EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
2025, Vol. 18, Issue 3, Article Number 6025
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

Portfolio Optimization Using CART and Genetic Algorithm

Cheibetta Ahmed Baba[1,*], Abdou Ka Dionque[1]

[1] *Department of Applied Mathematics, Faculty of Science and Technology, Gaston Berger University of Saint-Louis, Saint-Louis, Senegal*

Abstract. In this article, we present an approach based on the approximation of data by segmentation using the CART algorithm and Genetic algorithms, with a view to constructing an optimal portfolio-extract of financial assets. This approach generates a surplus of financial gains in terms of costs, and improves performance by reducing computing loads. This is a three-stage process: the first is to represent the financial series using a piecewise linear approximation obtained by CART, where the trend change points are optimally determined by segmentation of the series into periods of economic regime change. In the second stage, once the segments have been determined, each segment is represented in the plan (yield, volatility).We then develop an algorithm that selects the top ten financial assets in the overall portfolio, known as the extracted portfolio, in terms of the best ratio between return and VaR. In the third step,we apply heuristic optimization algorithms based on Genetic Algorithms and Particle Swarm Optimization to the extracted portfolio . The aim of this algorithm is to optimize the weights that will minimize the VaR and maximize the value of the extracted portfolio.

2020 Mathematics Subject Classifications: 91G10, 90C59, 62H30, 62M10, 68W50

Key Words and Phrases: Time series, portfolio management, Classification and Regression Trees , decision tree, genetic algorithm, Particle Swarm Optimization.

1. Introduction

Portfolio optimization has always been a central issue in finance. In this context, Markowitz pioneered the mean-variance model in 1952, which uses the variance of portfolio returns from their mean as a measure of risk to facilitate optimal portfolio selection. However, this model has been widely criticized, particularly for its use of variance as a measure of risk, the quadratic of the calculations, as well as the objective function and the normality of returns on financial assets. These criticisms have led to various attempts to improve the model, or to the development of new models. In this context,

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v18i3.6025

Email addresses: cheibetta_ahmed88@yahoo.com (Ch. Ahmed Baba), abdou.diongue@ugb.edu.sn (A. Ka Dionque)

several models have been developed to lighten the computational load and linearize the optimal portfolio choice problem,of which the Sharpe [1] and Stone models [2] are the best known.Alternatively, authors such as Konno, Konno and Yamazaki [3], and Zenios, Pang and Speranza [4], have suggested that portfolio risk should be calculated in linear rather than quadratic form, to create linear programming models for selecting an optimal portfolio, while incorporating asymmetrical risk criteria to simplify optimization. Other researchers, such as Pedersen and Stchell, have extended a family of risk functions by introducing a class that encompasses most existing risk measures. However, when Markowitz [5] established what is now recognized as modern portfolio theory, little effort has been made to explore improved methods for analyzing the price data used in its mean-variance framework. Since then, several advances in financial mathematics have enabled us to better understand these data series, leading to models that are better adapted to reality. Bayesian parameter estimation approaches based on historical data, such as those by Garlappi et al [6] , Avramov and Zhou [7] and Lai et al [8], have been widely cited and applied recently. The factor models of Fama and French [9] have also significantly improved on conventional pricing models based on mean-variance optimization. Selecting assets for a portfolio is an issue shared by a large number of financial market professionals. financial markets. The aim is to determine, from among the assets available for investment, the capital allocation that makes up the best possible portfolio.Since the future is uncertain, asset returns must be considered as a random vector, and "best" is generally understood as taking into account both the expected gain and the risk taken.This objective can be simplified within the framework of modern portfolio theory [5] which formulates the selection problem as a minimization of the estimated variance of the portfolio under the expected return constraint. However, the optimization implies estimating the expectation of the asset return vector and its variance-covariance matrix. A wide variety of approaches have been proposed and practiced by market players to make this estimate. Three main methodological trends can be identified : fundamental analysis, technical analysis and quantitative analysis. For our approach, we have proposed the (CART) algorithm for piecewise linear approximation, where trend change points are determined optimally by segmenting the financial series into periods of economic regime change.then we consider that each segment will be represented in the plan (yield, volatility), by choosing the ten best assets in this portfolio, called portfolio-extract, with the best ratio between expected returns and VaR, using the algorithm we've developed. we then apply heuristic optimization algorithms based on Genetic Algorithms and Particle Swarm Optimization to the extracted portfolio .. This work is organized as follows. In section 1, we present the data representation methods. Some elements of the approximation method are discussed in section 2. In section 3, we deal with the choice of method. In section 4, we present the method for selecting the ten best assets using the algorithm we have developed, and the Heuristic Optimization Algorithms using genetic algorithms and particle swarm optimization to the extracted portfolio . Finally, a numerical application is considered to describe the performance of our approach.

2. Representation Methods

In this study, we work with historical daily financial data from multiple asset classes: equities, indices, interest rates, credit instruments, currencies, and commodities. To meaningfully exploit this raw information, a structured representation of time series is essential. We chose to represent these time series using piecewise linear approximation, a method that involves dividing the series into homogeneous segments and modeling their behavior locally using simple functions (e.g., linear). This approach is particularly suited to the nature of financial time series, which typically exhibit stable regime phases interspersed with trend changes. Thus, this method not only simplifies the series but also highlights the underlying economic regimes, providing a solid foundation for analysis, interpretation, and portfolio optimization.

2.1. Piecewise Approximation: Principles and Justification

Piecewise approximation is based on the assumption that time series can be divided into homogeneous periods — that is, subintervals where price behavior remains relatively stable. This assumption is economically justified: financial assets evolve in environments that change over time (monetary policy shifts, macroeconomic shocks, crises, recoveries, etc.), leading to changes in trends or volatility in prices.

This type of method has attracted significant interest in the field of financial analysis (see [10, 11]), as it enables local modeling of the signal $y(t)$ over k distinct segments. Each segment is defined by:

$$\forall t \in [t_{i-1}, t_i], \quad y(t) = f_i(t) + \varepsilon_i(t),$$

where f_i is a deterministic function (typically affine), and ε_i is a centered noise. This representation offers several major advantages:

- It reduces data dimensionality by summarizing each segment with a small number of interpretable parameters (slope, duration, variance).
- It provides a structured and economically meaningful view of price evolution, whereas more global or mathematical approaches (filtering, wavelets, regularization) often lack interpretability.
- It facilitates regime break detection by identifying trend change points, which is essential for financial decision-making.

The main challenge of this method lies not in estimating the segments themselves, but in detecting the breakpoints — a well-known problem in statistics ([12, 13]). Recent work ([14–16]) has proposed effective solutions, even in the presence of dependent data and an unknown number of breakpoints. Three main segmentation approaches exist ([12]):

- **Sliding window**: progressive construction of segments.

- **Top-down**: recursive division of a global segment.
- **Bottom-up**: recursive merging of elementary segments.

2.2. Why This Method?

The choice of piecewise linear approximation is based on the search for a balance between accuracy, simplicity, and interpretability :

- **Economic interpretability**: Unlike other methods (frequency filtering, wavelets, regularization), this approach allows direct reading of trends, market reversals, and volatility. Each segment can be interpreted as an identifiable economic or financial regime.
- **Relevance to financial data**: It is widely acknowledged that financial series go through periods of stability, interspersed with abrupt or gradual regime changes. Segmentation allows explicit identification of these breaks ([17]).
- **Intuitive modeling**: Using linear or exponential functions to model each segment is natural in finance, and allows each period to be characterized by its trend and volatility.
- **Complexity control**: The method allows the number of segments to be adapted to the desired level of granularity. A penalization criterion can be added to avoid over-segmentation and maintain an economically usable representation.

2.3. Indexing of financial series

We suggest modeling financial asset time series by estimating local trends. Thus, we adopt the following model for the price series $y(t)$ of a specific asset: there is a function f and a centered noise. $\varepsilon(t)$ such that

$$y(t) = f(t) + \varepsilon(t),$$

Where f is a piecewise affine function, which may not be continuous, and that ε is locally stationary on each segment of f. We will use linear modeling throughout this document, but the use of a piecewise exponential function, which is also common in finance, does not constitute an additional difficulty. Notice this model can be considered as a variant of the Ornstein-Uhlenbeck processes, defined that by

$$dy(t) = -\theta(r - y(t))dt + \sigma dB(t),$$

where $B(t)$ is Brownian motion. We distinguish ourselves from these processes by assuming the existence of trend changes in $f(t)$ and more over we do not define the noise distribution a priori. As mentioned previously, the goal of piecewise representation is to efficiently identify homogeneous periods, i.e., breakpoints.This involves determining, at each point,

whether it represents a change in trend or simply a fluctuation due to volatility. Our method will therefore have to divide the series into a number k, unknown in advance, of segments, $I_k = [t_{k-1}, t_k]$. Furthermore, we want to control the complexity of the segmentation, since a very fine division would offer a good approximation of the data but would not provide relevant information on trend and noise. It is therefore necessary to penalize the criterion, in adequation with the representation to the data, by adding a complex criterion, such as the number or size of the segments.

2.4. Formulation of an adequation with the representation of the data

Suppose we have n data points $\{y_0, y_1, ...y_n\}$ $0 = t_0 < t_1 < t_2, ... < t_k = 1$ determining the segments I_k The time can thus be indexed by J_k $(j = 0, ..., k)$, and we note $\varepsilon_j = \varepsilon(j/n)$ We observe a noisy signal, composed of linear segments:

$$\forall (j = 0, ..., k), \quad y_j = f(j/n) + \varepsilon_j.$$

The target function $f^\star$ is entirely determined by its values at the breakpoints: for each $j \in 1, ..., k$,

$$\forall t \in [t_{j-1}, t_j], f^\star(t) = f^\star(t_{j-1}) + \frac{f^\star(t_j) - f^\star(t_{j-1})}{t_j - t_{j-1}}$$

with $f(t_i) = \lim_{t \to t_i t < t_i} f(t)$ .We denote by F the set of piecewise linear functions in which we will look for the solutions. The quality of an estimated function $f^\star \in F$ is assessed using its mean square error, or quadratic risk:

$$R(f^\star) = \mathbb{E}_y \left[\int_0^1 \left(f^\star(t) - f(t) \right)^2 \right] dt.$$

This risk is estimated statistically from the data by calculating the average of the square of the residuals, also called empirical risk:

$$R(f^\star) = \Sigma_{j=0}^n (y_j - f^\star(j/n))^2.$$

To control the complexity of the solution, we add a penalty proportional to the number k of segments in $f^\star$. The final criterion is therefore the empirical risk adjusted by a regularization:

$$R_{pen}(f^\star, \lambda) = \Sigma_{j=0}^n (y_j - f^\star(j/n))^2 + \lambda K,$$

where $\lambda > 0$ is a regularization parameter measuring the trade-off between approximation and complexity. Finally, the optimal function is the one which minimizes the regularized empirical risk:

$$f^\star = argmin_{f \in H} R_{pen}(f^\star, \lambda).$$

When the segmentation is known, the problem reduces to estimating the best line for each interval, I_i from the data. Thanks to the additivity of risk, it is enough to decompose the empirical risk into k parts:

$$f^\star = \Sigma_{j=1}^k L(I_j), \quad with \quad L(I_k) = \Sigma_{j=1}^n (y_j - f^\star(j/n))^2$$

The minimization of each $L(I_k)$ is carried out using a regression line estimated by the least squares method.The function f is defined on each I_k by

$$\forall t \in I_k, f^\star(t) = \alpha_k + \beta_k t.$$

Let $|I_k|$ represent the number of data points in the interval I_k. We define the following statistical quantities for the data:

$$\bar{t_k} = \frac{1}{|I_k|} \sum_{j:J/n, n \in I_k} J/n,$$

$$\bar{y_k} = \frac{1}{|I_k|} \sum_{j:J/n, n \in I_k} y_j,$$

$$\gamma_k = \frac{1}{|I_k|} \sum_{j:J/n, n \in I_k} (J/n - \bar{t_k})(y_j - \bar{y_k}),$$

$$s_k^2 = \frac{1}{|I_k|} \sum_{j:J/n, n \in I_k} (J/n - \bar{t_k})^2,$$

$$u_k^2 = \frac{1}{|I_k|} \sum_{j:J/n, n \in I_k} (y_j - \bar{y_k})^2.$$

The regression coefficients are:

$$\alpha_k = \bar{y_k} - \beta_k \bar{t_k}. \quad and \quad \beta_k = \gamma_k / s_k^2,$$

and the error resulting from the regression is written:

$$L^\star(I_k) = |I_k|(u_k^2 - \gamma_k^2 / s_k^2)$$

3. Segmentation by trees

With this least squares regression criterion, we suggest splitting the series using CART. The CART (Classification and Regression Trees) algorithm, developed by Breiman et al [18], is a well-known method for recursive binary tree partitioning. This is a top-down partitioning, which starts with a single segment and results in a more detailed partition. This process consists of two steps.

3.1. Tree construction

The root of the tree corresponds to the full interval, $(I_{0.1})$ characterized by the error $L^\star(I_{0.1})$ coming from the least squares regression on this interval.The objective is to find the best possible division,$(I_{0.1})$ into two intervals $(I_{1.1})$ and $(I_{1.2})$ which constitute the

child nodes. The division criterion is the error improvement between the root and child nodes, that is:

$$\triangle \{L\}(I_{0.1}, I_{1.1}, I_{1.2}) = L^{\star}(I_{0.1}) - L^{\star}(I_{1.1}) - L^{\star}(I_{1.2}).$$

We select the breaking point which maximizes this improvement. At each iteration of the algorithm, it is therefore a question of exhaustively calculating this criterion $\triangle \{L\}$ for each terminal node (undivided interval) and each possible division, then choosing the division offering the $\triangle \{L\}$ maximum "that is, the points at which there is a change in trend". We will note $I_{j+1,2k-1}$ and $I_{j+1,2k}$ the child nodes of $I_{j,k}$ We thus obtain a series of P divisions, classified in descending order of improvement, with their corresponding division points $\{x_1....x_p\}$associated, which are generally not in chronological order. At this step, the complexity parameter λ is not taken into account, because the objective is to build the entire tree until obtaining the finest partition. In practice, an additional stopping criterion is defined, such as the minimum size m of the intervals. Thus, in each iteration, only splits where child nodes containing at least m elements are considered. This stopping condition helps reduce execution time by immediately eliminating unwanted divisions.

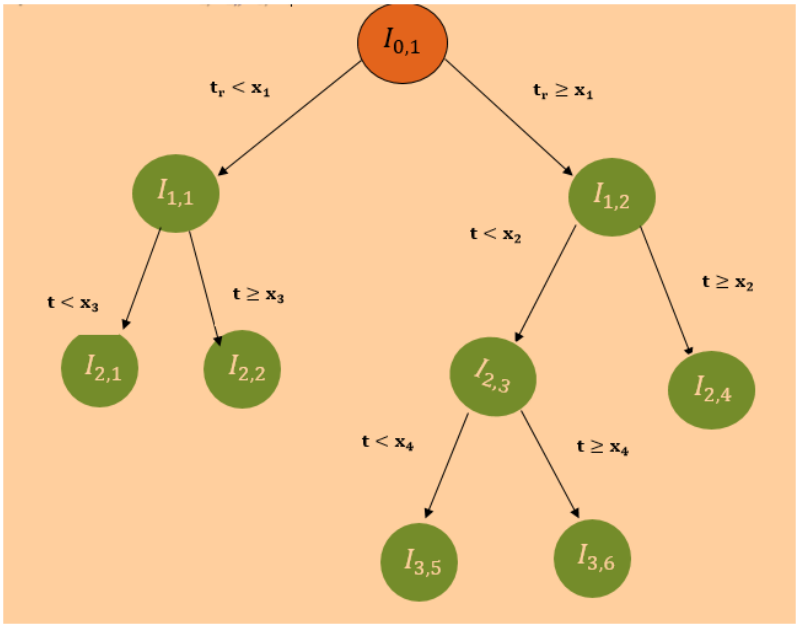

Figure 1: Construction of the decision tree using the CART method with four partitioning thresholds X_1, X_2, X_3, X_4.

3.2. Pruning branches

From the fully constructed tree $\mathcal{T}$, this step aims to find the optimal subtree $\mathcal{T}^{\star}$ minimizing the regularized empirical risk:

$$\hat{R}_{pen}(\acute{T}, \lambda) = \Sigma_{\mathrm{I}\in\mathcal{L}(\acute{\mathcal{T}})} L(I) + \lambda |L(\mathcal{T})|,$$

where $L(\acute{T})$ consists of the leaves (terminal nodes) of the tree $\acute{T}$ and $|L(\acute{T})|$ the number of leaves. The algorithm calculates the number $K \leqslant P$ of divisions to keep, ranked in decreasing order of error improvement. We fix $(\triangle \{L\}_1, ..., (\triangle \{L\}_p)$the improvements of the error for successive divisions, and, by convention , $\triangle \ \min_{k_0 \leq k \leq p}[-\Sigma \triangle \{L_k\} + \lambda(K+1)]$. The break points (if $K > 0$,) are then the $\{x_1, ..., x_p\}$ rearranged in ascending order of their value.

The complete tree construction provides great flexibility for pruning, which can be evaluated without having to repeat the first step. An alternative complexity criterion can consist of a minimum improvement threshold $\triangle \{L\}$ in the manner of optimization procedures.

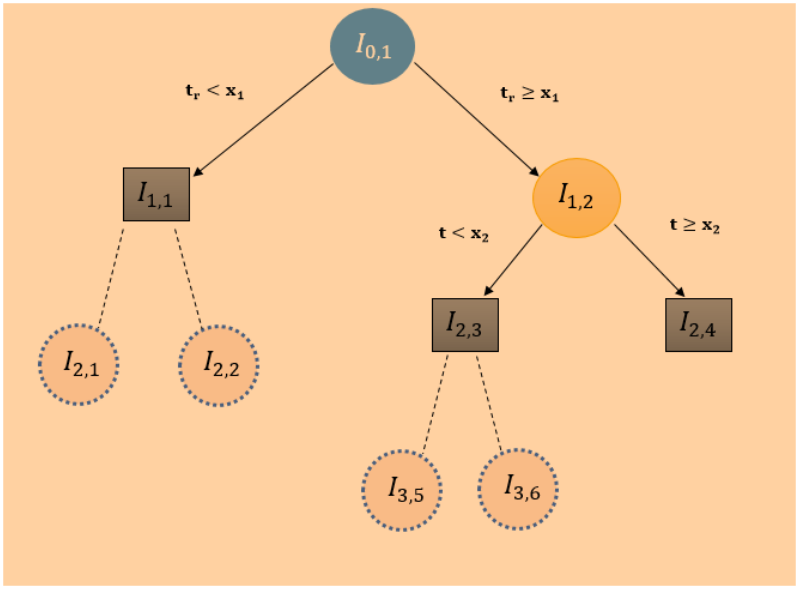

Figure 2: Pruning using the CART method: partition composed of $I_{1,0}, I_{2,3}$, and $I_{2,4}$.

3.3. Input

Let $\{t_1, ..., t_{k-1}\}$ and the homogeneous segments $\{I_1, ..., I_k\}$, associated, the duration $\{\tau_1, ..., \tau_k\}$ of each homogeneous period, the slope $\{\beta_1, ..., \beta_k\}$ for each segment, as well as the variance of the residuals $\{\nu_1, ..., \nu_k\}$ on each segment, estimated by the $L^\star(I_{0.1})$ higher order moments, if the size of the segments allows it. To work with comparable quantities between assets, it is common in the financial field to consider the trend (average return) and volatility, defined respectively as the slope and the standard deviation in relation to the price, per unit of time.

$$r_k = \frac{1}{\tau_k} \frac{\beta_k}{\bar{y_k}},$$

et

$$\sigma_k = \frac{1}{\tau_k} \frac{\sqrt{\nu_k}}{\bar{y_k}}$$

We will also refer to each segment as a diet.
We will also refer to each segment as a diet. We examine, from a global perspective, a set of D assets chosen to represent the market, at a date t, and an observation window T on the data, i.e. the interval [t-T + 1, t]. For each asset d, we propose to perform the segmentation and keep only the last $\mathring{k}$ segments to reflect the recent behavior and control the final dimension of the data. We therefore select several vectors of size $\mathring{k}$:
•. Trends $r_t^d = (r_{k-\mathring{k+1}}, ...r_k)_{\acute{d}}$,
•. Volatilities $\varrho_t^d = (\sigma_{k-\mathring{k+1}}, ...\sigma_k)_{\acute{d}}$ and
•. durations $\tau_t^d = (\tau_{k-\mathring{k+1}}, ...\tau_k)_{\acute{d}}$ that we decide to omit at first. The final descriptive vector at date t is given by:

$$\mathbf{X}_T = (r_t^1, ..., r_t^D, \sigma_t^1, ..., \sigma_t^D)$$

4. Optimization procedure

Before accessing the optimization procedure, we define some useful concepts relating to portfolio optimization:

4.1. Profitability and Value Portfolio

We call the return r_t of a stock obtained by investing in a stock, the ratio between the price of the stock at time t and its price at $t-1$, plus the income (dividends) received during the period $[t-1; t]$:

$$r_t = \frac{c_t - c_{t-1} + d_t}{c_{t-1}}$$

, Where :

- c_t: The price of the stock i at the end of the period t.
- d_t: Dividend income at the end of the period.

The expected return of a stock for a period T is given by:

$$\overline{r_i} = \frac{1}{T}\sum_{t=1}^{T} r_{it}$$

. The profitability of a portfolio consisting of the expected returns of k stocks r_i, where $i = 1, \ldots, k$, is:

$$\overline{R_p} = \sum_{i=1}^{k} x_i \overline{r_i}$$

, where $x_1, \ldots, x_n$ are the proportions of the investor's wealth allocated to shares i respectively ($i = 1, \ldots, k$). For a portfolio containing k stocks, its value is given by:

$$V = \sum_{i=1}^{k} x_i \cdot V_i$$

where $x_1, \ldots, x_k$ are the quantities of the investor's wealth allocated to the shares respectively, and V_i is the value of the i^{th} action. The change in value will obey the following relationship:

$$\Delta V(x) = \sum_{i=1}^{k} x_i \Delta V_i$$

.

4.2. Risk Portfolio

The risk of a financial asset is the uncertainty regarding the value of that asset at a future date. Variance, mean absolute deviation, semi-variance, VaR and CVaR are ways of measuring this risk.

4.2.1. Variance of a portfolio

To compute the variance of a portfolio, we need to get the variance of each asset into it. The variance of an individual asset measures the dispersion of its returns relative to the average of returns. It quantifies the degree to which an asset's returns vary over time. A high variance indicates that returns are widely dispersed, meaning the asset is riskier, while a low variance indicates returns are more concentrated around the mean, suggesting lower risk. Mathematically, the variance (σ^2) is calculated as follows:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (R_i - \bar{R})^2,$$

When n the number of periods (or observations), R_i represents the return on the asset in period i, and $\bar{R}$ is the average of the asset's returns, calculated, as

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} R_i.$$

This formula measures the average of the squares of the differences between each return and the average of the returns. The variance of a portfolio of financial assets takes into account not only the variance of returns for each asset, but also the covariance between assets. The formula for the variance of a portfolio composed of n assets is given by:

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j \sigma_{ij},$$

Where w_i is the proportion of the asset i in the portfolio. , σ_i^2 is the variance of the asset i,and σ_{ij} is the covariance between assets i and j. The first sum represents the variance contribution of individual assets, while the second sum represents the effect of covariance between assets, which can reduce the overall portfolio risk. In summary, diversification across assets can decrease total portfolio variance.

4.2.2. VaR of a portfolio

Value at Risk (VaR) is defined as the maximum potential loss in the portfolio value of financial instruments with a given probability over a certain horizon. Simply, it indicates how much a financial institution can lose with a certain probability over a given period [19]. The VaR of assets, for a period t and a probability level q, is defined as the expected loss amount such that this amount over the period $[t-1;t]$ should not be greater than the VaR with probability q.

$$P[P_{t,t-1} \leq \text{VaR}(t,q)] = q,$$

Where

- $P_{t,t-1} = P_t - P_{t-1}$: represents the loss ("Loss") and is a positive or negative random variable;
- t: is the horizon associated with the VaR;
- q: is the probability level, typically 95

VaR depends on three elements:

- the distribution of profits and losses of the portfolio valid for the holding period.
- the confidence level;
- the holding period of the assets.

If the distribution of assets follows a normal distribution,the VAR is given by result:

$$P[P_{t,0} \leq \text{VaR}(t,\alpha)] = \alpha \implies P\left(\frac{P_{t,0} - E(P_{t,0})}{\sigma(P_{t,0})} \leq \frac{\text{VaR}(t,\alpha) - E(P_{t,0})}{\sigma(P_{t,0})}\right) = \alpha$$

$$\implies \frac{\text{VaR}(t,\alpha) - E(P_{t,0})}{\sigma(P_{t,0})} = z_\alpha \implies \text{VaR}(t,\alpha) = E(P_{t,0}) + \sigma(P_{t,0}) \cdot z_\alpha$$

For a portfolio of n assets with weights $w_1, \ldots, w_k$, if we assume that the value of this portfolio V follows a multivariate normal distribution, then the Value at Risk (VaR) of the portfolio is given by:

$$\text{VaR}(t,\alpha) = -N\mu_t + z_\alpha \cdot \sqrt{N_t \Omega_t \Sigma}$$

Where:

$$\mu_t = E(\Delta V), \quad \Omega_t = \sigma_t(\Delta V), \quad N = (w_1, \ldots, w_k), \quad z_\alpha \text{ is the } \alpha\text{-quantile.}$$

The first step in the optimization process is to calculate the expected return of each asset and its VaR, storing the results in a matrix called `MAT`.
We have developed an algorithm that selects the best-performing assets in terms of score (i.e., the highest expected return and the lowest VaR). Considering the sub-portfolio extracted from the full asset set, we then proceed to apply an optimization algorithm based on Genetic Algorithms (GA), which we define as follows.

4.3. Heuristic Optimization Algorithms

4.3.1. Genetic Algorithms (GA)

Genetic Algorithms[20] are optimization methods developed by Holland, based on the principles of biological evolution. They operate on a population of fixed size, consisting of candidate solutions called *chromosomes*. Each chromosome is composed of elements called *genes*, which encode a potential solution to the problem at hand. Genetic Algorithms are iterative search procedures for finding optimal solutions. At each iteration, known as a *generation*, a new population is created, maintaining the same number of chromosomes. This new population is composed of chromosomes that are better adapted to their environment, as evaluated by a selection (fitness) function. Over successive generations, the population evolves towards the optimum of this function.
The algorithm begins by randomly generating an initial population of individuals. To evolve from generation k to generation $k+1$, three genetic operations are repeatedly applied to the individuals in generation k: selection, crossover, and mutation.

- **Selection:** This operation selects the best-performing chromosomes based on the objective function, preserving those with higher fitness.
- **Crossover:** This operation produces two new "child" chromosomes by combining two selected "parent" chromosomes.
- **Mutation:** This operation introduces random modifications by inverting one or more genes in a chromosome[21].

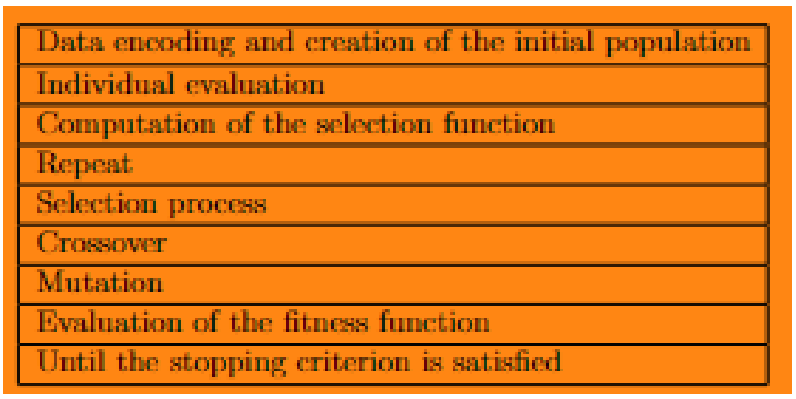

Figure 3: Illustration of the operations in a basic genetic algorithm

4.3.2. Application of the genetic algorithm

The algorithm considered in this article is freely implemented by [22],and begins with the maximization of the ratio of two functions. The numerator represents the change in value of the sub-portfolio while the denomotor is the associated risk, when respecting

certain additional constraints. Subsequently, it aims to dynamically maximize the value of the sub-portfolio. The value obtained must be higher than that of the first step, and the resulting VaR must be lower than that established initially, while taking into account other constraints.
In summary, the steps of the algorithm:

1- Initialization

The population consists of chromosomes consisting of n genes representing w_i ($i = 1, \ldots, n$) numbers, where w_i is the amount of wealth invested in the 'active i. This population is initially generated randomly using real values.

2- Evaluation function

The next step consists of evaluating the chromosomes resulting from the previous operation using an evaluation function (fitness function), which is essential in the context of genetic algorithms. The fitness functions applied in this study are as follows:

i) $f = \dfrac{\Delta V(x)}{\sigma^2}$

Under the following conditions:

$\overline{R_p}_{\text{initial}} \leqslant \overline{R_p}_{\text{optimal}}$ and $\sigma^2_{\text{optimal}} \leqslant \sigma^2_{\text{initial}}$

ii) $f = \dfrac{\Delta V(x)}{\text{VaR}(\alpha, t)}$

Under the following conditions:

$\overline{R_p}_{\text{initial}} \leqslant \overline{R_p}_{\text{optimal}}$ and $\text{VaR}(t, \alpha)_{\text{optimal}} \leqslant \text{VaR}(t, \alpha)_{\text{initial}}$

3- Selection operations

After the population evaluation operation, the best chromosomes are selected using Roulette selection, which associates a selection probability, denoted P_i, to each chromosome f. For the maximization problem:

$$P_i = \frac{f_i}{\sum_{j \in \text{Pop}} f_j},$$

where f_i is the value of the individual's fitness function i, which measures the quality of the candidate solution with respect to the problem objective, and $\sum_{j \in \text{Pop}} f_j$ is the sum of the fitness values of all individuals in the population. Each chromosome is reproduced with a certain probability. Some chromosomes will be "no longer" reproduced while other "bad" ones will be eliminated.

4- Crossing operations

After using the selection method to choose two individuals, we apply the one-point crossover operator between these individuals. This operator divides each parent into two parts at a randomly chosen position. Child 1 is made up of the first part of the first parent and the second part of the second parent, while child 2 is made of the second part of the first parent and the first part of the second parent.

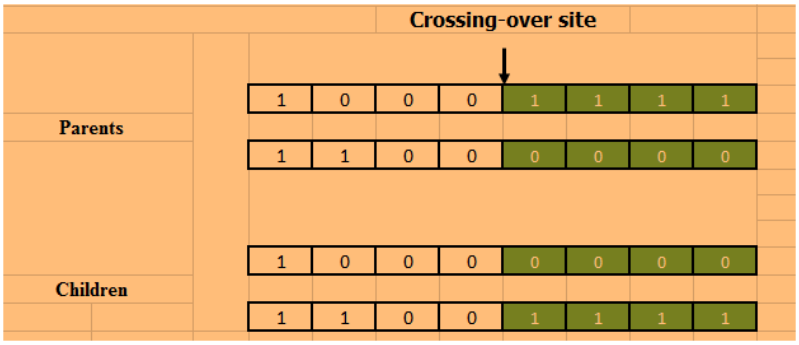

Figure 4:

5-Mutation operator

The mutation operator [21] is an operator that allows an genetic algorithm to reach all points in the state space in a susceptible way, without traversing them all in the solution process.It which allows the convergence of genetic algorithms towards the global optimum. This operation consists of randomly replacing a gene in the chromosome by a random value. This can be chosen in the vicinity of the value initial. It is generally used for discrete problems. The figure below illustrates this mechanism well.

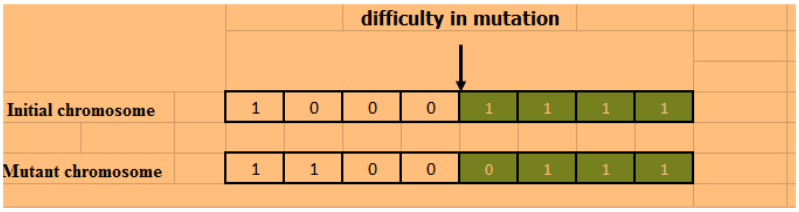

Figure 5:

6- Convergence Conditions

At this level, the final generation is considered. If the result is favorable, then the optimal chromosome is obtained. Otherwise, the evaluation and reproduction steps are repeated for a certain number of generations, until a population convergence criterion is reached.

4.3.3. Particle Swarm Optimization

The **Particle Swarm Optimization (PSO)** algorithm is an optimization method inspired by the collective behavior of birds in flight or fish schooling. Each potential solution (called a *particle*) explores the search space by moving according to both its own experience and that of other particles (i.e., the best solution found by the swarm) [23].

Each particle updates its **position** (current solution) and **velocity** based on:

- its own **personal best position** found so far,
- the **global best position** found by the swarm,
- **random factors** that help diversify the search.

PSO is often used for complex continuous optimization problems, such as **portfolio allocation**, due to its simplicity, robustness, and speed. We applied it to the ten best-performing assets that were also evaluated using the genetic algorithm.

5. Results and discussion

The information in question aggregates daily historical prices of assets from various categories.We, then examine a portfolio composed of 21 financial assets from different stock exchanges, acquired daily between 01/01/2018 and 01/01/2024. After applying the procedure described above, we obtain the results of the selected assets 'MSFT', 'PFE', 'V', 'MRK', 'AAPL', 'JNJ', 'MA', 'GOOGL', 'T', and 'UNH'. illustrated in the following figures :

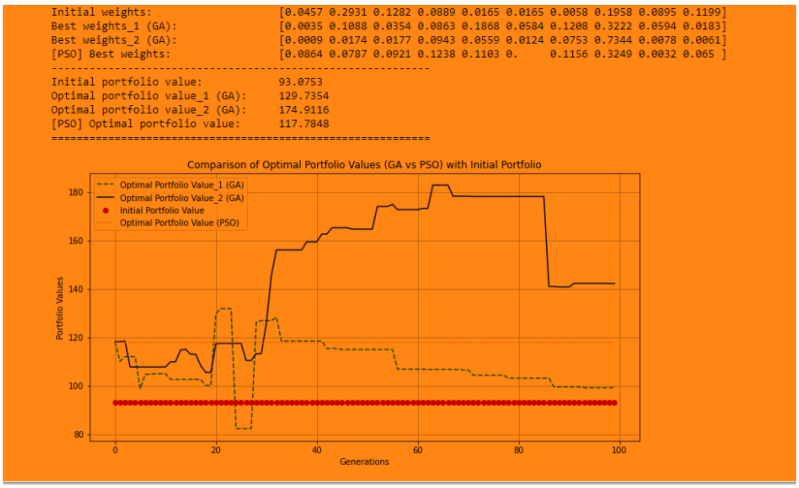

Figure 6: Result

Interpretation of Portfolio Optimization Results

Figure 6 provides the interpretation of portfolio optimization results. The dotted and continuous lines illustrated the optimal portfolio values for the GA with best weight 129.7354 expected and 174.9116 respectively while the circle represents the initial portfolio value.

- **Initial portfolio value: 93.0753**

This represents the *initial value* of the portfolio before any optimization . It is assumed here that the initial weights were assigned arbitrarily or equally.

- **Optimal portfolio value_1 (alg_1): 129.7354**
- **Optimal portfolio value_2 (alg_2): 174.9116**

These are two variants of optimized portfolios using the genetic algorithm:

- alg_1 shows an improvement of approximately

$$\frac{129.7354 - 93.0753}{93.0753} \approx 39.3\%$$

- alg_2 performs even better, with a gain of about

$$\frac{174.9116 - 93.0753}{93.0753} \approx 87.9\%$$

This suggests that alg_2, which relies on VaR, is the most effective.

- **[PSO] Optimal portfolio value: 117.7848**

The portfolio optimized using the PSO algorithm shows a gain of:

$$\frac{117.7848 - 93.0753}{93.0753} \approx 26.5\%$$

Performance Summary Comparison

Method	**Optimized Value**	**Gain (%)**
Initial Portfolio	93.0753	—
GA_1	129.7354	+39.3%
GA_2	174.9116	+87.9%
PSO	117.7848	+26.5%

Table 1 presents results from the three optimization algorithms in terms of standard value and gain. Among them, genetic algorithm with best weight 174.9116 provides an interesting performance with the best final portfolio value rate gain: 87.9Despite the

promising results, our work has certain limitations, the absence of comparison of our results with other methods. In addition, more optimization algorithms are considered in this study; we plan to compare with others like Differential Evolution (DE).
In risk management, optimization algorithms are often investigated to determine under certain risk measure the best investment. Note that genetic algorithms are well used. In this work, different algorithms are considered to compare their performance using the two risk measures (variance and VaR)and different stock exchange data. The genetic algorithm performs better compared (specially alg_2) to the particle swarm. This result is consistent with some previous studies like [24, 25].

Conclusion

In this article, we presented an approach for the optimal selection of a mining portfolio. This approach first involves a detailed description of the financial market by segmenting the data in order to distinguish the regimes of each financial asset, using a CART-based algorithm. Then, we select the assets corresponding to the identified market regime called portfolio — extraction using the algorithm we developed. Next, we apply dynamic optimization algorithms to this mining portfolio, namely the genetic algorithm, which is divided into two stages: the first stage aims to maximize a first ratio under constraints, while the second seeks to maximize a second ratio. We also apply the Particle Swarm Optimization (PSO) algorithm for comparison. The results obtained are generally satisfactory, demonstrating that optimization using the genetic algorithm is highly effective, and also confirming that risk measurement via Value-at-Risk (VaR) is superior to that via variance. Moreover, the genetic algorithm outperforms the PSO algorithm.

References

[1] W. F. Sharpe. A linear programming algorithm for mutual fund portfolio selection. *Management Science*, 13:499–510, 1967.
[2] B. Stone. A linear programming formulation of the general portfolio selection problem. *Journal of Quantitative Analysis*, 6:107–123, 1973.
[3] H. Konno and H. Yamazaki. A mean absolute deviation portfolio optimisation model and application to tokyo stock market. *Management Science*, 37:519–531, 1991.
[4] M. G. Speranza. Linear programming models for portfolio optimisation. *Finance*, 14:107–123, 1993.
[5] H. Markowitz. Portfolio selection. *Journal of Finance*, February 1952.
[6] Lorenzo Garlappi, Raman Uppal, and Tan Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach, 2005. Sauder School of Business Working Paper.
[7] Doron Avramov and Guofu Zhou. Bayesian portfolio analysis. *Annual Review of Financial Economics*, 2(1):25–47, 2010.
[8] Tze Leung Lai, Haipeng Xing, and Zehao Chen. Mean-variance portfolio optimization

when means and covariances are unknown. *Annals of Applied Statistics*, 5(2A):798–823, 2011.

[9] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

[10] Tak-Chung Fu, Fu-Lai Chung, and Chak-Man Ng. Financial time series segmentation based on specialized binary tree representation. In *Conference on Data Mining*, 2006.

[11] Tak-Chung Fu, Fu-Lai Chung, Robert Luk, and Chak-Man Ng. Representing financial time series based on data point importance. *Engineering Applications of Artificial Intelligence*, 21:277–300, 2008.

[12] Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. Segmenting time series: A survey and novel approach. In *Data Mining in Time Series Databases*, volume 57, pages 1–22. 2004.

[13] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, 1993.

[14] Azadeh Khaleghi and Daniil Ryabko. Locating changes in highly dependent data with unknown number of change points. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

[15] Azadeh Khaleghi and Daniil Ryabko. Nonparametric multiple change point estimation in highly dependent time series. In *Proceedings of the 24th International Conference on Algorithmic Learning Theory (ALT–13)*, 2013.

[16] Azadeh Khaleghi and Daniil Ryabko. Asymptotically consistent estimation of the number of change points in highly dependent time series. In *ICML, JMLR Workshop and Conference Proceedings*, volume 32, pages 539–547, 2014.

[17] R. Zhang. *Apprentissage statistique en gestion de portefeuille*. PhD thesis, Télécom ParisTech, 2014.

[18] Leo Breiman, Jerome H. Friedman, Robert A. Olshen, and Charles J. Stone. *Classification and Regression Trees*. Chapman and Hall, 1984.

[19] Philippe Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 2nd edition, 2000.

[20] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.

[21] E. Lutton. *Algorithmes génétiques et Fractales*. PhD thesis, Université Paris XI Orsay, 1999. Dossier d'habilitation à diriger des recherches.

[22] M. El Hachlou, Z. Guennoun, and F. Hamza. Stocks portfolio optimization using classification and genetic algorithms. *Applied Mathematical Sciences*, 6(94):4673–4684, 2012.

[23] Z.-H. Zhan, J. Zhang, Y. Li, and Y.-J. Gong. Evolutionary optimization in uncertain environments: A survey. *IEEE Transactions on Evolutionary Computation*, 25(2):188–211, 2021.

[24] G. A. Flores-Fernandez, M. Jimenez-Carrion, F. Gutierrez, and R. A. Sanchez-Ancajima. Genetic algorithm and lstm artificial neural network for investment portfolio optimization. *Universidad Nacional De Piura*, 2024. Reçu le 10 janvier 2024 ; révisé le 17 février 2024 ; accepté le 8 avril 2024 ; publié le 29 juin 2024.

[25] A. Mahmoudil, L. Hashemi, M. Jasemi, and J. Pope. A comparison on particle swarm optimization and genetic algorithm performances in deriving the efficient frontier of stocks portfolios based on a mean-lower partial moment model. *Wiley Research Article*, 2020.