PORTFOLIO DESIGN AND SIMULATION
USING EVOLUTION-BASED STRATEGY

1. Introduction

Nowadays ever increasing attention is being paid to methods based on the principle of evolution. Evolutionary Computation has become a subject of general interest with regard to the capacity to solve complex optimization problems in science and technology [01], [17], [14], [03] and [09].

This paper presents an evolutionary approach to financial economics, more precisely to the optimization portfolio problem. It consists of minimizing, for a given
level of the expected portfolio return, the value of the corresponding risk indicator. Since currently available analytical solutions were designed in restricted contexts, giving up restrictive assumptions would require completely new efficient algorithms which cannot be developed in the framework of classical methods.

Our approach combines the power of genetic algorithms ([04], [13]) used to generate artificial trading experts, to the opportunities provided by Evolution Strategies which leads to the optimization of portfolio structures where individual trading experts' advice is integrated. The algorithm presented here is the result of extensive research in the application of artificial intelligence to stock trading, details of which are documented in [06] and [08].

The paper is structured as follows. In section 2, a quick overview of the Markowitz portfolio theory is presented. Section 3 describes our approach to portfolio optimization, taking into account real market constraints. An evolutionary algorithm solving a given problem is presented in section 4. The approach is evaluated using real financial time series in section 5. The paper ends with some concluding comments.

2. Overview of Portfolio Theory

2.1. Introduction

The main goal of investors is to achieve optimal allocation of funds among various financial assets. Searching for an optimal stock portfolio, characterized by random future returns, seems to be a difficult task and is usually formalized as a risk-minimization problem under a constraint of expected portfolio return. Portfolio risk of is often measured in terms of the variance of returns but many other risk criteria have been proposed in the financial literature.

Portfolio theory can be traced back to the [12] seminal paper and it is presented in an elegant way in [05] or [16].

2.2. Portfolio Optimization Problem

Consider a financial market in which \( N \) risky assets are traded; let \( \mathbf{R}' = (R_1, R_2, ..., R_N) \) be the square-integrable random vector of their returns. Denote as \( \mathbf{r} = \mathbf{E}\mathbf{r} \) the vector of expected returns and \( \mathbf{V} \) the corresponding covariance matrix which is assumed positive definite. A portfolio is a vector \( \mathbf{x} \in \mathbb{N} \) verifying \( \mathbf{x}'\mathbf{1} = 1 \) where \( \mathbf{1} \) is a \( N \)-component vector of ones. Hence \( x_i \) is the proportion of wealth invested in the \( i \)-th asset. Denote as \( \mathcal{X} \) the set of all portfolios; for each \( \mathbf{x} \in \mathcal{X} \), we define \( \mathbf{R}_x = \mathbf{x}'\mathbf{R} \) as the portfolio return and then \( \mathbf{x}'\mathbf{r} = \mathbf{E}\mathbf{r}_x \) is the portfolio expected return.
For a fixed level $e$ of expected return, $X_e = \{x \in X : x'r = e\}$ is the set of all portfolios leading to the desired expected return $e$. The optimization problem is then to find such that:

$$Risk(\bar{x}) = \min \{Risk(x) : x \in X_e\}$$

where $Risk(.)$ is the risk indicator (variance of returns in the Markowitz theory).

### 2.3. Capital Asset Pricing Model

As individuals choose optimal portfolios according to their criteria, equilibrium stock prices arise on the financial market. The most celebrated equilibrium model, the so-called CAPM, was introduced by W. Sharpe, J. Lintner and J. Mossin in the mid-sixties [19], [10], [15]. The essential result is based on the perfect market assumption (no transaction costs, no restrictions on short selling and perfect divisibility of stocks to be traded) along with homogeneous expectations, meaning that investors possess the same information and interpret it in the same way. In other words, $r$ and $V$ are common knowledge. It implies that all investors with the same value of $e$ invest in the same portfolio. When a risk-free asset is introduced, the portfolio of risky assets in which agents invest does not depend on $e$; it is the same for all economic agents; the desired expected return is obtained through the investment in the risk-free asset.

### 2.4. Efficient Frontier

A portfolio is called an efficient portfolio if it realizes the minimum variance among the portfolios having the same expected return. The set of efficient portfolios (when $e$ varies) is called the efficient frontier.

In other words, a portfolio $x$ is efficient if and only if it is the solution to:

$$\min \{x'Vx : x \in X\}$$

under the constraints:

$$x'r = e,$$
$$x'1 = 1,$$

where $e$ stands for the desired expected return.

Using the standard method of Lagrange's coefficients, the following solution is obtained:

$$x = \frac{1}{D} [BV^{-1}1 - AV^{-1}r] + e \frac{1}{D} [CV^{-1}r - AV^{-1}1],$$

where

$$A = r'V^{-1}1 \quad B = r'V^{-1}r \quad C = 1'V^{-1}1 \quad D = BC - A^2$$
If a risk-free asset is traded, generating a known return \( r_0 \), the optimal solution, independent of \( e \) and called the market portfolio, becomes:

\[
x_m = \frac{V^{-1}(r - r_0 1)}{\text{Tr}(V^{-1}(r - r_0 1))}.
\]

Moreover, the market equilibrium is characterized by:

\[
r_k - r_0 = \beta_k (E R_m - r_0),
\]

where \( R_m = x_m' R \) is the return on the market portfolio and \( \beta_k \) is defined as

\[
\beta_k = \frac{\text{Cov}(R_k, R_m)}{\text{Var}(R_m)}.
\]

3. Evolutionary Approach to Portfolio Optimization

3.1. Basic Concepts

In spite of its wide diffusion in the professional and academic worlds, the CAPM is often criticized for its artificial assumptions. Although it is an interesting theoretical model, its practical applications may often misfire.

In our previous work on the differential evolution applied to the problem of portfolio optimization [08], some artificial assumptions of the CAPM were rejected. More precisely, several operational constraints were introduced such as the imperfect divisibility of stocks, the existence of proportional transaction costs (at a rate \( c \)) and the restrictions on short selling.

In this paper, this approach is extended by considering an alternative measure of risk, emphasizing the downside risk, the semivariance\(^1\) of returns, which was first suggested in the initial work of Markowitz.

In the previous section, a portfolio was defined as proportions of wealth invested in various stocks. In our approach, the stock quantities are considered so as to take into account real market conditions such as transaction costs. For example, \( x = (40, 30, 5, 25)' \) means that the individual possesses 40 units of the first stock, 30 units of the second, and so on.

\(^1\) A semivariance of a random variable \( X \) is defined as

\[SVar(X) = E((X - EX)_-^2),\]

where

\[(X - EX)_- = \begin{cases} 0 & 0 \leq X - EX \\ X - EX & X - EX < 0 \end{cases}\]
3.2. Artificial Trading Experts

For the purposes of this research, it is also assumed that artificial trading experts for each stock are based on technical analysis rules discovered by the genetic algorithm. In general, technical analysis assumes that future trends can be identified as a more or less complicated function of past prices. Using a trading rule is a practical way of identifying trends which, in turn and generate buying and selling signals.

Let $S$ be the set of technical analysis trading rules used to take a trading decision on the market. Let $M$ denotes the cardinality of $S$. On the basis of past prices, each rule generate a signal: to sell, to hold or to buy. For the sake of simplicity of computing these decisions will be replaced with real numbers $0.0$, $0.5$ and $1.0$ respectively.

In the approach, an expert $e = (e_1, e_2, ..., e_M)$ is an $M$-dimensional binary vector. A $i$-th coordinate of the expert is equal to 1, if and only if the expert uses the $i$-th rule in the decision process to generate a buying or selling advice. Thus, there are $2^M$ possible experts, but only a few of them are usually efficient.

For example, $e = 001101$ means that the expert $e$ generates an advice on the basis of rules numbered 3, 4 and 6.

In order to generate an expert advice, an arithmetic average $\bar{d}$ of active rules decisions is calculated as follows:

$$\bar{d} = \frac{\sum_{i=1}^{M} e_i d_i}{\sum_{i=1}^{M} e_i},$$

where $d_i$ denotes the decision of the $i$-th rule. Next, the obtained number $\bar{d}$ is transformed to a decision, i.e. a number $0.0$, $0.5$ or $1.0$. This can be done by means of a valuation function $f$ and an earlier chosen threshold $s \in [0.00, 0.50]$ as follows:

$$f(\bar{d}) = \begin{cases} 
0.0, & \bar{d} \leq s \\
0.5, & s < \bar{d} < 1 - s \\
1.0, & s \leq \bar{d}
\end{cases}$$

Finally, an advice given by the expert is equal to $f(\bar{d})$.

For example, $e = 001101$, two rules lead to buy and one leads to do nothing. Hence $\bar{d} = 0.8333$. The final decision is to buy the stock as long as $1 - s \leq 0.8333$, where $s$ denotes the earlier chosen threshold.

The threshold can be referred to as the risk aversion coefficient of the expert. For low levels of $s$ the probability of doing nothing is high because the interval $[s, 1-s]$
is large. Consequently, the strategy is conservative and the expert does not transact frequently. If the opposite is true, i.e., if \( s \) is near to 0.50, almost all decisions will be to buy or to sell.

Artificial trading experts are daily generated for each stock in the considered portfolio according to the process described in [07]. Decisions of these experts constitute the soul of a trading process presented in next section.

3.3. Trading Process

Let \( \mathbf{a}_t = (\mathbf{a}^{(1)}_t, \mathbf{a}^{(2)}_t, ..., \mathbf{a}^{(N)}_t) \) denotes the vector of expert advices at the end of day \( t \). Let \( S = \{ 1, 2, ..., N \} \) and

\[
S^{(b)}_t = \{ i \in S: \mathbf{a}^{(N)}_t = 0.0 \}, \\
S^{(h)}_t = \{ i \in S: \mathbf{a}^{(N)}_t = 0.5 \}, \\
S^{(s)}_t = \{ i \in S: \mathbf{a}^{(N)}_t = 1.0 \},
\]

\( S^{(b)}_t \) is the set of stocks which experts advice to buy, \( S^{(h)}_t \) is the set of stocks which experts advice to hold and \( S^{(s)}_t \) is the set of stocks which experts advice to sell. Superscripts (b), (h), (s) are abbreviations of 'buy', 'hold' and 'sell' respectively.

At the beginning of day \( t+1 \), a non negative number of stocks of each stock from \( S^{(b)}_t \) will be bought and a non negative number of stocks of each stock from \( S^{(h)}_t \) will be bought. Let \( \Delta \mathbf{x}_t = (\Delta \mathbf{x}^{(1)}_t, \Delta \mathbf{x}^{(2)}_t, ..., \Delta \mathbf{x}^{(N)}_t) \) denotes the vector made up of numbers of traded stocks.

For example, \( \Delta \mathbf{x}_t = (10, 20, 0, -12)' \) means that 10 stocks of first stock are bought, 20 stocks of second stock are bought and 12 stocks of fourth stock are sold.

Certainly, the following constraints should be fulfilled:

\[
\Delta \mathbf{x}^{(i)}_t > 0, \text{ for } i \in S^{(b)}_t, \\
\Delta \mathbf{x}^{(i)}_t = 0, \text{ for } i \in S^{(h)}_t, \\
\Delta \mathbf{x}^{(i)}_t < 0, \text{ for } i \in S^{(s)}_t
\]

Moreover, a budget constraint as presented below should be fulfilled.

\[
\sum_{i \in S^{(b)}_t} (1+c) \cdot p_i^{(t)} \cdot \Delta \mathbf{x}^{(i)}_t \approx \sum_{i \in S^{(h)}_t} (1-c) \cdot p_i^{(t)} \cdot \Delta \mathbf{x}^{(i)}_t,
\]

where \( \mathbf{p}_t = (p^{(1)}_t, p^{(2)}_t, ..., p^{(N)}_t) \) denotes the vector of opening prices at day \( t \). This condition come from the idea of self financing, which is discussed in the next section.

The process begins with a portfolio \( \mathbf{x}_0 \) at time \( t_0 \). Let \( X^{(1)}_t \) be a space consisting of all portfolios, which can be obtain from \( \mathbf{x}_0 \) at time \( t_1 \) according to the process presented above. The purpose is to find a portfolio \( \mathbf{x}_1 \in X^{(1)} \) minimizing the risk factor (i.e. semivariance) among the space \( X^{(1)}_t \). By repeating this process, a sequence of trading decisions, which constitutes an investor strategy, can be obtained.
3.4. Idea of Self Financing

One of the main assumptions in this approach is an idea of self financing. All funds are invested at the beginning of the trading process and while the process is running, funds can neither be added nor withdrawn. However, small amounts of money can be used to fulfill the equality as defined in the previous section.

The important question is what to do in the case where $S_t^{(b)} = \emptyset$ or $S_t^{(s)} = \emptyset$. If $S_t^{(b)} = \emptyset$, because obtained funds cannot be invested elsewhere. Similarly, if $S_t^{(s)} = \emptyset$, there will be no trading, because no funds are obtained.

However, a special risk-free asset is introduced which allows to store funds obtained in selling operations and makes it possible to realize buying operations where there has been luck in selling transactions. In order to avoid a situation where all funds of the risk-free asset are invested on the first date, a threshold, which limits the percentage of funds available for investing on one day, is defined.

3.5. Financial Time Series

The approach has been validated using real data from the Paris Stock Exchange (Euronext). Every day, for every stock, the opening, maximum, minimum and closing prices are available, as are the transaction volume and the market index value (CAC40). In performance calculation the market index is used as a proxy for the market portfolio.

4. Evolution Strategy

This approach is based on Evolution Strategies, which are described in detail in [18] and [09]. In this section the modification introduced to the standard evolution strategy is presented.

In this approach, a portfolio is encoded as a real valued vector of dimension $N$, where $N$ denotes the number of stocks included in the portfolio.

To evaluate the generated portfolios, various objective functions can be used. In the designed prototype, several functions, based on expected return and risk factors, are implemented. The available objective functions are the following:

$$F_1(x) = \frac{1}{1 + \epsilon_1 \cdot SVar(R_x)}$$

$$F_2(x) = \frac{1}{1 + \epsilon_1 \cdot SVar(R_x) + \epsilon_2 \cdot |\beta_x - \beta_x^*|}$$
where $x_0$ denotes the initial portfolio, given by the user, $R_i$ stands for the market return and $\beta_x$, $\beta_{x0}$ stand for the $\beta$ coefficient of the considered portfolio $x$ and the initial portfolio $x_0$ respectively. The factors $\epsilon_1, \epsilon_2, \epsilon_3$ are used to tune the algorithm and to adjust the importance of each component of the objective function. The objective functions refer to some heuristics using parameters such as the $\beta$ coefficient. By introducing the difference between the $\beta_x$ of the generated portfolio and the $\beta_{x0}$ of the portfolio of reference, we penalize the portfolio having $\beta_x$ far away from $\beta_{x0}$ of the reference. Nevertheless, the performance of a solution is defined in terms of expected return and risk of the portfolio on a test period as was mentioned in previous sections.

There are several methods of generating an initial population. The simplest method is random generating with uniform probability. It consists of $\mu$-times random choosing of an individual from the search space. The probability of choosing an individual should be the same for every individual in the search space.

The second method uses an initial portfolio given by a user as algorithm parameters. An initial population is chosen from the neighborhood of the given portfolio. It is done by generating a population of random modifications of the initial solution.

In the algorithm, common evolution operators such as reproduction and replacement are used.

In the process of reproduction, population of size $\mu$ generates $\lambda$ descendants. Each descendant is created from $\rho$ ancestors. Reproduction consists of three parts: parent selection, recombination and mutation, repeated $\lambda$ times.

Parent selection consists of choosing $\rho$ parents from a population of size $\mu$. There are several commonly used methods of parent selection. The simplest method is random choosing with uniform probability. One of the most popular methods is random choosing using the "roulette wheel", which means that the probability of choosing an individual is proportional to its value of the objective function.

Recombination consists of generating one descendant from $\rho$ parents chosen earlier. The recombination operators described in [18], such as no recombination, global intermediary recombination, local intermediary recombination and uniform crossover are incorporated into the system.
The approach uses a self-adaptive mutation which is presented in [02] and [18]. The parameters of the mutation are encoded in an individual together with a representation of the portfolio.

Each generated descendant has to undergo a process of verification in order to satisfy several constraints. An individual is accepted if the portfolio that it represents can be obtained in accordance with the trading process from the initial portfolio. In other cases, the individual is rejected, and the process of reproduction is repeated. As a result of this verification, offspring are obtained according to the trading process and the idea of self-financing is fulfilled.

In the replacement process, a new population of size $\mu$ is chosen from an old population of size $\mu$ and its $\lambda$ descendants.

The simplest method of replacement is deterministic selection. According to this method, from $(\mu+\lambda)$ individuals, i.e. from the union of an old population and its offspring, $\mu$ best survivals are chosen. But every individual can survive no more than $\kappa$ generations in history.

Apart from deterministic selection, the tournament selection can be used. To start with, $\tau$ individuals are randomly chosen from the union of an old population and its offspring. From these $\tau$ individuals, the best one is chosen for the new population. By repeating this process $\mu$ times a new population is obtained.

Termination criteria include several conditions. The first condition is defined by the acceptable level of valuation function value. The second is based on the homogeneity of population, defined as a minimal difference between the best and the worst portfolio. The third condition is defined as a maximal number of generations. The algorithm stops when one of them is satisfied. Readers interested in the programming aspects of the evolution-based strategy, can find more details in [09].

5. Case Study

The test concerns the Paris Stock Exchange, particularly the titles belonging to its market index the CAC40. In this test, the 40 stocks are tracked over a period of about 4 years beginning January 2, 1997. Each stock time series contains the open, close, lowest and highest price, the trading volume and the value of the index at close of trading.

In our approach, an artificial trading expert is generated for each stock of the considered portfolio based on the genetic algorithm described in [07]. Expert trading decisions are respected in the portfolio evolution.

All experiments are carried out with the same financial parameters. Transaction costs are fixed at 0.25 percent of the transaction value. The sell limit equals 50%, i.e. during each transaction no more than 50 percent of the current number of stocks can be sold. Due to this limit, the trading risk is reduced; it is impossible to
sell out all stocks at once. Moreover, no more than 50% of the capital of the risk-free asset can be invested at one time. The decision threshold, which is used to determine artificial trading expert decisions, is contained between 0.25 and 0.45. These values indicate reasonable, i.e. risk-aversive strategies. Moreover, an initial portfolio may be arbitrarily defined in the prototype. However, the initial capital of the risk-free asset is always equal to 100 000 Euros.

Two general types of tests have been carried out. The first one refers to a portfolio constituted with 10 stocks randomly chosen among the stocks of the CAC40 index. The purpose of the test was to evaluate the algorithm efficiency for medium portfolios. The second type of tests refers to a portfolio consisting of all 40 stocks of the CAC40. The purpose of this was to compare the performance of each computed portfolio with that of the market portfolio approximated by the index return.

Each test was repeated several times during different time periods to avoid bias. In addition, different initial portfolios were used. In this paper, the detailed results are not presented because of the large amount of data, but the interested reader can find an extended report in [11].

By selecting an initial portfolio and carrying out evaluations over the test period for each day of the test period, the optimal portfolio was discovered. The calculated portfolio for the next day should be the optimal one, according to the constraints defined by expert decisions and the principle of self-financing. Moreover, according to the heuristics, the $\beta$ coefficient of these portfolios will be relatively stable as compared to its initial value. In addition, the performance of the result has been evaluated on the basis of expected return and risk, the latter being defined as the semivariance of the portfolio return. Finally, the profit obtained in the suggested trading process had a significant impact.

In order to assess the results, the final profit was compared with the profit achieved by the Buy-and-Hold (B&H) strategy, which consists in keeping the initial portfolio unchanged during the whole test period. In most cases, the suggested strategy outperforms the simple B&H strategy. Although it is tightly linked to the test period and current trends of the market, repeating these experiments several times on different test periods confirms the quality and the efficiency of the proposed approach.

A brief summary of some performed tests is presented below:

<table>
<thead>
<tr>
<th>Titles</th>
<th>Length of test period</th>
<th>Number of tests</th>
<th>Number of results outperforming B&amp;H</th>
<th>Number of results outperforming index</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20 days</td>
<td>10</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>20 days</td>
<td>10</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>20 days</td>
<td>10</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>20 days</td>
<td>10</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>20 days</td>
<td>10</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>
Each case refers to a different initial portfolio and test period. Every test has been repeated several times and the obtained results were compared with the B&H strategy and the market index rate each time. Certainly, there is no connection between outperforming B&H and index, but outperforming the index seems to be a more difficult task than beating B&H.

Unfortunately, the simulation of the trading process of the portfolio consisting of all 40 stocks did not turn out to be better than the market portfolio. However, further research on configuring algorithm parameters should lead to an improvement of this score.

It is worth noting that the quality of the obtained results depends on the quality of the artificial trading experts generated earlier (see the detailed study of the expert performance in [07]). Moreover, the current market situation is also important, because when the prices of a large number of stocks are increasing, the obtained results are usually satisfactory, but they do not outperform the Buy-and-Hold strategy. In the inverse case, when prices of a large number of stocks are decreasing, the obtained profit will not be very high, but it is usually higher than the profit of the Buy-and-Hold strategy. In other words, the general appreciation of the performance of the approach cannot be easily assessed.

6. Conclusions

In this paper the evolutionary approach to the problem of portfolio optimization has been presented. The goals and constraints of the problem have been presented and an algorithm based on Evolution Strategies has been proposed. The approach rejects some artificial assumptions used in theoretical models such as perfect divisibility of stocks, and introduces transaction costs and other risk measures such as the semivariance. The approach has been evaluated and validated using real data from the Paris Stock Exchange.

In order to evaluate this approach, the obtained investment strategy has been compared with the Buy-and-Hold strategy. To reduce the time period bias on performance, several time series have been selected. The results have demonstrated that our evolutionary approach is capable of investing more efficiently than the simple Buy-and-Hold strategy.

The evolutionary approach in stock trading is still in an experimentation phase. Further research is needed, not only to build a solid theoretical foundation in knowledge discovery applied to financial time series, but also to implement an
efficient validation model for real data. The presented approach seems to constitute a practical alternative to classical theoretical models.

References

PORTFOLIO DESIGN AND SIMULATION USING EVOLUTION-BASED STRATEGY

ABSTRACT

In this paper an evolutionary algorithm to optimize a stock portfolio is presented. The method, based on Evolution Strategies, uses artificial trading experts discovered by a genetic algorithm. This approach is tested on a sample of stocks taken from the French market. Results obtained are compared with the Buy-and-Hold strategy and a stock index. Presented research extends evolutionary methods on financial economics worked out earlier for stock trading.