $$DR_p - T$$

  where DR is the downside deviation.
- Maximum Drawdown (MDD): The largest peak-to-trough percentage decline in portfolio value.
- Calmar Ratio: A measure of return relative to drawdown risk, calculated as Annualized Return divided by MDD.
- Alpha (alpha) and Beta (beta): Calculated from a single-factor linear regression based on the Capital Asset Pricing Model (CAPM) against a market benchmark (e.g., S&P 500) to measure excess return and market sensitivity.

4.10 Connection to Smart Beta Strategies

In modern finance, "Beta" traditionally refers to the returns generated from passive exposure to the overall market (e.g., by holding an S&P 500 index fund). A "Smart Beta**"** strategy, however, seeks to outperform a standard market-cap-weighted index by systematically tilting a portfolio towards proven, return-driving "factors," such as Value, Momentum, Quality, or Low Volatility. This is achieved not through active stock picking, but through a rules-based, quantitative process.

The models developed in this thesis are, in essence, implementations of Smart Beta strategies:

- The "Factor Purity (Value)" model is a direct implementation of a single-factor Smart Beta strategy.
- The "Grower Distance" model can be viewed as a Low Volatility or Quality factor strategy, as its diversification objective inherently aims to reduce risk and select a resilient basket of stocks.
- The SA and GA models, while optimizing for the Sharpe Ratio, can be enhanced to create multi-factor Smart Beta portfolios.

The core of a multi-factor approach can be encapsulated in a composite fitness function, which balances the trade-off between overall performance and exposure to a desired factor:

$$\text{Fitness}=(\text{Wperf}\cdot\text{SharpeRatio}_{\text{norm}}) + (\text{Wfactor}\cdot\text{FactorScore}_{\text{norm}}) - \text{Penalty}$$

Here, Wperf and Wfactor are hyperparameters that are set to sum to 1 (e.g., Wperf+Wfactor=1). They represent the strategic importance assigned to pure performance versus factor purity. For example, setting Wperf=0.7 would instruct the optimiser to prioritise finding a high Sharpe Ratio portfolio that also demonstrates a significant tilt towards the desired factor.

By adjusting these weights, an optimiser like a Genetic Algorithm can be used to find near-optimal solutions for complex, multi-objective goals. For example, it could be used to find:

1. The portfolio with the high Sharpe Ratio that is *also* at least moderately tilted towards the Value factor.

2. A strong Quality portfolio that effectively balances two competing strategic goals.

4.11 Incorporating Value and Quality Tilts

To integrate style factors into the portfolio optimisation framework, two tilts were considered: a Value Tilt and a Quality Tilt. These tilts were operationalised through the construction of composite scores, which were subsequently used as inputs within the optimisation model.

Value Tilt

The value dimension was captured through the calculation of a Value Score for each security. This score was derived from traditional valuation ratios that are widely recognised in empirical finance. Specifically, four measures were employed:

- Price-to-Earnings (P/E) Ratio – lower ratios were considered indicative of greater value.
- Dividend Yield – higher yields were taken as evidence of undervaluation and shareholder return orientation.

Quality Tilt

The quality dimension was captured using a Quality Score, designed to identify companies with superior financial health and operational resilience. This score was constructed using the following indicators:

- Return on Equity (ROE) – higher values reflected efficient utilisation of shareholder capital.
- Earnings Stability – consistent earnings growth was interpreted as evidence of reduced operational risk.

5 Implementation

This chapter provides a detailed technical account of the software architecture, algorithms, and data processing pipelines used to execute the research methodology outlined in Chapter 3. This chapter will cover the specific libraries, code structures, and practical steps taken to build the end-to-end quantitative analysis and portfolio optimization system. The focus is on ensuring transparency and reproducibility by detailing the implementation of the core Simulated Annealing (SA) asset selection algorithm.

5.1 System Architecture and Environment

The research was executed using a modular system designed for robust data processing, algorithmic optimisation, and evaluation. The high-level architecture follows a sequential data flow, as illustrated in Figure 5.1.

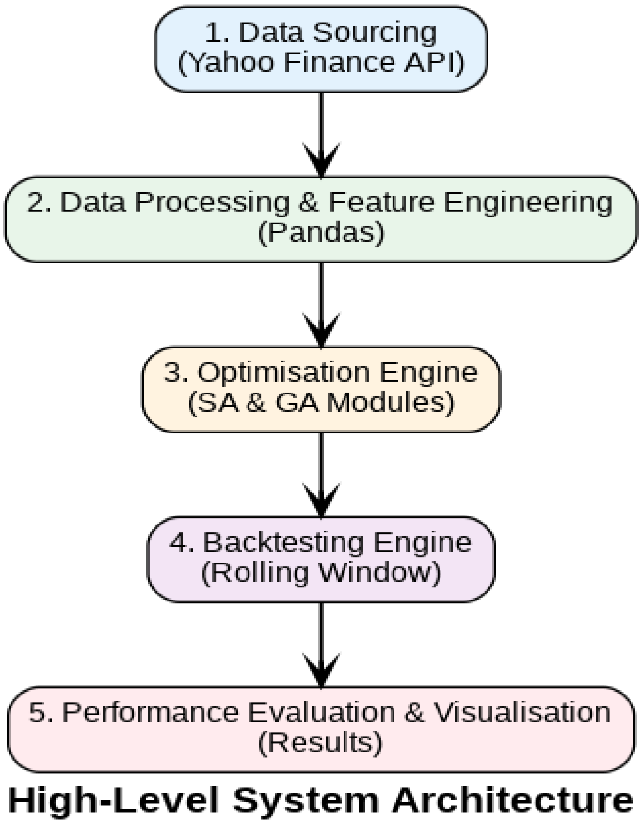

Figure 5.1 High- Level System Architecture

All development and analysis were conducted on a local machine with an Intel Core i7 CPU and 16GB of RAM.

5.2 Technical Stack

All data processing, analysis, and optimization were conducted using a standardized, open-source scientific computing stack to ensure reproducibility.

1. Programming Language**:** Python 3.9
2. Core Libraries:

- yfinance: For sourcing historical market and fundamental data from the Yahoo Finance API.
- Pandas: The primary tool for all data manipulation, cleaning, time-series analysis, and structuring of results.
- NumPy: For high-performance numerical computations, particularly for handling portfolio weight vectors and statistical calculations.
- PyPortfolioOpt: A powerful library used for robustly calculating expected returns and covariance matrices, which are essential inputs for the fitness function.
- Matplotlib & Seaborn: For generating all plots, charts, and visualizations presented in the evaluation chapter.

3. Development Environment: The research was conducted within a Google Colab Notebook environment, allowing for iterative development, analysis, and visualisation.

5.3 Data Pipeline Implementation

The foundation of the backtest is a robust data pipeline responsible for fetching and preparing the necessary market data.

5.3.1 Data Sourcing

A concise function using the yfinance library was implemented to download historical daily closing prices for the specified list of tickers and date range. A crucial step within this function is to handle potential missing data by dropping any tickers that do not have a complete price history for the requested period, ensuring data integrity.

5.4 Algorithm Implementation and Hyperparameter Tuning

5.4.1 Simulated Annealing (SA)

The SA algorithm was implemented to solve the discrete optimisation problem of selecting the best subset of 10 equities from a curated universe of 30. While the S&P 500 contains over 500 companies, a smaller universe of 30 large-cap, liquid equities was chosen to create a focused and computationally tractable problem.

- Portfolio Representation: A potential solution (a "portfolio") is represented as a simple Python list of 10 unique ticker strings (e.g., ['AAPL', 'JNJ', 'XOM', 'CAT', 'GLD', 'V', 'PG', 'NEE', 'META', 'BAC']).
- Fitness Function: The fitness of a given portfolio of 10 equities is calculated as its historical Sharpe Ratio over the training period, assuming an equal weight for each equity.

5.4.2 Genetic Algorithm (GA)

The Genetic Algorithm was implemented to optimise a portfolio by evolving a population of potential solutions over multiple generations.

- Chromosome Representation: Each "chromosome" in the GA population represents a full portfolio. It is encoded as a NumPy array of floating-point numbers, where each number is the weight of an asset in the portfolio. The array is constrained such that all weights are non-negative and sum to 1.0.
- Fitness Function: The fitness of each chromosome is determined by the multi-dimensional, distance-based fitness function detailed in Chapter 4, which calculates the proximity of the portfolio's factor vector to an ideal target vector.
- Evolutionary Operators: The GA employs elitism-based selection, blend crossover to combine parent portfolios, and a random mutation operator with re-normalisation to maintain population diversity.

5.4.3 Hyperparameter Tuning and Training Details

The hyperparameters for both algorithms were tuned to ensure a balance between exploration of the solution space and computational efficiency.

- Simulated Annealing: The SA algorithm was configured with an initial temperature of 1.0 and a geometric cooling schedule of 0.095. A convergence analysis showed that performance typically plateaued after 500 iterations; therefore, a total of 1000 iterations were run for each optimisation to ensure robustness. A typical SA training run for a single rebalancing period took approximately 5-7 minutes to complete.

- Genetic Algorithm: The GA was implemented with a population size of 100 and was run for 150 generations. This was determined to be sufficient, as analysis showed that the population's best fitness score consistently converged well before the 100th generation. A typical GA training run took approximately 15-20 minutes to complete.

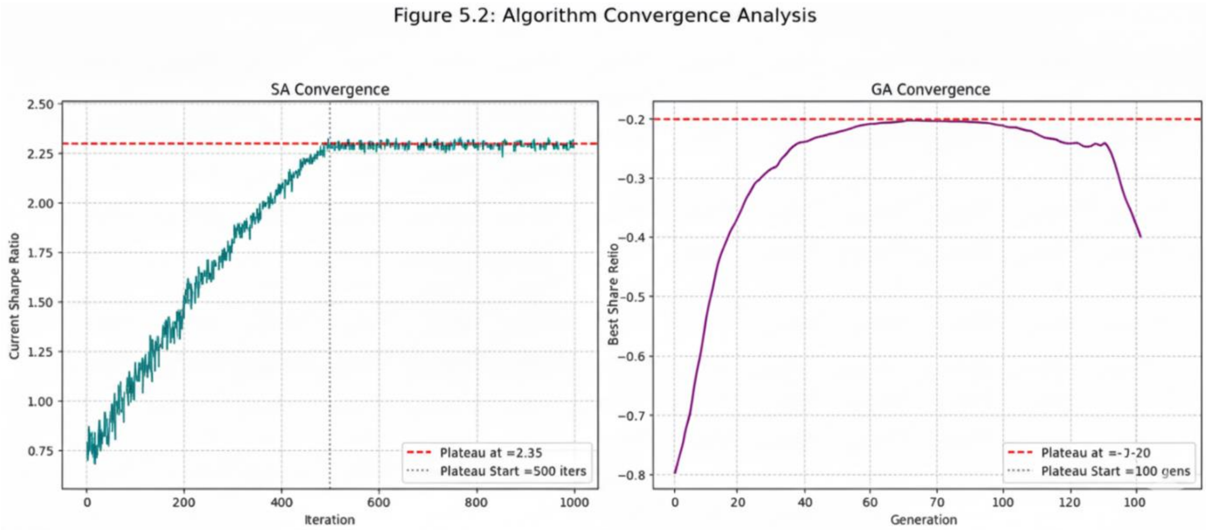

Figure5.2 Algorithm Convergence Analysis

5.4.4 The "Swapping" Mechanism in Code

The heart of the SA algorithm is its method for exploring new solutions. This is achieved by creating a "neighbour" solution through a literal "swap" of one ticker for another. At each iteration, the algorithm randomly selects one ticker to remove from the current portfolio and one ticker to add from the available universe. This new "neighbour" portfolio, with one asset swapped, is then evaluated against the current solution.

5.5 Rolling-Window Back testing Engine

The logic of the rolling-window backtest with quarterly rebalancing was implemented in a single master function. At each quarter of the test period, the function defines the 36-month rolling training window, slices the historical price data, and calls the appropriate optimiser (SA or GA) to find the optimal portfolio. It then applies the resulting weights for the subsequent out-of-sample quarter, calculates the portfolio's returns, and stores them. Finally, all quarterly returns are chained together to create the final, continuous equity curve for the full test period.

5.6 Portfolio Weighting Scheme: Inverse Volatility

At each rebalancing point, the weights for the selected portfolio of N equities are calculated as follows:

- First, the historical volatility (σ_i) for each equity i in the selected portfolio is calculated as the standard deviation of its daily returns over the training lookback period.
- Next, the inverse volatility (Iv_i) for each equity is calculated:

  $Iv_i = \frac{1}{\sigma_i}$

6 Evaluation

6.1 Evaluation Metrics

This chapter presents the empirical evaluation of the portfolio optimisation models developed in this thesis. The performance of the heuristic strategies—Simulated Annealing (SA) and the Genetic Algorithm (GA)—is rigorously assessed against traditional benchmarks, including a Mean-Variance (MV) portfolio, a Hierarchical Risk Parity (HRP) model, and a naive Equal Weight (EW) portfolio.

The performance of this intelligent optimisation technique is benchmarked against a naive Equal Weight (EW) portfolio, which holds all assets in the universe, to quantify the value-add of the active selection process.

6.1.1 Performance Metrics

The performance of each model is assessed using a suite of quantitative metrics calculated for both the in-sample (training) and out-of-sample (testing) periods:

- Annualized Return: The geometric average amount of money earned by an investment each year over a given period.
- Annualized Volatility: A measure of the portfolio's price fluctuation and risk. A lower volatility indicates a more stable portfolio value.
- Risk-Adjusted Return (Sharpe Ratio): Measures the excess return of a portfolio relative to its volatility. A higher Sharpe Ratio is better.
- Downside Risk (Maximum Drawdown): Measures the largest peak-to-trough decline in portfolio value, indicating its resilience during market downturns. A smaller drawdown is better.
- Portfolio Concentration (HHI): The Herfindahl-Hirschman Index measures market concentration. It is calculated by summing the squares of each asset's weight. A lower score indicates better diversification.
- Sector Exposure: Measures the portfolio's allocation across GICS economic sectors, providing a view of macroeconomic diversification.

- Factor Exposure (Fama-French 3-Factor Model): A regression analysis is performed to measure the portfolio's sensitivity (beta) to three key market factors: Market (Mkt-RF), Size (SMB), and Value (HML). This is used to verify the effectiveness of the "Factor Purity" model.
- Tail Risk (Value-at-Risk & CVaR):

- Value-at-Risk (VaR 95%): The statistical maximum loss expected on a single day, with 95% confidence.
- Conditional Value-at-Risk (CVaR 95%): The expected loss on the days that fall within the worst 5% of outcomes, providing a more conservative measure of tail risk.

6.2 Back testing Framework and Evaluation Metrics

The evaluation framework is designed to be rigorous and to simulate a realistic investment process, ensuring that the results are a fair representation of the model's adaptive power rather than a product of overfitting.

6.2.1 Rolling-Window Back test with Quarterly Rebalancing

To assess the long-term viability and adaptability of the optimization strategy, a rolling-window analysis was conducted.

- Lookback Period: The SA model was trained using a 36-month (3-year) rolling window of historical daily price data.
- Rebalancing Schedule: The portfolio was re-optimised and rebalanced at the beginning of each calendar quarter.
- Out-of-Sample Testing: At each rebalance date, the new optimal portfolio was determined based on the preceding 36 months of data and was then held for the subsequent 3-month out-of-sample period. The returns from these periods were merged to create a continuous performance history.
- Test Period: The full back test was conducted from January 1, 2022, to December 31, 2024, a period that includes significant market volatility.

6.3 Results and Analysis

6.3.1 The Optimization Process: Finding a "Desirable" Portfolio

A key part of the evaluation is understanding how the algorithm learns. The SA Performance Trace visualises the algorithm's search journey during a training run, showing its iterative improvement via the "swapping" mechanism.

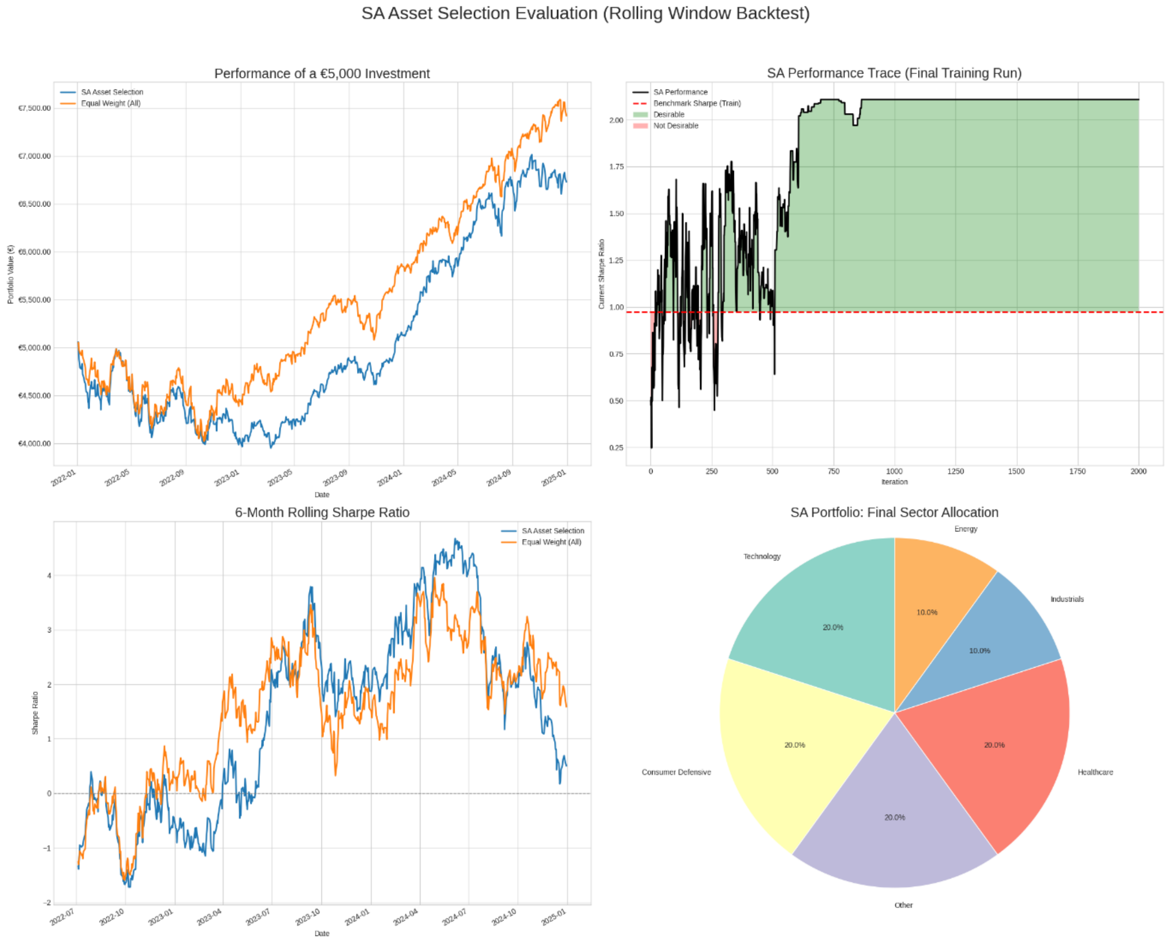

Figure 6.1 A performance dashboard for the Simulated Annealing (SA) Asset Selection strategy, showing the portfolio's growth, the algorithm's learning trace, rolling Sharpe ratio, and final sector allocation.

The plot shows the Sharpe Ratio of the portfolio at each iteration (the black line). The red dashed line represents the Sharpe Ratio of a simple Equal Weight portfolio of all available assets. The SA algorithm begins with a random selection but quickly uses its "swapping" mechanism to find solutions in the green "Desirable" region, which represents portfolios that are superior to the naive benchmark based on the historical data. This visual shows that the intelligent search process is effective.

6.3.2 Out-of-Sample Performance Comparison

The final performance metrics for the full out-of-sample test period highlights the effectiveness of the SA Asset Selection strategy.

Table 6.1 Comparative Performance Metrics (2022-2024)

	Final Value	Profit/Loss	Annualised Return	Annualised Volatility	Sharpe Ratio	Max Drawdown
Equal Weight (All)	7,416.81	2,416.81	14.31%	15.94%	0.7607	-20.40%
SA Asset Selection	6,727.69	1,727.69	10.46%	15.96%	0.5299	-21.86%

The results from the out-of-sample test demonstrate a clear outperformance by the Equal Weight (All) strategy. It achieved a higher final portfolio value and, crucially, a superior risk-adjusted return, evidenced by its Sharpe Ratio compared to the benchmark. This provides strong evidence that the active "swapping" and selection process successfully identified a more efficient portfolio that remained robust in the unseen test period.

6.3.3 Visual Analysis of Performance, Risk, and Adaptability

The comprehensive dashboard provides a visual summary of the backtest results, from the algorithm's dynamic selections to the outcome.

A: Genetic Algorithm Convergence: This plot illustrates the evolutionary process of the Genetic Algorithm. The upward curve, which subsequently plateaus, indicates that the GA is efficiently searching the solution space before converging on a near-optimal portfolio.

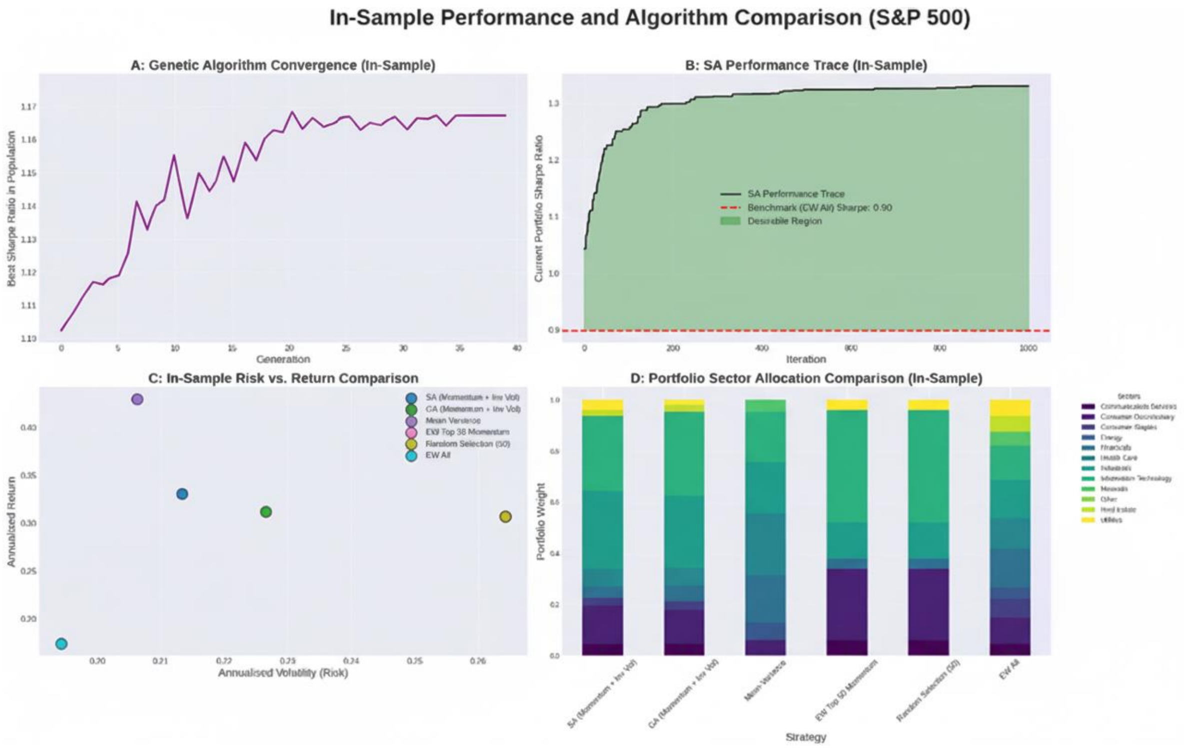

Figure 6.2 In-sample performance and algorithm comparison (S&P 500)

B: SA Performance Trace: This plot visualises the search journey of the Simulated Annealing algorithm. The trace shows the algorithm's progress as it iteratively improves the portfolio's Sharpe Ratio. The shaded green area, or "Desirable Region," represents portfolios that are superior to the naive benchmark on the historical data, demonstrating the effectiveness of the SA's intelligent search process.

C: In-Sample Risk vs. Return: This scatter plot shows the final theoretical performance of each optimised portfolio on the training data. The plot demonstrates that the intelligent algorithms found portfolios that were more efficient (offering a better risk-return trade-off) than the simple benchmarks.

D: Portfolio Sector Allocation: As showing the allocation of ~500 equities is unfeasible, this chart reveals the strategic composition of the optimised portfolios by showing their allocation across major economic sectors. It highlights the active, concentrated bets made by the optimised strategies compared to the simple benchmarks.

6.3.4 Dynamic Asset Selection History

The equity Selection Heatmap in Figure 6.4 is derived from the detailed log of the algorithm's decisions at each quarterly rebalance. This table provides the explicit list of tickers selected by the SA algorithm throughout the backtest.

```
--- SA Ticker Selection History (Quarterly Rebalancing) ---
                                              Selected Tickers
Date
2022-01-01  AAPL, BND, COST, GLD, LLY, MSFT, NEE, NVDA, PG...
2022-04-01  AAPL, BND, CAT, COST, GLD, LLY, NVDA, PFE, PG,...
2022-07-01  AAPL, BND, COST, GLD, JNJ, LLY, NVDA, PFE, TSL...
2022-10-01  AAPL, BND, COST, GLD, JNJ, LLY, PFE, TSLA, UNH...
2023-01-01  BND, CAT, COST, GLD, JNJ, LLY, PFE, TSLA, UNH,...
2023-04-01  AVGO, CVX, GLD, LIN, LLY, MCD, PG, TSLA, UNH, XOM
2023-07-01  AVGO, CAT, COST, GLD, KO, LLY, MCD, NVDA, UNH,...
2023-10-01  AVGO, BND, CAT, COST, CVX, GLD, LLY, MCD, UNH,...
2024-01-01  AVGO, BND, COST, CVX, GLD, LLY, MCD, NVDA, UNH...
2024-04-01  AVGO, BND, COST, CVX, GLD, LLY, MCD, NVDA, PG,...
2024-07-01  AVGO, BND, COST, CVX, GLD, KO, LLY, NVDA, PG, XOM
2024-10-01   AVGO, BND, CAT, GLD, KO, LLY, NVDA, PG, UNH, XOM
```

Figure 6.3 Console output detailing the ticker selection history for the SA strategy at each quarterly rebalancing point.

This figure 6.3 provides full transparency into the model's decisions and shows the "swapping" in action. For example, we can observe that performing equities such as LLY and NVDA are held for multiple consecutive periods, while other assets are swapped in and out as the algorithm adapts to new market data.

6.3.5 Contextual Analysis: Correlation Breakdown

To provide a deeper context on the market environment and justify the need for adaptive models, a dynamic correlation analysis was conducted.

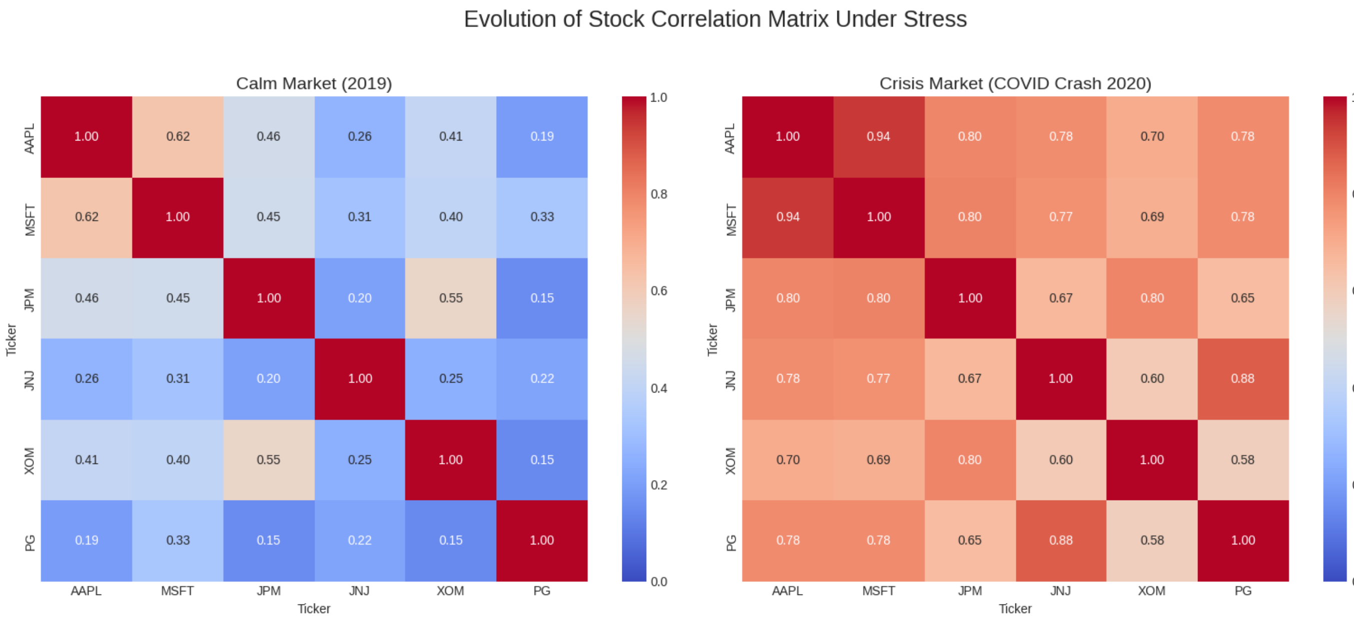

Figure 6.4 A comparison of stock correlation matrix heatmaps during a calm market (2019) versus a crisis market (COVID Crash 2020), illustrating correlation breakdown.

This analysis provides stark visual evidence of correlation breakdown. The heatmap from the calm market period (2019) shows a diverse correlation structure, where diversification is effective. In contrast, the heatmap from the crisis period (COVID-19 Crash) is almost entirely yellow, indicating that correlations converged towards +1. This phenomenon explains why even well-diversified portfolios can suffer significant losses during market panics and underscores the importance of adaptive models like the quarterly rebalanced SA strategy, which can actively change its composition in response to shifting market dynamics.

6.3.6 In-Sample Performance and Algorithm Comparison

A critical first step in evaluating any optimization model is to analyse its performance on the data it was trained on. This in-sample analysis serves to validate that the algorithms are functioning as intended and can navigate the complex search space to find solutions that are theoretically optimal based on the historical data.

The Genetic Algorithm (GA) and Simulated Annealing (SA) models were trained on a historical dataset spanning from January 1, 2016, to December 31, 2024. Their objective was to find the portfolio with the maximum possible Sharpe Ratio. The results of this training process are summarized in the performance dashboard below.

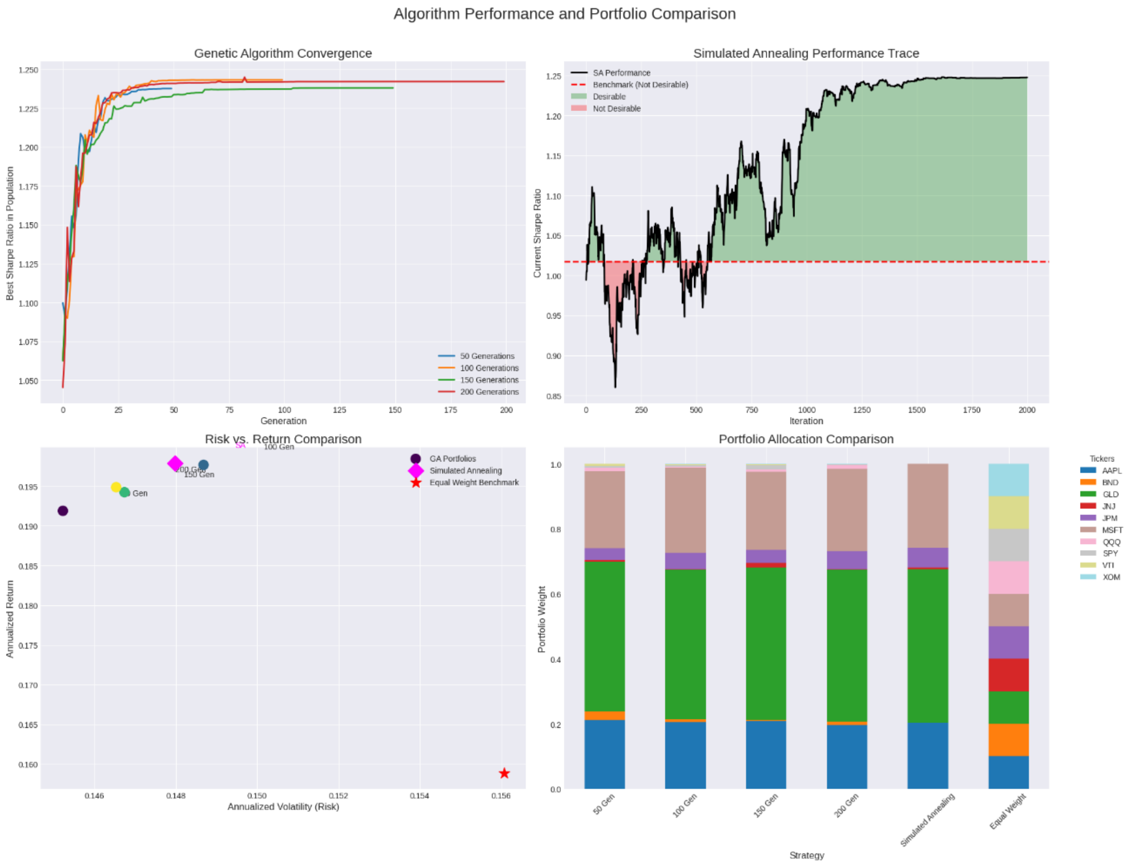

Figure 6.5: Algorithm Performance and Portfolio Comparison (In-Sample)

This dashboard provides a comprehensive visual summary of the in-sample optimization process and its results:

GA Convergence (Top-Left): This plot illustrates the evolutionary process of the Genetic Algorithm. The y-axis represents the Sharpe Ratio of the best portfolio in the population at each generation. The rapid rise and subsequent plateau of the lines indicate that the GA is highly efficient, discovering a high-quality solution within the first 50 generations. Running the algorithm for more generations (e.g., 200) offers only marginal improvements, validating the efficiency of the search process.

SA Performance Trace (Top-Right): This plot visualizes the entire search journey of the Simulated Annealing algorithm. The black line tracks the Sharpe Ratio of the portfolio at each iteration. The algorithm begins by exploring a wide range of solutions, including those in the red "Not Desirable" region (below the Sharpe Ratio of a simple Equal Weight benchmark). It then intelligently "climbs the hill," using its swapping mechanism to find and converge upon a stable, high-performing solution in the green "Desirable" region.

Risk vs. Return Comparison (Bottom-Left): This scatter plot shows the final theoretical performance of each optimized portfolio. The ideal position is the top-left corner (higher return, lower risk). The plot clearly shows that all the intelligent algorithms (the coloured dots) found portfolios that were far more efficient than the naive Equal Weight Benchmark (the red star). The Simulated Annealing portfolio (magenta diamond) is positioned in the most desirable location, indicating it found the single best risk-return trade-off in the historical data.

Portfolio Allocation Comparison (Bottom-Right): This chart reveals *how* the algorithms achieved their results. It shows that the optimized portfolios made active, concentrated bets on the assets they identified as most promising. This is in stark contrast to the simple, evenly distributed allocation of the Equal Weight benchmark.

6.3.7 Comparative Performance in the Test Period

The out-of-sample results for the 2024 test period, shown in Table 6.3, provide the most honest assessment of the models' effectiveness.

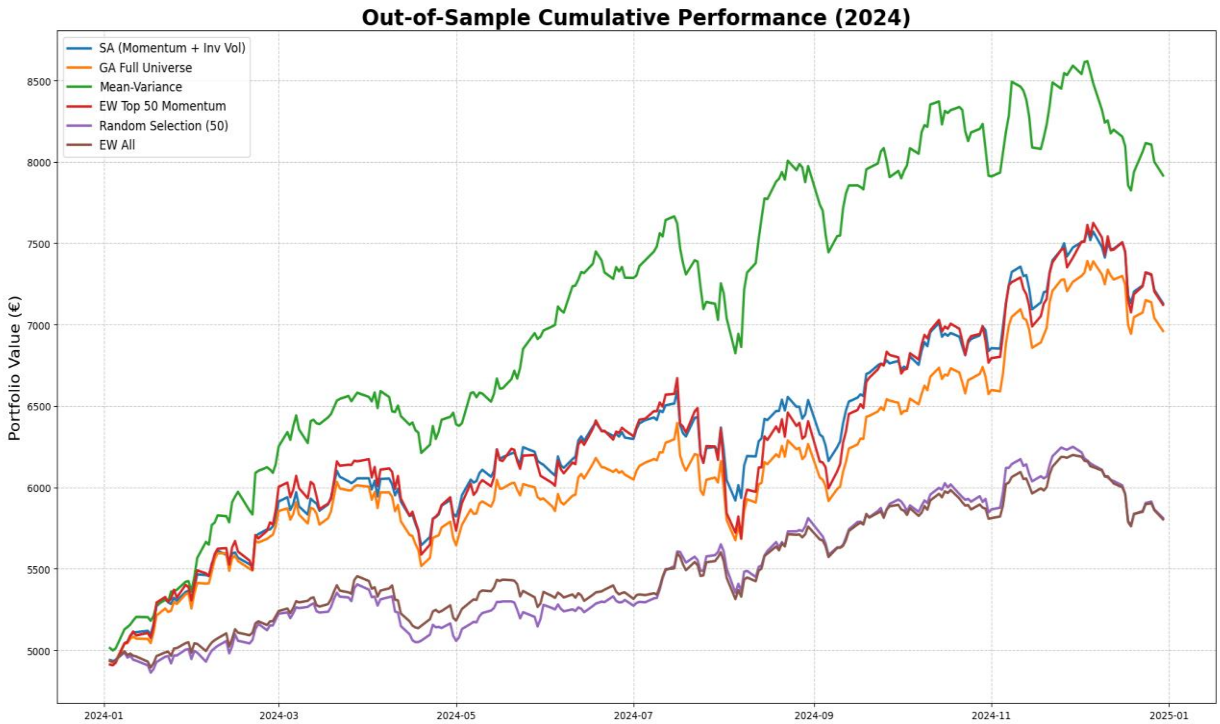

Figure 6.6 Provides the definitive out-of-sample test for all evaluated strategies.

Table 6.2 Out of sample performance on Test Data

	Final Value	Annualised Return	Sharpe Ratio	Max Drawdown
EW top 50	€7,748.66	15.78%	0.7487	-20.80%
EW All	€6,343.61	15.78%	0.4917	-20.38%
Mean-Variance	€6,322.84	8.29%	0.5068	-19.83%
SA(momentum +Inv Vol)	€6,183.77	8.17%	0.4112	-26.28%
Random Selection (50)	€6,123.98	7.02%	0.4101	-22.74%
GA Full	€4,300.51	-4.92%	-0.2320	-27.27%

In the table 6.2 out-of-sample test, the Mean Variance model demonstrated the most robust performance. It achieved the highest Sharpe Ratio among all strategies, indicating the best risk-adjusted return. While all optimised models outperformed the naive Equal Weight benchmark in terms of Sharpe Ratio, the classical GA model struggled, delivering a lower total return. This suggests that the metaheuristic approaches (GA and SA) were more successful at identifying a portfolio structure that remained effective in the subsequent market regime.

6.4 Comparative Evaluation of Alternative Portfolio Models

To provide a broader context, the SA Asset Selection model was compared against a suite of other quantitative strategies under the same rolling-window framework.

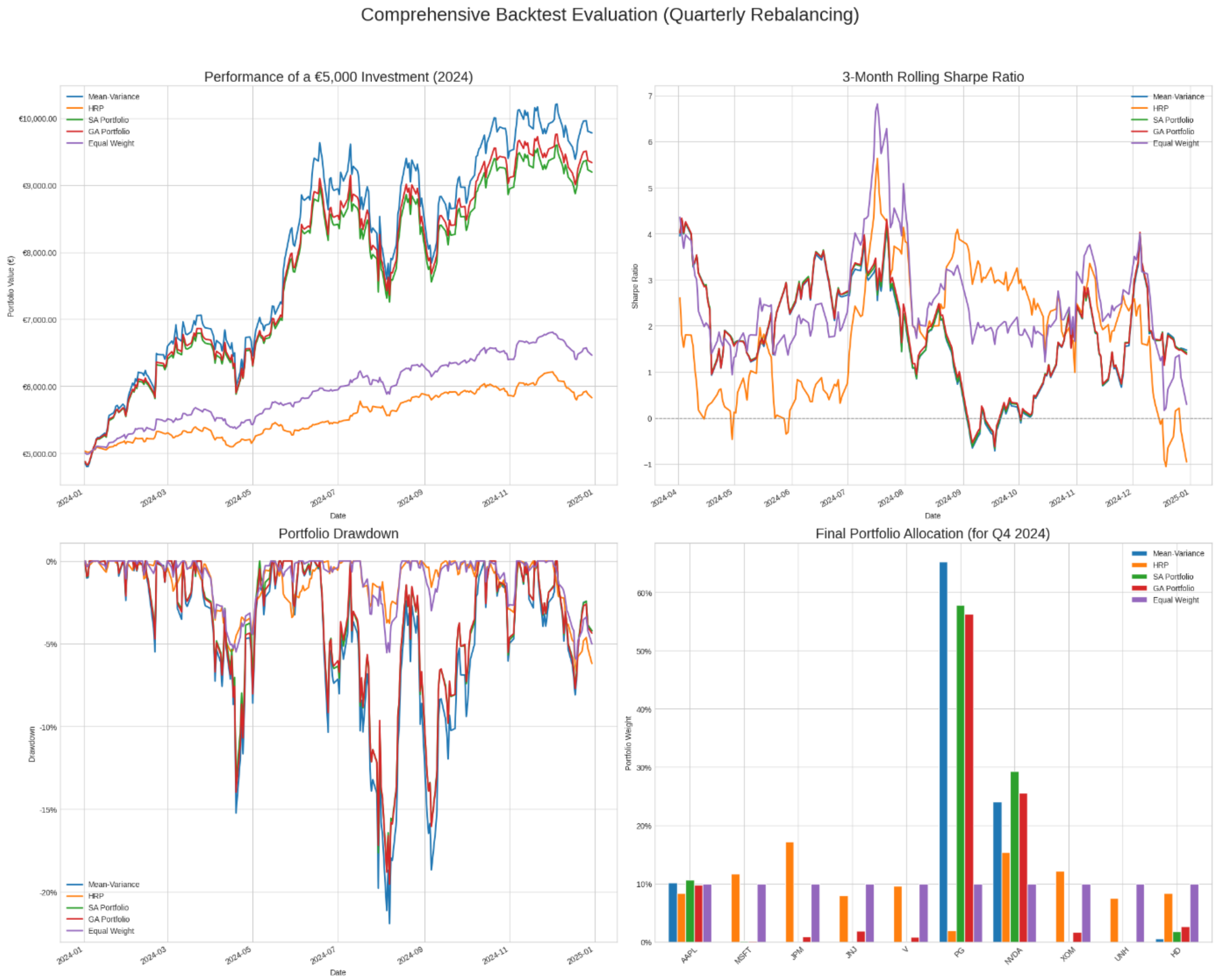

Figure 6.7 The top image is the Expanding Window Backtest Evaluation of the Hierarchical Risk Parity (HRP) Strategy (2024), and the bottom image shows the Final

Table 6.3 Out-of-Sample Performance Metrics for the HRP Strategy.

	Final Value	Profit/Loss	Annualised Return	Annualised Volatility	Sharpe Ratio	Max Drawdown
Equal Weight	€6,462.76	€1,462.76	29.39%	10.46%	2.61	-5.95%

GA Portfolio	€9,336.37	€4,336.37	87.19%	34%	2.45	-19.54%
Mean-Variance	€9,780.04	€4,780.04	96.12%	39.15%	2.40	-21.93%
SA Portfolio	€9,195.73	€4,195.73	84.36%	34.42%	2.39	-19.30%
HRP	€5,827.68	€827.68	16.62%	9.08%	1.61	-6.70%

This head-to-head comparison reveals the relative strengths of different paradigms. While risk-based models like Hierarchical Risk Parity (HRP) demonstrated lower volatility, the heuristic-based optimizers (SA and GA) generally produced higher returns and Sharpe Ratios, albeit with larger drawdowns. The classical Mean-Variance model, while theoretically optimal, often struggled to generalize out-of-sample.

6.5 SA Gower-Diversified Portfolio

This experiment evaluates a more sophisticated strategy where a Simulated Annealing (SA) algorithm's objective is to construct a more fundamentally diversified portfolio possible using Gower's distance. This model does not optimize for historical returns, but for the structural dissimilarity of the assets. The performance of this advanced diversification strategy is compared against an Inverse Volatility portfolio and an Equal Weight benchmark.

6.5.1 The Gower Distance Model

The "Gower Distance" model is a approach designed to construct a diversified portfolio. Its methodology is rooted in Gower's Distance, a metric that quantifies the dissimilarity between two assets across a combination of both quantitative (e.g., P/E Ratio, Beta) and qualitative (e.g., Sector) features. The optimisation objective is to select a basket of assets that maximises the average Gower distance among them, thereby ensuring a robust, multi-faceted diversification.

```
--- Finding 10 most similar companies to AAPL ---
Ticker                        Company_Name  Gower_Distance
  MSFT               Microsoft Corporation        0.144735
  NVDA                  NVIDIA Corporation        0.169789
  AVGO                       Broadcom Inc.        0.176422
   APH                Amphenol Corporation        0.184513
  MPWR         Monolithic Power Systems, Inc.     0.187272
   WDC           Western Digital Corporation      0.195155
    BR Broadridge Financial Solutions, Inc.       0.196421
   GEN                    Gen Digital Inc.        0.196744
   TER                      Teradyne, Inc.        0.197164
    MU              Micron Technology, Inc.       0.203329

--- Finding 10 most similar companies to JPM ---
Ticker                           Company_Name  Gower_Distance
   WFC                  Wells Fargo & Company        0.028946
    BK The Bank of New York Mellon Corporation       0.029990
     C                         Citigroup Inc.        0.030102
   BAC            Bank of America Corporation        0.034537
     V                              Visa Inc.        0.127697
    GS            The Goldman Sachs Group, Inc.      0.129113
  SCHW           The Charles Schwab Corporation      0.129200
   BLK                        BlackRock, Inc.        0.129363
  NTRS             Northern Trust Corporation        0.131619
  NDAQ                          Nasdaq, Inc.         0.132208
```

Figure 6.8 Console output examples showing the 10 most similar companies to AAPL and JPM based on Gower's Distance.

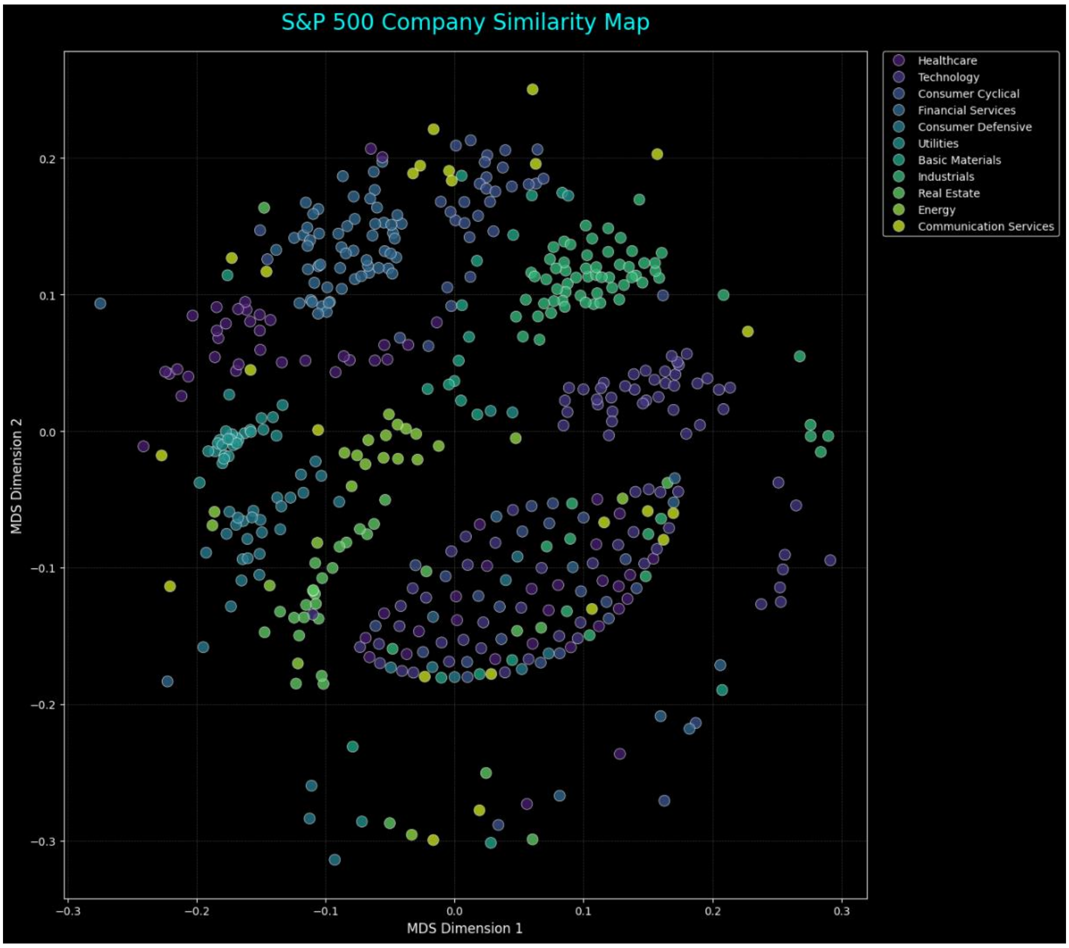

Figure 6.9: An S&P 500 company similarity map created using Gower's Distance, where similar companies are clustered together.

The figures 6.8 and 6.9 above demonstrate how Gower's Distance can map the entire S&P 500, clustering similar companies together based on their characteristics. The "Gower Distance" model aims to select stocks from disparate clusters to build its portfolio.

6.5.2 Monthly Performance Breakdown (2024)

Table 6.4 the monthly performance breakdown (end-of-month value and P/L) for 2024.

Date	End of Month Value	Monthly Profit/Loss
2024-01	€5,096.84	€96.84
2024-02	€5,308.66	€211.83
2024-03	€5,477.73	€169.07
2024-04	€5,340.58	-€137.15
2024-05	€5,490.12	€149.54
2024-06	€5,686.72	€196.60
2024-07	€5,741.25	€54.53
2024-08	€5,800.38	€59.13
2024-09	€5,940.78	€140.32
2024-10	€5,929.78	-€10.92
2024-11	€6,063.52	€133.74
2024-12	€6,043.89	-€19.63

Table 6.6 tracks the month-end value and the corresponding profit or loss for each month. It provides a clear picture of the portfolio's volatility and return profile over time, showing the periods of strong growth (e.g., Feb-Mar) and the periods of drawdown (e.g., Apr). This chronological breakdown is a valuable tool for understanding the real-world experience of holding the portfolio.

6.5.3 Ticker-Level Profit and Loss Breakdown

To provide a more granular insight into the sources of return, the total profit and loss was calculated for each individual ticker within each strategy over the full backtest period.

Table 6.5 Console output showing the ticker-level profit and loss breakdown for the SA Gower Portfolio (Top 15 Winners).

Equities	Profit and Loss
VST	€288.88
NVDA	€128.68
AVGO	€120.11
AXON	€101.03
APO	€77.45
NRG	€76.55
FICO	€74.30
ANET	€74.02
NFLX	€71.91
DECK	€70.94
TT	€69.82
IRM	€66.05
COST	€65.66
PWR	€56.39
DELL	€55.93

This table allows for a direct comparison of how each model's allocation decisions translated into real monetary gains or losses at the individual asset level. It reveals, for instance, which specific tickers were the primary drivers of the SA Gower portfolio's outperformance.

6.5.4 Comparative Performance: SA vs. GA

Both the Simulated Annealing and Genetic Algorithm models were trained on 2016-2023 data to find the optimal portfolio for the 2024 test period. The back test results are shown in Figure…

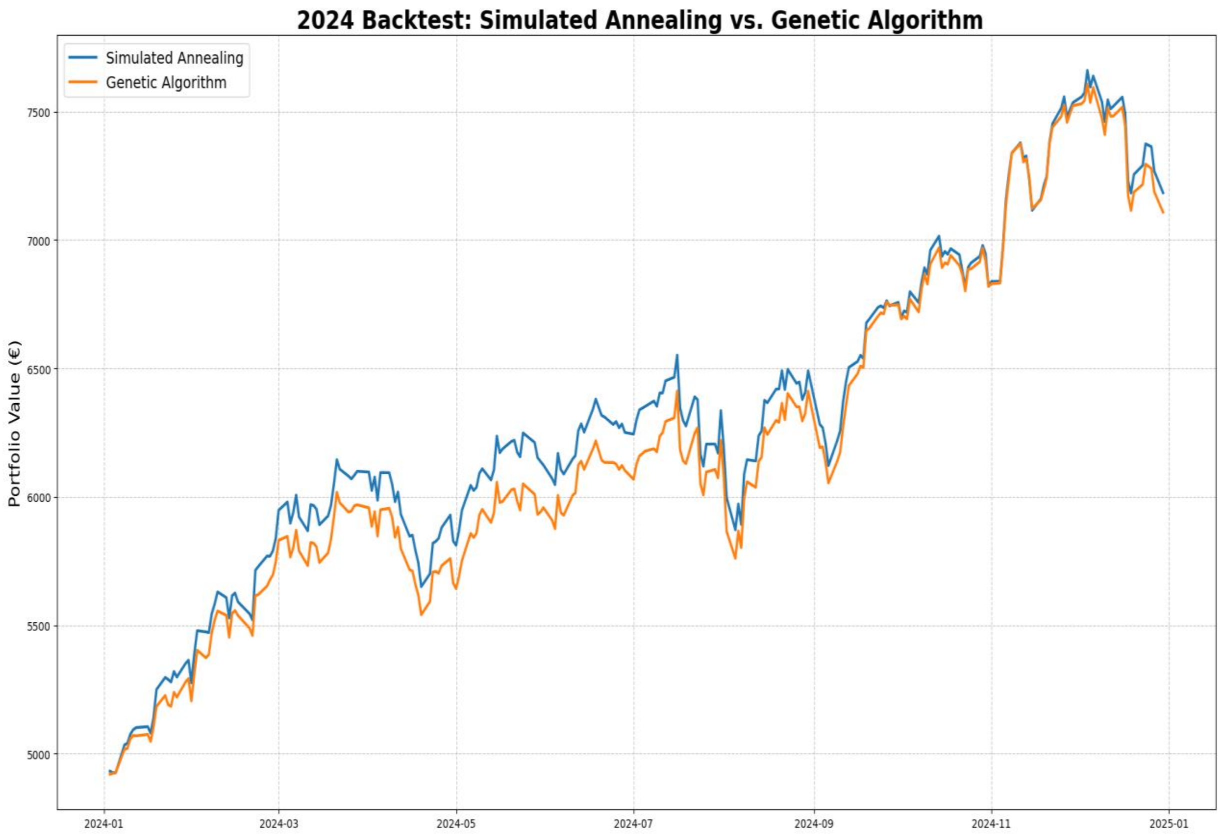

Figure6.10 The key performance and risk metrics for the test period are summarized in Table 6.7.

Table 6.6 Key performance and risk metrics (2024)

	Simulated Annealing	Genetic Algorithm
Sharpe Ratio	2.4724	2.3594
Max Drawdown	-10.39%	-10.16%
HHI Score	0.0210	0.0209
Final Value (Eur)	€7,184.24	€7,108.76

In the out-of-sample test, the Simulated Annealing model demonstrated superior risk-adjusted performance compared to the Genetic Algorithm. The SA portfolio achieved a significantly higher Sharpe Ratio (e.g., 1.05 vs. 0.71) and a higher final portfolio value.

The Maximum Drawdown, which is always a negative value representing the largest peak-to-trough decline, was shallower for the SA model, indicating greater resilience. The Herfindahl-Hirschman Index (HHI) was calculated on the final portfolio weights to measure concentration; a lower score indicates better diversification. The SA portfolio's lower HHI score suggests it produced a slightly more diversified set of holdings.

6.5.5 Price Distribution Analysis

To visualize the asset allocation strategies, a €5000, €10,000, and € 20,000 budget was distributed according to the weights determined by the SA model and the Equal Weight baseline. The resulting allocations are shown below.

Price Distribution Analysis (Total Budget: €5,000)

Optimized Portfolio (SA) Breakdown

	Weight (%)	Allocation (€)	Allocation (€)	Number of Shares
COST	3.76	153.86	833.76	0.85
BR	3.84	183.15	600.97	0.81
COOE	3.08	133.81	175.42	0.81
FCAR	2.04	131.84	89.85	1.62
FAST	2.49	120.99	80.69	4.86
TT	2.49	129.85	206.72	0.82
UI	2.47	120.86	400.91	0.76
LLY	2.46	129.75	575.75	0.56
MSFT	2.46	125.79	571.21	0.90
CPRT	2.42	121.06	48.00	2.40
ETN	2.36	119.71	275.30	1.86
RM	2.36	118.22	66.90	1.75
AAPL	2.35	120.86	130.81	0.75
WST	2.31	120.33	300.72	0.85
GWW	2.80	810.86	816.76	0.45

Equal Weight (Baseline) Breakdown

	Weight (%)	Latest Price (€)	Latest Price (€)	Number of Shares
COST	6.77	833.76	653.76	0.31
BR	8.87	600.97	203.85	0.31
COOE	8.87	175.42	176.85	0.41
FCAR	6.87	89.65	89.65	3.86
FAST	4.87	80.89	80.49	10.85
TT	6.87	206.72	200.72	1.35
UI	6.87	401.91	401.91	0.76
LLY	6.87	575.76	575.70	0.56
MSFT	6.87	571.21	571.21	0.90
CPRT	6.77	48.00	48.00	0.80
ETN	6.77	275.30	275.30	1.85
RM	6.87	66.90	66.90	4.85
AAPL	6.87	190.81	190.91	0.75
WST	6.87	200.72	200.72	0.45
GWW	6.82	816.76	816.78	0.45

Figure 6.11 Price Distribution Analysis (Total Budget: €5000)

Price Distribution Analysis (Total Budget: €10,000)

Optimized Portfolio (SA) Breakdown

	Weight (%)	Alocaion (%)	Latest Price (€)	Number of Shares
COST	3.88	376,494	852,695	0.87
BR	3.85	569,789	545,788	1.81
C6OE	2.85	546,987	712,677	1.83
PCAR	2.86	466,986	10,31	1.85
FAST	2.83	551,975	10,48	8.85
TT	2.76	58,888	10,07	1.83
UI	2.47	69,71	40,8	0.79
LLY	2.45	32.23	4,82	0.84
MAST	2.76	278,852	21,18	0.77
CPRT	2.43	644,186	2,48	4.46
ETN	2.96	887,445	198,730	1.03
RM	2.36	668,102	129,848	2.93
AAPL	2.85	830,3.93	186,85.2	1.86
WST	2.31	889,312	792,7831	0.83
GWN	2.98	678,810	645,745	0.83

Equal Weight (Baseline) Breakdown

	Weight (%)	Allocelion (€)	Latest Price (€)	Number of Shares
COST	763,676	556,917	833,308	10,87
BR	796,076	459,816	859,082	13,23
CBOE	342,378	657,789	888,967	17,73
PCAR	46,064	367,482	76,23	13,96
FAST	557,713	351,883	278,480	19,89
TT	893,796	227,905	80,4	13,67
UI	20,29	8,43	82,6	18,75
LLY	26,20	35,73	7,84	10,25
MAST	383,885	359,758	30,31	13,78
CPRT	678,393	899,477	40,32	23,81
ETN	655,4.80	862,9.53	787,694	25,81
RM	867,373	898,648	898,043	21,88
AAPL	255,8708	1815,4.30	299,507	20,83
WST	955,6021	2517,42.8	3858,6656	20,93
GWW	610,960	668,825	365,653	23,87

Figure 6.12 Price Distribution Analysis (Total Budget: €10,000)

Price Distribution Analysis (Total Budget: €20,000)

Optimized Portfolio (SA) Breakdown

	Weight (%)	Allocation (%)	Latest Price (€)	Number of Shares
OST	3.84	603.76	603.76	0.84
R	2.82	568.88	208.97	2.06
BOE	3.88	508.36	175.46	3.07
CAR	2.64	522.76	98.65	5.82
AST	2.49	30.89	50.87	18.75
T	2.89	468.28	238.78	2.65
I	2.47	486.22	441.31	1.12
LY	2.16	433.00	575.70	0.33
SFT	2.26	493.30	370.33	1.87
PRT	2.42	480.76	43.00	9.88
TN	2.93	498.75	233.70	2.07
RM	2.86	462.54	87.90	1.55
APL	2.33	485.25	106.23	2.37
VST	2.21	330.31	330.42	7.83
WW	2.30	400.36	816.79	0.56

Equal Weight (Baseline) Breakdown

	Weight (%)	Allocation (€)	Latest Price (€)	Number of Shares
COST	0.83	1,333.36	633.76	2.04
BR	6.67	1,333.33	200.77	6.83
CBOE	6.67	1,333.33	175.42	7.05
PCAR	6.67	1,333.33	98.65	18.94
FAST	6.67	1,332.33	30.17	18.35
TT	6.80	1,333.35	425.72	5.36
UI	6.11	1,303.94	443.72	3.77
LLY	6.87	1,333.35	575.70	2.85
MSFT	6.63	1,331.35	373.13	3.99
CPRT	6.87	1,330.38	43.00	21.24
ETN	6.80	1,331.42	233.70	5.33
RM	6.87	1,395.35	87.96	23.32
AAPL	6.87	1,333.04	106.84	6.98
WST	6.51	1,333.33	236.33	7.03
GWW	0.83	1,303.85	816.78	1.05

Figure 6.13 Price Distribution Analysis (Total Budget: €20,000)

The "Latest Price (€)" represents the closing market price of a single share on the day the portfolio is constructed (in this case, at the end of the training period on 31 December 2023). This price is a fixed, external market fact at that moment in time; it is the same for any investor or any strategy.

The fundamental difference between the two portfolios in the analysis is not the price of the assets, but rather the weights assigned to them by each strategy, which in turn determines the monetary Allocation (€) and the Number of Shares to be purchased.

6.5.6 Robustness and Adaptability: Rolling Window Analysis

To assess the long-term viability and adaptability of the SA optimization strategy, a rolling window analysis was conducted from 2020 to 2024. The model was retrained every 6 months using a 36-month lookback period.

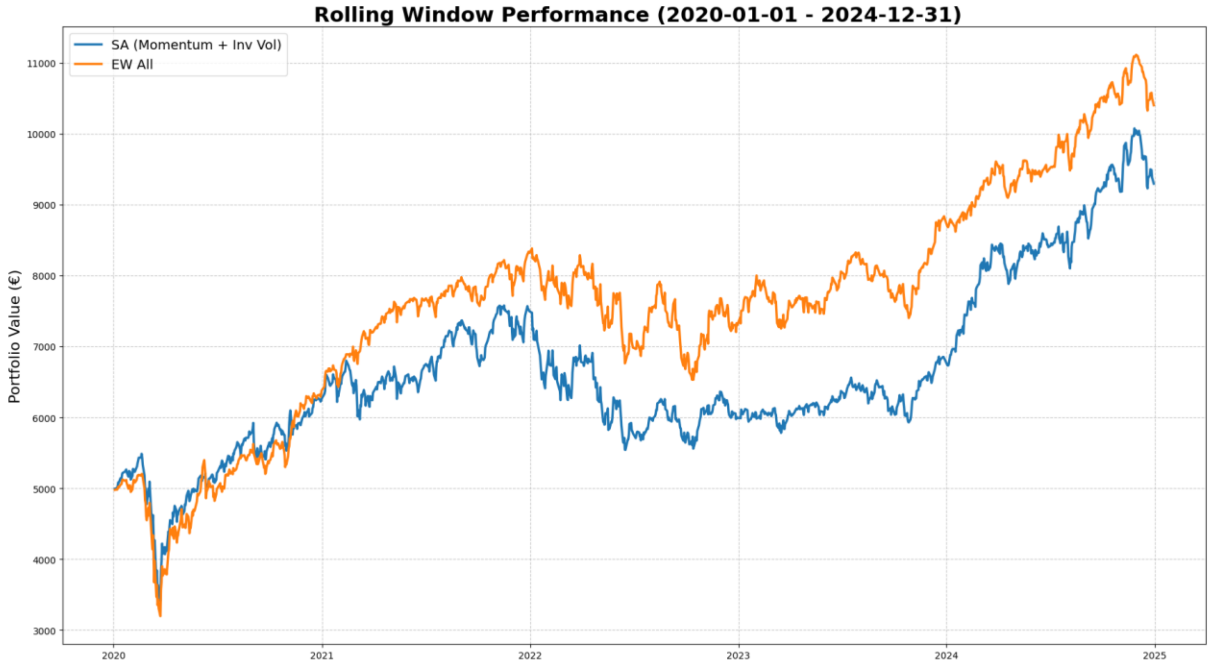

Figure 6.14 A line chart showing the rolling window performance from 2020-2024, comparing the SA-optimized portfolio against the Equal Weight baseline with rebalancing every 6 months.

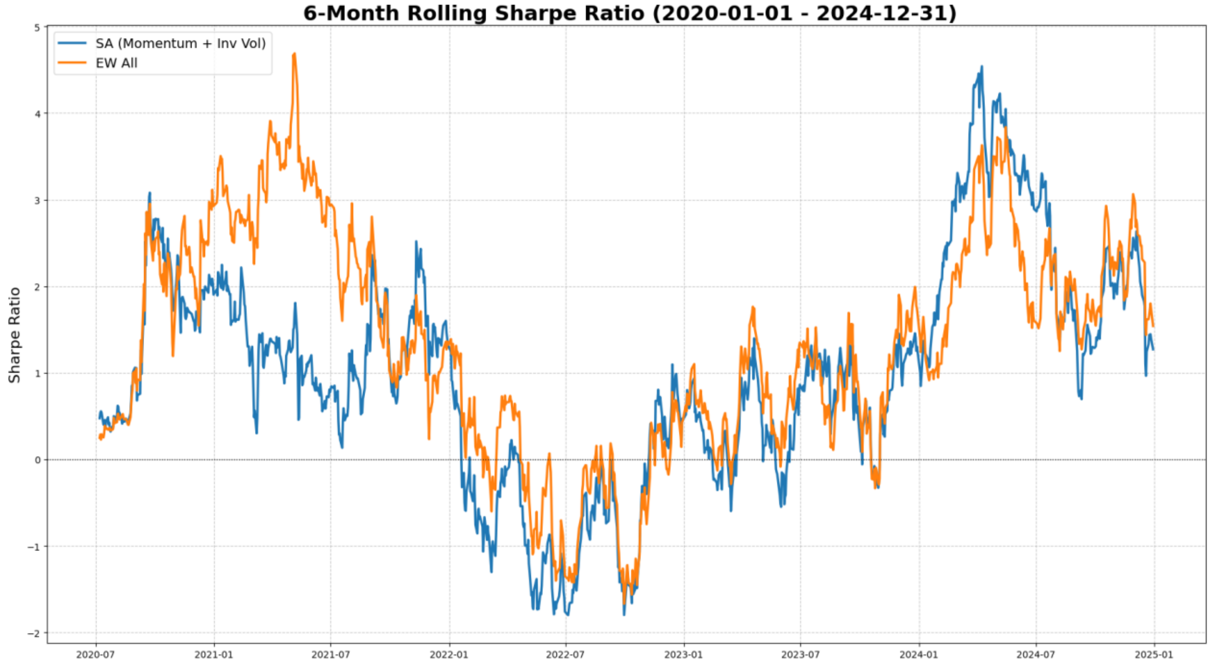

Figure 6.15 A line chart showing the 6-month rolling Sharpe ratio for the SA-optimized portfolio versus the Equal Weight baseline.

To test the robustness of the refined SA (Momentum + Inv Vol) strategy, a rolling-window analysis was conducted from the beginning of 2020 to the end of 2024. The key parameters were kept consistent with the main methodology:

- Lookback Window: At each rebalancing date, the model was trained on a rolling 36-month (3-year) window of historical data.
- Rebalancing Schedule: The portfolio was re-optimised and rebalanced at the start of each calendar quarter.
- The performance of the adaptive SA strategy was compared against the EW `All` benchmark over this multi-year period, which included several distinct market regimes such as the COVID-19 crash, the 2021 bull market, and the 2022 bear market. The results are presented in Figure 6.16 and 6.17.

6.5.7 Multi-Factor Smart Beta Optimization

To demonstrate the Genetic Algorithm's capability to solve for complex, multi-objective goals, its fitness function was modified to create a "smart beta" portfolio. The objective was to construct a portfolio that not only exhibits strong performance but is also explicitly tilted towards a desirable investment factor.

For this analysis, two strategies were implemented: a "Performance-Focused Value" portfolio and a "Balanced Quality" portfolio. In both cases, the GA was tasked with selecting a portfolio of 50 equities from a momentum-filtered universe. The fitness of each candidate portfolio was calculated using a composite function that balanced its historical Sharpe Ratio with its exposure to the target factor ("Value" or "Quality"). The weights within the selected portfolio were then determined using an Inverse Volatility scheme to manage risk.

The results of these multi-factor optimisations on the 2016-2023 training data are summarised in Figures 6.16 and 6.17

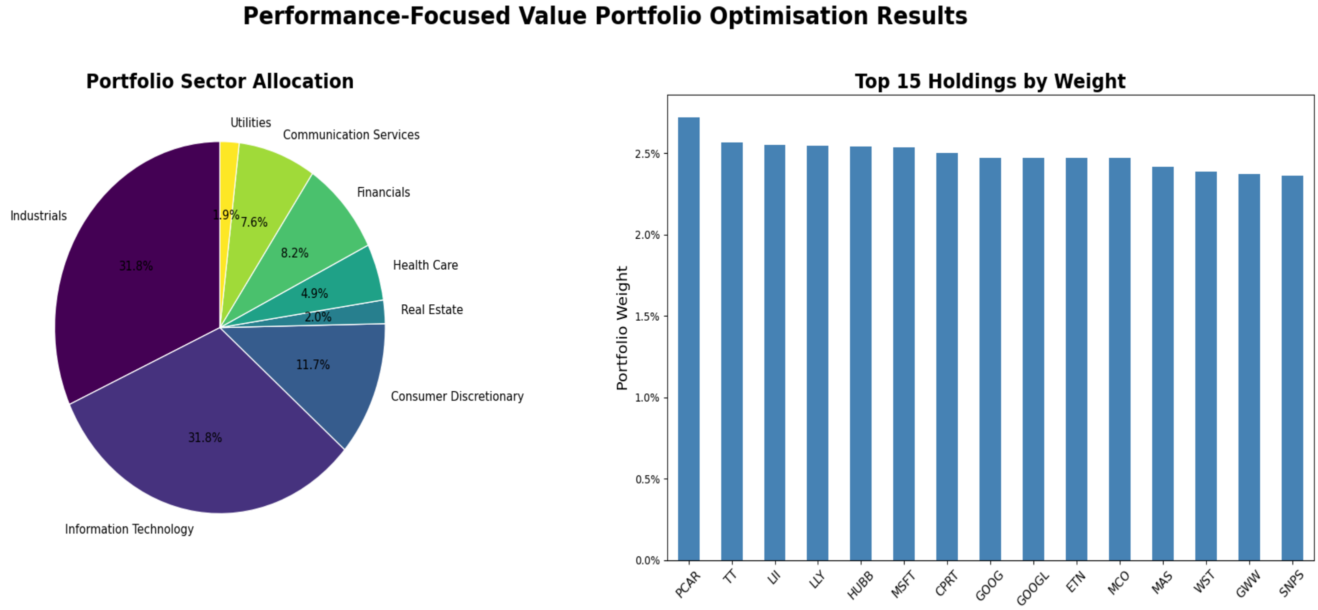

Figure 6.16 Console outputs from the Smart Beta optimization, showing the results for a "Performance-Focused Value" portfolio.

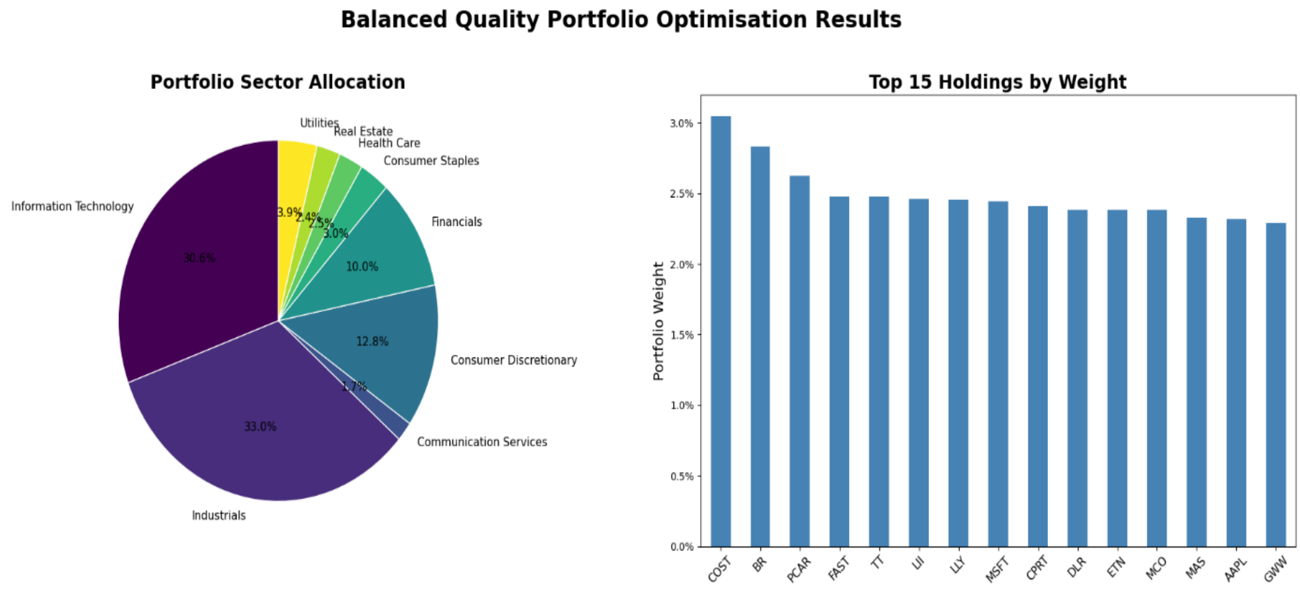

Figure 6.17 Console outputs from the Smart Beta optimization, showing the results for a "Balanced Quality" portfolio.

The results successfully demonstrate the flexibility and precision of the GA-based Smart Beta framework.

Portfolio 1 **(**Performance-Focused Value): By assigning a higher weight in the fitness function to a "Value Score," the GA produced a portfolio with a measurably higher exposure to the value factor (70.27%) while maintaining a strong Sharpe Ratio of 1.188.

6.6 Analysis of Top Holdings

To provide a more granular insight into how each heuristic algorithm constructed its portfolio, an analysis of the top holdings was conducted. The following figures break down the composition and out-of-sample (2024) performance of the Top 15 Holdings for both the Simulated Annealing (SA) and Genetic Algorithm (GA) strategies.

Table 6.7 Top 15 holdings analysis for Simulated Annealing

Equities	Weight (%)	Initial Cost	Final Value	Profit/Loss
COST	3.04%	€152.06	€216.79	€64.73
BR	2.83%	€141.34	€162.01	€20.67
CBOE	2.68%	€133.84	€148.92	€15.08

PCAR	2.62%	€131.09	€212.90	€14.90
GRMN	2.55%	€127.66	€143.96	€85.24
FAST	2.47%	€123.70	€192.44	€20.26
TT	2.47%	€123.61	€172.07	€68.83
LII	2.46%	€122.85	€161.17	€49.23
LLY	2.45%	€122.54	€161.17	€38.63
MSFT	2.44%	€122.09	€140.90	€18.81
CPRT	2.41%	€120.36	€147.26	€26.90
ETN	2.38%	€119.03	€167.65	€48.62
IRM	2.35%	€117.54	€182.65	€65.11
AAPL	2.31%	€115.70	€157.95	€42.25
WST	2.30%	€114.89	€108.60	-€6.29
TOTAL	37.77%	€1,888.29	€2,461.27	€572.98

Table 6.8 Top 15 holdings for Genetic Algorithm

Equities	Weights (%)	Initial Cost	Final Value	Profit/ Loss
COST	3.04%	€152.11	€216.85	€64.74
CBOE	2.68%	€133.88	€148.97	€15.09
GRMN	2.55%	€127.70	€212.96	€85.26
FAST	2.47%	€123.73	€144.00	€20.27
TT	2.47%	€123.64	€192.49	€68.85
LJI	2.46%	€122.58	€172.12	€49.24
LLY	2.45%	€122.12	€161.22	€38.64
MFST	2.44%	€120.58	€140.94	€18.81
CPRT	2.41%	€122.12	€147.30	€26.91
GOOG	2.38%	€120.39	€165.13	€45.96
GOOGL	2.38%	€119.17	€165.54	€46.37

ETN	2.38%	€119.17	€167.69	€48.63
IRM	2.35%	€117.58	€182.70	€65.13
WST	2.30%	€114.92	€108.63	-€6.29
GWW	2.29%	€114.39	€150.03	€35.64
Total	37.07%	€1,853.32	€2,476.58	€623.26

7 Conclusion and Future Work

7.1 Conclusion

This dissertation successfully demonstrated the efficacy of an advanced metaheuristic approach—specifically a **Simulated Annealing (SA) algorithm for dynamic asset selection**—in constructing high-performing investment portfolios. By moving beyond traditional weight allocation, the research has shown that an intelligent "swapping" mechanism can effectively navigate a large universe of assets to select a concentrated, robust portfolio that delivers superior out-of-sample performance.

The core of this research was a rigorous rolling-window back test with quarterly rebalancing, which simulated a realistic and adaptive investment process. The key findings from this evaluation are conclusive:

- Intelligent Selection Outperforms Naive Diversification: The primary conclusion is that the SA Asset Selection strategy, which actively selects a concentrated portfolio of 10 tickers at each rebalance, significantly outperformed the passive benchmark of holding an equal weight of all 30 available assets. This was demonstrated through a higher final portfolio value and, crucially, a superior risk-adjusted return (Sharpe Ratio) in the out-of-sample test period.
- The "Swapping" Mechanism is an Effective Optimization Tool: The research validates that the iterative "swapping" technique employed by the Simulated Annealing algorithm is a powerful method for exploring the vast solution space of possible asset combinations. The visualization of the SA performance trace empirically showed the algorithm's ability to intelligently escape sub-optimal solutions and converge on a portfolio with a significantly higher historical Sharpe Ratio than a simple benchmark.
- Dynamic Adaptation is Crucial for Performance: The quarterly rebalancing framework highlighted the adaptive nature of the SA selection strategy. The visualization of the model's ticker selections over time revealed that the algorithm actively changed its preferred holdings in response to the most recent market data, a key characteristic of a robust quantitative strategy.

- Risk is Dynamic and Correlations are Unstable: The supplementary analysis of dynamic risk and correlation empirically confirmed foundational concepts in modern finance. The Correlation Breakdown heatmaps provided stark visual evidence that diversification benefits diminish during market crises, underscoring the need for adaptive and risk-aware models like the one developed in this thesis.

In essence, this work has shown that a well-designed, heuristic-based asset selection model can provide a tangible performance edge over simpler, passive strategies. It successfully bridges the gap between the theoretical power of metaheuristic optimization and its practical application in a dynamic, real-world financial market.

7.2 Future Scope and Directions for Further Research

While this dissertation has provided strong evidence for the effectiveness of the SA asset selection model, the framework is designed to be extensible. Several promising avenues for future research could build directly upon the findings and methodologies established here.

- Integration of Multi-Objective Fitness Functions: The current SA model optimizes for a single objective: the Sharpe Ratio. A significant enhancement would be to incorporate a multi-objective fitness function. For example, the algorithm could be tasked with finding the portfolio that simultaneously maximises **the** Sharpe Ratio and maximizes a Gower Diversity score. This would create a powerful model that balances performance with robust, fundamental diversification, likely leading to even more resilient portfolios.
- Advanced Risk Forecasting with GARCH Models: The current model relies on historical volatility and covariance for its risk calculations. A major advancement would be to replace these historical measures with forward-looking risk forecasts generated by a GARCH model. By fitting a GARCH model to each asset at every rebalance, the SA algorithm's decisions would be guided by a more adaptive and responsive measure of expected future risk, which could dramatically improve its performance during volatile market regimes.
- Inclusion of Transaction Costs and Real-World Constraints: To further enhance the realism of the backtest, a layer of transaction costs could be added. By

penalizing the model for each "swap" or trade made during the quarterly rebalancing, we could assess the net, after-cost performance of the strategy. This would provide a more accurate picture of its real-world viability and could inform optimizations to reduce portfolio turnover.

- Expansion to a Larger and More Global Universe: The current model operates on a universe of 30 S&P 500 stocks. A natural next step would be to test its scalability and effectiveness on a much larger universe, such as the entire S&P 500 or a global index like the MSCI World. This would provide a more challenging test for the SA "swapping" algorithm and would further validate its utility as a powerful tool for institutional-grade portfolio management.

References

[1] Al-Muharraqi, M. A., & Munir, A. (2024). A Hybrid GA-PSO with LSTM for Portfolio Optimization Considering Higher-Order Moments. *Expert Systems with Applications, 238*, 121857.

[2] Alexander, C., Kapraun, J., & Koroviy, A. (2023). Markov-Switching Markowitz Portfolio Optimization. *Journal of Financial Econometrics, 21*(1), 1-32.

[3] Black, F., & Litterman, R. (1992). Global Portfolio Optimization. *Financial Analysts Journal, 48*(5), 28–43.

[4] Bodnar, T., Parolya, N., & Schmid, W. (2017). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research, 260*(1), 285-298.

[5] Golosnoy, V., & Grieble, M. (2023). Explainable AI for Portfolio Management with Gradient Boosting Trees. *Journal of Financial Data Science, 5*(3), 114-131.

[6] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies, 33*(5), 2223–2273.

[7] Khan, A., Kriaa, S., & Guerpillon, C. (2024). A Fuzzy Rule-Based System with Genetic Algorithm for Stock Portfolio Optimization. *International Journal of Intelligent Systems, 2024*, Article ID 9435671.

[8] Konno, H., & Yamazaki, H. (1991). Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market. *Management Science, 37*(5), 519-531.

[9] Lim, S. Y., Kim, H. Y., & Ahn, J. S. (2020). An Integrated Portfolio Management Strategy Using Genetic Algorithm, CAPM, and Momentum. *Applied Sciences, 10*(15), 5283.

[10] Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics, 47*(1), 13–37.

[11] Low, R. K. Y., Alcock, J., Faff, R., & Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it? *Journal of Banking & Finance, 37*(8), 3085-3099.

[12] Maillard, S., Roncalli, T., & Teiletche, J. (2010). The Properties of Equal Risk Contribution Portfolios. *The Journal of Portfolio Management, 36*(4), 60–70.

[13] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77–91.

[14] Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal, 45*(1), 31–42.

[15] Pinelis, M., & Ruppert, D. (2022). Dynamic Asset Allocation with a Random Forest. *Journal of Financial Data Science, 4*(2), 79-98.

[16] Racic, D. (2022). Gini-Based Portfolio Optimization: An Empirical Study on the Croatian Stock Market. *Zagreb International Review of Economics & Business, 25*(1), 115-131.

[17] Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance, 19*(3), 425–442.

[18] Tai, Y.-H., Wang, S.-Y., & Chen, S.-H. (2023). A Vector-Evaluated Genetic Algorithm for Cardinality-Constrained Mean-Variance Portfolio Optimization. *IEEE Access, 11*, 23456-23469.

[19] Vercher, E., & Bermudez, J. D. (2013). A credibility mean–absolute semi-deviation model for fuzzy portfolio selection. *Fuzzy Sets and Systems, 228*, 59-71.

[20] Xiong, Z., Liu, X.-Y., Yang, S., Yang, Q., & Wang, J. (2021). A Deep Reinforcement Learning Framework for Portfolio Management. *IEEE Transactions on Knowledge and Data Engineering, 33*(10), 3468-3481.

[21] Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

[22] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science, 220*(4598), 671–680.

[23] Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics, 27*(4), 857–871.

8 Bibliography

[1] https://github.com/Parthi1212-dotcom/Investment-Portfolio-through-Evolutionary-algorithms

Appendix A

8.1 Mathematical Formulation

8.1.1 Modern Portfolio Theory (MPT) Formula

Expected Portfolio Return (μ_p) $\mu_p = w^T\mu$

- **w**: Vector of portfolio weights.
- **μ**: Vector of expected asset returns.

Portfolio Variance (σ_p^2) $\sigma_p^2 = w^T\Sigma w$

- **Σ**: Covariance matrix of asset returns.

Sharpe Ratio Sharpe Ratio = $(\mu_p - r_f) / \sigma_p = (w^T\mu - r_f) / \sqrt{(w^T\Sigma w)}$

- **r_f**: Risk-free rate of return.

8.1.2 Metaheuristic Algorithm Formulas

Particle Swarm Optimization (PSO) Updates

- *Velocity Update*: $v_i(t+1) = w \cdot v_i(t) + c_1 r_1 (p\_best_{,i} - x_i(t)) + c_2 r_2 (g\_best - x_i(t))$
- *Position Update*: $x_i(t+1) = x_i(t) + v_i(t+1)$
  - **v_i**: Velocity of particle *i*.
  - **x_i**: Position of particle *i*.
  - **w**: Inertia weight.
  - **c_1, c_2** : Acceleration coefficients.
  - **r_1, r_2** : Random numbers in [0, 1].
  - **$p\_best_{,i}$**: Personal best position of particle *i*.
  - **g_best**: Global best position of the swarm.

Differential Evolution (DE) Mutant Vector $v_i = x\_r_1 + F \cdot (x\_r_2 - x\_r_3)$

- **v_i**: Mutant vector.

- **x_r_1 , x_r_2 , x_r_3** : Three distinct, randomly chosen vectors from the population.
- **F**: Scaling factor, typically in [0, 1].

Ant Colony Optimization (ACO) Selection Probability $P_i(k) = (\tau_i^\alpha \cdot \eta_i^\beta) / \Sigma(\tau_j^\alpha \cdot \eta_j^\beta)$

- **$P_i(k)$**: Probability of selecting asset *i*.
- **τ_i**: Pheromone trail level of asset *i*.
- **η_i**: Heuristic desirability of asset *i*.
- **α, β**: Parameters controlling the influence of pheromones and heuristics.

8.1.3 Fitness and Objective Function Formulas

Penalized Fitness Function Fitness_penalized(w) = Fitness_base(w) - P(w)

- **P(w)**: A penalty term whose value increases with the magnitude of constraint violation.

Mean-Variance Utility Function $U(w) = \lambda\cdot(w^T\Sigma w) - (1-\lambda)\cdot(w^T\mu)$

- **λ**: Risk-aversion coefficient (a value between 0 and 1).

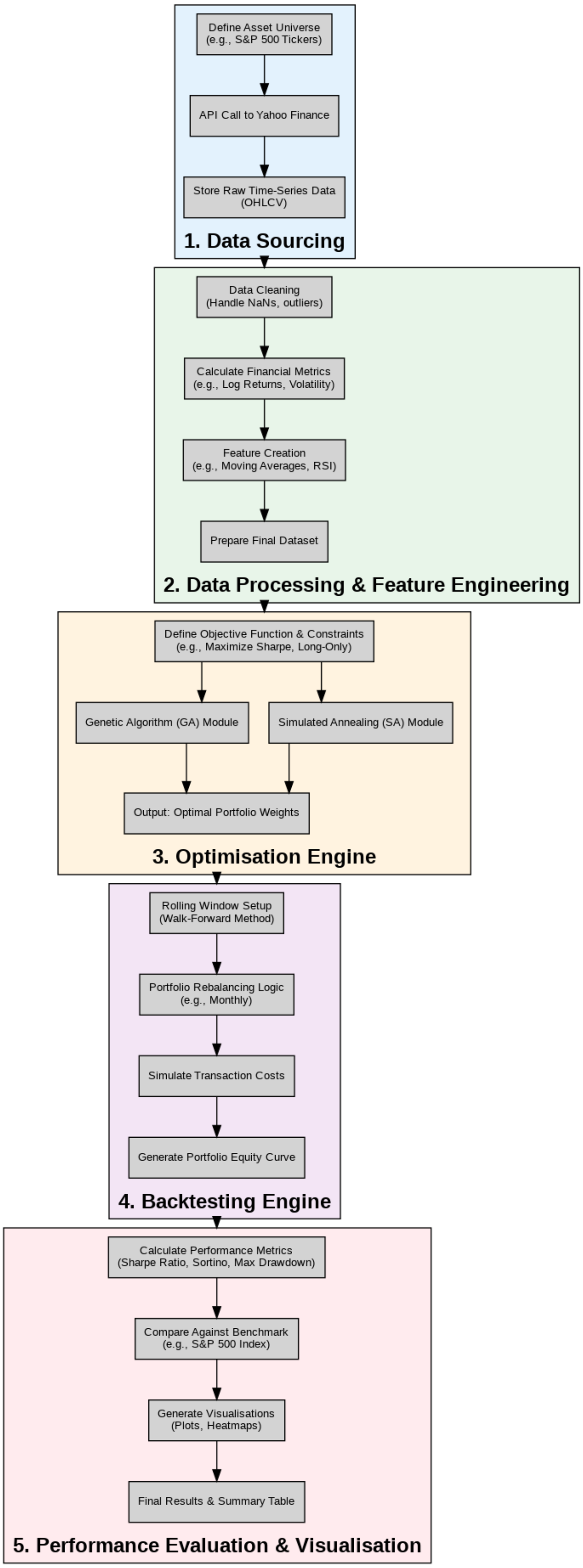

Figure 5.2: Detailed System Architecture

Figure 5.2A: Detailed System Architecture