ON THE PERFORMANCE OF A HYBRID GENETIC ALGORITHM: APPLICATION ON THE PORTFOLIO MANAGEMENT PROBLEM

Vassilios Vassiliadis, Vassiliki Bafa, George Dounias, Management and Decision Engineering Laboratory, Department of Financial & Management Engineering, University of the Aegean, Greece
Email: v.vassiliadis@fme.aegean.gr, v.bafa@yahoo.gr g.dounias@aegean.gr , http://fidelity.fme.aegean.gr/decision/

Abstract. In this study, a hybrid intelligent scheme which combines a genetic algorithm with a numerical optimization technique is applied to a cardinality-constrained portfolio management problem. Specifically, the objective function aims at maximizing the Sortino Ratio with a constraint on tracking error volatility. What is more, results from the proposed algorithm are compared with other financial and intelligent heuristics, such as financial rule-of-thumbs and simulated annealing. In order to obtain a better insight on the hybrid's behavior, out-of-sample results are shown. The contribution of this work is twofold. Firstly, some useful conclusions regarding the performance of the proposed hybrid algorithm are drawn, based on experimental simulations. Secondly, some basic points, based on the comparison between the proposed algorithm and the benchmark heuristics, are highlighted. Finally, concerning the cardinality-constrained optimization problem, financial implications are discussed in some extent.

Keywords: Genetic Algorithm, Portfolio Optimization, financial heuristics, hybrid algorithm, evolutionary mechanisms

1. Introduction

Nowadays, the portfolio optimization problem is of crucial important, for many reasons. Firstly, there is a portion of investors who seek good investment opportunities such as investing to a number of stocks from a certain market, rather than investing to a single stock or the market as a whole. Investing to a single stock usually incurs high risk. On the other hand, investing to the stock market, considering as the portfolio defined by the market, incurs high transaction costs. Secondly, the portfolio optimization problem becomes even more complex, if multiple, and sometimes conflicting, objectives or many real-world constraints are considered. As a result, many investors aim at finding high-quality, near-optimum combinations of stocks, which satisfy certain objectives. In this point, it is worth mentioning that both the concept and notion of portfolio management problem were first introduced by Markowitz in his nominal work (Markowitz, 1952). However, in Markowitz's portfolio selection problem, the objective is to minimize the portfolio's risk, with a constraint to the portfolio's expected return. Recently, other objectives and constraints, which correspond to the current investment needs, have been of great interest.

A number of various methodologies have been applied to the portfolio optimization problem, which can be divided into two discrete phases. In phase one, a combination of assets has to be determined. In doing so, several heuristics techniques such as financial rules-of-thumb, metaheuristics algorithms such as Tabu Search and other intelligent techniques have been applied. After the combination of assets has been determined, the amount of capital invested in each of these assets (portfolio's weights) has to be computed. Traditional methodologies from statistics and mathematics (e.g. non-linear programming algorithms) have been implemented in order to calculate the portfolio's weights. However, one drawback of these techniques is that the possibility of getting trapped in local, low-quality, optimum areas is considerable. In order to overcome this obstacle, intelligent metaheuristics from the field of Artificial Intelligence (AI) may be used. AI comprises of several methodologies whose main characteristic is that each of them employs certain intelligent heuristics, mostly based on the way natural systems work and evolve, in order to solve problems from various domains such as industrial application, financial problems etc.

In this study, a hybrid genetic algorithm is applied in order to solve a complex portfolio optimization problem. More specifically, a certain type genetic algorithm is used in order to find good combination of assets, and the LM algorithm is applied so as to find optimal weights for the portfolios. The objective is to maximize the financial ratio Sortino, which takes into account both the concept of expected return and risk. Moreover, there's a constraint, which refers to the tracking error volatility of the constructed portfolios. The objective function of the problem, as well as the constraint on the tracking error volatility, is non-linear. The main aim of this study is to highlight the effectiveness of the proposed hybrid scheme both in terms of solutions' quality and computational effort required. In order to provide some useful insights regarding the performance of the genetic algorithm, some benchmark methodologies are applied to the problem at hand.

This paper is organized as follows. In section 1, some introductory comments regarding both the methodology and the application domain are presented. In section 2, some basic studies, regarding the application of genetic algorithms in the portfolio optimization problem, are shown. In section 3, several methodological issues are discussed in brief. In section 4, the mathematical formulation of the problem is presented. In section 5, results from experimental simulations are provided. Also, a detailed analysis on these results is given. Finally, in section 6, some concluding remarks and future work directions are provided.

2. Literature Review

Several studies have been conducted regarding the application of genetic algorithms in the portfolio optimization problem. In this section, a selection of standard works in the field is presented in brief. In any case, this analysis is not considered to be exhaustive, but only representative of the application of genetic algorithms in portfolio optimization.

In their study, [Branke, Scheckenbach, Stein, Deb & Schmeck 2009] proposed a hybrid scheme comprising of principles from evolutionary algorithms and the critical line algorithm (for parameter quadratic programming) with the aim of solving complex portfolio optimization problems with nonlinear constraints. The task of the MOEA is to define a set of subset in the solution space. Then, for each subset found, the critical line algorithm generates a set of optimal portfolios. Authors deal with the classical mean-variance optimization problem and their dataset comprised of Hang Seng 31, S&P 98, Nikkei 225 and S&P's 500 markets. Cardinality constraint was set to 4 and 8 assets. Parameters for the evolutionary algorithms were: population size of 250 and 30 generations. Results are promising regarding the superiority of the proposed methodology.

In their paper, [Maringer & Kelleler 2003] proposed a hybrid algorithm consisting of the Simulated Annealing algorithm and several mechanisms from Evolutionary Programming. The optimization problem was the classical mean-variance Markowitz's approach. Cardinality constraint was set to 9 and 39, for two datasets, namely the FTSE100 and DAX30 indices. In order to find the optimum solution, the population size was set to 100 and the number of generations to 750. First tests led to promising results and supported the findings for the algorithm presented in their study.

In another work, [Streichert, Ulmer & Zell 2003] dealt with the classical Markowitz portfolio optimization problem, using principles of evolutionary strategies. Specifically, two extensions to evolutionary algorithms are introduced. First, a problem specific representation of evolutionary algorithms is introduced. Second, in the previous algorithm, a local search mechanism is introduced in order to enhance the performance of the scheme. As far as the application domain is concerned, data from the Hang Seng 31 stock market are used. Cardinality was set to 2, 4 and 6 assets, while the parameters of the evolutionary scheme were: population size of 500, tournament group size 8, crossover probability 1 and mutation probability 0,01.

In their paper, [Chen & Hou 2006] applied a modified genetic algorithm, which has the ability to efficiently solve combinatorial optimization problems. The main characteristic of the algorithm is the specific combination encoding scheme and genetic operators, both of which are designed for solving combination optimization problems. This method is applied to the portfolio optimization problem with the aim of maximizing the portfolio's expected return with a built-in constraint on the portfolio's risk. Regarding the application domain, data from the Taiwan Stock Exchange were used. The parameters of the genetic algorithm were: population size of 200, crossover probability 1 and mutation 0,05. Cardinality constraint was set to 20 assets. The experimental results demonstrate the feasibility and effectiveness of the combination GA for the integer portfolio optimization problem.

[Chen, Hou, Wu & Chang 2009] applied a specific type of genetic algorithms to the investment portfolio problem. The objective of the problem is to maximize portfolio's expected return, under a constraint in the portfolio's risk. Dataset comprised of stocks from the Dow Jones Industrial Average. The portfolios were formed using 9 months of historical data and they were adjusted every 3 months. The parameters of the genetic algorithm were: population size of 200, crossover probability 1, mutation probability 0,05 and number of generations 500.

All in all, it can be seen that several types of genetic algorithms, from standard to more advanced, have been applied to the portfolio optimization problem. In most cases, the formulation of the problem refers to the classical Markowitz's mean-variance optimization. However, there have been studies, which incorporated non-linear real-world constraints. The applicability and effectiveness of genetic algorithms in the portfolio optimization problem is quite obvious.

3. Methodology

In this study, a variant of the genetic algorithm is proposed, combined with a mathematical optimization tool, for solving the portfolio management problem. The standard genetic algorithm was firstly proposed by Holland [Holland 1992]. Their main characteristics lie in the concept of evolutionary process. More particularly, as in the real world, genetic algorithms apply the mechanisms of selection, crossover and mutation in order to evolve the members of a population, through a number of generations. The ultimate goal is to reach a population of good-quality. In order to assess the quality of each member of the population, the concept of fitness value is introduced. All in all, genetic algorithms are computationally simple procedures, though powerful in their search for the optimal solution. Moreover, they are not limited by any assumptions regarding the search space, thus enabling them to apply quite capable strategies.

In practice, the genetic algorithm operates upon a number of possible solutions called population. Each member in the population is called chromosome, representing a solution to the problem. In this study, each combination of assets (portfolio), along with their corresponding weights, comprises a chromosome. Therefore, the population is a collection of the 'best-so-far' portfolios. At first, the population is randomly initialized. Then, for each generation (epoch) of the algorithm, the chromosomes evolve by combining with each other, using the crossover and mutation operators. As the number of generations increases, better quality solutions are kept in the population. However, due to the fact that the size of the solution space is vast, there is a considerable probability of stagnation, meaning that after a certain number of generations the evolution process may halt, i.e. the genetic algorithm is stuck to a local optima region. This could be a potential problem for the performance of the genetic algorithm. It is important to note, in this point, that the evaluation of each chromosome is based on the value of the objective function of the optimization problem at-hand (fitness value).

In the previous paragraph, the concepts of evolutionary operators, i.e. selection, crossover and mutation, were mentioned. Crossover and mutation are applied to already selected (good) members of the population so as to produce 'children' with better solution-characteristics (assets/weights) and therefore better fitness value. However, the main point of focus in this study lies in the mechanism of selection, which refers to the procedure of picking members of the existing population in order to produce descendants (new portfolios). Particularly, the basic aim is the assessment of how the choice of a specific selection strategy affects the performance of the generic algorithm. Three alternative selection operators are implemented: a) selection of *N-best* members from the existing population, b) application of roulette-wheel process for selection, c) application of tournament selection process. According to the first method, only the best solutions from the existing population are considered for the evolution of the process. The second method (roulette wheel) can be described as follows: a probability of selection, based on the fitness value, is assigned to each member of the existing population. The mathematical formula which calculates this probability is:

$$Prob(fitness\ value) = \frac{\frac{1}{e^{objf_i}}}{\sum_{i=1}^{n} \frac{1}{e^{objf_i}}} \qquad (1)$$

Finally, the third approach (tournament selection) works as follows: firstly, all members of the existing population are split into *n-groups*, randomly. Then, from each group, the best member is chosen.

In order to determine the weights of portfolio's assets found by the genetic algorithm, a local search non-linear programming technique, namely the Levenberg-Marquardt method, is applied. In this case, the solution space is continuous and has upper and lower limits, defined by the floor and ceiling constraints. For a given combination of assets, the aim of the Levenberg-Marquardt method is to find a vector of weights which minimizes the given objective function under certain constraints.

4. Portfolio Optimization Problem

The portfolio optimization problem deals with finding the optimal combination of assets and their corresponding weights, as well. Harry M. Markowitz, with his seminal paper [Markowitz 1952], established a new framework for the study of portfolio optimization. In his classical problem, the objective of the investor is to minimize the portfolio's risk, whereas imposing a tight constraint in portfolio's expected return.

Nowadays, more complex formulations of the portfolio optimization problem are tackled. Non-linear objectives and constraints are introduced to the classical formulation, in a way that real-world situations are reflected. The objective of the portfolio optimization problem is to maximize a financial ratio, namely the Sortino ratio [Kuhn 2006]. Sortino ratio is based on the preliminary work of Sharpe (Sharpe ratio) [Sharpe 1994], who developed a reward-to-variability ratio. The main concept was to create a criterion that takes into consideration both assets's expected return and volatility (risk). However, in recent years, investors started to adopt the concept of "bad volatility", which considers returns below a certain threshold. Sortino ratio considers only the volatility of returns, which fall below a defined threshold.

Also, in this work, there is a constraint on tracking error volatility, i.e. a measure of the deviation between the portfolio's and benchmark's returns, is imposed. This restriction refers to a passive portfolio management, namely index tracking, which aims at constructing a portfolio using assets from a (stock) market in a way that attempts to reproduce the performance of the market itself. Passive portfolio management, as a concept, is adopted by investors who believe that financial markets are efficient, i.e. it is impossible to consistently beat the market.

The formulation of the financial optimization problem is presented below:

$$Maximize\ Sortino\ Ratio = \frac{E(r_P) - r_f}{\theta_0(r_P)} \qquad \textbf{(2)}$$

s.t.

$$\sum_{i=1}^{N} w_i = 1 \qquad \textbf{(3)}$$

$$0 \le w_i \le 1 \qquad \textbf{(4)}$$

$$K = N \qquad \textbf{(5)}$$

$$\sqrt{Var(r_P - r_B)} \le H \qquad \textbf{(6)}^{1}$$

where,

$E(r_P)$, is the portfolio's expected return

r_f, is the risk-free return

$\theta_0(r_P)$, is the volatility of returns which fall below a certain threshold and equals

$$\theta_0(r_P) = \sqrt{\int_{-\infty}^{0} (r_f - r_P)^2 * f(r_P) dr_P} \qquad \textbf{(7)}$$

w_i, is the percentage of capital invested in the ith asset

K, is the maximum number of assets contained in a portfolio (cardinality constraint)

r_B, is the benchmark's daily return

H, is the upper threshold for the tracking error volatility[2]

[1] The constraint on tracking error volatility is incorporated into the objective function, using a penalty term (0.8)

[2] H equals 0,0080

$f(r_P)$, is the probability density function of the portfolio's returns. Assuming that portfolio's returns follow a normal distribution, the probability density function can be defined as: $f(r_P) = \frac{e^{\frac{-(r_P - E(r_P))^2}{2*\sigma^2}}}{\sigma * \sqrt{2*\pi}}$

5. Experimental results

In order to extract some useful conclusions regarding the performance of the hybrid scheme, a number of independent simulations have been conducted. In each simulation, the configuration settings of both the algorithm and the problem have been properly adjusted.

The dataset comprised of 93 daily returns, corresponding to the period 04/01/2010 – 29/05/2010, of 49 stocks of the FTSE/ASE40 Index. In this point, it has to be mentioned that all stocks of the Index have been taken into consideration (even those stocks corresponding to firms which have been excluded the Index). The reason for doing this is to eliminate the effect of survivorship bias[3].

In what follows, the configuration settings of both the hybrid algorithm and the portfolio management problem are presented (Table 1). As it can be seen, a range of values for the number of generations and the cardinality of the portfolio are used.

Parameters for Genetic Algorithm	
Population	200
Generations	20/30
Crossover Probability	0,90
Mutation Probability	0,35
Number of best members for selection (for *n-best* members selection)	20
Number of groups (for tournament selection)	20
Number of members in each group (for tournament selection)	10
Parameters for optimization problem	
[Floor Ceiling] – Constraint	[-1 1]
Cardinality	10/20

Table 1. Configuration settings

In table 2, the main findings of the simulations are presented. Specifically, percentiles of the distribution of the independent runs are presented. If the percentile of variable X is a in 0.95 confidence level, then there is a probability of 95% that X will get values larger than a. So, it is preferable for the percentiles to have large values, indicating a distribution that is shrunk as far right as possible. What is more, benchmark results from other heuristics are presented, in order to provide some means of comparison. The following financial heuristics have been used: constructing portfolios with a) maximum Sortino ratio, b) maximum cumulative return, c) maximum Sharpe ratio, d) maximum expected return with a penalty term in tracking error volatility and e) random choice of assets. The weights are calculated applying the LMA method.

Based on the results, the following basic points can be pointed out. First of all, the roulette wheel process yields the worst results in all cases. This mechanism lies in the random selection process in a great extend. Thus, somebody

[3] Tendency for failed companies to be excluded from performance indices mainly because they no longer exist. This effect often causes the results of the studies to skew higher because only companies which were successful enough to survive until the end of the time period of the study are included.

could state that the selection of members from the existing population is implemented through a 'pure' random process. On the other hand, it seems that the selection of *N-best* members for re-production is the best selection operator. Preliminary results indicate that although this mechanism is biased to pick good members from the population, it does not stuck to local optima regions, compared to the alternative mechanisms. Distribution of simulation results is denser and is located more to the right. Finally, we could remark that the hybrid scheme out-performs the financial rules.

Percentiles	**0,05**	**0,50**	**0,95**
Hybrid Algorithm			
Cardinality: 10			
Generations: 20			
N-best	1,9424	2,2061	2,5134
Roulette Wheel	1,3118	1,6167	1,9677
Tournament Selection	1,8482	2,1742	2,5935
Generations: 30			
N-best	1,9685	2,4310	2,6940
Roulette Wheel	1,5755	1,7760	2,1270
Tournament Selection	1,9720	2,3090	2,7317
Cardinality: 20			
Generations: 20			
N-best	2,2554	2,6970	3,3285
Roulette Wheel	1,8660	2,1488	2,4879
Tournament Selection	2,1896	2,6218	3,0311
Generations: 30			
N-best	2,4437	2,8810	3,4704
Roulette Wheel	2,0992	2,3512	2,7197
Tournament Selection	2,3770	2,7617	3,3117

Financial Rules	**Best solution**		
	Cardinality: 10	*Cardinality: 20*	
Sortino ratio	0,6272	0,8068	
Cumulative return	0,5909	0,7795	

Sharpe ratio	0,5515	0,4784	
Expected return[4]	0,7136	0,7747	
Random asset selection	0,3024	1,6182	

Table 2. Simulation results

In table 3, bootstrapping is applied to the original dataset with the aim of producing a number of different scenarios. The reason for that is to examine the performance of both the hybrid algorithm and the financial heuristics in 'unknown data', by providing statistics of the distribution of results[5] [Gilli & Winker 2008].

In order to apply the specific sampling method, we considered that the cardinality constraint defines, in a way, a unique optimization problem, i.e. results from the 10-asset problem cannot be compared directly to the 20-asset problem (where the distribution of results is more acceptable). Also, for each cardinality the 'global'-best portfolio found in the hybrid scheme is used for implementing the resampling technique. In essence, 500-scenarios for the stocks returns-series were produced based on the original dataset. By doing this, the unique characteristics of the original dataset's distribution are kept. Then, for each scenario, the best portfolio was applied and a value for the objective function was calculated. As a final result, the percentiles of the objective function's distribution are presented in each case. For the 10-asset portfolio optimization problem, the 'global'-best portfolio was found when the number of generations was set to 30 and tournament selection was applied. For the 20-asset portfolio optimization problem, the hybrid scheme yielded the 'best' results in 20 generations and when selection of *N-best* members from population was applied. Results indicate that the hybrid scheme out-performs the financial rules in 'unknown' data.

Percentiles	**0,05**		**0,50**		**0,95**	
Cardinality	*10*	*20*	*10*	*20*	*10*	*20*
Hybrid Algorithm	2,1161	2,3176	2,9547	3,7487	4,0191	6,7839
Sortino ratio	0,1754	0,3885	0,6201	0,8068	1,0506	1,2000
Cumulative return	0,1970	0,4338	0,5923	0,7820	1,0365	1,3816
Sharpe ratio	0,1840	0,2719	0,5425	0,4784	0,9809	0,7187
Expected return	0,3673	0,4086	0,7050	0,7747	1,1259	1,2258
Random asset selection	0,0152	1,0740	0,3105	1,6182	0,6277	2,1613

Table 3. Results from bootstrapping

6. Conclusions

In this study, the performance of a hybrid intelligent scheme was analyzed. More specifically, the proposed technique comprised a genetic algorithm and the LMA algorithm. This hybrid scheme was applied to a portfolio optimization problem. The GA component aimed at finding high-quality combination of assets, whereas the LMA algorithm computed the optimal weights. This paper focused on the application of alternative mechanisms for selection, which is a main component of the genetic algorithm. Three different mechanisms were applied: *N-best*

[4] Tracking error volatility constraint included

[5] By applying bootstrapping in the original dataset, the produced scenarios of returns retain the properties of the original dataset. This is a more 'fair' approach than applying the portfolios found to another dataset (there is not any forecasting component in the proposed methodology).

members, roulette wheel and tournament selection. The selection operator plays a vital role in the process of portfolio construction.

Results from experimental simulations indicate that the application of the *N*-best mechanism is acceptable, compared to the other two strategies. In order to apply the specific mechanism, the only thing to be considered is the N-best portfolios in each generation. The important matter is that no random selection component is implemented. So, there is a 'clear' bias towards the best solutions. However, this is not always the case, because in several studies the issue of getting stuck in local optima regions, when considering only the best-so-far solutions, hinders the optimization process. Another point of this study is that the hybrid intelligent scheme out-performed the financial rules implemented, in all cases. Nevertheless, in order to assess the performance of the proposed techniques in 'unknown' data, we applied a method of resampling from the original dataset so as to produce new scenarios of the stocks' returns-series. Afterwards, the best portfolios were applied to the produced data series, and the distributions of results were calculated. The hybrid scheme yielded better results, compared to the financial heuristics, in this case also.

However, in order to obtain some better insight regarding the overall performance of the hybrid scheme, the following basic future research directions are proposed. First of all, the proposed hybrid intelligent scheme should be compared with other hybrid intelligent algorithms, whose main characteristic (and advantage) is the application of good searching strategies. Complex search strategies could be able to capture any trends and patterns in the dataset. Another interesting direction is the implementation of an intelligent trading system, whose main components could be: a) a hybrid nature-inspired algorithm for portfolio optimization, b) a set of intelligent trading rules (for application in the validation phase) and c) re-balancing of the portfolio in specific (maybe pre-determined) time intervals.

References

Branke, J., Scheckenbach, B., Stein, M., Deb, K. & Schmeck, H., Schmeck, H. 2009, 'Portfolio Optimization with an Envelope-based Multi-objective Evolutionary Algorithm', European Journal of Operational Research, pp. 684-693

Chen, J. S. & Hou, J. L. 2006, 'A Combination Genetic Algorithm with Applications on Portfolio Optimization', In Advances in Applied Artificial Intelligence, Berlin, Germany, vol. 4031, pp. 197-206

Chen, J. S., Hou, J. L., Wu, S. M. & Chang, C. Y. W. 2009, 'Constructing investment strategy portfolio by combination genetic algorithms', Expert Systems with Applications, vol. 36, pp. 3824-3828

Gilli, M. & Winker, P. 2008, 'A review of heuristic optimization methods in econometrics, Swiss Finance Institute Research Paper, no. 08-12,

Holland, J. H. 1992, 'Genetic Algorithms', *Scientific American*, pp. 66-72

Jeurissen, R. 2005, 'A hybrid genetic algorithm to track the Dutch AEX-Index', Bachelor thesis, Informatics & Economics, Faculty of Economics, Erasmus University of Rotterdam

Kuhn, J. 2006, 'Optimal risk-return tradeoffs of commercial banks and the suitability of probability measures for loan portfolios', *Springer Berlin*

Maringer, D. & Kellerer, H. 2003, 'Optimization of Cardinality Constrained Portfolios with a Hybrid Local Search Algorithm', OR Spectrum, vol. 25, no. 4, pp. 481-495

Markowitz, H. 1952, 'Portfolio Selection', *The Journal of Finance*, vol. 7, no. 1, pp. 77-91

Sharpe, W. F. 1994, 'The Sharpe ratio', *Journal of Portfolio Management*, pp. 49-58

Streichert, F., Ulmer, H. & Zell, A. 2003, 'Evolutionary Algorithms and Cardinality Constrained Portfolio Optimization Problem', Selected papers of the International Conference on Operations Research, Heidelberg, pp. 3-5