Another Approach for Portfolio Selection using Evolutionary Programming for the Mexican Stock Exchange

CÉSAR A. COUTIÑO
Departamento de Computación
ITESM Campus Ciudad de México
Calle del Puente 222, Tlalpán, D. F.
MÉXICO

JOSÉ TORRES-JIMÉNEZ
Departamento de Computación
ITESM Campus Cuernavaca
A.P. 99-C, Cuernavaca, Morelos
MÉXICO

Abstract: **-** This paper presents an approach for portfolio selection using evolutionary programming as a tool for optimization. The goal is to find the mix of stocks that minimize risk expressed as standard deviation for a certain expected return. Two alternatives approaches are developed (Hillclimbing and Random) to measure the performance of the modified genetic algorithm.

Key-Words: **-** Evolutionary Computing, Finance, Genetic Algorithms, Optimization, Portfolio Selection, Investment Analysis.

1 Introduction

The objective behind portfolio selection is to find the best combination of stocks to get the highest level of return at a certain level of risk. The Markowitz model offers a process to attain the portfolios which are located at the efficient frontier[1][2]. They are the mix of assets that offer the highest expected return and the minimum risk measured as a standard deviation.

Markowitz's work on portfolio selection was conducted during the 50s and concluded with the publication of Portfolio Selection on 1959. The assumption behind his work as well as this approach is that investors are risk averse, which means accepting higher risk only if they get higher expected return. The previous assumption leads to the mix of selected stocks which as an average give the highest return at the lowest risk, this is called diversification.

This paper presents an approach using evolutionary programming to find the portfolio that at a certain level of expected return tries to reduce the risk associated with it. In other words, portfolio selection is presented as an optimization problem where risk is the variable to be minimized using a modified genetic algorithm[3].

This article is organized as follows: Section 2 establishes the problem as well as the mathematical formulation. Section 3 presents the evolutionary programming implementation using real data from the Mexican Stock Exchange. Section 4 demonstrates two alternative approaches (hillclimbing and random selection) to measure the performance of the proposed solution. Section 5 illustrates the conclusion and future work of this research.

2 Problem Formulation

In this section we present the mathematics of mean-variance efficient sets, which will be needed to find portfolios at the efficient frontier. The complete solution can be consulted at Campbell 1997 [4].

We will have N risky assets with mean μ and covariance matrix Ω. Assume that the expected returns of at least two assets differ and that the covariance matrix is of full rank. ω_a is defined as the $(Nx1)$ vector of portfolio weights for an arbitrary portfolio a with weights summing to unity. Portfolio a has mean return

$$\mu_a = \omega_a^{'}\mu \qquad (1)$$

and variance

$$\sigma_a^2 = \omega_a^{'}\Omega\omega_a \quad (2)$$

The covariance between any two portfolios a and b is $\omega_a^{'}\Omega\omega_b$. Portfolio p is the minimum-variance portfolio of all portfolios with mean return μ_p. The solution of $\min_{\omega} \omega'\Omega\omega$ subject to $\omega'\mu = \mu_p$ and $\omega' i = 1$ is

$$\omega_p = g + h\mu_p \quad (3)$$

where g and h are $Nx1$ vectors,

$$g = \frac{1}{D}[B(\Omega^{-1}i) - A(\Omega^{-1}\mu)] \quad (4)$$

$$h = \frac{1}{D}[C(\Omega^{-1}\mu) - A(\Omega^{-1}i)] \quad (5)$$

where:

$A = i^{-1}\Omega^{-1}\mu$

$B = \mu^{'}\Omega^{-1}\mu$

$C = i^{'}\Omega^{-1}i$

$D = BC - A^2$

i = vector of ones.

Based on the previous results, our objective is to minimize the variance subject to the number of stocks of the portfolio. Therefore, equation (2) is used as the fitness function for the three approaches.

2.1 The Evolution Algorithm Solution

In order to implement this problem we use a modified genetic algorithm (GA).

2.1.1 Representation

We use natural representation as a chromosome. The portfolio is a vector of integers where each number represents an asset (Fig. 1). The length of the vector depends on the requirements of the investor as well as the expected return.

3	23	45	35	42

Fig. 1 Portfolio representation

Due to the fact that portfolio cannot have duplicated stocks; an algorithm is implemented to prevent this problem.

2.1.2 Initial population

The initial population is created at random. For this problem we use 500 chromosomes.

2.1.3 The fitness function

The fitness function will be the variance of the portfolio represented by $\sigma_a^2 = \omega_a^{'}\Omega\omega_a$.

2.1.4 Genetic operators

Three genetic operators are used for this approach in the following order: tournaments, crossover and mutation. The tournaments are between three chromosomes and are performed within each generation.

The crossover point is randomly selected and a modified portfolio is used to prevent duplicated stocks. This operator is applied based on the crossover probability.

Finally, the mutation operator alters only one gene (stock) within the portfolio.

2.1.5 Parameters

We use the parameter listed in table 1.

Parameter	**Value**
Population Size	500
Generations	150
Prob. of mutation	0.1
Prob. of crossover	0.85

Table 1. Parameters of the GA

2.2 The Hillclimbing Solution

The hillclimbing solution is implemented based on the following algorithm:

procedure hillclimbing
begin
  t=0
  select a portfolio Pa at random
  evaluate portfolio Pa
  repeat
    generate portfolio Pb by modifying one stock of Pa at random
    evaluate portfolio Pb
    if $\sigma_{Pb} < \sigma_{Pa}$
      then Pa=Pb
  until t=MAX
end

For this procedure MAX=75,000 to have the same opportunities as the genetic algorithm.

2.3 The Random Solution

The random approach is based on the following algorithm:

procedure random
begin
  t=0
  select a portfolio Pa at random
  evaluate portfolio Pa
  repeat
    generate portfolio Pb at random
    evaluate portfolio Pb
    if $\sigma_{Pb} < \sigma_{Pa}$
      then Pa=Pb
  until t=MAX
end

where MAX=75,000.

2.3 The Mexican Stock Exchange

The Mexican Stock Exchange (La Bolsa) has about 180 companies. For this evaluation we work with 60 stocks and the data is from 300 days from September 01, 2000 to November 14, 2001. This sample was restricted to 300 days because of market limitations.

The following assumptions were taken:

- Investors are risk averse.
- Closing price is taken for each day.
- Short selling is allowed

3 The Experiments

The experiments were conducted on a Compaq Armada E500, 64MB of memory and running windows 98. On the other hand the algorithms were developed using Java.

We calculated 10 portfolios for each approach (hillclimbing, random and GA) and the number of stocks ranged from 5 to 15.

The search space for each portfolio can be seen on table 2.

Portfolio size	Search space
5	8.E+10
6	3.E+13
7	1.E+16
8	4.E+18
9	2.E+21
10	1.E+24
11	5.E+26
12	3.E+29
13	2.E+32
14	1.E+35
15	9.E+37

Table 2: Search space

The performance of the genetic algorithm can be seen on fig. 2. The average risk is drawn versus the portfolio size.

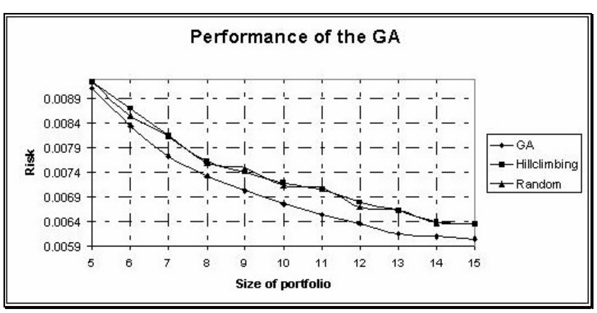

Fig. 2: Performance of the GA

These results can also be seen on table 3.

Portfolio size	GA	Hillclimbing	Random
5	0.009169	0.00952298	0.0094428
6	0.008362	0.00887185	0.00881433
7	0.007731	0.00841355	0.00848648
8	0.007408	0.00796331	0.00796775
9	0.007144	0.00765491	0.00767698
10	0.006852	0.0073886	0.00733448
11	0.006661	0.00715514	0.00719321
12	0.006454	0.00693802	0.00685832
13	0.006297	0.00674813	0.00672226
14	0.006173	0.00656254	0.00656181
15	0.006094	0.00647069	0.00640206

Table 3: Performance of the GA

In order to be able to evaluate each pair of evaluations the rate of relative difference between the GA and the Hillclimbing can be seen in fig. 3.

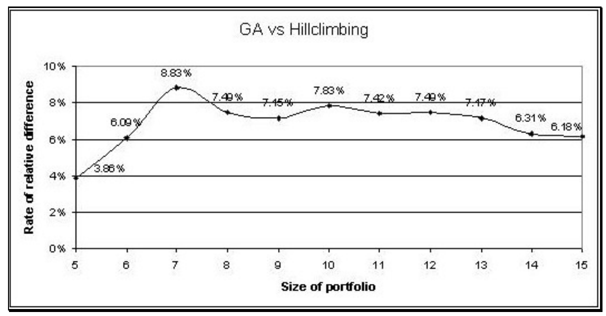

Fig. 3: Rate of relative difference between GA and Hillclimbing

The rate of relative difference between GA and Random can be seen on fig. 4.

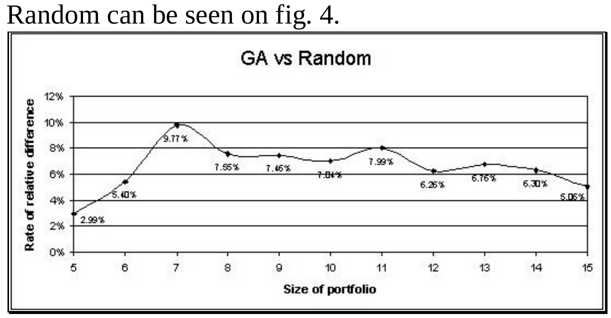

Fig. 4: Relative difference between GA and Random

Finally, the rate of relative difference can be seen on table 4.

Portfolio size	GA vs Hillclimbing	GA vs Random
5	3.86%	2.99%
6	6.09%	5.40%
7	8.83%	9.77%
8	7.49%	7.55%
9	7.15%	7.46%
10	7.83%	7.04%
11	7.42%	7.99%
12	7.49%	6.26%
13	7.17%	6.76%
14	6.31%	6.30%
15	6.18%	5.06%

Table 4: Rate of relative difference

4 Conclusions

According with the results the GA performed better compared to the Hillclimbing and Random experiments. The GA found the best portfolios between 7 and 11 stocks.

Even though, the difference between the GA and the others can be seen marginal, the result of the GA offers a superior approach to obtain the expected return with a better standard deviation.

On the other hand, the marginal difference can possible be explained to the relative stability of the Mexican Market during the past 300 days, however future work will use different windows of information.

Finally, future work will be developed using others approaches such as simulated annealing and tabu search.

5 Acknowledgements

This paper has been greatly improved by the comments of Alejandro Tavera and José Luis de la Cruz from the ITESM Campus Ciudad de México.

References:

[1] S. Baglioni et al, Evolutionary Multiperiod Asset Allocation, *Proceedings of the Genetic and Evolutionary Computation Conference*, 2000, pp. 597-604.

[2] H. Markowitz, *Portfolio Selection. Efficient Diversification of Investment*, 2nd Edition, Blackwell, 1959.

[3] Z. Michalewicz, *Genetic Algorithms+Data Structures=Evolution Programs*, Springer, 1999.

[4] J. Campbell et al., *The Econometrics of Financial Markets*, Princeton University Press, 1997.