Parallelisation of Modern Portfolio theory

Challa Samarasimha Reddy- AE17B024

Guide: Prof. P Sarith Sathian

Abstract

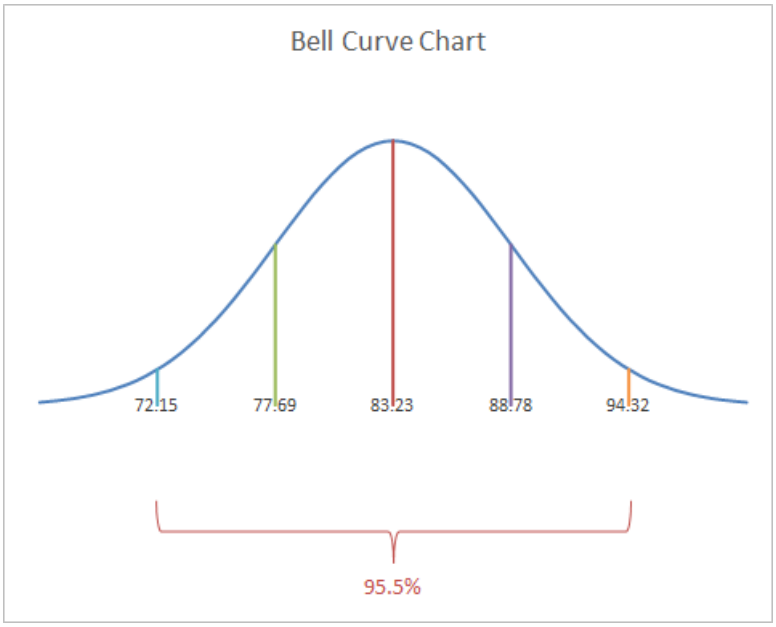
The aim of this term paper is to analyse the Modern Portfolio theory which was discovered by Dr. Harry Markowitz in 1952 from a high performance computing perspective. Modern Portfolio theory with Genetic Algorithm based serial optimization code has been chose to paralellize and to yeild better results in lower times. The paper starts with explaining the finance terms and statistics related to modern portfolio theory. The key finding of the modern portfolio theory is that to be able to get an idea of how to diversify risk in the assets that an investor is investing in. MPT uses statistics and Historical data to form an optimized portfolio.

1 Introduction

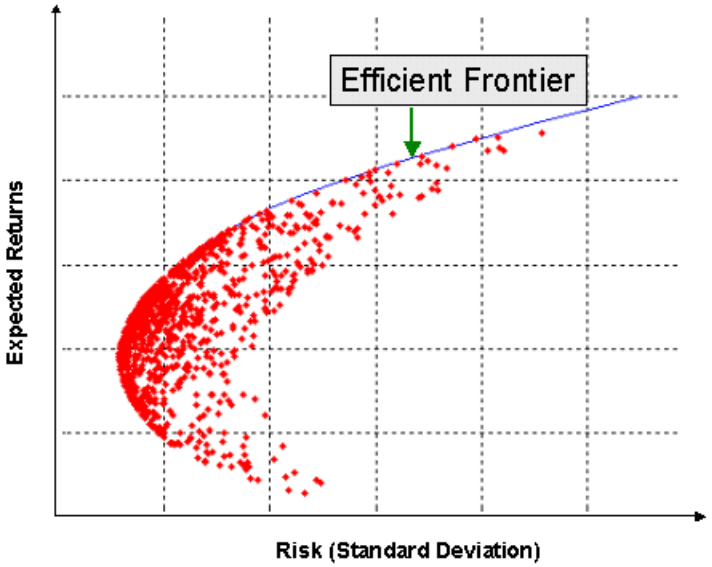
Modern Portfolio Theory is a theory on how risk-averse investors can construct portfolios to maximize expected return based on a given level of risk. It tried to formulate a standardized framework for evaluating risk and return at the time when it was absolutely random. It uses the assumption of Central Tendency Theory stating as the sample of data increases in size, the observations will increasingly concentrate around a central value. The definition of Standard deviation is that it measures magnitude of the uncertainty or volatility of a potential outcome. Risk and Volatility are similar things in the field of finance. Thus standard deviation is being used as the measure of risk in MPT. The Bell curve chart below makes us understand about the confidence intervals in the Central Tendency Theory. From it, we can say that **95.5 %** of all the values will fall under 2 S.D variation about the mean. The Theory of diversification of assets can also be commented using this theory. Correlation, which means measure of how two things move relative to one another. For example, stocks of Microsoft and Oracle are very likely to move up and down at the same time that means they are positively correlated. Risk in investment market is of two types, Diversifiable risk and Non Diversifiable risk. Non Diversifiable risk is the systematic risk inherent to the entire market or market segment. Diversifiable risk is the inherent in a specific company or industry like the airline industry now. From this logic,

Total risk of Asset A & B = Tot. Risk of A + Tot. Risk of B + Correlation of $(A\&B)$

which means by choosing assets that are negatively correlated we can actually decrease our total risk.



For an universal portfolio containing all the possible assets and if we try to form a Expected return vs Risk graph. It will look something like this and the dark line is called efficient frontier which is the optimal solution to minimize the risk for a given expected return.



2 Problem Statement Formulation

The Goal of this term paper is to obtain a set of weights using which an investor can invest in the chosen assets to form an optimized portfolio. The performance of a portfolio is being measure in terms of Sharpe Ratio. Sharpe Ratio is nothing but the average return earned in excess of the risk-free rate per unit of volatility or total risk.

$$\text{Sharpe Ratio} = (R_p - R_f)/\sigma_p$$

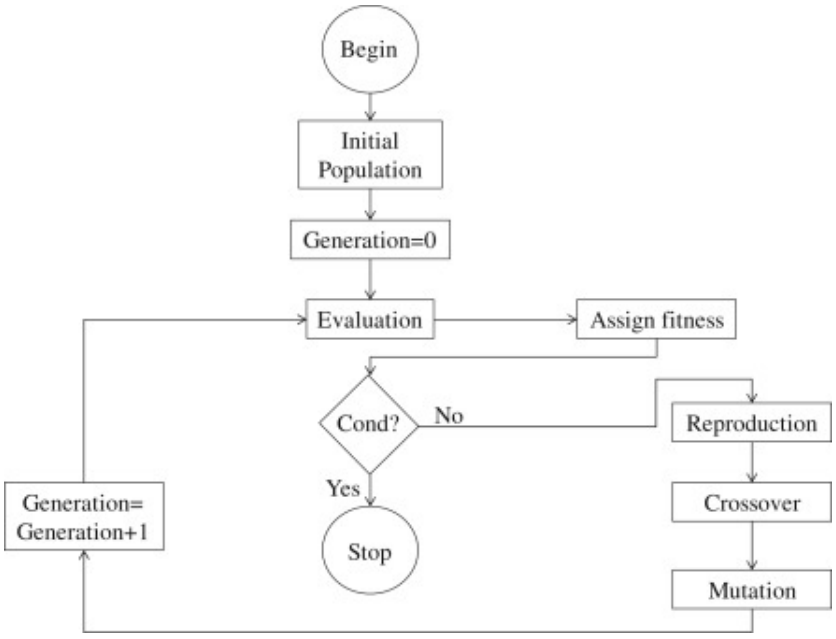
Where R_p is Return of portfolio, R_f is risk free rate, σ_p is Standard deviation of the portfolio's excess return

3 Need for Paralellization

The Optimization technique that has been chosen to use is the Genetic Algorithm technique. Genetic Algorithm is a search based optimization technique based on the principles of Genetics and natural selection. In Genetic Algorithm technique, we form a population of possible solutions to the given problem. These solutions then undergo recombination and mutation, producing new children and the process is repeated for thousands of generations. Each entry of solution is given a fitness value (For MPT, Fitness function is nothing but the Sharpe Ratio which we are trying to optimize). Fitter individuals are given a higher chance to mate and yield more fitter values. The greater the number of generations that we want to iterate, the higher the computation load. Also since we know that Modern Portfolio Theory uses historical performance of data to predict the performance in the future. The more data we have the more precise the prediction can get. The Input data that is chosen is the Adjusted closing value (Adj. closing value) of the asset each day. To tackle the above issues and to decrease the run time, it definitely needs parallelization.

4 Serial Code explanation

The flow chart for Genetic Algorithm is as shown in the figure below.



The packages used in this study are Numpy and Pandas. Pandas are used since the data is very huge and can cause issues while handling and sending

through processes. The data required here (Adjusted closing values of assets every trading day) for analysis of assets like to calculate the mean returns and co-variance matrix, has been taken from Yahoo finance and pandas data reader is used to collect the data. Here, Population is the set of possible solutions. The number of population will always remain same in Genetic Algorithm Technique. Fitness function in this case will be the Sharpe ratio which we are trying to optimize. Evolutionary principles are the most important important in GA. The better we define them, the close we can get to the optimum value soon. Evolutionary principles used in this study are

1. **Elitism**: The process of considering to have top performers in next generation also is called Elitism. For our case, a reasonable elitism rate of 0.25 has been chosen.

2. **Selection**: The process of selecting parents from the population. Parents are chosen based on the fitness value i.e, Sharpe ratio in our case. The remaining population that means other than elites and parents, will be discarded.

3. **Crossover**: The process of parents mating to form children with their qualities is called Crossover.

4. **Mutation**: The process of inducing changes in the DNA of children to not repetition happen in the process is called mutation. Mutation is very much important and cannot be neglected in the evolution process.

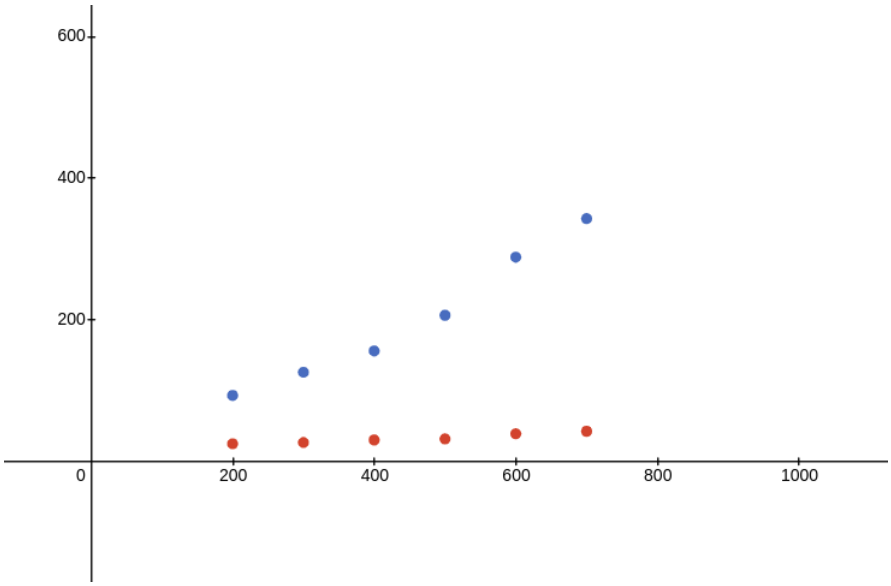
5. **Next Generation**: Finally the next generation is formed by combining elites, parents and children after mutation.

5 Performance comparison between serial and Parallel code

The following are being considered to compare the performance of serial code and parallel code.

1. **Elapsed times for serial and parallel codes for different cases of number of generations:**

   Below is the graph where the number of generations are varied along x-axis and elapsed times(in sec) are plotted for serial and parallel codes along y-axis. A standard population of 200 has been chosen to observe the variation.



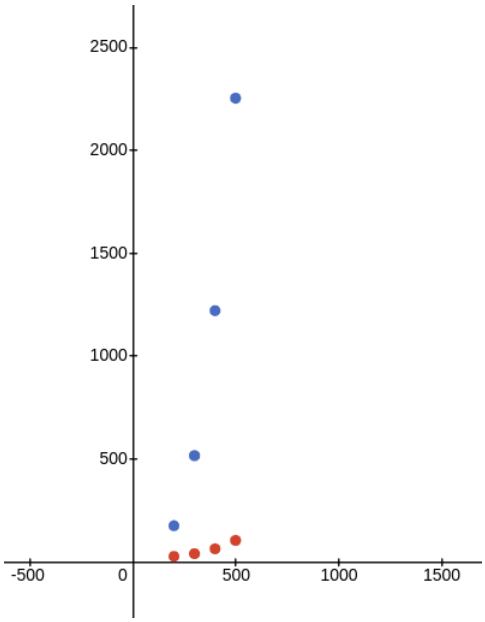
For the above graph we can also see the speed ups as shown below

S.no	Population	Generations	serial time(s)	Parallel time(s)	Speed up
1	200	200	92.546	24.222	3.820
2	200	300	125.387	26.073	4.809
3	200	400	155.358	29.567	5.254
4	200	500	205.836	31.1348	6.611
5	200	600	287.952	38.418	7.496

From the table we can see that as the number of generations are increasing, the speed is not same and thus paralellization helps very much when it comes to very large data otherwise might take hours to execute.One of the reasons for increase in the time taken is the data reading part from web.

2. **Elapsed times for serial and parallel codes for different cases of number of population:**

   The number of generations has been chosen as 400 to observe the variation of elapsed times along y-axis with varying population sizes along x-axis here. To show that in case of large sizes of populations and number of generations, the serial run time will be hugely impacted.



We can see the massive rise in computation time in case of serial code with increase in population size. To understand better,speed ups in this case are shown below which can make us understand the importance of paralellization.

S.no	Population	Generations	serial time(s)	Parallel time(s)	Speed up
1	200	400	174.408	25.3448	6.881
2	300	400	515.165	38.446	13.399
3	400	400	1219.492	62.395	19.544
4	500	400	2252.119	102.971	21.871

6 Results from Modern Portfolio Theory

The chosen assets to obtain a set of weights with an optimized Sharpe ratio are 'HDFCBANK' , 'RELIANCE' , 'INFY' , 'TCS' , 'ITC' , 'AXISBANK' (Completely random and any number can be chosen). The solution set after 400 generations with 400 populations is as shown below. The optimum Sharpe ratio is 1.248

Asset	Weightage
HDFC BANK	$8.49 * 10^{-7}$
RELIANCE	0.324
INFY	0.467
TCS	0.208
ITC	$1.605 * 10^{-7}$
AXIS BANK	$2.501 * 10^{-7}$

Disclaimer: All the research done is for educational purpose only and no type of Investment advices have been provided. MPT has a number of limitations as it is using only historical data to predict outcomes. Instead the study done can be used as a standard framework to understand correlation and diversify their portfolio.