Portfolio optimisation with multi-objective genetic algorithms

André Sebastián Cóndor *

October 2022

Abstract

Markowitz's approach to portfolio optimisation focuses on a bi-objective optimization problem in which return is maximized and investment risk is minimized. Within metaheuristics, multi-objective genetic algorithms become a very useful tool to deal with this specific problem. In this report, two different portfolio optimisation algorithms are proposed, explained and compared. The empirical attainment function and statistical tests are used to compare and better understand the behavior of both algorithms. The purpose of this report is to give the Financial Decision-Maker (FDM) a real tool to increase the sustenance of the decision-making process and improve it's outcomes.

Keywords: Portfolio optimisation, computational finance, etaheuristics, multi-objetive genetic algorithms.

1. Introduction

A portfolio is a collection of investments owned by an individual. A portfolio main objective is to generate profits through an investment strategy. The central focus of the strategy is to correctly allocate wealth among the group of assets which conform to the portfolio (Elton and Gruber, 1997).

Portfolio theory is a well-researched field of financial asset management. There are different approaches to portfolio investment strategy, but most of them converge in a rigorous mathematical analysis of the assets under consideration.

In general, the investor must take into account two vital factors when making investment decisions on portfolios. The first point is the risk of the investment. The second factor is the return on investment, i.e. the gain received for taking the risk.

This report aims to introduce modern portfolio theory and present two genetic algorithms that optimise investment portfolios by maximising return and minimising risk. The two algorithms will be compared, the results will be discussed, and conclusions and recommendations will be formulated.

2. Background

A stock is a security that represents the ownership of a part of the issuing corporation (Hayes, 2022). Each stock offered on the market carries an investment risk and return.

When making an investment, the Financial Decision-Maker (FDM) assumes the risk that the value invested may decrease or, alternatively, increase. The fluctuation in the value of the stocks may be due to general market conditions (market risk) or to independent decisions by the corporation issuing the asset (idiosyncratic risk). In return for taking on the investment risk, the FDM receives a return for investing money in a particular stock.

The FDM has to correctly analyse the return and risk generated by each investment in order to decide which stock to invest in. In general, the FDM has two perspectives to deal with this problem: to take more risk expecting more profit in the short term or to adopt a conservative behaviour and bet on a lower risk and profit in the short term in order to be compensated in the long term. The FDM has to make a decision based on his or her judgement and personal preferences.

Investing in a single stock exposes the FDM to various factors that can affect the investment value incredibly rapidly and affecting the investment strategy with relative ease. Diversification combines stocks with different characteristics following a mathematically backed logic to reduce overall investment risk. Diversification uses a portfolio to allocate wealth among a group of stocks in a way that reduces risk and increases return compared to a single stock investment.

Markowitz (1952) initiated the modern portfolio theory, the focus is on achieving an overall balance of risk and return, rather than concentrating on the risk of each stock in the portfolio. Markowitz aims to achieve a bi-objective optimisation increasing the portfolio return and decreasing the risk (Aouni et al., 2014). Markowitz's portfolio model provides de FDM with an efficient frontier from where it is possible to make decisions depending on individual risk return preferences (Markowitz, 1952).

The following section will explain the financial and mathematical principles of the formulation of the portfolio optimisation problem.

3. Portfolio optimisation

Since a portfolio is a set of stocks, the portfolio weights $\rho = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ are the percentage compositions of each stock in a portfolio of n stocks. The sum of all portfolio weights must be 1, since out of a total investment amount, wealth is divided into a set of assets that make up a portfolio.

$$\rho = \sum_{i=1}^{n} \alpha_i = 1 \tag{1}$$

Return β is the change of the asset value v over time t. For calculating the return in a certain time t, the difference in value is divided over the initial value of the asset.

$$\beta_t = \frac{v_t - v_{t-1}}{v_{t-1}} \tag{2}$$

In order to compute portfolio return λ, it is first needed to multiply α_n for each correspondent stock value percentage change β_t.

$$f(\lambda) = \sum_{i=1}^{n} \alpha_i \cdot \beta_i \tag{3}$$

[*] *E-mail address:* andre.balladares.22@ucl.ac.uk (A.S. Cóndor).

Returns spread around the mean is measured by variance σ^2, it is used as a measure of risk. As the stocks that are part of a certain portfolio are correlated, this relationship will influence the risk levels of the investment. Then, a portfolio risk σ^2 can be computed as:

$$g(\sigma^2) = [\rho]^T \begin{bmatrix} var(\sigma_1^2) & \dots & cov(\sigma_n, \sigma_1) \\ \dots & \dots & \dots \\ cov(\sigma_n, \sigma_1) & \dots & var(\sigma_n^2) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{bmatrix} \tag{4}$$

Variance is considered by Markowitz as a measure of risk, justifying it utilization in this report.

4. Implemented algorithm

In addition to the above clarifications, in the implementation of the algorithm in Python, certain theoretical observations in the financial field were considered.

The returns β of each share were annualized to 36 months, knowing that the historical data extracted covers from October 5 2019 to October 5 2022.

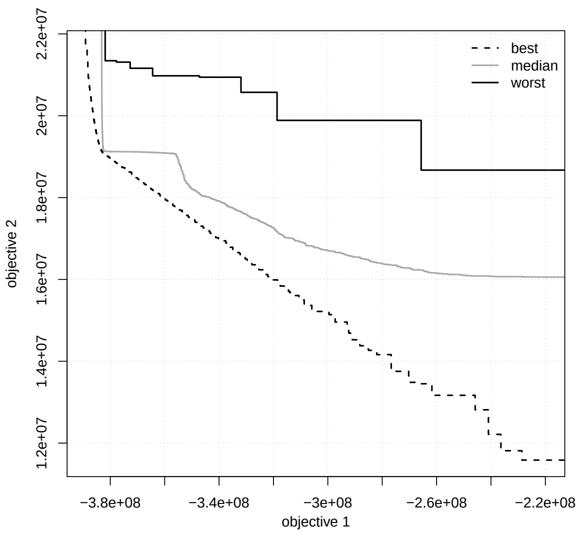

Fig. 1. EAF of Algorithm 1. The best and the median solution concluded by this algorithm share the same best point in $\mathcal{X}$.

The volatility was annualized σ^2 of the stocks was multiplied by $\sqrt{250}$ [1]. The algorithms have the constraint that the sum of all the weights must be equal to 1, following Eq. 3.

Inside the algorithms, the volatility σ^2 is treated as the standard deviation, i.e. risk $= \sqrt{\sigma^2}$.

The portfolio consists of 20 stocks from different corporations and industries, all of them relevant in the UK financial market.

The genetic algorithms were crafted using the 0.6.0 version of pymoo library. The implementations use a reference point based version of the Non-dominated Sorting Genetic Algorithm, so called R-NSGA-II. This algorithm selects the individuals frontwise. In certain situations, the front needs to be split due to the impossibility of certain individuals to survive.

[1] Conventionally, a year is considered to have 250 trading days.

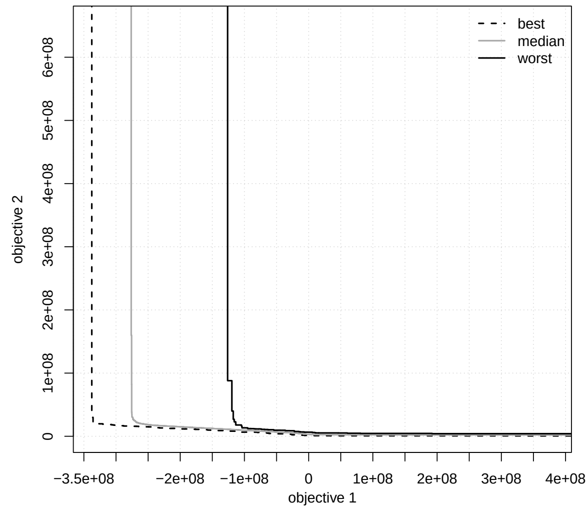

Fig. 2. EAF of Algorithm 2. The differences between the three plotted solutions are more accentuated when maximizing the return.

When this disruption occurs, the solutions are picked by a rank computed as the euclidean distance to each reference point provided by the FDM, so the closest solutions to a reference point receive a rank of 1 (Blank and Deb, 2020).

For both algorithms, an attempt was made to maximize the return and minimize the risk of the portfolio. By programming requirements of pymoo library of only minimizing values, the returns of the portfolio were multiplied by -1 to minimize the negative of these values, as it is the same as maximizing the return. However, there were certain significant differences that will be explained below.

4.1. *Algorithm 1*

The first algorithm maximizes the return and minimizes the risk of the portfolio, so its a multi-objective evolutionary algorithm (MOEA). It has two constraints, the first is that the sum of all the weights must be equal to 1 and that the return must be equal or greater than 20 %, $\lambda \geq 0.2$.

The mathematical formulation for this optimisation problem is as follows:

$$\begin{aligned} \min_{\lambda,\, \sigma} \; & f(\lambda),\; g(\sigma) \\ \text{s.t. } & \rho = 1 \\ & \lambda \geq 0.2 \end{aligned} \tag{5}$$

As the algorithm optimises two objectives, it receives two points in a horizontal array as one reference point. The first number represents $f(\lambda)$ and the second number is $g(\sigma)$. In this case, the program was developed to acquire two references points, these points were:

$$\delta_1 = \begin{bmatrix} 0.05 & 0.2 \\ 0.07 & 0.7 \end{bmatrix}$$

4.2. *Algorithm 2*

The second algorithm maximizes the return and minimizes the risk of the portfolio. It has only one constraint, the sum of

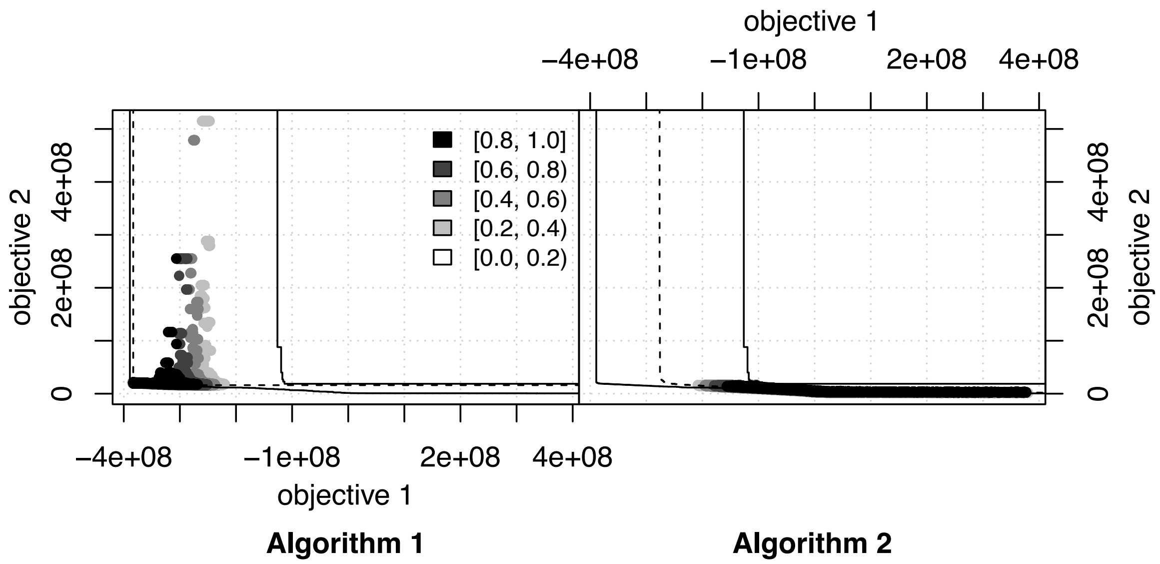

Fig. 3. EAF differences between Algorithm 1 and Algorithm 2.

all the weights must be equal to 1.

The main difference with the other algorithm is the integration of uncertainty in the value of risk and return. First, statistical distributions were tested on both the λ and σ data using the fitter library. The results showed that the return, λ, follows a uniform distribution (τ) with a mean of $5.203e^4$ and a standard deviation of 0.383. The risk, σ, has also a uniform distribution (ψ) but with a mean of $4.248e^{-05}$ and a standard deviation of 0.019.

Then, in the calculation of both objectives, the respective function was introduced to obtain random numbers corresponding to the mentioned distributions, and in this way take into account the mobility of the financial markets.

The mathematical formulation for this optimisation problem is as follows:

$$\begin{aligned} &\min_{\lambda,\,\sigma} f(\lambda)+\tau,\ g(\sigma)+\psi \\ &\text{s.t. } \rho = 1 \end{aligned} \tag{6}$$

Maintaining the same order and logic from the first algorithm, the two references points passed to this new algorithm were:

$$\delta_2 = \begin{bmatrix} 0.2 & 0.02 \\ 0.4 & 0.07 \end{bmatrix}$$

In the next section, the results of both algorithms will be analysed and compared.

5. Algorithms comparison

Diaz and López-Ibáñez (2021) mention that a MOA generates a set of solutions called *Pareto-optimal set* conformed by all the solutions that are not dominated by any other solution. The image of the *Pareto set* in objective space is the Pareto-front. Having access to the Pareto-front, the FDM can choose preferred solution(s).

In a multi-objective problem, n functions are being optimised simultaneously. This means that any solution from the MOEA is contained in the $\mathcal{X}$ solution space as a n-dimensional vector inside the objective space $\mathbb{R}^n$, i.e. $\vec{h}: \mathcal{X} \to \mathbb{R}^n$. An objective vector $\vec{v} \in \mathbb{R}^n$ of an $x \in \mathcal{X}$ solution is defined by $\vec{v} = \vec{h}(x) = (v_1 = h_1(x), \ldots, v_n = h_n(x))$ (Diaz and López-Ibáñez, 2021).

Assessing the performance of MOEAs is a complex task due to the stochasticity of meta-heuristics, the diversity of optimisation problems to be solved and incomparability of fronts in terms of Pareto optimality (Diaz and López-Ibáñez, 2021).

To facilitate the comparison of MOEAs, many methods have been proposed through the literature. Particularly, this report will focus on the *empirical attainment function*, since it is a way to compare the MOEAs' performances in $\mathbb{R}^n$.

5.1. *Empirical attainment function (EAF)*

Fonseca and Fleming (1996) explained the attainment surface as a border that separates $\mathbb{R}^n$ into two areas: attained[2] and not attained vectors. The attained surfaces enable summarizing the behaviour of an algorithm (López-Ibáñez et al., 2010).

López-Ibáñez et al. (2010) explain that the empirical attainment function (EAF) summarizes the outcome of various runs of a certain algorithm, and simultaneously, is capable to distinguish where a certain algorithm outperforms another.

For plotting the EAF, the eaf package was utilized over R through RStudio. Each of the two algorithms were executed 400 times, each time using a different seed. Then, the results were saved with the correct format in order for eaf package to plot the attainment functions.

Fig. 1 illustrates the EAF of the first algorithm. The first objective, in the horizontal axis, is the negative of the return which is being minimized. The vertical axis represents the risk of the portfolio, which is also minimized. This figure present three lines corresponding to the best, the worst and the median solutions found by the algorithm. It can be seen that the algorithm does a good job of maximizing the return, but the risk is not minimized as much as the first objective. It can even be observed that the best solution and the mean share an opti-

[2] A vector $\vec{p}$ attaines or dominates another vector $\vec{q}$, iff $p_i < q_i \forall i$.

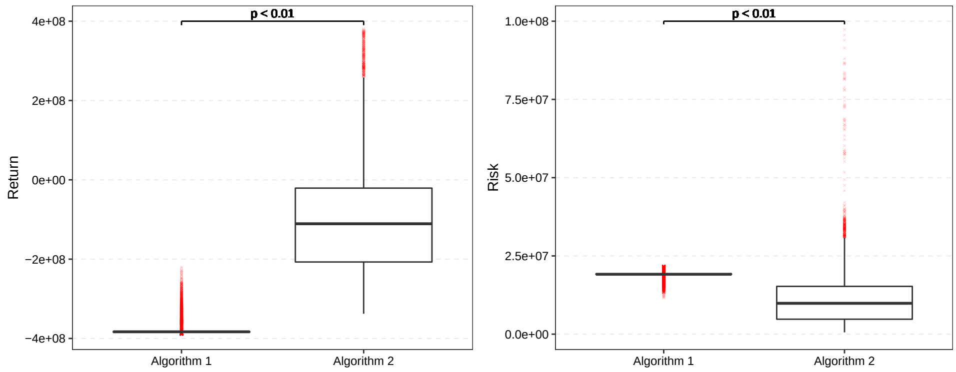

Fig. 4. EAF differences between Algorithm 1 and Algorithm 2.

mal point. On the other hand, the worst solution is very close to minimizing the second objective found by the best solution.

The EAF of the second algorithm is displayed at Fig. 2. In this algorithm, the three solutions drawn almost share the minimized value of the second objective. The differences are in the first objective. There is a great disparity in objective one between the best and the worst solution. The mean of the algorithm is closer towards smaller values of the negative of the return.

Knowing the behaviour of the two algorithms, it is possible to visually compare them to find places where one algorithm is better than another. Fig. 3 exposes the differences between both algorithms and the regions in which one is better than the other. Certain areas of the EAF are painted with a colour scale from light gray to black, this means that the darker the coloured region, the more likely the plotted algorithm is to find better solutions compared to another algorithm.

In this particular case, starting with the first algorithm, it is possible that there are darker points to the left of the horizontal axis, that is the minimization of the negative of the return. As the return decreases, algorithm one is less likely to find better results in that area of the EAF compared to algorithm 2.

In contrast, algorithm two evidently performs better within the minimization of the second objective. The fact that the second algorithm performs better with the second target may be due to the given reference points. While for the first algorithm, the references emphasized return maximization, algorithm two's reference points tended toward minimizing investment risk.

Lastly, the results of the objectives are compared separately using the box plots presented in Fig. 4. For the return, algorithm one finds the best solution with values that are very close to each other. Algorithm two has a higher mean for the negative of the return, and also its results are more dispersed. A Mann–Whitney U test was performed between the values obtained from the two algorithms using the mwu function from the pingouin library, the test resulted in $U_1 = 119436$ and a p-value of $p_1 = 0$, assuring that the solutions from the algorithms are different.

Speaking of portfolio risk, algorithm one offers more concentrated answers, but its mean is above the mean of algorithm 2. The second algorithm finds better solutions to minimize risk, although more dispersed than the first algorithm, however its third quartile is less than the first quartile of the first algorithm. Another Mann–Whitney U test was performed between the values obtained from the two algorithms regarding risk, the test resulted in $U_2 = 381184272.5$ and a p-value of $p_2 = 0$, confirming that the results are different.

Reproducibility. All codes, datasets and scripts used for this report are fully available at the following repository.

6. Conclusions and future work

In conclusion, investing in a stock portfolio is a bi-objective optimization problem in which the aim is to maximize the portfolio's return while reducing the investment risk.

The objective of this report is to give the Financial Decision-Maker (FDM) a couple of genetic algorithms as a method to make investment decisions based on technological and mathematical models in which the best possible solution is sought, including the uncertainty of the financial markets.

Two algorithms were presented, each with different characteristics. The analysis concluded that the results differ between the two, each one getting better options than the other, either increasing the return or decreasing the risk. Comparisons between algorithms were made using the empirical attainment function and statistical tests.

The explanations given in this report are a basis for the FDM to make decisions depending on their tolerance for risk and their expectation of return.

As future work, studying how the results change depending on the genetic algorithm used (R-NSGA-II was used in this report) can lead to interesting conclusions. For example, certain approaches could be found that are better to maximize the return or, failing that, to minimize the risk and later combine them to find even better solutions.

Acknowledgement

Thanks to all who helped in the production of this report with their knowledge, experience, ideas and recommendations.

References

Aouni, B., Colapinto, C., and La Torre, D. (2014). Financial portfolio management through the goal programming model: Current state-of-the-art. *European Journal of Operational Research*, 234(2):536–545. https://doi.org/10.1016/j.ejor.2013.09.040.

Blank, J. and Deb, K. (2020). pymoo: Multi-objective optimization in python. *IEEE Access*, 8:89497–89509. https://doi.org/10.1109/ACCESS.2020.2990567.

Diaz, J. E. and López-Ibáñez, M. (2021). Incorporating decision-maker's preferences into the automatic configuration of bi-objective optimisation algorithms. *European Journal of Operational Research*, 289(3):1209–1222. https://doi.org/10.1016/j.ejor.2020.07.059.

Elton, E. J. and Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of banking & finance*, 21(11-12):1743–1759. https://doi.org/10.1016/S0378-4266(97)00048-4.

Fonseca, C. M. and Fleming, P. J. (1996). On the performance assessment and comparison of stochastic multiobjective optimizers. *International Conference on Parallel Problem Solving from Nature*, 1141(1):584–593. https://doi.org/10.1007/3-540-61723-X_1022.

Hayes, A. (2022). Stocks: What they are, main types, how they differ from bonds. *Trading Skills*. https://www.investopedia.com/terms/s/stock.asp.

López-Ibáñez, M., Paquete, L., and Stützle, T. (2010). Exploratory analysis of stochastic local search algorithms in biobjective optimization. 1:209–222. https://doi.org/10.1007/978-3-642-02538-9.

Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91. https://doi.org/10.2307/2975974.