



A GENETIC APPROACH TO PORTFOLIO SELECTION

Salvatore Arnone,¹ Andrea Loraschi,² Andrea Tettamanzi³

Abstract: This paper suggests the opportunity of taking advantage of genetic techniques in order to approach Portfolio Selection following conceptual models that handle risk in a non-trivial way.

We investigate the possibility of applying genetic techniques to this problem. It turns out that Portfolio Selection lends itself to a distributed genetic implementation.

Finally we illustrate an implementation of some of the ideas presented in this paper and give some preliminary experimental results supporting the feasibility and the advantages of the proposed approach.

Key words: *Genetic techniques, portfolio selection, risk handling*

Received: September 15, 1993

Revised and accepted: October 1, 1993

1. Introduction

Choosing an optimal portfolio weighting of assets when their future rate of return is uncertain may be seen as a problem of minimizing the uncertainty for a given level of the portfolio expected return. Modern Portfolio Theory (MPT) calls this uncertainty risk and measures it as the standard deviation or the variance of the probability distribution of future returns. In this framework the Portfolio Selection Problem (PSP) is solved using quadratic programming methods.

Other more recent approaches that require less simplistic assumptions, such as those based on the idea of "downside-risk", are not very easy to deal with using deterministic combinatorial optimization methods.

The notion of risk as the variance of returns is at the foundation of MPT. Yet, Harry Markowitz, acknowledged as one of the first developers of MPT, pointed out [9] that variance

... often leads to unsatisfactory predictions of [investor] behavior.

¹Salvatore Arnone

Etnoteam S.p.A., Via Adelaide Bono Cairoli 34 - 20127 Milano, ITALY

²Andrea Loraschi

SIGE Consulenza S.p.A., C.so Matteotti 4/6 - 20121 Milano, ITALY

³Andrea Tettamanzi

Università degli Studi di Milano, Dip. Scienze dell'Informazione,

Via Comelico 39/41 - 20135 Milano, ITALY. E-mail: tettamanzi@hermes.mc.dsi.unimi.it

Realizing that the investor perceives risk only on the downside part of the return distribution, he proposed the use of "semivariance", defined as the weighted average of the squared distances of the returns *below* the mean of the distribution, as an index of risk. In mathematical notation:

$$\text{Semivar} = \sum_{R_i < E[R]} p_i (R_i - E[R])^2$$

where R_i is one of the elements of the portfolio return distribution, p_i its attached probability and $E[R]$ the expected (mean) return. However, because of the inherent difficulties in carrying out the necessary calculations, he built his model on variance. Choosing an optimal portfolio became a question of variance minimization for a given level of expected return. Markowitz showed that as the level of risk increases, the expected return attached to optimal portfolios draws a convex non-decreasing curve (*efficient frontier*). That is, to attain a greater expected return, an investor must accept a higher exposure to risk.

Minimizing variance is equivalent to minimizing semivariance *if the return distribution is symmetric around the mean*. Distributions of returns, though, are more often skewed than symmetric: if, for example, a distribution is highly skewed to the left, variance will greatly underestimate the risk of realizing a return below the mean.

1.1. Other indices of risk

Variance has the advantage of lending itself naturally to analytical treatment, but shows serious drawbacks when used as an index of risk. As Domar and Musgrave remarked in a seminal article [6]:

Of all the possible questions which the investor may ask, the most important one, it appears to us, is concerned with the probability of a loss. This is the essence of risk.

Thus semivariance might be regarded as an alternative candidate, but its reliance on the concept of distance from the mean (the mean changes from distribution to distribution) may lead to the unwanted result of discarding an asset which should be preferred to others because it could offer *with certainty* higher realized returns even if it shows a relatively higher semivariance.

A new approach, which only recently has been partially explored by the literature in the field, defines risk as a measure related to the part of a return distribution below a certain target of return. Adapting the formula of semivariance to the concept of "subjective threshold below which the investor judges risky the investment" leads to the *target semivariance* expression:

$$\text{T.S.} = \sum_{R_i < \tau} p_i (\tau - R_i)^2$$

This measure of *downside risk* [3], [8] is usefully generalized as

$$\text{DSR}(\tau) = \sum_{R_i < \tau} p_i (\tau - R_i)^q$$

The parameter q may be seen as a coefficient of risk aversion in the sense that if

- $q = 0$, the above expression equals the probability of realizing a return below the target (*shortfall probability*);
- $q = 1$, it shows the expected value of the returns below the target (*target shortfall*);
- $q = 2$, it is equivalent to the *target semivariance*;
- $q > 2$, the formula attaches increasing weight to the returns below the target, i.e. the higher the parameter, the more the investor dislikes downside-risk.

A straightforward substitution of *DSR* in place of variance within the MPT context raises computational problems: this formula does not lend itself to analytical treatment and the classic deterministic algorithms used for optimization purposes are no longer suitable.

2. The Case for a Genetic Approach

When deterministic algorithms for optimization fall short, stochastic methods may become attractive; they are even more interesting if they can be guided by some heuristics. Genetic Algorithms (GA) [7], [5] bring together a stochastic background and a heuristic search strategy.

The well-known ability of GAs to explore enormous search spaces, together with their flexibility with respect to the problem to be solved, make a genetic approach to the problem stated above look promising and interesting.

In the following section we propose a reformulation of PSP suitable for a GA implementation.

3. Problem Reformulation

The Portfolio Selection Problem can be viewed as a convex combination parametric programming problem [12], in that we want to minimize risk while maximizing expected return; on the efficient frontier a larger expected return corresponds to a greater risk. This can be expressed as a two-objective optimization problem:

$$\begin{aligned} & \min_{\mathbf{w}} \{\text{Risk}(\mathbf{w})\} \\ & \max_{\mathbf{w}} \{\text{Return}(\mathbf{w})\} \\ \text{subject to} & \quad \sum_i w_i = 1 \\ & \quad w_i \geq 0. \end{aligned}$$

These two objectives can be parametrized to yield a convex combination parametric programming problem with objective

$$\min_{\mathbf{w}} \{\lambda \text{Risk}(\mathbf{w}) - (1 - \lambda) \text{Return}(\mathbf{w})\}$$

where parameter λ is a trade-off coefficient ranging between 0 and 1. When $\lambda = 0$ the investor disregards risk and only seeks to maximize expected return; when $\lambda = 1$ the risk alone is being minimized, whatever the expected return.

Since there is no general way to tell which particular trade-off between risk and return is to be considered the best, solving a PSP means finding a whole range of optimal portfolios for all the possible values of the trade-off coefficient; the investor will thus be able to choose the one he or she believes appropriate for his/her requirements.

4. Encoding and Operators

Three different approaches to the constraint problem in genetic algorithms have been proposed [10]:

1. using penalty functions;
2. using appropriate decoders and/or repair algorithms;
3. designing a GA with new data structures and operators to suit the problem.

The most effective strategy for handling constraints in GAs consists, as long as this is feasible, in choosing an appropriate encoding and reproduction procedure such that illegal solutions are never generated or, even better, such that any possible chromosome encodes for a legal solution (case 3 above). This relieves the algorithm from having to check new individuals for constraint satisfaction.

The encoding we propose is intuitive and has the advantage that only legal solutions can be represented: the share of each asset is encoded as an integer between 0 and a positive constant g_{max} . The actual weight of each share is computed as the ratio of its encoded value to the total sum of all the encoded values in the chromosome.

We propose a variation of uniform cross-over as a genetic cross-over operator suitable for the problem: let γ and κ be two parent chromosomes and TOT_γ and TOT_κ the total sums of all the genes that make up, respectively, γ and κ ; furthermore suppose that the i -th gene in γ , g_i has been chosen to be passed onto the offspring: then the value of the i -th gene in the offspring is:

$$h_i = \min \left(g_{max}, g_i \frac{TOT_\gamma + TOT_\kappa}{2TOT_\gamma} \right).$$

We propose a mutation operator which alters a gene either by incrementing or decrementing it by one. This ensures that a chromosome undergoing a mutation will not experience abrupt changes (mutation should not disrupt schemata).

While preserving schemata, this mutation strategy still involves a certain randomness, for it implies that a chromosome with a high total sum is less sensitive to mutations, whereas a chromosome with a small total sum feels the effect of a mutation more strongly.

5. Guidelines for a GA for PSP

The structure itself of PSP suggests the global architecture of a suitable GA: many solutions being actually needed—one for each trade-off coefficient considered. A natural way to achieve that in a genetic setting is to have several distinct populations evolve for a number of trade-off coefficient values. The greater this number is, the higher will be the resolution with which the investor will be able to explore the efficient frontier.

Because it is likely that slight variations in the trade-off coefficient do not significantly worsen a good solution, a straightforward way to speed-up the evolution process is to allow migration or cross-over between individuals belonging to adjacent or neighbouring populations, though occasional exchanges of genetic information over longer distances should not be ruled out.

The above approach lends itself quite naturally to a parallel or distributed machine implementation [11]; moreover it appears to be “scalable” with respect to the resolution required by an investor, that is the amount of detail in efficient frontier exploration.

While this population topology has often been suggested for performance purposes [13], in the PSP case it can be viewed as the most natural way to keep control over species differentiation in an evolving population where different species are explicitly requested by the problem.

On the other hand, a single population, comprising a number of species (one for each value of the trade-off coefficient), might be let evolve and select spontaneously optimal recombination strategies. This could be achieved by inserting additional genetic material into the genotype to implement *restricted mating* techniques. These additional genes would undergo genetic operators just like other genes. Likewise one could attach to each genotype other “meta-” informations such as a mutation probability or a cross-over proneness coefficient, to name just two possibilities.

The idea of giving each chromosome complete control over itself instead of using global, centralized parameter settings, leads to the design of what is called an *evolution strategy* [2], [1]. Since the investor could actually be interested in concentrating the search efforts for values of the trade-off coefficient that lead to better results (with respect to some decision criterion), each genotype of the initial population could be generated belonging to a given species with a probability proportional to the investor's interest in that species. As spontaneously selected recombination strategies do not guarantee the size of each species to stay constant during the evolution, some action should be taken in order to prevent some species from eventually disappearing. This can be achieved by allowing new genotypes to replace in the population only genotypes of their own species.

6. Applications and Experimental Results

A decision support system for portfolio selection in a downside-risk framework which implements most of the ideas presented in this paper is being developed by two of us (Loraschi and Tettamanzi) on behalf of SIGE Consulenza S.p.A. of I.M.I.,

a prominent Italian financial group. The system, named DRAGO (for Downside Risk asset Allocation by means of Genetic Optimization), is based on Whitley's GENITOR system [14], which uses a steady-state breeding strategy and elitist selection.

6.1. Implementation

One subpopulation (a species) is allocated for each value of the trade-off coefficient; individuals from a species are allowed to mate only with individuals from the same or an adjacent species, and new individuals always replace the worst individual of their own species, but under all other respects the population is treated as a single entity.

The fitness of an individual is computed as a distance from the best individual of its species, following the convention of the GENITOR system in which the most fit individual has the lowest worth, so that every individual that is best with respect to each trade-off coefficient gets the same worth of zero.

The worth is used to rank the population so that the reproduction probability for an individual is assigned on the basis of its ranking. Note therefore that these probabilities are neither proportional to the raw fitness of genotypes, nor to their worth. As noted in [4], since this results in a reduced selection pressure it seems to prevent premature convergence. It should also be pointed out, for what concerns this topic, that experiments have given strong evidence that DRAGO performs best (in speed as well as in accuracy) when reproduction probability is made uniform over the whole population. In this case all the selection pressure derives from discarding the worst individuals at each generation.

The user can specify the risk aversion coefficient q that is to be used for computing the downside risk function, the number of different values for the trade-off coefficient and the investment time horizon over which risk and return are to be estimated.

The downside risk can be computed either against a benchmark or against a fixed threshold supplied by the user; optionally, an upper constraint can be placed on the number of assets that are to make up the final portfolios.

6.2. Experiments

We tested DRAGO on a small PSP with 15 assets plus a benchmark, all expressed in Italian Lire, using monthly historical series from January 1985 to March 1993 (amounting to 98 monthly returns for each asset and the benchmark) to estimate expected returns and risks with an investment horizon of one month.

Fig. 1. shows four snapshots of the evolution of a population of 200 individuals over 10000 "generations" (here generation means the birth of a new individual and the death of the worst one). The population is segmented into 20 species, the mutation rate is set to 0.5 and the risk aversion coefficient q is set to 2 (Target Semivariance is the measure of risk).

Note that the space of risk and return is not the original problem space (which in this case has 15 dimensions), but only a convenient space to view solutions; thus, although the initial population is uniformly distributed over the original problem space, it does not look so in Fig. 1.

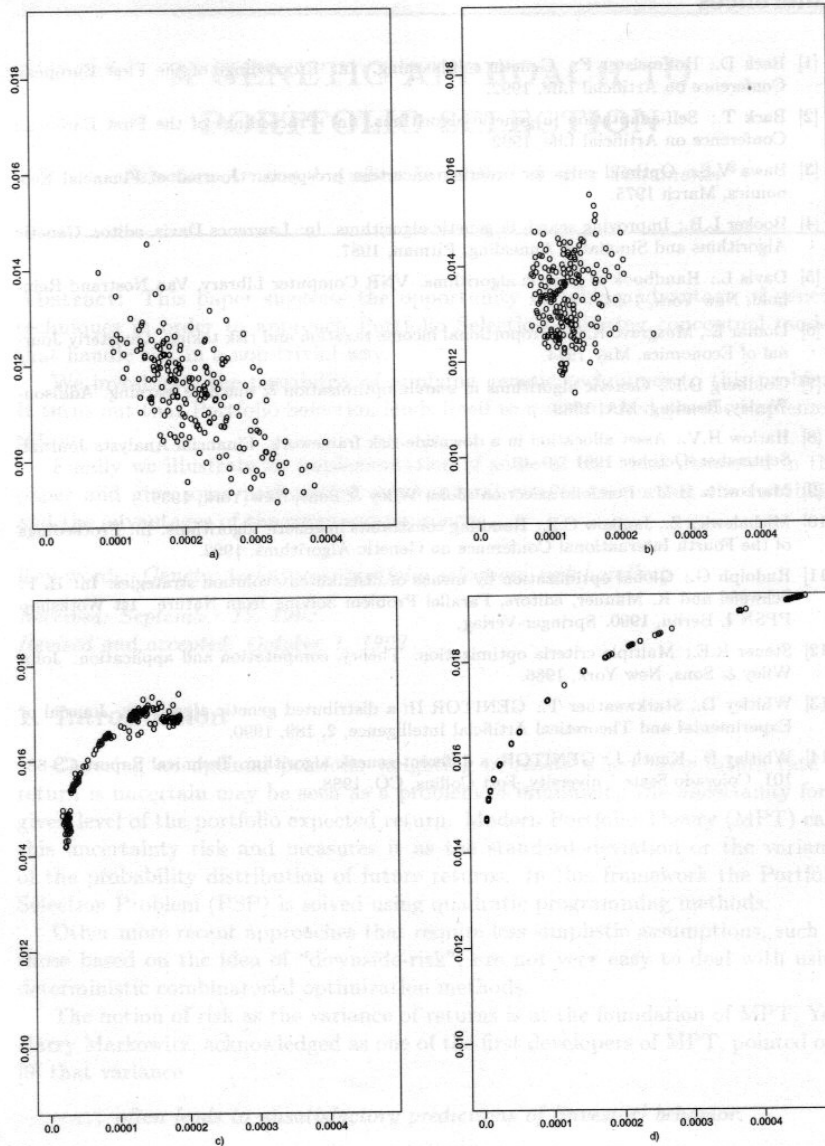


Fig. 1. Evolution of a population of size 200: a) initial population; b) after 300 generations; c) after 2000 generations; d) after 10000 generations (final population). Individuals are plotted according to their downside risk (x-axis) and return (y-axis).

References

- [1] Back D., Hoffmeister F.: Genetic self-learning. In: Proceedings of the First European Conference on Artificial Life, 1992.
- [2] Back T.: Self-adaptation in genetic algorithms. In: Proceedings of the First European Conference on Artificial Life, 1992.
- [3] Bawa V.S.: Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, March 1975.
- [4] Booker L.B.: Improving search in genetic algorithms. In: Lawrence Davis, editor, *Genetic Algorithms and Simulated Annealing*, Pitman, 1987.
- [5] Davis L.: *Handbook of genetic algorithms*. VNR Computer Library, Van Nostrand Reinhold, New York, 1991.
- [6] Domar E., Musgrave R.A.: Proportional income taxation and risk taking. *Quarterly Journal of Economics*, May 1944.
- [7] Goldberg D.E.: *Genetic algorithms in search, optimization & machine learning*. Addison-Wesley, Reading, MA, 1989.
- [8] Harlow H.V.: Asset allocation in a downside-risk framework. *Financial Analysts Journal*, September/October 1991, 30-40.
- [9] Markowitz H.M.: *Portfolio selection*. John Wiley & Sons, New York, 1959.
- [10] Michalewicz Z., Janikow C.Z.: Handling constraints in genetic algorithms. In: Proceedings of the Fourth International Conference on Genetic Algorithms, 1992.
- [11] Rudolph G.: Global optimization by means of distributed evolution strategies. In: H. P. Schwefel and R. Männer, editors, *Parallel Problem Solving from Nature*. 1st Workshop PPSN I, Berlin, 1990. Springer-Verlag.
- [12] Steuer R.E.: *Multiple criteria optimization: Theory, computation and application*. John Wiley & Sons, New York, 1986.
- [13] Whitley D., Starkweather T.: GENITOR II: a distributed genetic algorithm. *Journal of Experimental and Theoretical Artificial Intelligence*, 2, 189, 1990.
- [14] Whitley D., Kauth J.: GENITOR: a different genetic algorithm. Technical Report CS-88-101, Colorado State University, Fort Collins, CO, 1988.

6.2. Experiments

We tested DRACO on a small ESP with 15 assets plus a benchmark, as expressed in Italian Lire, using monthly historical series from January 1983 to March 1993 (amounting to 96 monthly returns for each asset and the benchmark) to estimate expected returns and risks with an investment horizon of one month. Fig. 1 shows four snapshots of the evolution of a population of 200 individuals over 10000 "generations" (here generation means the birth of a new individual and the death of an old one). The population is represented by a horizontal bar chart, where the length of the bars indicates the fitness of the individuals. The population starts with a wide distribution of fitness values, but after a few generations, it converges to a narrow distribution of high fitness values. The population size remains constant at 200 individuals throughout the evolution.