❖❖ Scientific
❖❖ Research

The Constrained Mean-Semivariance Portfolio Optimization Problem with the Support of a Novel Multiobjective Evolutionary Algorithm

K. Liagkouras, K. Metaxiotis

DSS Lab., Department of Informatics, University of Piraeus, Piraeus, Greece.
Email: kliagk@unipi.gr, kmetax@unipi.gr

ABSTRACT

The paper addresses the constrained mean-semivariance portfolio optimization problem with the support of a novel multi-objective evolutionary algorithm (n-MOEA). The use of semivariance as the risk quantification measure and the real world constraints imposed to the model make the problem difficult to be solved with exact methods. Thanks to the exploratory mechanism, n-MOEA concentrates the search effort where is needed more and provides a well formed efficient frontier with the solutions spread across the whole frontier. We also provide evidence for the robustness of the produced non-dominated solutions by carrying out, out-of-sample testing during both bull and bear market conditions on FTSE-100.

1. Introduction

Portfolio optimization is the process of choosing the assets and their proportions, so that it is attained the maximum profitability for the risk undertaken. MOEAs techniques applied to the portfolio selection problem have become increasingly popular relatively recently. Not only because they provide a fast and reliable way of calculating computationally demanding financial models but also why revolutionized the financial modeling research field itself by developing innovative algorithmic approaches for solving difficult financial problems that in many cases cannot be solved with exact methods. Over the past years researchers developed several approaches for the solution of the portfolio optimization problem with the use of MOEAs. Most of the early work on MOEAs adopts the unconstrained Markowitz Mean – Variance (MV) model [2, 4]. Obviously, the practical usefulness of such a model is limited to academic purposes and cannot address the complexity and multiple real world constraints faced by portfolio managers. More recent studies, incorporate a number of constraints into the optimization model, but do not examine how these constraints affect the evolutionary search process and the efficient frontier formulation. Additionally, still the majority of the studies use variance [3] for the quantification of portfolio risk although it's well known undesirable mathematical properties [1].

Our paper proposes a bi-objective return semivariance portfolio optimization model that incorporates a number of real world constraints such as cardinality constraints, floor and ceiling constraints, non-negativity constraint and budget constraint and analyzes their effects on the efficient frontier formulation. Finally, we provide empirical evidence for the robustness of the proposed methodology by performing, out-of-sample experiments during both bull and bear market conditions on FTSE-100.

The paper is organized as follows. Section 2 provides a formal introduction of the proposed multi-objective portfolio optimization model and the various constraints applied. Section 3 provides an analytical description of the proposed MOEA model, focusing especially on the applied techniques for the efficient exploration of the search space. Finally, section 4 evaluates the robustness of the proposed algorithm by carrying out a number of out-of-sample tests during both bull and bear market conditions.

2. Problem Definition

The Portfolio optimization problem seeks to optimize two conflicting objectives, portfolio return and risk (**Table 1**) subject to a set of constraints. In multi-objective optimization, and particularly when the objectives are conflicting, there is not a single solution that can opti-

The Constrained Mean-Semivariance Portfolio Optimization Problem with the Support of a
Novel Multiobjective Evolutionary Algorithm

23

mize all objectives simultaneously. Thus, in reality we are trying to find good compromises or "trade-offs" between the different objectives. This "trade-off" between the different objectives was expressed for first time by Francis Edgeworth and generalized by Vilfredo Pareto. Nowadays, is known as Pareto optimality. *Let Ω be the search space. Consider 2 objective functions f_1, f_2 where $f_i : \Omega \to \Re$ and $\Omega \subset \Re$.*

Table 1. The Portfolio Optimization Problem

Optimize	$f(w) = (f_1(w), f_2(w))$
Maximize portfolio return	$f_1(w) = \sum_{i=1}^{m} w_i \overline{r}_i$
Minimize portfolio risk	$f_2(x) = \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j s_i s_j \rho_{ij}$

where:

Decision variables $w = (w_1, .., w_m)$ subject to $w \subset \Omega$ and $m = 100$ equal to the number of stocks in FTSE-100. Rate of return of assets $r_1, r_2, .., r_m$.

ρ_{ij} is the correlation between asset i and j and $-1 \le \rho_{ij} \le 1$.

s_i, s_j represent the semi-standard deviation of stocks returns i and j respectively. The formula is

$$S_i = \sqrt{SV_i} = \frac{1}{n} \sum_{r_t < T_R}^{n} (r_t - T_R)^2,$$

where sv_i represents the semivariance of security i. T_R represents the target return, equal to zero in this occasion and $r_t - T_R$ the difference between realized return and target return.

2.1. Constraints to the Problem

Budget constraint or summation constraint $\sum_{i=1}^{m} w_i = 1$

requires all portfolios to have non-negative weights ($0 \le w_i \le 1$ $i = 1, .., m$) that sum to 1. The non-negativity constraint indirectly implies that short selling is not allowed.

Floor and ceiling constraints $a_i \le w_i \le b_i, \forall = 1, 2, ..., n$. Where a_i = the minimum weighting that can be held of asset i ($i = 1, ..., n$), b_i = the maximum weighting that can be held of asset i($i = 1, ..., n$) and $0 \le a_i \le b_i \le 1 \forall = 1, 2, ..., n$

Cardinality constraint $C_{min} \le \sum_{i=1}^{m} q_i \le C_{max}$, where:

C_{min} = the minimum number of assets that a portfolio can hold

C_{max} = the maximum number of assets that a portfolio can hold

$$\begin{cases} q_i = 1, \text{ for } w_i > 0 \\ q_i = 0, \text{ for } w_i = 0 \end{cases}$$

2.2. Pareto Optimality Definitions

We make use of Pareto optimality framework in order to determine the solutions in this multi-objective optimization problem. In particular we use the following definitions:

Definition 1 (Pareto dominance)
Consider a maximization problem. Let x, y be two decision vectors (solutions) from Ω. Solution x dominate y (also written as $x > y$) if and only if the following conditions are fulfilled: (i) The solution x is no worse than y in all objectives $f_i(x) \ge f_i(y), \forall = 1, 2, ..., n$ (ii) The solution x is strictly better than y in at least one objective $\exists j \in \{1, 2, ..., n\} : f_j(x) \succ f_j(y)$. If any of the above conditions is violated the solution x does not dominate the solution y.

Definition 2 (Global Pareto-Optimal set)
The non-dominated set of the entire feasible search space Ω is the global Pareto-optimal set.

Definition 3 (Local Pareto-Optimal set)
Let $S \subseteq \Omega$ be a subset of the search space. All solutions which are not dominated by any vector of S are called non-dominated with respect to S. Solutions that are non-dominated with respect to S, $S \subseteq \Omega$ are called local Pareto solutions or local Pareto regions.

Definition 4 (Strong dominance)
A solution x strongly dominates a solution y, $x > y$, if solution x is strictly better than solution y in all m objectives.

Definition 5 (Weak dominance)
A solution x weakly dominates y, if $f_i(x) \ge f_i(y), \forall = 1, 2, ..., n$ and there is no $x \in \Omega$ such that $f_i(x) \succ f_i(y), i = 1, 2, ..., n$

Definition 6 (Non dominated solutions)
A solution x is called non-dominated if x is either a strong non-dominated or a weak non-dominated.

Figure 1 we provide a schematic illustration of the dominance in a bi-objective optimization (portfolio return and risk), where a higher value is better for $f_1(w) =$ "*Return objective function*" and a lower value is better for $f_2(w) =$ "*Risk objective function*"

In this example solution A_1 strongly dominates solution A_2, A_3 and B since the $f_1(w)$ and $f_2(w)$ of solution A_1 are strictly better than the $f_1(w)$ and $f_2(w)$ of solution A_2, A_3 and B respectively. Solution A and C are weakly dominated by solution A_1 where the value of one of the objectives functions of the dominated solution is equal to the corresponding value of solution A_1.

3. The Proposed Methodology

This section is dedicated to the presentation of the proposed MOEA. The n-MOEA is summarized in the following steps.
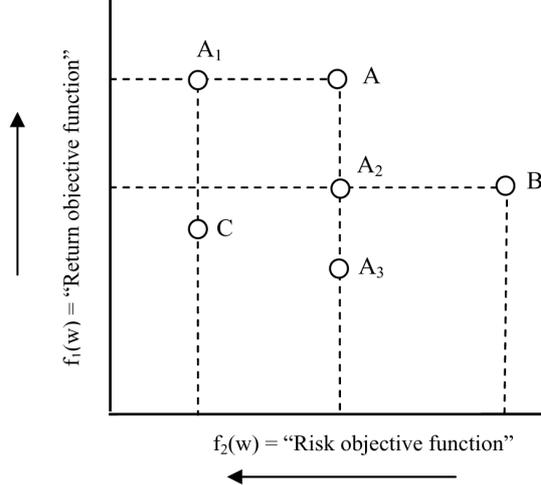
Algorithm_n-MOEA()

Figure 1. Pareto optimization in mean-risk portfolio optimization problem.

1) INITIALIZATION of the population

2) WHILE termination criterion not fulfilled

3) EVALUATE each portfolio of the population using Pareto Optimality

4) ARCHIVE add the Local non-dominated portfolios into the archive

5) EVALUATE_ARCHIVE according to Pareto Optimality conditions

5a. DOMINATED_SOLUTIONS removed from ARCHIVE

5b. NON_DOMINATED remain into the ARCHIVE as

Global Non-dominated solutions

6) REPRODUCTION_FITNESS assign reproduction fitness to the Global non-dominated portfolios according to the Euclidian distance between them.

7) CREATE_NEW_POPULATION new population is created probabilistically from the Global non-dominated solutions in the ARCHIVE according to the fitness assigned to each portfolio.

8) MUTUTION operation applied to the new population to variate the portfolios

9) CROSSOVER operation applied to the new population to preserve population diversity.

10) GO TO step 2

Below we provide analytical description of the key elements of the proposed algorithm.

3.1. Reproduction Fitness Assignment

As soon as, we identify the Global non-dominated solutions, we arrange them into the Cartesian system. Thus, for any solution into the system we know the two neighbor solutions (portfolios). Please, note that the most leftward and rightward points (E and D in this occasion) have one neighbor solution. Next, we calculate the Euclidian distance for each Global non-dominated solution (portfolio) in the Archive (see **Figure 2**).

$$where:\quad d_{AB} = \sqrt{(return_B - return_A)^2 + (risk_B - risk_A)^2}$$

$$d_{AC} = \sqrt{(return_A - return_C)^2 + (risk_A - risk_C)^2}$$

Generally, the distance between two points $x = (x_1,..., x_n)$ *and* $y = (y_1,..., y_n)$ *in Euclidean n-space is given by:*

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

We assign reproduction fitness to each Global non-dominated solution. The reproduction fitness points sum up to 1% or 100%. The reproduction fitness is assigned according to the Euclidian distance of the solution from its neighbor solutions. Thus, the higher the Euclidean distance of the solution, the higher gets the assigned reproduction fitness. The reproduction fitness assigned to the Global non-dominated Archive solutions will guide the exploratory process.

3.2. Generation of New Population

Then, probabilistically are selected the solutions (portfolios) that will be included in the new population. Obviously, the higher fitness solutions have more chances to be picked up. Depending on the size of population and the number of Global non-dominated solutions in the Archive, many copies of Solution A may be reproduced in the new population. For example if the population size is 50 and the non-dominated solutions in the Archive are 10, and assuming that solution A had much higher fitness (see **Figure 3**) than any other solution in the Archive, then we would expect solution A to be reproduced more
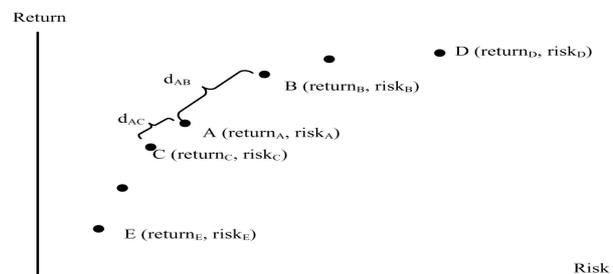

Figure 2. Calculation of the Euclidian distance between the Global non-dominated solutions for the assignment of reproduction fitness.
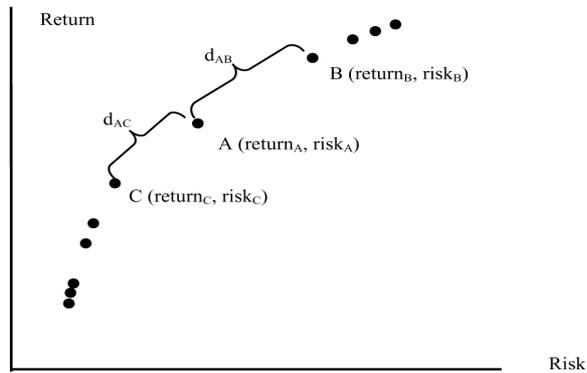

Figure 3. Global non-dominated solution A is assigned a higher reproduction fitness, due to its higher Euclidian distance from the rest non-dominated solutions.

than 5 times (average reproduction) in the new population due to the higher reproduction fitness.

3.3. The Mutation Mechanism

The number of mutants is determined by the portfolio size and the mutation rate. For determining which stocks will exist the portfolio is calculated the Inverse Fitness. The Inverse Fitness is calculated as follows:

Inverse Fitness = Maximum Stock Fitness
– Stock Fitness

where: *Maximum Stock Fitness*: *The stock with the highest fitness among the stocks on FTSE-100 and Stock Fitness*: *The fitness of each individual stock that is included in the portfolio.*

That means that the highest fitness stock of FTSE-100 has an Inverse Fitness of zero. As soon as we calculate the portfolio's Inverse Fitness, we probabilistically select the stocks to exit the portfolio. Obviously, the highest fitness stock will not exit the portfolio as has an Inverse Fitness of zero. Stocks with higher inverse fitness and thus lower fitness are more probable to be removed from the portfolio. This is an elitism mechanism that gives probabilistic advantage to the higher fitness stocks to remain into the portfolio.

Having decided which stocks will exit the portfolio, and then we have to determine which stocks will enter the portfolio as mutants. In this case all the stocks of the index are taken into account, according to their fitness. Again, cardinality constraints and lower and upper bounds are taken into account during this process. Taking stocks' fitness into account means that stocks with higher fitness are more probable to enter the portfolio as mutants.

3.4. The Crossover Mechanism

At this stage, the population already has been variated through the mutation process. In order to maintain diver-

sity in the population and avoid convergence to a single solution we make use of the crossover mechanism. The crossover between the various solutions (portfolios) is random, but within the maximum and minimum crossover rates that we have specified at an earlier stage. For example if we specify maximum crossover rate = 90% and minimum = 70%, that means that randomly will be selected a crossover rate within these limits let's say 80%. A crossover rate 80% simply means that 80% of portfolio A will be crossover with 20% of portfolio B and 20% of portfolio A will be crossover with 80% of portfolio B. Please note that the pairs to be crossover are selected randomly.

3.5. Efficient Frontier Formulation

Next, we present the process of efficient frontier formulation by quoting experimental results based on historical data of FTSE-100. For the first experiment, as a training set we use 63 daily historical observations of FTSE-100 from 30-Nov-2011 till 29-Feb-2012. The configuration of the n-MOEA for this experiment is: population size 70, maximum number of generations 200, floor constraint 5%, ceiling constraint 31%, Cardinality constraint 20 stocks, crossover rate between 70% and 90% and mutation rate 20%. Below, we provide two figures with the Efficient Frontier formulation at different stages of the algorithm execution.

Figure 4 displays the exploratory process of n-MOEA. It is evident even from **Figure 4(a)** at generation 50 that the algorithm pusses the exploratory process towards the left and upward corner of the figure where the most efficient solutions reside *i.e.* solutions that command higher return and lower risk. At generation 200 as it is evident from **Figure 4(b)** the region surrounding the efficient frontier has been turned solid grey, meaning that given the constraints imposed has been exhausted the possibility to move the efficient frontier further towards the left and upward corner of the figure.

4. Robustness of n-MOEA

In relevant literature there are different views about robustness. Some scholars define robustness as the consistency of results between different runs of the algorithm; others define robustness as the insensitivity of solution to small changes in the decision variables. We will examine the robustness of the proposed algorithm from a practical point of view. Specifically, we will test the robustness of the obtained solutions in out-of-sample environments. For this experiment we use a training set of 63 daily historical observations of FTSE-100 from 30-Nov-2011 till 29-Feb-2012. The configuration of the
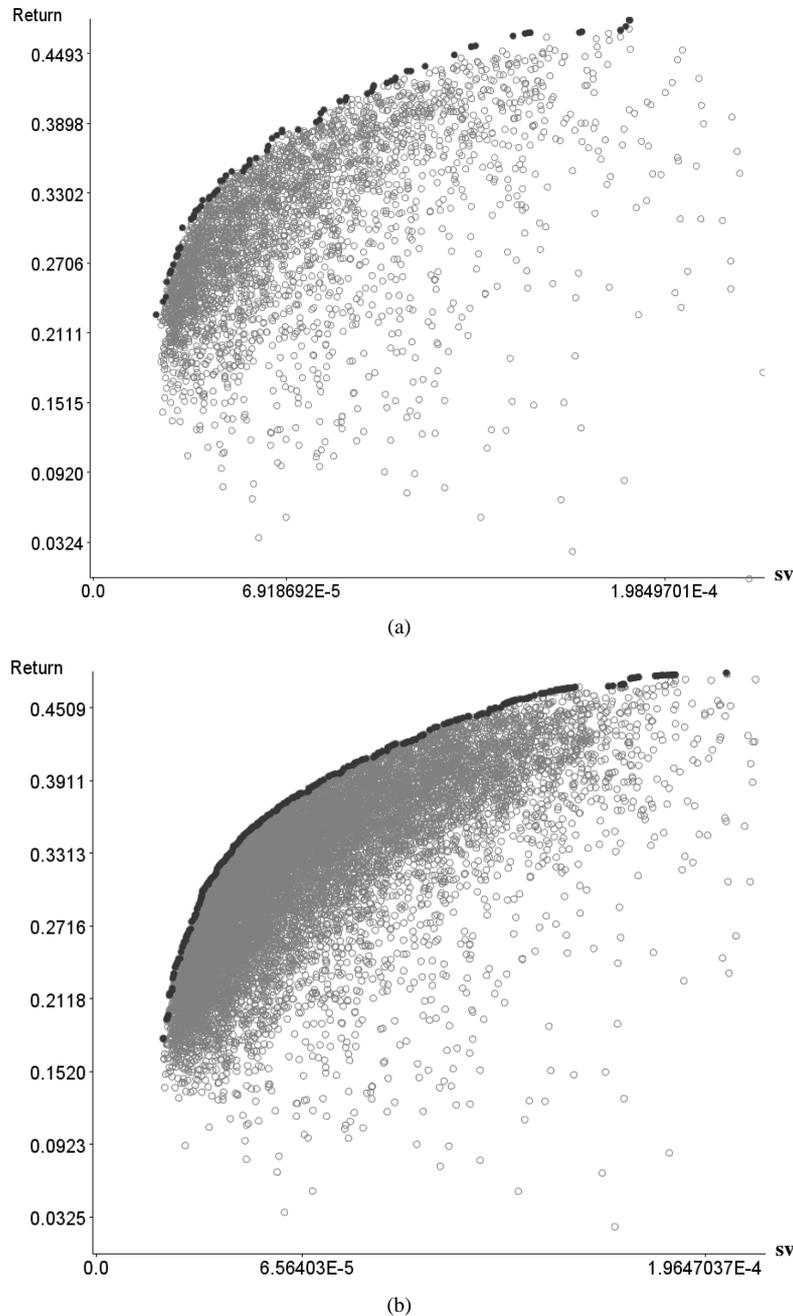
(a)

(b)

Figure 4. Exploratory process of n-MOEA.

n-MOEA for this experiment is: population size 70, maximum number of generations 200, floor constraint 5%, ceiling constraint 31%, Cardinality constraint 20 stocks, crossover rate between 70% and 90% and mutation rate 20%. The objective of this experiment is to test whether the non-dominated solutions emerged during the in-sample testing remain robust in out-of-sample testing. It is well known that portfolio managers of large funds in order to succeed satisfactory diversification and achieve better returns, keep in their portfolios stocks from every component of FTSE-100 index. However, this strategy can be followed only by big players and is associated with high administrative costs.
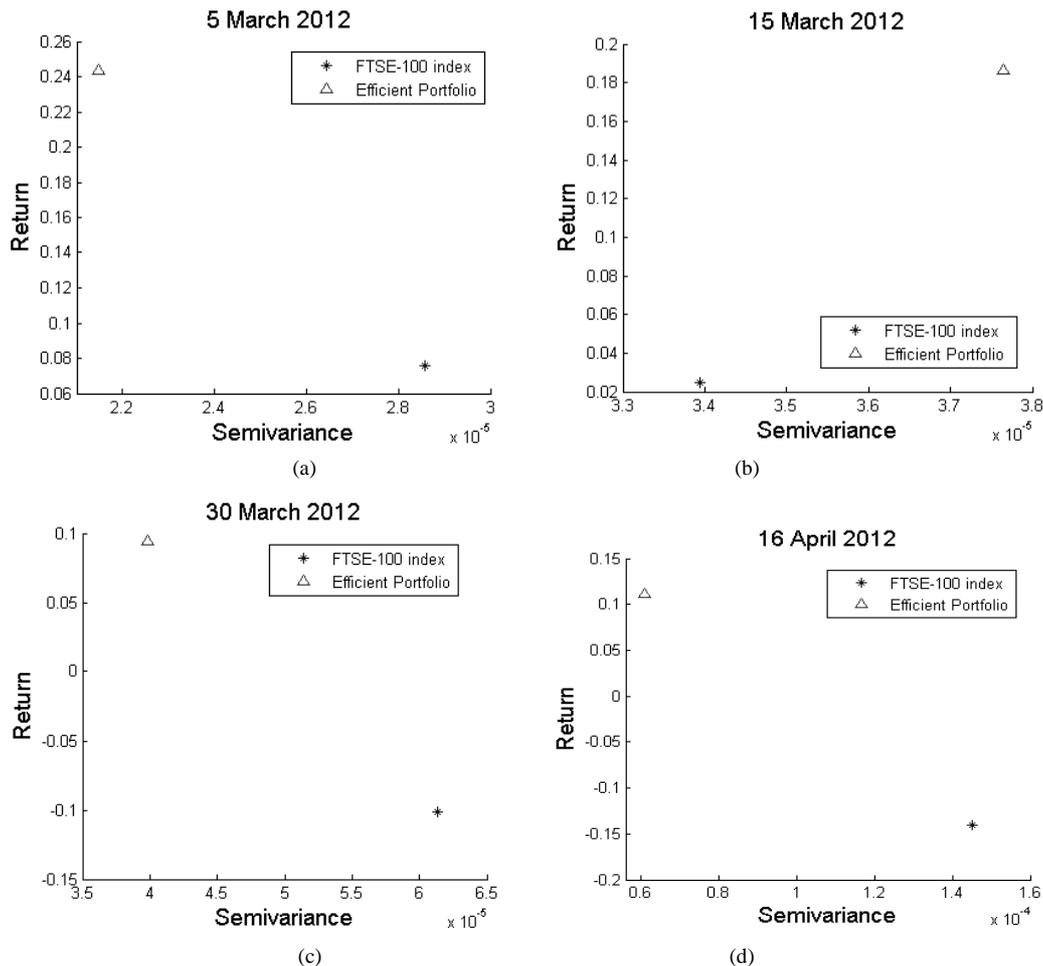
The purpose of this experiment is to find out if it is possible to outperform the FTSE-100 index performance by formulating an *efficient portfolio* that consist a small chunk of the whole index. If the experiment is successful it will mean that we are able to command combinations of higher return and lower risk than the FTSE-100 index with just a small proportion of the investment requirement and administrative costs associated with index portfolios. Please notice that during the out-of-sample

testing period of six months (March to August 2012) we do not rebalance our *efficient portfolio*. **Table 2** presents the *Efficient portfolio* and the corresponding stocks and weights as these emerged during the optimization process. For carrying out the test we separate the out-of-sample period of six months into twelve intervals. For the calculation of return and risk have been taken into account 30 observations, with ending observation the one that appear on the title of each graph of **Figure 5**. Thus, for example in **Figure 5(e)** for the calculation of return and risk of both *Efficient portfolio* and FTSE-100 index alike, are

taken into account 30 observations, with last observation the 27-April-2012 (same as the title of the graph) and first observation if we count 30 trading days backwards the 15-March-2012. From the **Figure 5** it is evident that *Efficient portfolio* strongly dominates FTSE-100 index return and risk combinations in 9 out of 12 cases. Even six months after its formulation, the *efficient portfolio*, strongly dominates FTSE- 100 index return and risk as it appears clearly in **Figure 5(k)**. In only, 3 situations (**Figures 5(b)**, **(i)** and **(l)**) *efficient portfolio* and FTSE-100 index are non-dominated which means that *efficient solution* is not strictly or weakly better than FTSE-100 index. It is worth mentioning that in none situation FTSE-100 index performance dominates *efficient portfolio* performance during this six months out-of-sample testing period. This is a strong indication of the robustness of the solution produced by the n-MOEA. Finally, we should highlight that the six months out-of-sample testing period includes both bull and bear market periods and as it is evident from **Figure 5**, the n-MOEA *Efficient portfolio* proved robust in both upward and downward market conditions.
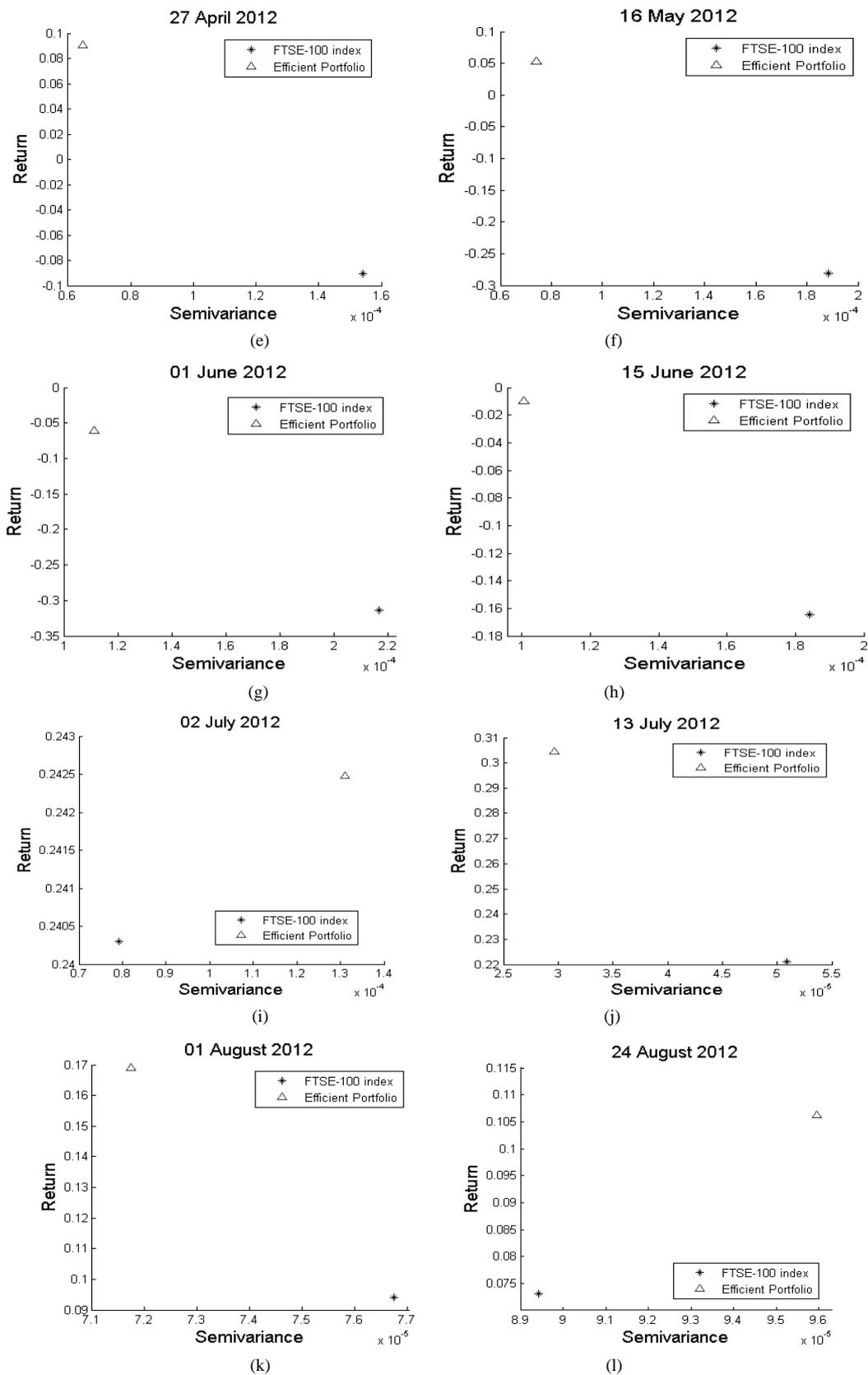
Table 2. The Efficient Portfolio and the corresponding stocks and weights

Efficient Portfolio	
stock	weight
ABF.L	29.81%
BNZL.L	30.64%
GFS.L	8.92%
SN.L	11.27%
DGE.L	19.36%

(a)

(b)

(c)

(d)

Figure 5. Out-of-sample testing of n-MOEA.

5. Conclusions

In this paper, we have proposed a novel MOEA for the solution of the constrained mean-semivariance portfolio optimization problem. The algorithm can handle successfully downside risk measures like semivariance and real world constraints and formulate well spread frontiers thanks to an elitism mechanism applied in the mutation process that give a reproduction advantage to the higher fitness stocks, and an exploratory mechanism that uses the notion that good solutions are more probable to be found near other non-dominated solutions. At the same time another mechanism calculates the Euclidian distance between the various non-dominated solutions in order to assign reproduction fitness. That way, we safeguard that the exploratory effort is concentrated where is needed more and the solutions are spread across the entire frontier. Also, the crossover process enhances the population diversity, by combining randomly two non-dominated solutions to produce two new off-springs. The robustness of the proposed algorithm is verified for short and mid-term by carrying out a number of out-of-sample tests during both bull and bear market conditions on FTSE-100.

REFERENCES

[1] P. Artzner, F. Delbaen, J. Eber and D. Heath, "Coherent measures of risk. Mathematical Finance,"*Mathematical Finance*, Vol. 9, No.3, pp.203–228. doi:10.1111/1467-9965.00068

[2] H. M. Markowitz, "Foundations of Portfolio Theory," *Journal of Finance*, Vol.46, No. 2, 1991, pp.469-477. doi:10.1111/j.1540-6261.1991.tb02669.x

[3] K. Metaxiotis and K. Liagkouras, "Multiobjective Evolutionary Algorithms for Portfolio Management: A comprehensive Literature Review," *Expert Systems with Applications 39*, Vol.39, No. 14, 2012, pp.11685–11698. doi:10.1016/j.eswa.2012.04.053

[4] F. Pasiouras, S. Tanna and C. Zopounidis, "The Impact of Banking Regulations on Bank's Cost and Profit Efficiency: Cross country Evidence," I*nternational Review of Financial Analysis*, Vol. 18, No. 5, 2009, 294 - 302. doi:10.1016/j.irfa.2009.07.003