Portfolio Optimisation

David Smith

February 2019

1 Introduction

Portfolio optimisation is a common optimisation problem in finance. It is concerned with investing a fixed amount of money into a portfolio of assets in an attempt to maximise return whilst minimising risk. A genetic algorithm is a suitable search heuristic for a multi-objective problem of this nature, where the objectives are the portfolios expected return and risk. The project consisted of implementing a genetic algorithm to optimise a portfolio based on these metrics.

2 Background

2.1 Modern Portfolio Theory

The project is based around Modern Portfolio Theory (MPT), a hypothesis about portfolio optimisation by Harry Markowitz. Markowitz theorised that risk-averse investors could construct portfolios to maximise return based on a given level of market risk. This model quantifies the two objectives, the expected return and risk, using weights which represent the portion of overall investment into each asset. The equations for risk and return can be seen below.

$$max \sum_{i=1}^{n} \mu_i w_i \quad (1)$$

$$min \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} w_i w_j \quad (2)$$

$$s.t \ \sum_{i=1}^{n} w_i = 1, 0 \leq w_i \geq 1, i = 1, 2, ..., n.$$

μ_i: expected return of asset i

w_i: portion of investment allocated to asset i

σ_{ij}: covariance between asset i and asset j

Equation 1 shows the portfolio return calculated from the sum of the products between an asset expected return and the weight of the respective relative investment. In Equation 2, the calculation for risk considers the covariance of the portfolio assets. The covariance considers how the returns of two assets vary over time, factoring in the relative investments in the assets. A high covariance indicates assets similarly affected by changes in the market.

The equations are constrained by the requirement of all weights being between zero and one with the total summing to one.

2.2 Multi-Objective Optimization and Pareto Optimality

Portfolio optimisation is a multi-objective problem which requires the objectives to be combined in some way. Pareto optimality describes a method of aggregating multiple fitness functions, each for its respective objective, into an overall metric for fitness. Search then yields a set of solutions which are non-dominated, i.e. no solution is better than any other solution in the set, in terms of overall fitness. This would form the basis of the fitness function used by the genetic algorithm.

2.3 Building a Portfolio

In order to be less exposed to changes in the market, it is industry standard to create a portfolio which contains a diverse range of stocks. There are 11 major sectors that make up the stock market. In order to create a strong portfolio, stocks from each sector were chosen. These were taken from NASDAQ and NYSE stock exchanges. The criteria was based on achieving as diverse a portfolio as possible, hence companies of various sizes were picked from every sector. These can be seen below in Table 1.

Sector	**Company**	**Stock Ticker**
Financials	JP Morgan & Chase, First Bancorp	JPM, FBNC
Utilities	El Paso Electric, Atmos Energy	EE, ATO
Consumer Discretionary	Amazon, Nike	AMZN
Consumer Staples	Hershey	HSY
Energy	British Petroleum	BP
Healthcare	Johnson & Johnson	JNJ
Industrials	Boeing, General Electric	BOE, GE
Technology	Apple, Netflix	APPL, NFLX
Telecom	Vodafone Group	VOD
Materials	Ball Corporation	BLL
Real Estate	HCP Inc	HCP

Table 1: Breakdown of chosen portfolio.

Data for each stock was downloaded from Yahoo Finance in the form of monthly returns between 2017 and 2018. The data was in the form of Extensible Time Series (XTS) objects. Once the portfolio had been created and the data in place, we could use the MPT equations to determine risk and expected return. Then, we could apply the genetic algorithm to determine the investment weightings.

3 Implementation

3.1 GA Package

The project involved use of a powerful evolutionary computation package for R called GA. The package proved highly capable of handling the problem, although population individuals had to be normalised outwith the GA library domain.

3.2 Fitness Function

The purpose of the fitness function was to provide a score for every population individual, based on the relevant objectives of risk and return. As seen in the previous section, the two objectives had opposing directions of desired optimisation. The GA package will attempt to maximise the fitness function by default, therefore the function must be created in such a way that a suitable balance between risk and return can be made. A sequence of objective weights was incorporated into the fitness function in order to vary the prioritisation given to each objective, aiming to produce Pareto optimality. The risk objective would also have to be altered to reward values tending towards zero. The fitness function can be seen below.

$$F(x) = w \cdot f_1(x) + (1-w) \cdot -(f_2(x))$$

$$f_1(x)\text{: portfolio return (1)}$$

$$f_2(x)\text{: portfolio risk (2)}$$

$$w\text{: objective weighting}$$

The portfolio risk f_2 was negated in order to reward lower risk values. The GA would be run for each objective weighting to represent a variation in which objective was prioritized.

The portfolio return $f_1(x)$ was calculated through summing the mean monthly return over the portfolio for the period. The portfolio risk $f_2(x)$ was calculated through considering the variance of the portfolio. These were calculated using the equations previously discussed.

3.3 Representation

The genetic algorithm representation consisted of populations of real-valued arrays. An individual in this population was an array of length equal to the number of assets in the portfolio. Each entry in the array represents the level of investment for its respective stock and was a real-number between 0 and 1. These limits were were established by the lower and upper bounds. These arrays had to be normalised outwith the GA package in order to be in line with the constraints stated in Section 3.

The objective weightings in the fitness function were a sequence between 0 and 1 across increments of 0.1. This meant the GA would run 11 times, which was deemed sufficient to show a variation in objective priority.

3.4 Run-Time Parameters

The crossover and mutation rates were kept constant at 0.80 and 0.05 respectively. This was in line with relevant literature on portfolio optimisation with genetic algorithms. The maximum iterations and population were kept at 1000 and 100 respectively, which provided a suitable spread of results. As the GA would be run iteratively over the objective weightings, we wanted the populations produced for each run to remain identical. Therefore a random number generator was used as an argument to replicate GA search results.

4 Results

The results of implementing the genetic algorithm can be seen below in Figures 1 - 4.

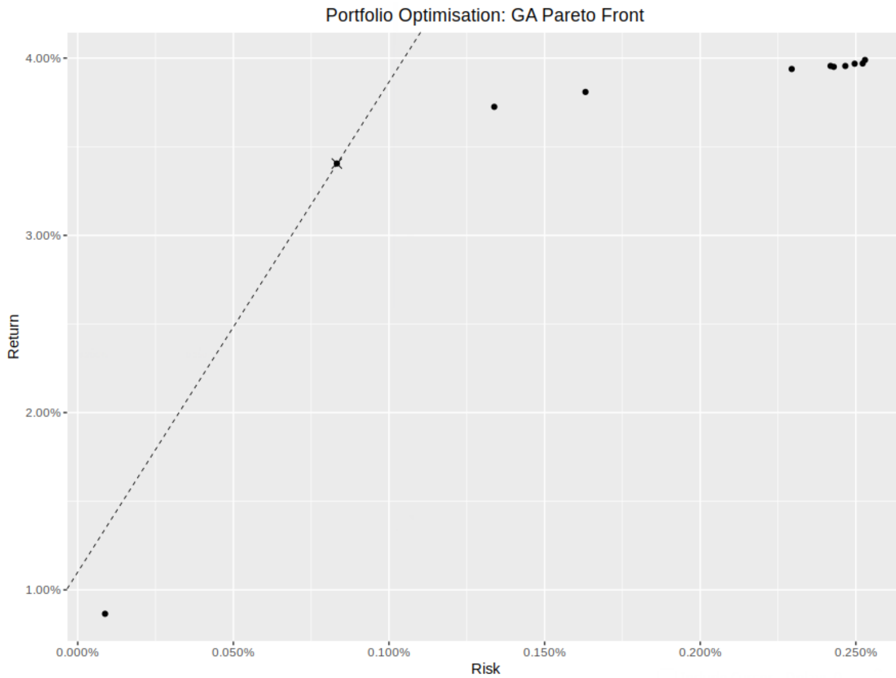


Figure 1: Genetic Algorithm (GA) Pareto front with capital allocation line. The point of intersection, marked by a cross, represents the chosen portfolio.

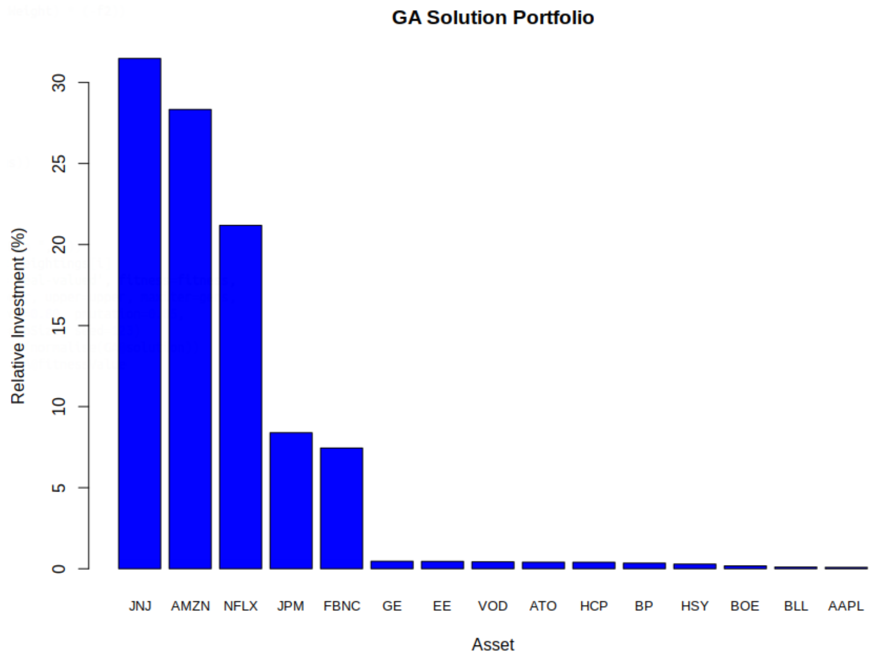
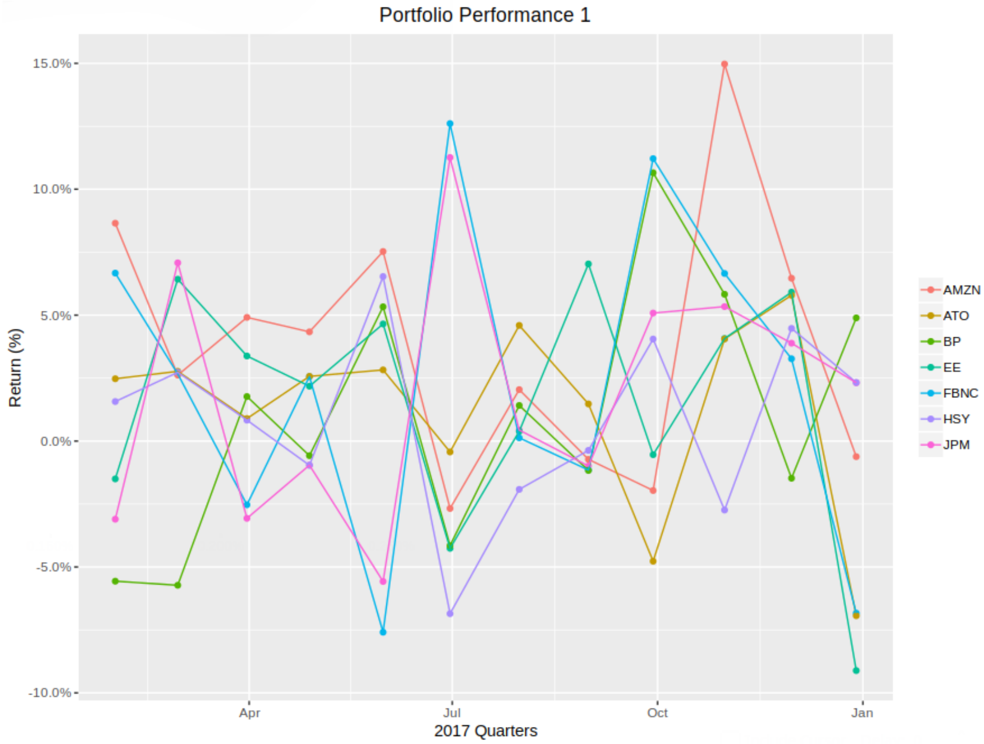


Figure 2: GA Solution portfolio breakdown highlighting main investments.

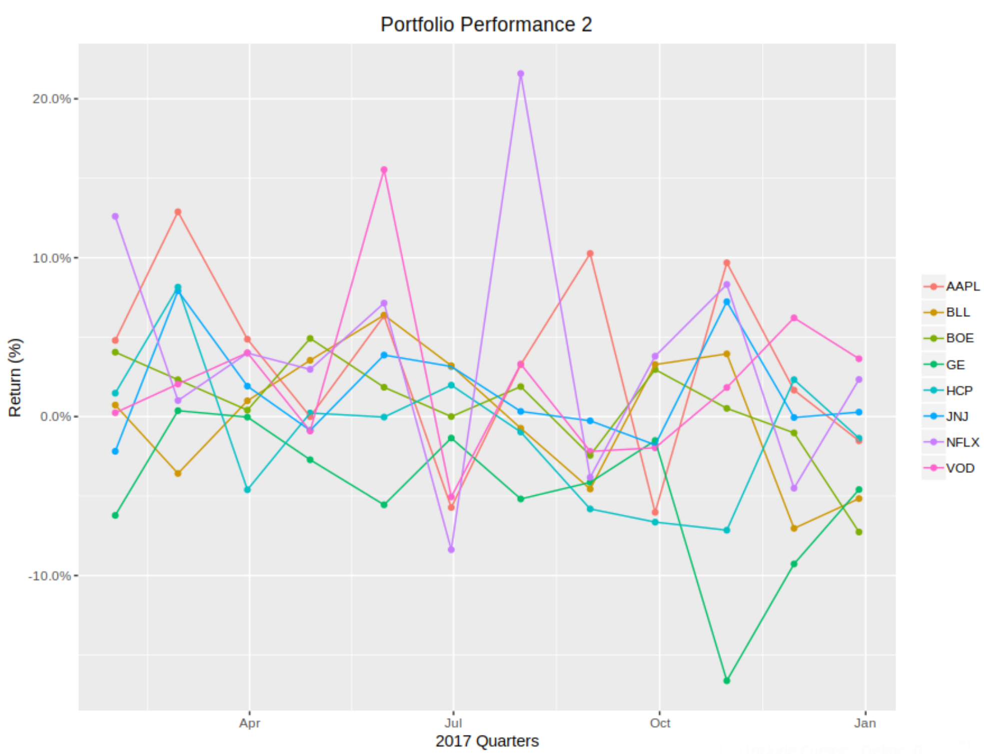
Figure 2 displays the Pareto front formed from the set of solutions produced by the GA. The dashed line represents the capital allocation line (CAL). A CAL is based on a "risk-free" asset, and is used to determine viable risk/return tolerances. A common low-risk security is a government bond, which tend to have a return of approximately 1%. The CAL therefore intersects the Pareto front where shown, helping to determine the chosen solution. The solution portfolio has an overall fitness of 0.003, based on an expected return of 3.4% and a risk of 0.83%. Overall, the fitness ranged from 0.00008 to 0.03.

Along with the CAL, there is additional rationale behind the choice of solution. It gives the lowest risk whilst providing a reasonable expected return. Choosing a solution with a greater return would incur a disproportionate increase in risk.

Figure 2 highlights the relative investments of the chosen solution. We see two assets, Johnson and Johnson (JNJ) and Amazon (AMZN), taking over 25% of the overall investment with a third, Netflix (NFLX), around 20%. The remaining sizable investments, JP Morgan Chase (JPM) and First Bancorp (FBNC), both between 5 and 10%, are the only other sizable investments. The remainder of the portfolio are under 1.5%. We can see in Figure 3 how this correlates with real-market performance over the period.



(a) Portfolio Performance over 2017



(b) Portfolio Performance over 2017

Figure 3: Stock market performance for portfolio assets from Yahoo Finance Data.

Figure 3 demonstrates the performance of the assets over the period the GA was trained on. We see large spikes in return for AMZN, FBNC, JPM, NFLX and VOD. With the exception of VOD, these increases largely occurred as the only increase in the portfolio at the time. For example, in July 2017 JPM and FBNC were the only assets to show a positive return. This indicates a low covariance for these high return stocks, making them viable investments. APPL exhibited high returns across 2017, however its behaviour was largely mirrored by the majority of the portfolio, demonstrating a high covariance. This is to be expected of a solution which is prioritising low risk. A solution prioritising return more strongly may have invested more in APPL, for example.

However, FBNC also exhibits a high covariance with certain stocks at times, notably with EE towards the end of 2017. However, the mean return for FBNC was 2%, double that of EE at 0.9%. Additionally, FBNC is the lowest of the 5 main investment allocations within the portfolio. BOE and BLL both exhibit small returns and closely track the middle of the portfolio variations, demonstrating very high covariance. Therefore the GA to allocate considerably low investment to them. This is to be expected from a solution favouring low risk.

This balance of risk and return was achieved through using weightings previously mentioned. A weighting of 0.2 achieved this solution, favouring lower returns for low risk. We can see the outcome of altering the level of prioritisation below in Figure 4.

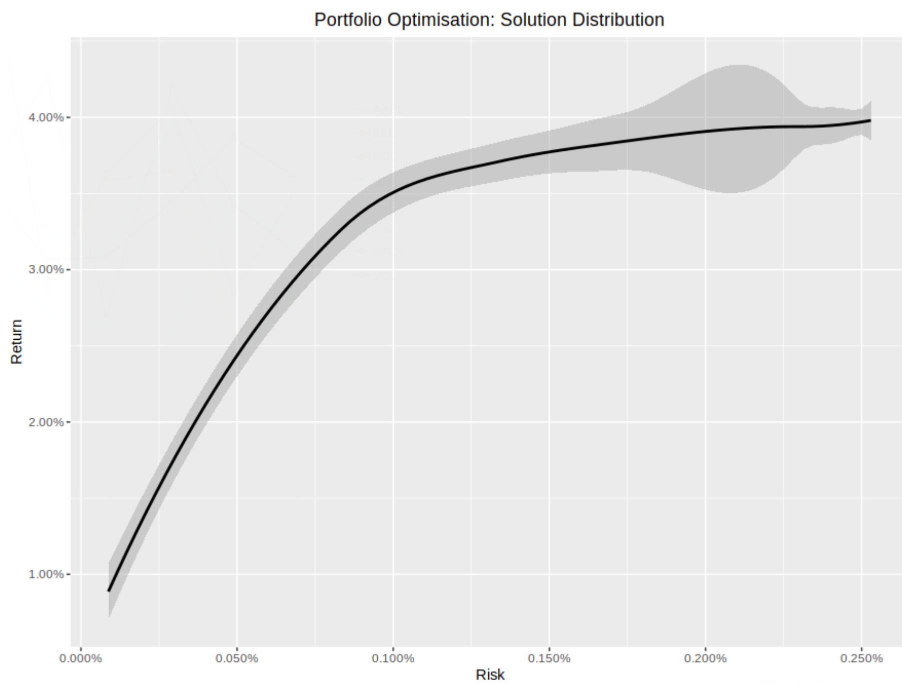


Figure 4: Distribution of solutions due to objective weight variation.

Figure 4 shows the distribution of GA solutions. The objective weight is increased moving from left to right. With a smaller weight, low risk with low returns were observed. As the weight was raised, the return and risk both increased. The solutions reach around 3.5% return and begin to plateau, with risk increasing at a greater rate. There is a high concentration of solutions around 4% return and 0.2% risk. This tells us the GA tended towards high return solutions as opposed to lower risk. Therefore the impact of altering the balance of risk and return was that solutions converged on an area of high risk and return.

5 Analysis with comparison

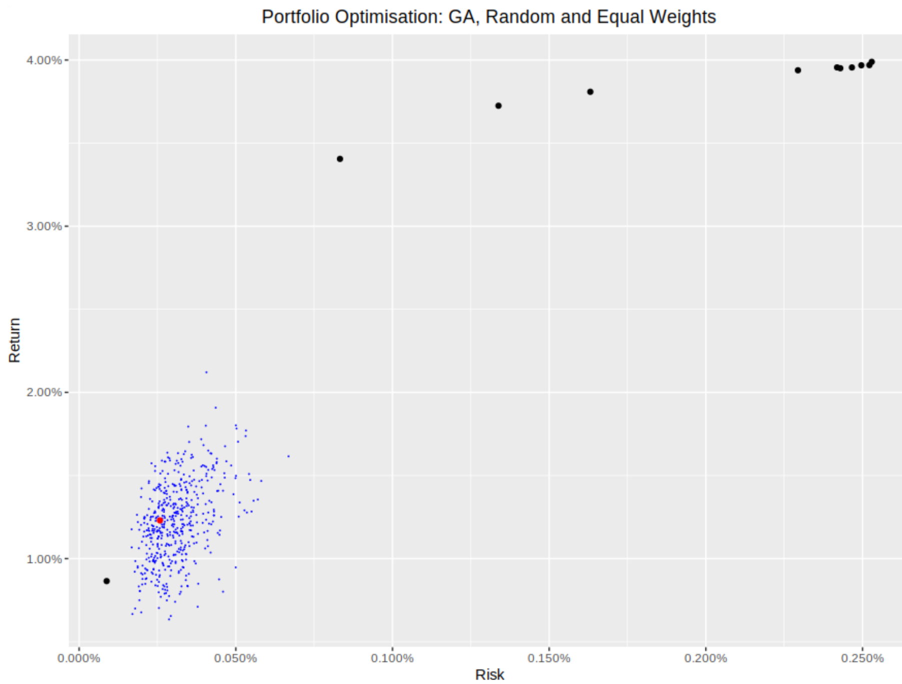


Figure 5: Genetic algorithm (black), randomly-generated solutions (blue) and an equally-distributed solution (red).

Figure 5 shows the produced solutions from the genetic algorithm compared with random and equally allocated investments. The distinct Pareto front of the GA is evident, showing a clear variation in risk and return levels. The random group has high proximity to the lower left of the front. This highlights the GAs tendency for the previously mentioned area of high risk and return in the top right hand corner. We can also see the solution of equally allocated investments for the portfolio. It is almost exactly centred within the random solutions, as to be expected from a solution showing no preference to either objective.

6 Future Performance

Data on the chosen portfolio performance over 2018 was downloaded and analysed. Using the GA solutions, we could evaluate how a selected portfolio, created around the assets designated by the GA, would actually have performed. A table showing a range of GA solutions compared with the actual portfolio performance can be seen below.

Portfolio	Expected Return	Risk
GA Low Risk	-1.20%	0.10%
GA Final	0.60%	0.50%
GA High Return	2.50%	1.10%
Actual	-0.40%	2.70%

Table 2: Performance of GA solutions, including chosen final solution, compared with actual performance.

Similar to the GA, the actual stock performance was calculated through taking the average of all mean monthly returns over 2018. The risk was calculated using a covariance matrix based on the monthly returns. Notably, the final GA solution outperforms the actual stock performance by 1%, with less than 20% of the risk. Within the GA, the spread of solutions produces expected results, with the High Return giving the greatest return but also greatest risk. The Low Risk solution gives the smallest risk but a negative return of 1.20%.