COMPARISON BETWEEN SINGLE AND MULTI OBJECTIVE GENETIC ALGORITHM APPROACH FOR OPTIMAL STOCK PORTFOLIO SELECTION

Nejc Cvörnjek
Faculty of Mechanical Engineering, University of Maribor, Slovenia
and
Faculty of Economics and Business, University of Maribor, Slovenia
nejc.cvornjek@gmail.com

Miran Brezočnik
Laboratory of Intelligent Systems, Faculty of Mechanical Engineering
University of Maribor, Slovenia
miran.brezocnik@um.si

Timotej Jagrič
Institute for Finance and Banking, Faculty of Economics and Business
University of Maribor, Slovenia
timotej.jagric@uni-mb.si

Gregor Papa
Computer Systems Department
Jožef Stefan Institute, Ljubljana, Slovenia
and
Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
gregor.papa@ijs.si

Abstract Portfolio selection is one of the most common problem in the field of finance. Many investors would like to allocate their funds in such way that ratio between return and risk will be as high as possible. Up to today, the problem has been solved with various approaches based on genetic algorithm technique and GA has proved to be suitable. In this paper we applied two different approaches based on genetic algorithm technique in order to solve the problem. First is single objective approach and second is multi objective one (NSGA-II). Results are showing that there is no significant difference between approaches.

Keywords: Computational finance, Genetic algorithm, NSGA-II, Portfolio optimization, Portfolio selection.

1. Introduction

Few decades ago finance was just one discipline inside of economy. In synergy with other science disciplines like engineering, mathematics, statistics, risk management, and computer science, finance is expanding rapidly. Today finance is independent, heavily interdisciplinary field in science with many sub-disciplines, such as portfolio management and computational finance.

Portfolio is a collection of assets desired to achieve diversification. There are different types of assets on the market. Most known assets are stocks, bonds, derivatives, commodities, etc. Portfolio can include assets of only one type as well as assets of different types. Stock portfolio is portfolio that contains only stocks. There can be any number of stocks in portfolio. By adding stocks in portfolio idiosyncratic risk can be reduced. With portfolios containing 40 or more stocks from different industries almost half of a whole risk can be eliminated. This is called diversification. Due to market risk, entire risk can never be eliminated [1].

Managing any portfolio can be a difficult task. Main goal of portfolio management is choosing best asset on the market and allocating investors capital among these assets in such proportions that there will be a maximum return along with a minimum risk. The fact which makes problem difficult is that return and risk are conflicting. Assets with high return would often have a high risk. Risk can be measured with different metrics such as variance, semi-variance, VaR, cVaR, etc.

Portfolio selection problem (PSP) is a quadratic programming (QP) problem. However, heuristic techniques could be used in optimal PSP. Among heuristic techniques genetic algorithm (GA) are very common. Shoaf and Foster demonstrated effectiveness of GA where they proved that GA has smaller time complexity than QP [10].

Until today, problem was solved with single and multi objective GA approaches. In this paper, we applied both, single as well as multi objective approach, in order to find optimal stock portfolio and compare results to see if there is any significant difference.

Paper is organized as follows. In Section 2 a problem of stock portfolio optimization is presented. In Section 3 both used techniques are described in detail. In Section 4 we give a brief description about related work. In Sections 5 a practical problem and methodology of work are presented. We show results and discuss about them in Section 6. Last section is conclusion.

2. Problem Presentation

State of the art of today's modern portfolio theory is the mean-variance model introduced by H. Markowitz [7] in 1952. Markowitz developed his mean-variance model (M-V model) where is assumed that there is a trade-off between return and risk. M-V model includes two parameters. First is mean which stands for expected return of portfolio. Expected return is mathematically described as

$$E\left(r_p\right) = \sum_{i=1}^{n} E\left(r_i\right) w_i \tag{1}$$

where $E\left(r_p\right)$ is an expected portfolio return, $E\left(r_i\right)$ is an expected return of i-th stock in portfolio and w_i is a proportion of i-th stock in portfolio.

A second parameter is variance which stands for risk. Portfolio variance can be computed by using the equation below

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{2}$$

where σ_p^2 is portfolio variance, w_i and w_j are weights of i-th and j-th stock, and σ_{ij} is covariance between i-th and j-th stock.

There are weight constraints in portfolio optimization problem. The basic model has two constraints

$$\sum_{i=1}^{n} w_i = 1 \tag{3}$$

and

$$0 \leq w_i \leq 1 \tag{4}$$

where $i, j = 1, \ldots, N$.

We must warn that the last constraint applies only when long positions are allowed. If short selling is allowed portfolio weights can be negative.

To improve basic model other constraints could be included. Typical constraints are constraints on cardinality, floor-ceiling, transaction costs, etc. More on constraints in PSP can be found in [5].

3. Genetic Algorithm

Genetic algorithm [4] (GA) is stochastic nonlinear optimization and search technique developed by J. Holland. GA is based on principles of nature. Those principles are natural selection, reproduction and mutation.

Each organism has his own fitness, which represents organism's ability to survive. The higher is the organism's fitness, the higher is the probability that organism would be selected for reproduction and for retaining organism into the next generation.

3.1 Simple GA

Multi objective problem can be easily converted into single objective problem. Most often used methods are weighted sum method, ϵ-constrained method, etc. [2].

Simple GA starts with randomly generated population P of size N. Population is a set of organisms. Each organism represent a possible solution of the problem. Then, we assign fitness to each organism, and expose them to evolutionary operations. First operation is natural selection. With natural selection best organisms in present generation are carried into the next generation without making any changes. Next operation is reproduction. Reproduction consists of two operations: Selection of parents and crossover. Operation stars with selection. Most often used is a tournament selection. There are k participants in the tournament, and organism with best fitness is a winner, which becomes a parent. When two parents are selected, crossover can happen with some probability P_c. If crossover is happened offspring is produced. The process of reproduction is repeated until new population of offsprings with size P is created. Next operation is mutation. Mutation is a random change in genetic material of organism and it occurs with some probability P_m.

Now new generation of organism is made. Procedure is repeated until number of generation or stopping criteria is reached.

3.2 NSGA-II

With multi objective approach instead of a single solution we get a whole set of solutions. This set is called a Pareto front. Every solution in set is not worse than other solutions. In multi objective approach we implemented NSGA-II algorithm, but there exist many multi objective approaches based on GA. NSGA-II was developed by Deb et. al [3] and has been proved as an efficient algorithm for multi objective optimization, with better time efficiency than other similar approaches.

In NSGA-II, first population of parents P_0 of size N is randomly generated. This population then produces a population of offsprings O_0 also of size N. Both populations are combined into one population R_0. Then population R_0 is transferred to non-dominate sorting procedure.

Non-dominated sorting is a procedure in which a rank or level is assigned to each organism. Organisms that are not dominated by any other organism have the best rank. Thus, organisms are removed from population, and procedure is repeated until all organisms in population have their ranks. Organism one is dominated by organism two if two conditions are satisfied. First, if organism two is strictly better in at least one criteria, and second, if organism two is not worse than organism one in any criteria.

Furthermore, new population of parents P_1 is made according to organism rank. Population gets filled with organisms with the same rank. When, by adding new front in population, its size exceeds, organisms in that front are selected with crowding distance. Crowding distance is a distance between neighboring organisms. For boundary organisms is assigned as $c = \infty$, but for other organisms is computed. Organisms with bigger crowding distance are added into population P_1 until size N for population P_1 is reached.

Now new generation of offsprings can be made. Procedure is repeated until the number of generation or stopping criteria is reached.

4. Related Work

Until today, fairly large amount of research has been done. Most of research is based on M-V model.

Problem can be solved by converting it into single objective approach. There are two popular approaches. First, used in [11], parameter λ is included and it stands for risk factor. Evaluation function is

$$maximise \quad (1-\lambda)\sum_{i=1}^{n} E\left(r_i\right) w_i - \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{5}$$

$$0 \leq \lambda \leq 1 \tag{6}$$

With iterating through λ efficient frontier of portfolios could be produced. If $\lambda = 1$ risk is disinterested and portfolio with maximum return could be found. If $\lambda = 0$ global minimum portfolio could be found because return is disinterested. The second approach used in [6] is a Sharpe ratio with risk free rate ignored. Function is maximization of a ratio between return and risk

$$maximise \ \frac{\sum_{i=1}^{n} E\left(r_i\right) w_i}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij}} \tag{7}$$

and this approach we are going to use.

In MOGA approaches Mishra et al. [8] compare four elitist approaches PAES, APAES, PESA and NSGA-II. In research they evaluated their performance through three metrics of Pareto front performance, S metric, δ metric, C metric. Results were demonstrating that NSGA-II is superior against other approaches. In [9] Mishra et al. compare PESA SPEA-II and NSGA-II. Their results were showing that NSGA-II significantly outperform other two approaches. Based on that we will compare simple GA and NSGA-II.

5. Problem Definition and Methodology

According to our case, we attempt to find optimal stock portfolio with two different approaches, depending on the size of portfolio, and then compare their results in order to see if there is any significant difference. To implement M-V model we need historical data on prices of each stock in portfolio. Data we used was observed within the period from 01.01.2013 to 01.01.2014. In this case, all stocks are a part of S&P 500 stock market index. Data were obtained from [12]. Abbreviation of stocks used on market are in Table 1. For algorithms implementation we used Python 2.7.5 and Python(x, y) 2.7.5.0 environment.

Table 1. Abbreviation of stocks.

Portfolio size	*Stocks included in portfolio*
5	CAD, TIF, AXP, NOC, FRX
10	CAD, TIF, AXP, NOC, FRX, AA, CVX, KO, F, GOOG
20	CAD, TIF, AXP, NOC, FRX, AA, CVX, KO, F, GOOG, GS, JEC, KSU, MCS, NVDA, PFE, TAP, PM, GPS, MHK

Sizes of portfolios were 5, 10 and 20 stocks. According to this, cardinality constraint was ignored. In both approaches we used the same parameters and they are shown in Table 2. Parameters were selected based on multiple runs. With these values performance was the best.

Generation sizes were 100, 250, 500 and 1000 generations. Each organism was encoded as vector of weights. This is showed in Figure 1.

Table 2. Parameters used in research.

Parameter	*Simple GA*	*NSGA-II*
Population size	50	50
Natural selection	0.05	/
Tournament size	2	2
Crossover rate	0.9	0.9
Mutation size	0.2	0.2

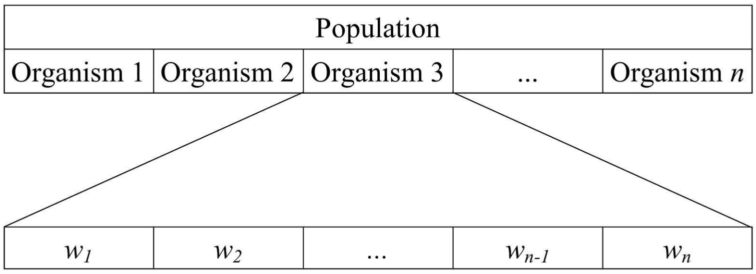

Figure 1. Organism encoding.

In reproduction process the tournament selection was used. Fitness function in single objective approach was a formula for Sharpe ratio, but risk free interest rate was ignored. Fitness function is defined as

$$maximize f_{(x)} = \frac{E(r_p)}{\sigma_p^2} \tag{8}$$

In multi objective approach we optimize

$$f_{(x)} = \begin{cases} maximize\ f_1(x) = \sum_{i=1}^{n} E(r_i)\, w_i \\ minimize\ f_2(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \end{cases} \tag{9}$$

6. Results

Results are given in Figures 2, 3, and 4. On every plot axes are labeled as Return and Variance. Return in case stand for expected return of portfolio $E(r_p)$ and Variance stand for portfolio's risk defined as σ_p^2.

In Figure 2 we include five stocks in portfolio. Four of them have positive return and one has negative return. With NSGA-II we get a Pareto front with all different and equivalent portfolios. Solution obtained with simple GA is on that front regardless to size of generations. Stocks included in five stocks portfolio are in Table 1.

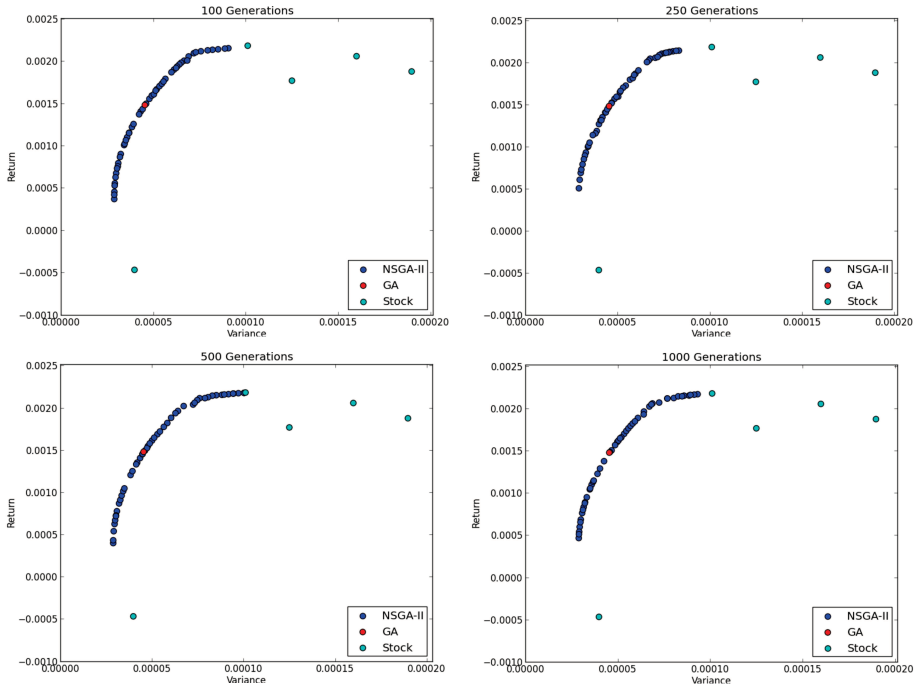

Figure 2. Results for five stocks in portfolio.

In Figure 3 we include ten stocks in portfolio. Nine of them has positive return and one with negative return and the lowest variance. Solution obtained with simple GA is again on the Pareto front in every generation size. Stocks included in ten stocks portfolio are in Table 1.

In Figure 4 we include twenty stocks in portfolio. Nineteen of them has positive return and one with negative return and the lowest variance. Again, solution obtained with simple GA is on the Pareto front regardless to generation size. Stocks included in twenty stocks portfolio are in Table 1.

We can say that results obtained with simple GA approach are comparable with results obtained with NSGA-II approach. Sometimes it even happened that a neighbor solution on the Pareto front was dominated by solution obtained with simple GA. We also made measurements of computation times for both techniques depending on portfolio size and number of generations. Simple GA needed significantly less time for its computation, regardless to number of generations or portfolio size. More details on computational are in Tables 3 and 4.

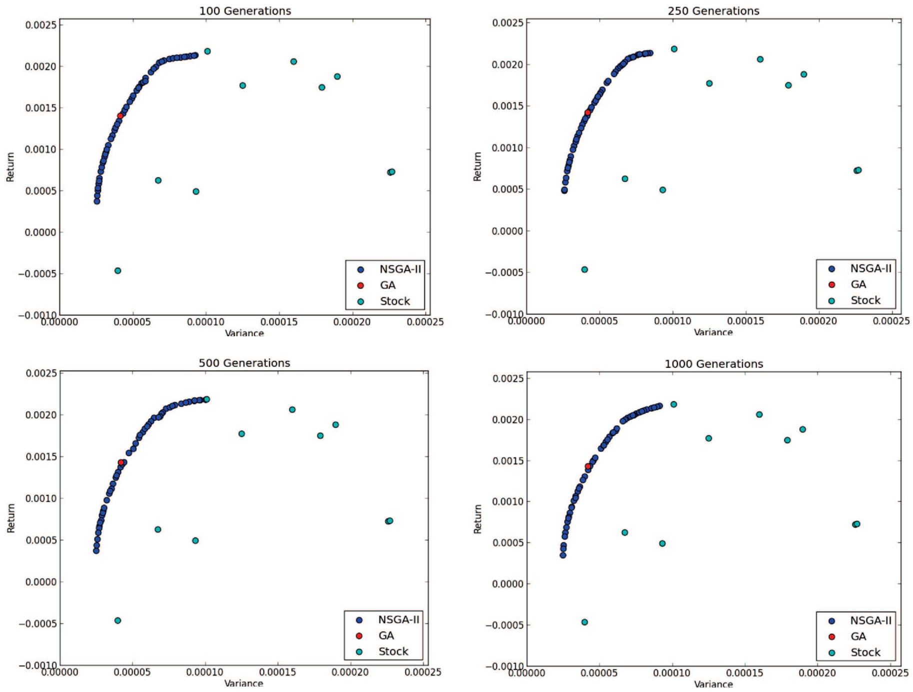

Figure 3. Results for ten stocks in portfolio.

Table 3. Computational times[a] of simple GA.

	Portfolio size		
Number of generations	*5*	*10*	*20*
100	0,62	0,7	0,83
250	1,55	1,78	2,02
500	3,25	3,43	4,04
1000	6,42	7,06	8,06

[a] All computational times are in seconds.

7. Conclusion

In this paper, we applied stock portfolio optimization problem. Purpose of portfolio optimization is to achieve highest possible return at known risk rate or vice versa. Because exact methods like QP are time complex we solve problem with GA.

We compare performance both, with simple as well as multi objective GA. In research we used simple GA and NSGA-II approach in order

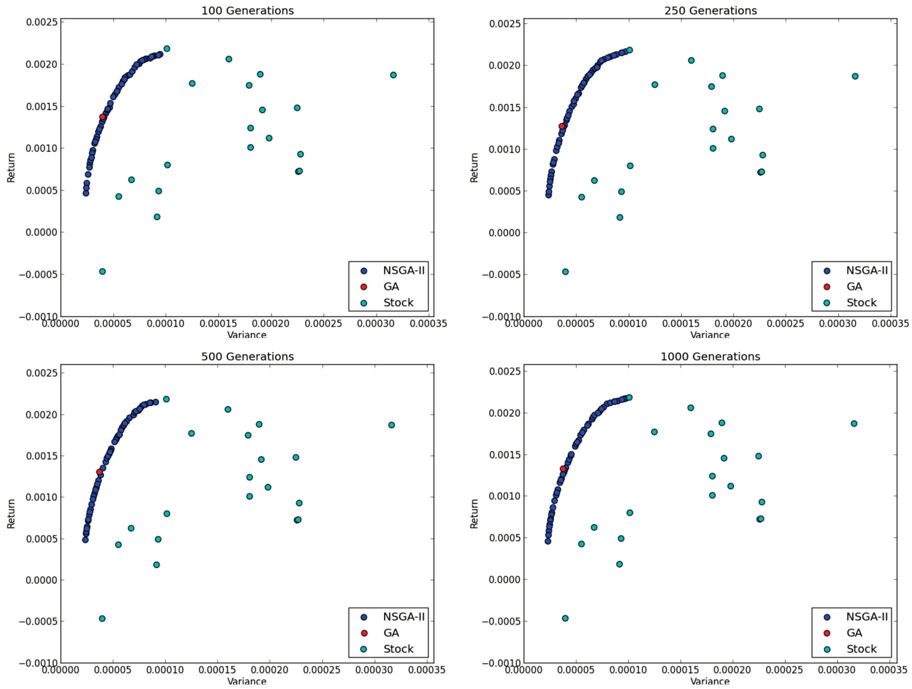

Figure 4. Results for twenty stocks in portfolio.

Table 4. Computational times[a] of NSGA-II.

	Portfolio size		
Number of generations	*5*	*10*	*20*
100	83,19	83,8	84,67
250	206,16	209,08	210,33
500	414,24	418,86	423,97
1000	827,01	841,79	857,36

[a] All computational times are in seconds.

to find optimal stock portfolio, according to Markowitz mean-variance model. Results show that even if NSGA-II is more complex algorithm, its performance was not significantly better than the performance of simple GA. On the contrary, sometimes simple GA solution dominate its neighbor on the Pareto front. Simple GA also had significantly lower computation time irrespective of portfolio size or number of generations.

There are still some open questions, like how approaches will perform with different risk metrics because in mean-variance model stocks with lower variance are favorized regardless to their returns, or how they will perform if we add some real world constraints in model. And that is a starting point for future work.

References

[1] E. Brigham and M. Ehrhardt. *Financial management theory and practice, 12th edition*, South-Western Cengage Learning, 2008

[2] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Chishester, 2001.

[3] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE T. Evolut. Comput.*, 6(2):182–197, 2002.

[4] R. L. Haupt and S. E. Haupt. *Practical Genetic Algorithm, 2nd edition*, Wiley, 2004.

[5] P. N. Kolm, R. Tütüncü, and F. Fabozzi. 60 years of portfolio optimization: Practical challanges and current trends. *Eur. J. Oper. Res.*, 234(2):356–371, 2014.

[6] C.-M. Lin and M. Gen. An effective decision-based genetic algorithm approach to multiobjective portfolio optimization problem. *Applied Mathematical Sciences*, 1(5):201–210, 2007.

[7] H. Markowitz. Portfolio selection. *J. Financ.*, 7(1):77–91, 1952.

[8] S. K. Mishra, G. Panda, S. Meher, R. Majhi, and M. Singh. Portfolio management assessment by four multiobjective optimization algorithm. In *Recent Advances in Intelligent Computational Systems (RAICS)*, 2011, pages 326–331.

[9] S. K. Mishra, G. Panda, S. Meher, and S. S. Sakhu. Optimal Weighting of Assets using a Multi-objective Evolutionary Algorithm. *Int. J. Recent Trends in Engineering*, 2(5):161–166, 2009.

[10] J. Shoaf and J. A. Foster. The efficient set GA for stock portfolios. In *Proc. IEEE World Congress on Computational Intelligence (WCCI)*, 1998, pages 354–359.

[11] Y. Xia, B. Liu, S. Wang, and K. K. Lai. A model for portfolio selection with order to expected returns. *Comput. Oper. Res.*, 27(5):409-422, 2000.

[12] `http://finance.yahoo.com/`