Machine Learning and Financial Engineering Techniques Applied to Quantitative Analysis

Matheus Torquato

March, 2021

Contents

1 Intro

This document is intended to demonstrate a few Machine Learning and Financial Engineering Techniques Applied to Quantitative Analysis. At the same time, some traditional financial concepts such as the Capital Asset Pricing and Markowitz Model are explored, this document also makes use of some statistical, data science and optimisation algorithms such as the Monte-Carlo simulation, correlation analysis and the use of constrained multi-objective optimisation for portfolio selection, for example.

Some other related analyses such as the Black-Scholes formula and time-series forecasting using Machine Learning predictors were left out of this document for the sake of brevity.

Even though no code is presented here, all studies, algorithms, results and images presented in this document are derived from software implementations coded in Python by the author of this portfolio and they are all available upon request (if they have not already been made available on GitHub.

2 Markowitz-Model Efficient Frontier

The following simulation shows results concerning the Markowitz-Model for Portfolio Optimisation according to the Efficient Frontier and Sharpe Ratio.

The efficient frontier theory was introduced by Nobel Laureate Harry Markowitz in 1952 and is a cornerstone of modern portfolio theory (MPT). The efficient frontier graphically represents portfolios that maximise returns for the risk assumed. Returns are dependent on the investment combinations that make up the portfolio. Every possible combination of assets that exists can be plotted on a graph, with the portfolio's risk on the X-axis and the expected return on the Y-axis. This plot reveals the most desirable portfolios.

The Sharpe Ratio $S(x)$ describes how much excess return you receive for the extra volatility you endure for holding a riskier asset. It is calculated according to the formula from Figure 1.

$$S(x) = \frac{(r_x - R_f)}{StdDev(r_x)}$$

where:

$x =$ The investment

$r_x =$ The average rate of return of x

$R_f =$ The best available rate of return of a risk-free security

$StdDev(x) =$ The standard deviation of r_x

Figure 1: Sharpe Ratio Formula

For this simulation, the used risk free rate was of 2.1%.

Given the following stocks from The Financial Times Stock Exchange 100 Index (FTSE100):

- LSE - London Stock Exchange Group PLC
- DGE - Diageo PLC
- OCDO - Ocado Group PLC
- EXPN - Experian PLC
- CARD - Card Factory PLC

30000 random long-only portfolio allocations were generated and their respective annualised returns, volatility (from the standard deviation) and Sharpe Ratios were calculated based on financial data (Adjusted Closing Price) collected from the *Yahoo Finance* platform from 01/01/2018 to 20/01/2020. Figure 2 displays the price of each one of the previously mentioned stocks during the analysed period and Figure 3 shows a snippet of the used historical data.

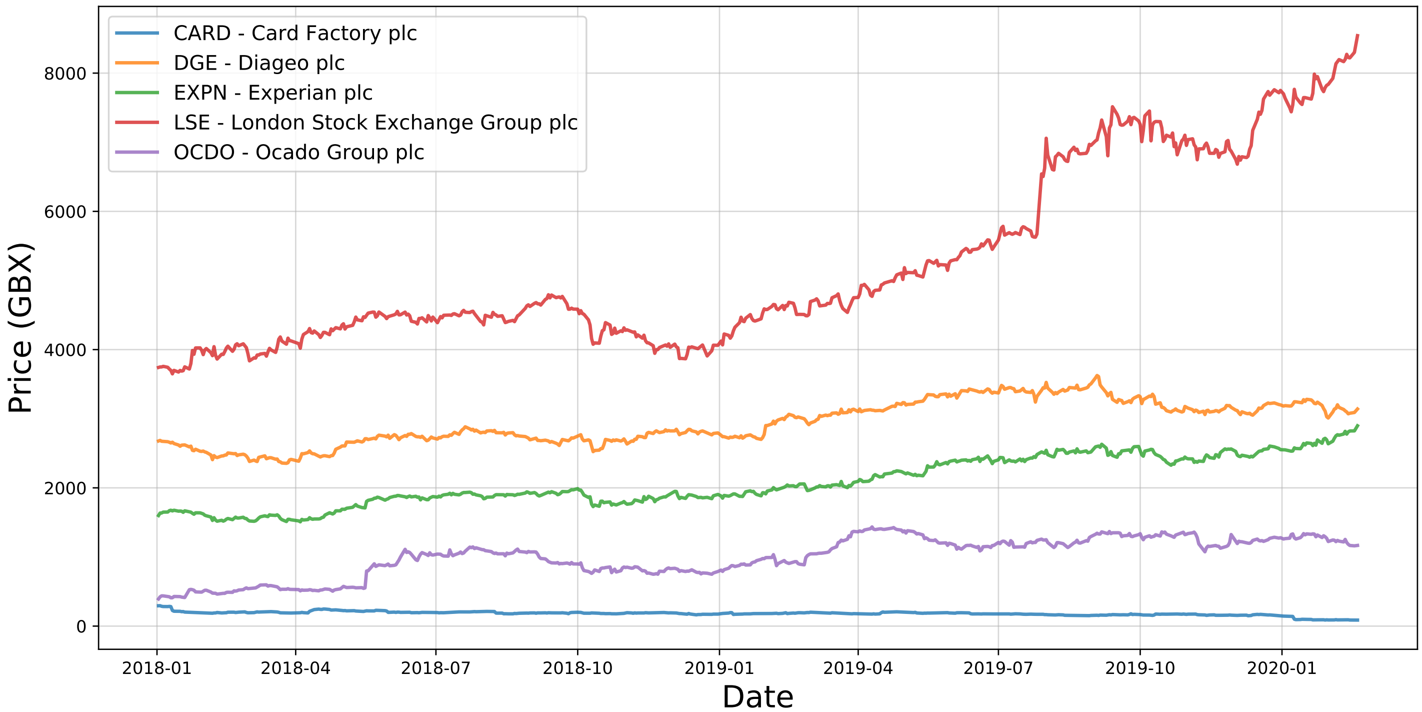

Figure 2: Historical Price Data

Considering that a hypothetical Portfolio A has an annual expected return of 20% and an annual standard deviation (volatility) of 14%, and that Portfolio B has an expected return of 20% and a standard deviation of 20%, Portfolio A would be deemed more efficient because it has the same expected return but lower risk. Both of these portfolios are illustrated in Figure 4.

In this same Figure, the 30000 generated portfolios are displayed together with the efficient frontier and the Portfolio with the highest Sharpe Ratio.

The dotted line from Figure 4 connect all of the most efficient portfolios, and this is known as the efficient frontier. Investing in any portfolio not on this curve is not desirable.

Among all of the generated portfolios it is possible to highlight three of them:

1. Highest Sharp Ratio Portfolio

Index	CARD.L	DGE.L	EXPN.L	LSE.L	OCDO.L
2019-10-02 00:00:00	162	3218.5	2477	7008	1284.5
2019-10-03 00:00:00	160	3279	2458	7168	1249
2019-10-04 00:00:00	160.2	3288.5	2532	7382	1284.5
2019-10-07 00:00:00	159.6	3328	2548	7452	1307
2019-10-08 00:00:00	157.4	3320	2558	7020	1285
2019-10-09 00:00:00	154.9	3355	2554	7248	1294.5
2019-10-10 00:00:00	159.6	3321.5	2508	7288	1298
2019-10-11 00:00:00	175	3210	2483	7300	1322
2019-10-14 00:00:00	171	3230.5	2455	7298	1302.5
2019-10-15 00:00:00	173.4	3164	2435	7208	1363.5
2019-10-16 00:00:00	172.7	3163	2411	7010	1350.5
2019-10-17 00:00:00	173.4	3146	2403	7044	1361
2019-10-18 00:00:00	172.3	3116.5	2371	7102	1330
2019-10-21 00:00:00	174.4	3096	2327	7074	1305.5
2019-10-22 00:00:00	172.9	3114.5	2347	7134	1294
2019-10-23 00:00:00	174.8	3123.5	2342	6936	1284
2019-10-24 00:00:00	174	3142	2389	6990	1317.5
2019-10-25 00:00:00	175	3121	2391	6818	1325
2019-10-28 00:00:00	171.5	3097	2420	7022	1350.5
2019-10-29 00:00:00	175.2	3105.5	2417	7046	1358.5
2019-10-30 00:00:00	171.9	3178	2446	7102	1322.5
2019-10-31 00:00:00	169.4	3164.5	2428	6954	1329

Figure 3: Historical Price Dataset

- **Return: 42.76%**
- **Sharp Ratio: 1.92**
- **Volatility: 21.16%**
- LSE - London Stock Exchange Group PLC: 0.02%
- DGE - Diageo PLC: 0.77%
- OCDO - Ocado Group PLC: 29.88%
- EXPN - Experian PLC: 46.95%
- CARD - Card Factory PLC: 22.38%

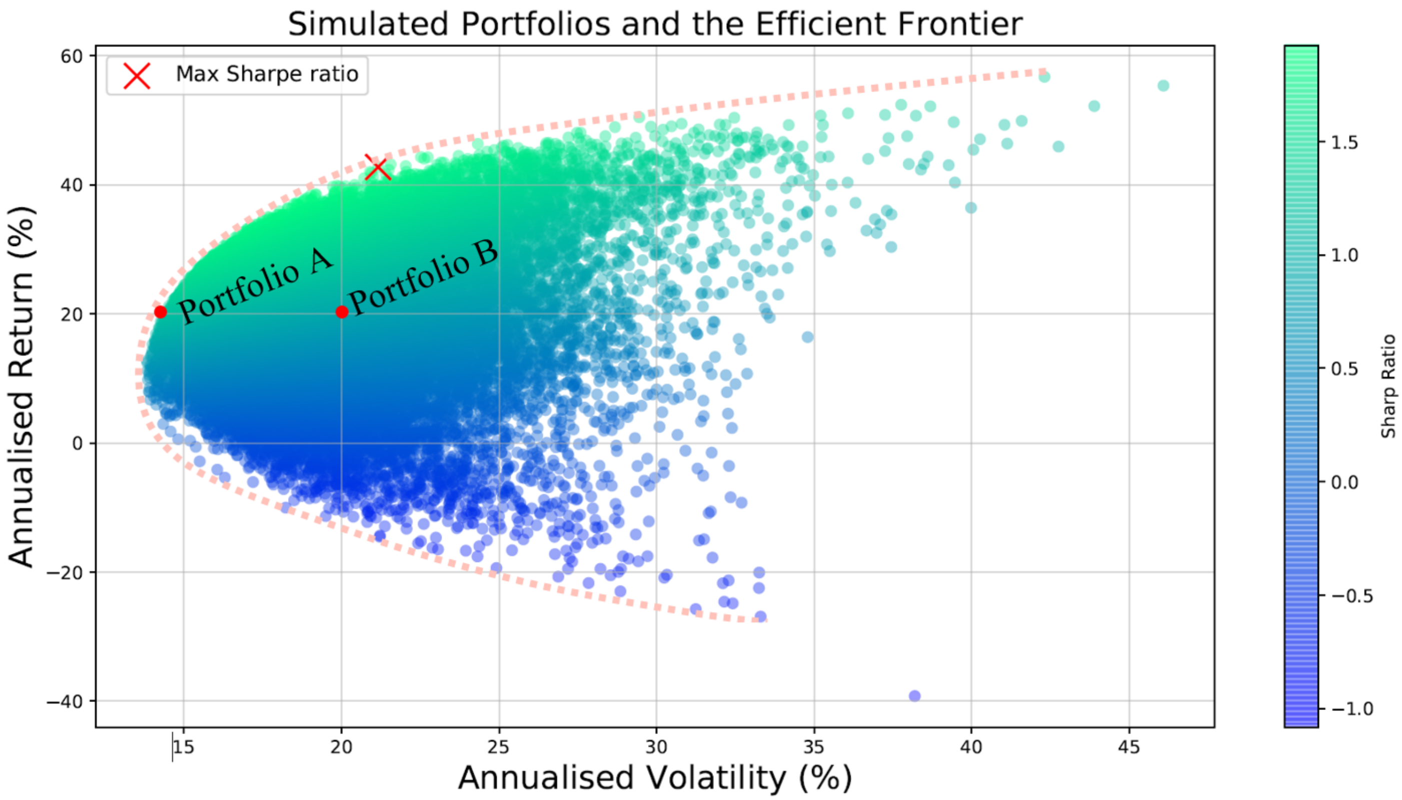

Figure 4: Efficient Frontier

2. Highest Return Portfolio
   - **Return: 56.74%**
   - **Sharp Ratio: 1.29**
   - **Volatility: 42.31%**
   - LSE - London Stock Exchange Group PLC: 0.47%
   - DGE - Diageo PLC: 4.97%
   - OCDO - Ocado Group PLC:0.80%
   - EXPN - Experian PLC: 16.97%
   - CARD - Card Factory PLC:76.78%
3. Lowest Volatility Portfolio

- **Return: 12.74%**
- **Sharp Ratio: 0.77**
- **Volatility: 13.83%**
- LSE - London Stock Exchange Group PLC:10.73%
- DGE - Diageo PLC: 53.76%
- OCDO - Ocado Group PLC: 18.68%
- EXPN - Experian PLC: 14.50%
- CARD - Card Factory PLC: 2.34%

3 Correlation Analysis

This Section simply presents a correlation study among all the stocks from the FTSE100 index. Stock correlation describes the relationship that exists between two stocks and their respective price movements. It can also refer to the relationship between stocks and other asset classes, such as bonds or real estate. Stock correlation is on a scale from -1 to 1 and is calculated by looking at a pair of stocks over time and figuring out their average movement.

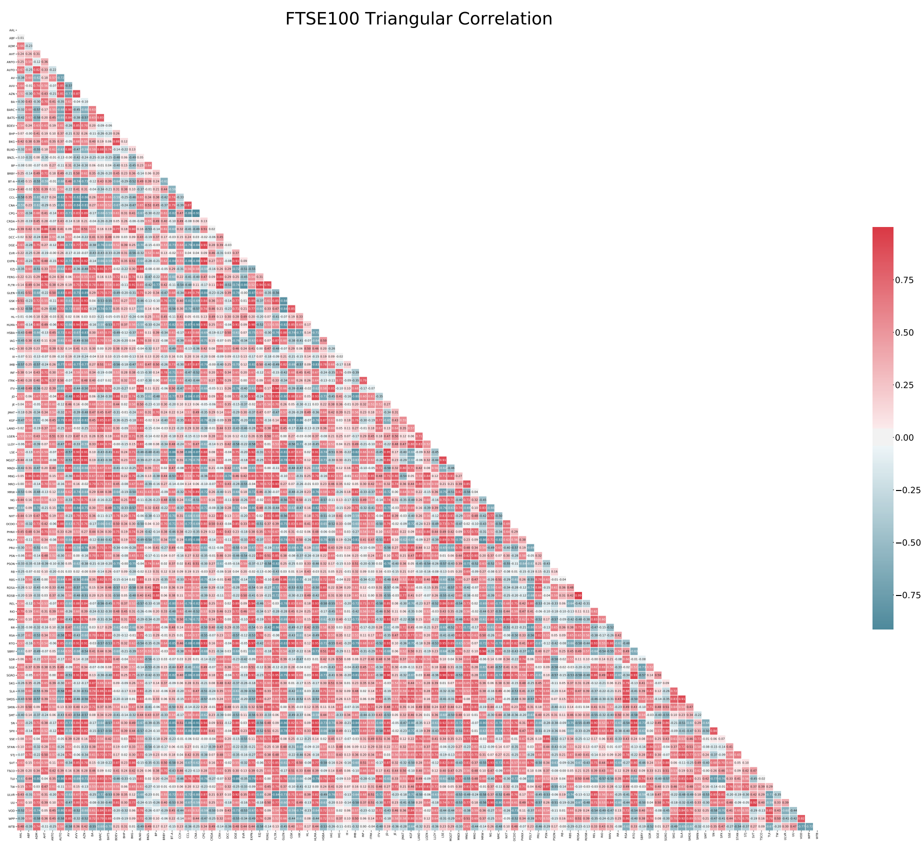

Figure 5: FTSE Triangular Correlation

Figure 5 shows the triangular correlation among all the stocks from the FTSE100 index.

Stocks can be positively correlated when they move up or down in tandem. A correlation value of 1 means two stocks have a perfect positive correlation. If one stock moves up while the other goes down, they would have a perfect negative correlation, noted by a value of -1. If each stock seems to move completely independently of the other, they could be considered uncorrelated and have a value of 0.

Now, individually analysing pairs of stocks, Figure 6 shows a pair of stocks with correlation value of 97.78% (0.97).

Figure 6: Example of High Positive Correlation

On the other hand, Figure 7 presents a pair of stocks with correlation value of -85.27% (-0.85). Finally, Figure 8 exemplifies a correlation close to 0 (0.04%) between NMC and MRO.

Figure 7: Example of High Negative Correlation

Figure 8: Example of Zero Correlation

4 Monte-Carlo Simulation

Monte-Carlo simulation is a computerised mathematical technique that allows people to account for risk in quantitative analysis and decision making. Monte Carlo simulation performs risk analysis by building models of possible results by substituting a range of values, a probability distribution, for any factor that has inherent uncertainty. It then calculates results over and over, each time using a different set of random values from the probability functions. Depending upon the number of uncertainties and the ranges specified for them, a Monte Carlo simulation could involve thousands or tens of thousands of recalculations before it is complete. Monte Carlo simulation produces distributions of possible outcome values.

Here, Monte-Carlo Simulations will be used for calculating Value at Risk (VaR). VaR calculates the maximum loss expected (or worst case scenario) on an investment, over a given time period and given a specified degree of confidence.

For the simulations here presented the time period used will be one year ahead of today's price and the confidence level level will be 95%. Again, two stocks from the FTSE100 index were randomly selected. This time: The Unilever Group (ULVR) and Carnival Corporation & PlC (CCL).

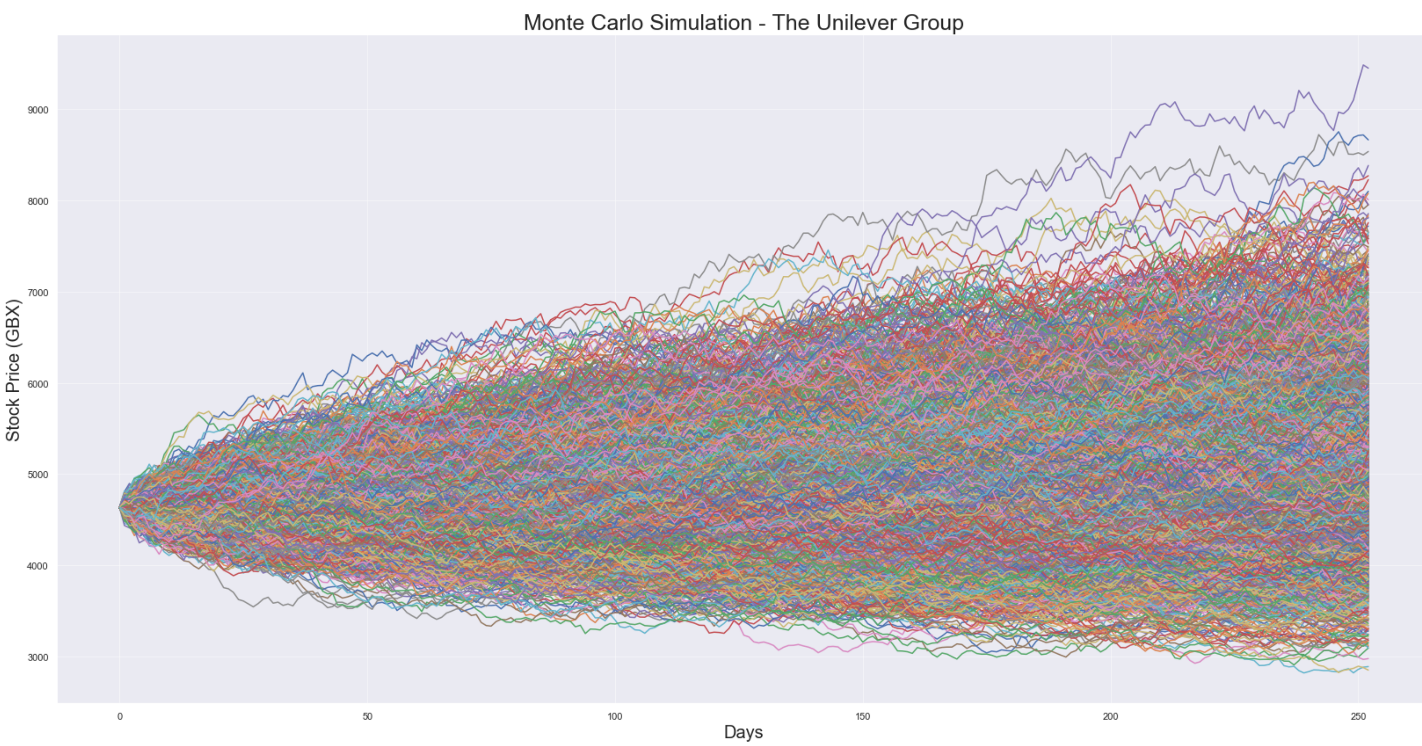

Figure 9: The Unilever Group - Monte-Carlo Simulation

The Monte-Carlo simulation from Figure 9 displays the 3000 simulated scenarios for the following 252 trading days of the The Unilever Group. The simulated returns from the Monte-Carlo simulation are based on the calculated Compound Annual Growth Rate (CAGR) and volatility from historical price data collected from the *Yahoo Finance* platform from 01/01/2018 to 20/01/2020, equivalent of what was shown in Figure 3.

From all of the simulated scenarios the following can be affirmed:

- **Company: The Unilever Group**
- Compound Annual Growth Rate (CAGR) = 9.68%
- Annual Volatility = 17.09%
- Current Price = GBX 4636.50
- Min Predicted Price = GBX 2853.32
- Max Predicted Price = GBX 9448.70
- Price Spread = GBX 6595.38
- Average Simulated Final Price = GBX 5113.63
- Average Simulated Return = 4.70%
- **Value at Risk at 95% Confidence = GBX 6693.60**

Taking the final price (simulated prices after 252 trading days) of each individual simulation from Figure 9 and plotting them together in a histogram produces Figure 10.

Figure 10 shows the distribution of the final price of The Unilever Group shares generated by the Monte Carlo simulation.

The same Monte-Carlo Simulation was performed to another company, now, Carnival Corporation & Plc, generating:

- **Company: Carnival Corporation & Plc**
- Compound Annual Growth Rate (CAGR) = -15.96%
- Annual Volatility = 27.22%
- Current Price = GBX 3086.00
- Min Predicted Price = GBX 1020.20

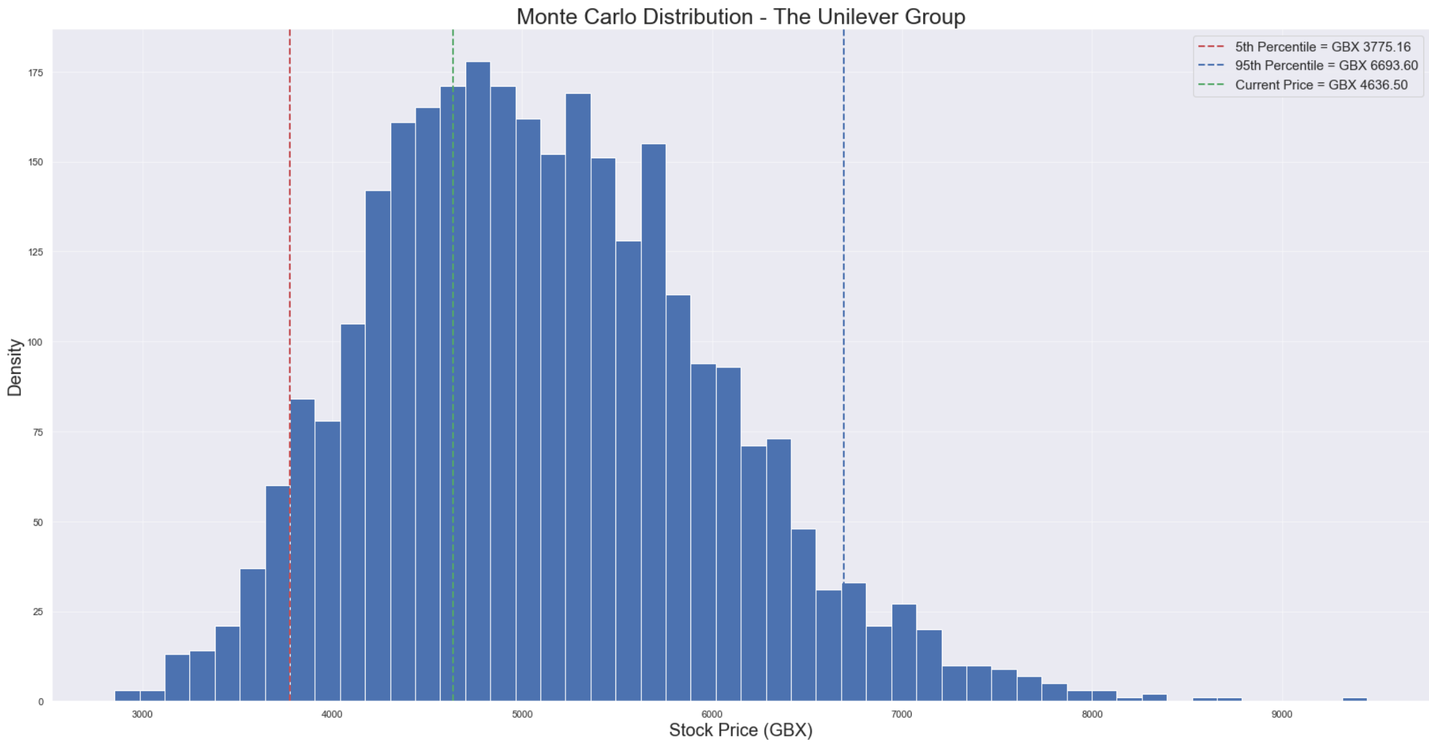

Figure 10: Final Price Distribution - Monte-Carlo Simulation

- Max Predicted Price = GBX 6550.34
- Price Spread = GBX 5530.14
- Average Simulated Final Price = GBX 2627.01
- Predicted Average Return = -7.28%
- Value at Risk at 95% Confidence = GBX 3922.31

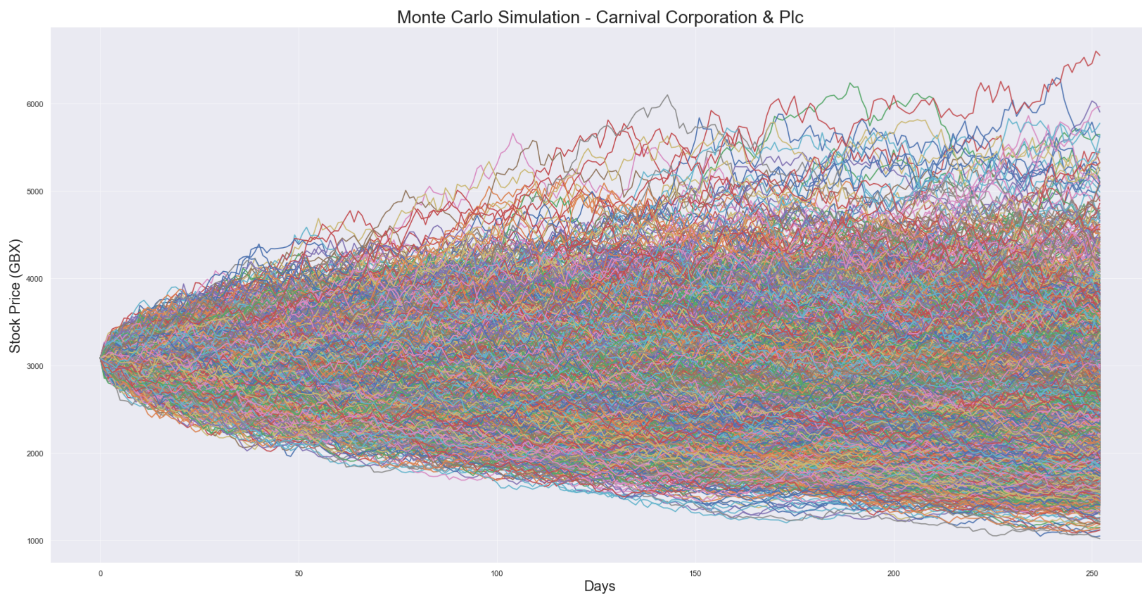

Figure 11: Carnival Corporation & PlC - Monte-Carlo Simulation

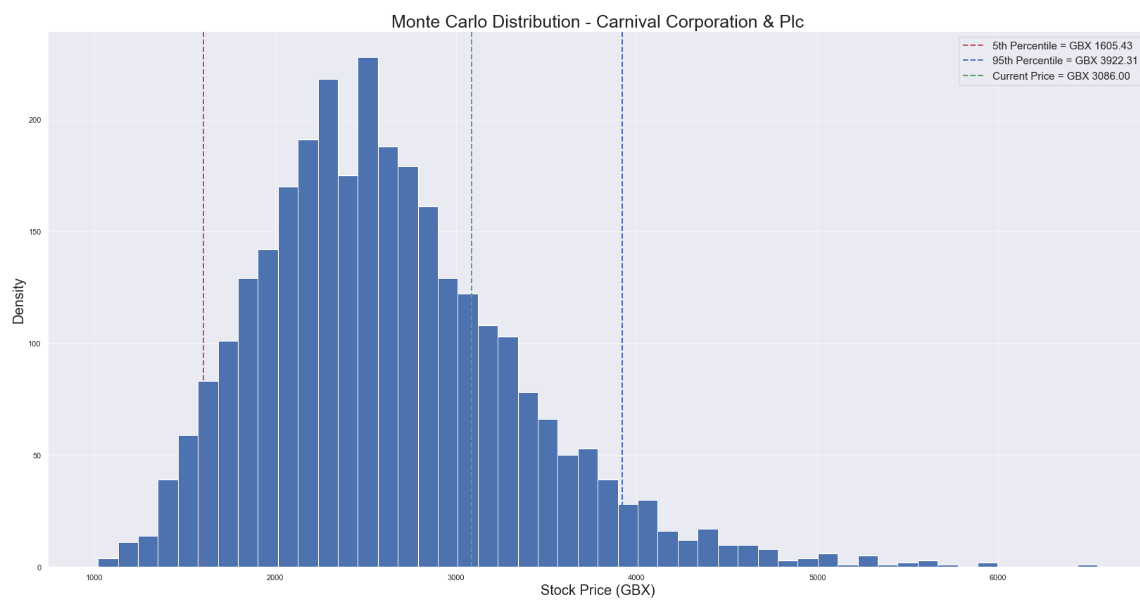

Figure 12: Final Price Distribution - Monte-Carlo Simulation

5 Bellman-Ford Algorithm for Forex Arbitrage

Forex arbitrage is a risk-free trading strategy that allows Forex traders to make a profit with no open currency exposure. The strategy involves acting on opportunities presented by pricing inefficiencies in the short window they exist. This type of arbitrage trading involves the buying and selling of different currency pairs to exploit any pricing inefficiencies. The act of exploiting the pricing inefficiencies will correct the problem so traders must be ready to act quickly with arbitrage strategies. For this reason, these opportunities are often around for a very short time.

The Bellman-Ford algorithm seeks to solve the single-source shortest path problem. It is used in situations where a source vertex is selected and the shortest paths to every other vertex in the graph need to be determined. After applying the Bellman-Ford algorithm on a graph, each vertex maintains the weight of the shortest path from the source vertex to itself and the vertex which precedes it in the shortest path.

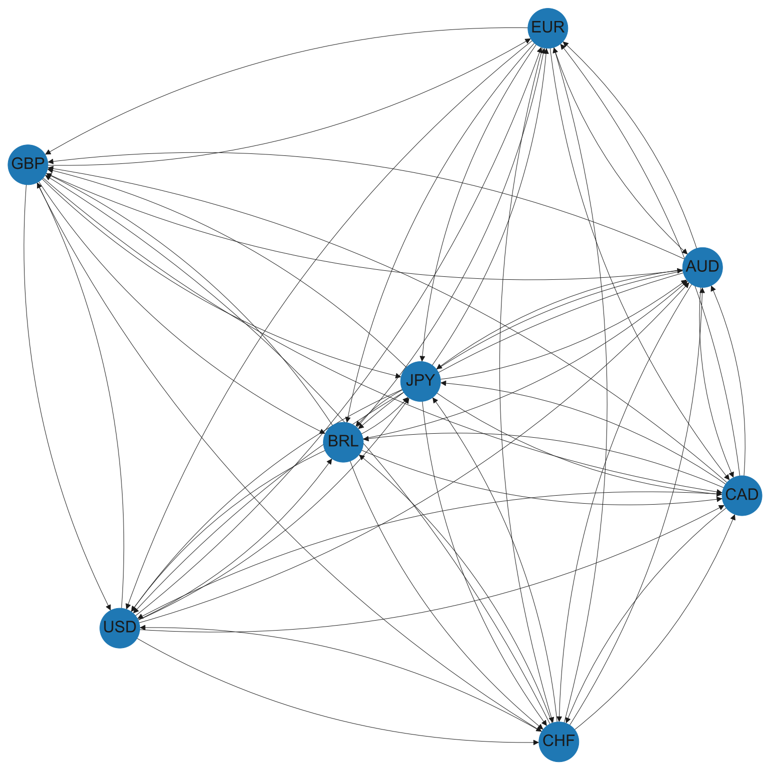

Figure 13: Currency Exchange Graph

Figure 13 shows a graph representing all the possible exchanges among the following currencies:

- GBP - British Pound
- EUR - European Euro
- USD - United States Dollar
- AUD - Australian Dollar
- BRL - Brazilian Real
- CAD - Canadian Dollar
- CHF - Swiss Franc
- JPY - Japanese Yen

Assuming the exchanges rates from Figure 14 the implemented Bellman-Ford Algorithm resulted in the following arbitrage opportunities:

Index	GBP	EUR	USD	BRL	CHF	CAD	JPY	AUD
GBP	1	0.82985	0.767243	0.17653	0.781477	0.578494	0.00699115	0.512411
EUR	1.20504	1	0.924556	0.212725	0.941708	0.697107	0.0084246	0.617475
USD	1.30337	1.0816	1	0.230084	1.01855	0.753991	0.00911205	0.66786
BRL	5.66476	4.7009	4.34625	1	4.42688	3.27703	0.0396032	2.90269
CHF	1.27963	1.0619	0.981786	0.225893	1	0.740258	0.00894608	0.655696
CAD	1.72863	1.4345	1.32628	0.305154	1.35088	1	0.0120851	0.885767
JPY	143.038	118.7	109.745	25.2505	111.781	82.7466	1	73.2942
AUD	1.95156	1.6195	1.49732	0.344508	1.5251	1.12896	0.0136436	1

Figure 14: Exchange Rates from 19/02/2020

Arbitrage Opportunity:
AUD → BRL → AUD → JPY → GBP → EUR → AUD
Profit of 4.353971405635093e-07%

Arbitrage Opportunity:
AUD → BRL → AUD → AUD
Profit of 7.872301921452163e-09%

Arbitrage Opportunity:
BRL → AUD → EUR → BRL → BRL

Profit of 1.1878256600539316e-08%

Arbitrage Opportunity:
AUD → BRL → AUD → JPY → USD → EUR → AUD
Profit of 2.501003422139547e-07%

Arbitrage Opportunity:
AUD → BRL → AUD → AUD
Profit of 7.872301921452163e-09%

Arbitrage Opportunity:
AUD → BRL → AUD → AUD
Profit of 7.872301921452163e-09%

Arbitrage Opportunity:
AUD → BRL → AUD → JPY → CHF → GBP → AUD
Profit of 4.3697400542441756e-07%

Arbitrage Opportunity:
AUD → BRL → AUD → AUD
Profit of 7.872301921452163e-09%

Arbitrage Opportunity:
AUD → BRL → AUD → JPY → CAD → EUR → AUD
Profit of 3.4936047654809954e-07%

Arbitrage Opportunity:
AUD → BRL → AUD → AUD
Profit of 7.872301921452163e-09%

Arbitrage Opportunity:
AUD → BRL → AUD → JPY → GBP → AUD
Profit of 4.34856559650143e-07%

Arbitrage Opportunity:
AUD → BRL → AUD → AUD
Profit of 7.872301921452163e-09%

It can be observed that of all the suggested arbitrage opportunities would produce profits under 0.00001%. This means that no real arbitrage opportunities were actually found and that are no pricing inefficiencies whatsoever. In a real scenario none of the above "opportunities" would be shown since they would need to pass through a threshold first in order to be actually considered as a arbitrage opportunity, something as $Profit \geqslant 0.2\%$. All of the above results were generated in a computing time of 3.54 milliseconds.

Now, in order to simulate a real arbitrage opportunity, one of the original exchange rates from Figure 14 was changed. The British Pound to European Euro exchange rate was intentionally changed from £1/€1.20504 to £1/€1.21709, a 1% increase. The new exchange rate Table is depicted in Figure 15

Index	GBP	EUR	USD	BRL	CHF	CAD	JPY	AUD
GBP	1	0.82985	0.767243	0.17653	0.781477	0.578494	0.00699115	0.512411
EUR	1.21709	1	0.924556	0.212725	0.941708	0.697107	0.0084246	0.617475
USD	1.30337	1.0816	1	0.230084	1.01855	0.753991	0.00911205	0.66786
BRL	5.66476	4.7009	4.34625	1	4.42688	3.27703	0.0396032	2.90269
CHF	1.27963	1.0619	0.981786	0.225893	1	0.740258	0.00894608	0.655696
CAD	1.72863	1.4345	1.32628	0.305154	1.35088	1	0.0120851	0.885767
JPY	143.038	118.7	109.745	25.2505	111.781	82.7466	1	73.2942
AUD	1.95156	1.6195	1.49732	0.344508	1.5251	1.12896	0.0136436	1

Figure 15: Modified Exchange Rate

Running the Bellman-Ford Algorithm again produced:

Arbitrage Opportunity:
AUD → BRL → AUD → JPY → GBP → EUR → AUD
Profit of 1.0000004356848393%

Arbitrage Opportunity:
AUD → BRL → AUD → AUD
Profit of 7.872301921452163e-09%

⋮

Again, a few opportunities with profits under 0.00001% were presented since no filters were applied, but the 1% artificial increase in the GBP/EUR exchange rate was identified as an Arbitrage Opportunity (AUD → BRL → AUD → JPY → GBP → EUR → AUD) producing an approximate profit of 1%.

6 Capital Asset Pricing Model (CAPM) for β Calculation

The Capital Asset Pricing Model (CAPM) describes the relationship between systematic risk and expected return for assets, particularly stocks. CAPM is widely used throughout finance for pricing risky securities and generating expected returns for assets given the risk of those assets and cost of capital.

The formula for calculating the expected return of an asset given its risk is as follows:

$$ER_i = R_f + \beta_i(ER_m - R_f)$$

where:

ER_i = expected return of investment

R_f = risk-free rate

β_i = beta of the investment

$(ER_m - R_f)$ = market risk premium

Figure 16: Expected Return Formula

Investors expect to be compensated for risk and the time value of money. The risk-free rate in the CAPM formula accounts for the time value of money. The other components of the CAPM formula account for the investor taking on additional risk.

The β of a potential investment is a measure of how much risk the investment will add to a portfolio that looks like the market. If a stock is riskier than the market, it will have a β greater than one. If a stock has a β of less than one, the formula assumes it will reduce the risk of a portfolio.

The goal of the CAPM formula is to evaluate whether a stock is fairly valued when its risk and the time value of money are compared to its expected return.

In statistics, linear regression is a linear approach used to model the relationship between a dependent variable and one or more independent variables and this is precisely what is used here for calculating the β value as per the Equation from 16. Here, for calculating the β value of a certain company share, the monthly returns of this share are used as the independent variable and the monthly percentage variations of the index in which this share is inserted is used as the dependent variable.

For example, using five years of historical data as displayed in Figure 17 and the monthly returns of the NYSE Composite Index in the same period will produce a β value of approximately 0.65 for Pfizer Inc. (NYSE) which is approximately the exact value. The result of the produced regression can be seen in Figure 18.

Figure 17: Historical Price - PFE - Pfizer Inc. (NYSE)

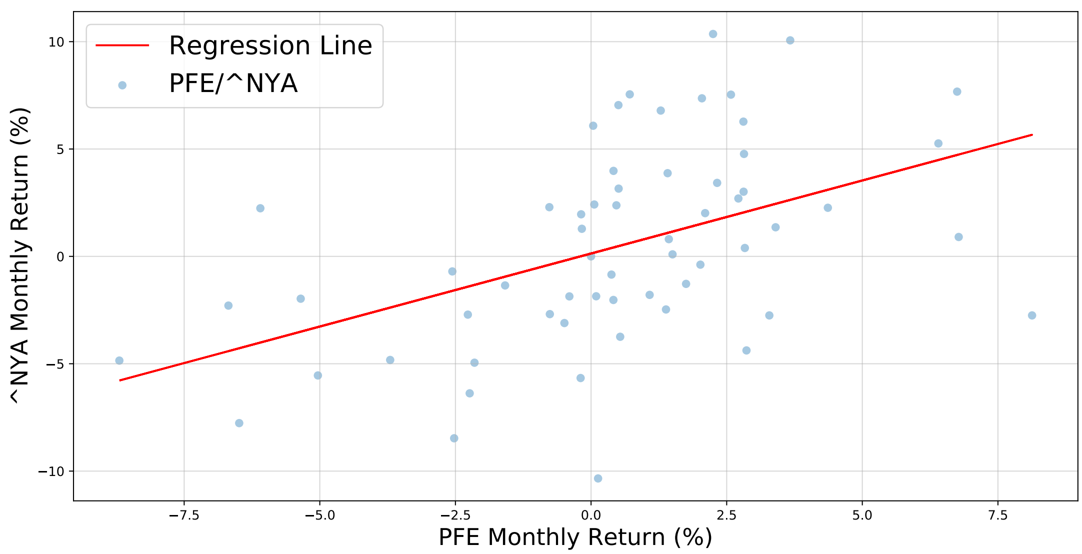

Figure 18: PFE - Pfizer Inc. (NYSE) Linear Regression

7 Portfolio Optimisation with Evolutionary Computing

In computer science and operations research, a genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimisation and search problems by relying on biologically inspired operators such as mutation, crossover and selection. The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.

In particular, in the fields of genetic programming and genetic algorithms, each design solution is commonly represented as a string of numbers (referred to as a chromosome). After each round of testing, or simulation, the idea is to delete the n worst design solutions, and to breed n new ones from the best design solutions. Each design solution, therefore, needs to be awarded a figure of merit, to indicate how close it came to meeting the overall specification, and this is generated by applying the fitness function to the test, or simulation, results obtained from that solution.

Here a GA was developed and applied to the Portfolio Allocation optimisation problem. A portfolio size of eight long-only stocks was defined and historical data was utilised in order to optimise the portfolio based on three targets:

1. Maximising Portfolio Return

2. Maximising Portfolio Sharpe Ratio

3. Minimising Volatility

These three targets formed the GA fitness function (or objective function).

Some of the optimised resulting portfolios are showm below and in Figure 19. The results were obtained based on financial data (Adjusted Closing Price) collected from the *Yahoo Finance* platform from 01/01/2018 to 20/01/2020.

- Portfolio 1:
  Return: 83.29%
  Sharp Ratio: 4.67
  Volatility: 17.39%
  JE - Just Eat plc: 21.69%
  MNG - M&G Plc: 0.39%
  BA - BAE Systems plc: 33.39%
  SDR - Schroders plc: 0.50%
  PSN - Persimmon Plc: 3.25%
  AV - Aviva plc: 0.00%

SN - Smith & Nephew plc: 15.94%
IAG - International Consolidated Airlines Group, S.A.: 24.98%

- Portfolio 2:
  Return: 77.86%
  Sharp Ratio: 4.37
  Volatility: 17.33%
  AVV - AVEVA Group plc: 21.69%
  MNG - M&G Plc: 0.10%
  BA - BAE Systems plc: 33.18%
  WTB - Whitbread PLC: 6.37%
  PSN - Persimmon Plc: 3.49%
  AV - Aviva plc: 0.00%
  SN - Smith & Nephew plc: 9.92%
  IAG - International Consolidated Airlines Group, S.A.: 24.98%

- Portfolio 3:
  Return: 77.00%
  Sharp Ratio: 4.14
  Volatility: 18.10%
  PSN - Persimmon Plc: 21.20%
  MNG - M&G Plc: 0.05%
  BA - BAE Systems plc: 33.41%
  WTB - Whitbread PLC: 6.62%
  POLY - Polymetal International plc: 3.47%
  DCC - DCC plc: 0.01%
  CRDA - Croda International Plc: 9.92%
  IAG - International Consolidated Airlines Group, S.A.: 25.08%

- Portfolio 4:
  Return: 76.73%
  Sharp Ratio: 3.68
  Volatility: 20.29%
  SGRO - SEGRO Plc: 1.01%
  AAL - Anglo American plc: 1.23%
  SGE - The Sage Group plc: 2.50%

SDR - Schroders plc: 6.52%
PSN - Persimmon Plc: 13.60%
MNG - M&G Plc: 19.30%
JE - Just Eat plc: 23.09%
IAG - International Consolidated Airlines Group, S.A.: 32.43%

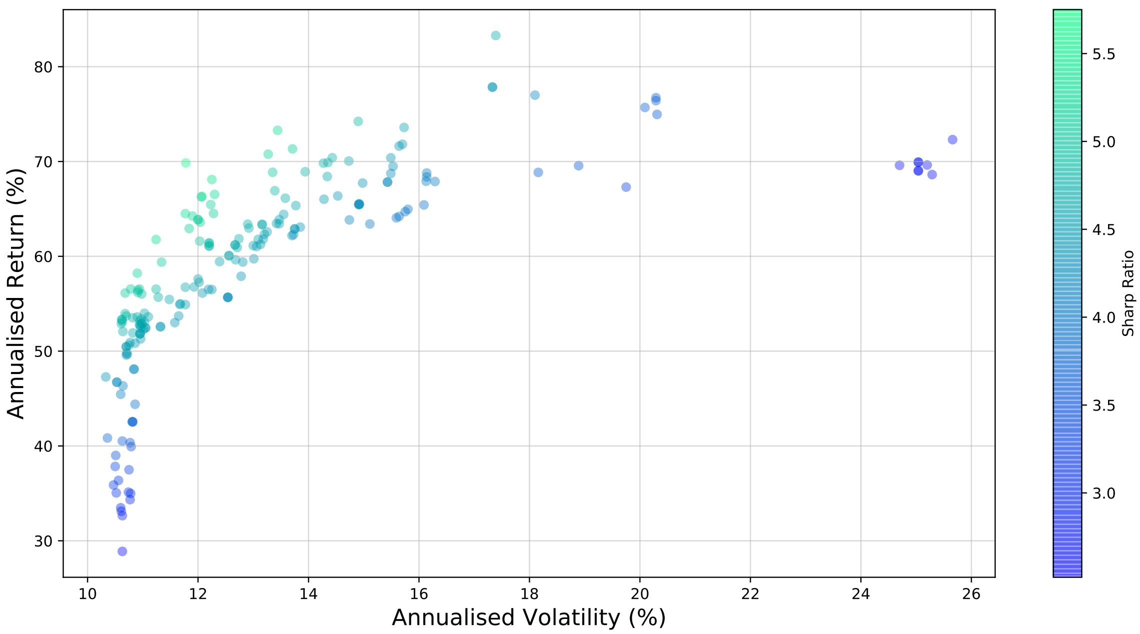

Figure 19: Generated Portfolios in the Genetic Algorithm

Finally, Figure 20 shows the evolution of the fitness function average value throughout the 10000 generations of the GA. This value is a linear combination of the Return, Sharpe value and $\frac{1}{Volatility}$. Thus, the higher the overall value, the better the portfolio performance.

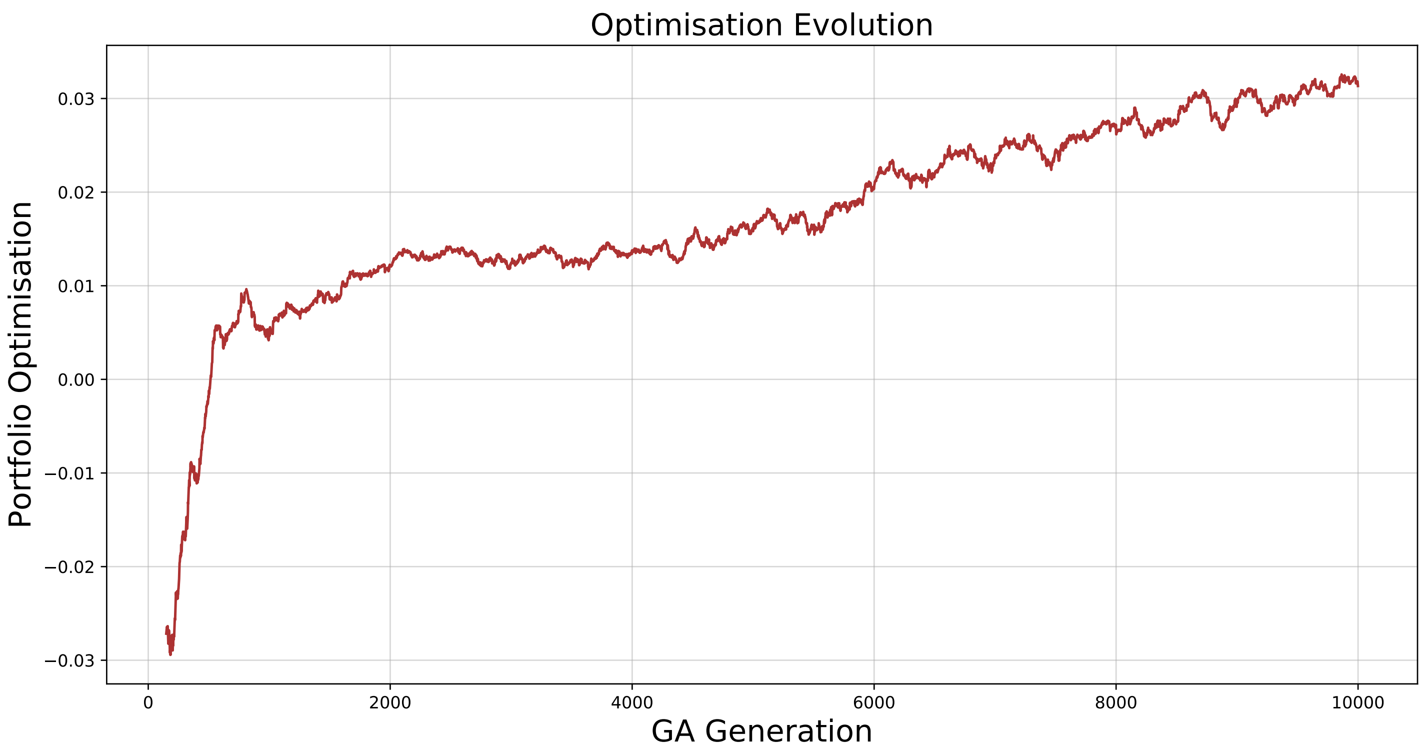

Figure 20: Evolution of the GA Fitness Function