Evolutionary Computations for Trading Systems

(SETTORE SCIENTIFICO-DISCIPLINARE SECS-S/06)

DOTTORANDO

Massimiliano Kaucic

COORDINATORE

Chiar.mo Prof. Maurizio Fanni
Università degli Studi di Trieste

TUTORE E RELATORE

Dott. Pietro Millossovich
Università degli Studi di Trieste

Indice

Elenco delle figure

Elenco delle tabelle

Acknowledgements

During the development of my thesis several persons and institutions collaborated directly and indirectly with my research. Without their support it would be impossible for me to finish my work. I want to start expressing a sincere acknowledgement to my advisor, Dr. Pietro Millossovich, for his guidance and encouragement. Special thanks go to Prof. Maurizio Fanni, who provided me the opportunity to embark on this journey. I owe special thanks to the Department of Economics and to the Department of Applied Mathematics for providing the resources for the development of my research. Also, I would like to express deep gratitude to Generali Investments and, in particular, to Fabrizio Barbini, Michele Morganti and Marco Giovannini, for their encouragements.

I am grateful to my family for supporting and motivating me. Above all, I thank God for the gift of life, friends and family.

1
Introduction and motivation

1.1 Preamble

Evolutionary computations, also called evolutionary algorithms, consist of
several heuristics, which are able to solve optimization tasks by imitating
some aspects of natural evolution. They may use different levels of abstrac-
tion, but they are always working on populations of possible solutions for a
given task. The basic idea is that if only those individuals of a population
which meet a certain selection criteria reproduce, while the remaining indi-
viduals die, the population will converge to those individuals that best meet
the selection criteria. If imperfect reproduction is added the population can
begin to explore the search space and will move to individuals that have an
increased selection probability and that hand down this property to their
descendants. These population dynamics follow the basic rule of the Dar-
winian evolution theory, which can be described in short as the "survival of
the fittest".
Although evolutionary computations belong to a relative new research area,
from a computational perspective they have already showed some promising
features such as:

- evolutionary methods reveal a remarkable balance between efficiency
  and efficacy;

- evolutionary computations are well suited for parameter optimisation;

- this type of algorithms allows a wide variety of extensions and con-
  straints that cannot be provided in traditional methods;

- evolutionary methods are easily combined with other optimization
  techniques and can also be extended to multi-objective optimization.

From an economic perspective, these methods appear to be particularly well suited for a wide range of possible financial applications, in particular in this thesis I study evolutionary algorithms

- for time series prediction;

- to generate trading rules;

- for portfolio selection.

1.2 Motivations

It is believed that asset prices are not random, but are permeated by a "melting pot" of interrelated effects. These interrelationships may result in assets mispricing, giving rise to potentially profitable opportunities. However, it is rarely possible to detect them through simple approaches, such as dividend discount models or even capital asset pricing theories. Rather, they require models that are able to capture the market complexities. To this end, in the last decades, researchers have employed intensive econometric and statistical modeling, that examine the effects of a multitude of variables, such as price-earnings ratios, dividend yields, interest rate spreads and changes in foreign exchange rates, on a broad and variegated range of stocks at the same time. However, this computationally intensive approach often evolves into a "black box" that spits out answers of which no one knows the questions. In fact, these models result in complex functional forms difficult to manage or interpret and, in the worst case, are solely able to fit a given time series but are useless to predict it. Parallelly to quantitative approaches, other researchers have focused on the impact of investor psychology (in particular, herding and overreaction) and on the consequences of considering informed signals from management and analysts, such as share repurchases and analyst recommendations. These theories are guided by intuition and experience, and thus are difficult to be translated into a mathematical environment.

Hence, the necessity to combine together these point of views in order to develop models that examine simultaneously hundreds of variables, including qualitative informations, and that have user friendly representations, is urged. To this end, the thesis focuses on the study of methodologies that satisfy these requirements by integrating economic insights, derived from academic and professional knowledge, and evolutionary computations.

1.3 Contributions of the thesis

The main task of this work is to provide efficient algorithms based on the evolutionary paradigm of biological systems in order to compute optimal trading strategies for various profit objectives under economic and statistical constraints. The motivations for constructing such optimal strategies are:

i) the necessity to overcome data-snooping and supervisorship bias in order to learn to predict good trading opportunities by using market and/or technical indicators as features on which to base the forecasting;

ii) the feasibility of using these rules as benchmark for real trading systems;

iii) the capability of ranking quantitatively various markets with respect to their profitability according to a given criterion, thus making possible portfolio allocations.

More precisely, I present two algorithms that use artificial expert trading systems to predict financial time series, and a procedure to generate integrated neutral strategies for active portfolio management.

The first algorithm is an automated procedure that simultaneously selects variables and detect outliers in a dynamic linear model using information criteria as objective functions and diagnostic tests as constraints for the distributional properties of errors. The novelties are the automatic implementation of econometric conditions in the model selection step, making possible a better exploration of the solution space on one hand, and the use of evolutionary computations to efficiently generate a reduction procedure from a very large number of independent variables on the other hand.

In the second algorithm, the novelty is given by the definition of evolutionary learning in financial terms and its use in a multi-objective genetic algorithm in order to generate technical trading systems.

The last tool is based on a trading strategy on six assets, where future movements of each variable are obtained by an evolutionary procedure that integrates various types of financial variables. The contribution is given by the introduction of a genetic algorithm to optimize trading signals parameters and the way in which different informations are represented and collected.

1.4 Structure of the thesis

The rest of the thesis is organized into three parts. The first part, titled *Background*, collects Chapters 2 and 3. Its purpose is to provide an introduction to search/optimization evolutionary techniques on one hand, and to the theories that relate the predictability in financial markets with the concept of efficiency proposed over time by scholars on the other hand. More precisely, Chapter 2 introduces the basic concepts and major areas of evolutionary computation. It presents a brief history of three major types of evolutionary algorithms, i.e. evolution strategies, evolutionary programming and genetic algorithms, and points out similarities and differences among them. Moreover it gives an overview of genetic algorithms and describes classical and genetic multi-objective optimization techniques. Chapter 3 first presents an overview of the literature on the predictability of financial time series. In particular, the extent to which the efficiency paradigm is affected by the introduction of new theories, such as behavioral finance, is described in order to justify the market forecasting methodologies developed by practitioners and academics in the last decades. Then, a description of the econometric and financial techniques that will be used in conjunction with evolutionary algorithms in the next chapters is provided. Special attention is paid to economic implications, in order to highlight merits and shortcomings from a practitioner perspective.

The second part of the thesis, titled *Trading Systems*, is devoted to the description of two procedures I have developed in order to generate artificial trading strategies on the basis of evolutionary algorithms, and it groups Chapters 4 and 5. In particular, chapter 4 presents a genetic algorithm for variable selection by minimizing the error in a multiple regression model. Measures of errors such as ME, RMSE, MAE, Theil's inequality coefficient and CDC are analyzed choosing models based on AIC, BIC, ICOMP and similar criteria. Two components of penalty functions are taken in analysis-level of significance and Durbin Watson statistics. Asymptotic properties of functions are tested on several financial variables including stocks, bonds, returns, composite prices indices from the US and the EU economies. Variables with outliers that distort the efficiency and consistency of estimators are removed to solve masking and smearing problems that they may cause in estimations. Two examples complete the chapter. In both cases, models are designed to produce short-term forecasts for the excess returns of the MSCI Europe Energy sector on the MSCI Europe index and a recursive estimation-window is used to shed light on their predictability performances. In the first application the data-set is obtained by a reduction procedure from a very

large number of leading macro indicators and financial variables stacked at various lags, while in the second the complete set of 1-month lagged variables is considered. Results show a promising capability to predict excess sector returns through the selection, using the proposed methodology, of most valuable predictors. In Chapter 5 the paradigm of evolutionary learning is defined and applied in the context of technical trading rules for stock timing. A new genetic algorithm is developed by integrating statistical learning methods and bootstrap to a multi-objective non dominated sorting algorithm with variable string length, making possible to evaluate statistical and economic criteria at the same time. Subsequently, the chapter discusses a practical case, represented by a simple trading strategy where total funds are invested in either the S&P 500 Composite Index or in 3-month Treasury Bills. In this application, the most informative technical indicators are selected from a set of almost 5000 signals by the algorithm. Successively, these signals are combined into a unique trading signal by a learning method. I test the expert weighting solution obtained by the plurality voting committee, the Bayesian model averaging and Boosting procedures with data from the the S&P 500 Composite Index, in three market phases, up-trend, down-trend and sideways-movements, covering the period 2000–2006.

In the third part, titled *Portfolio Selection*, I explain how portfolio optimization models may be constructed on the basis of evolutionary algorithms and on the signals produced by artificial trading systems. First, market neutral strategies from an economic point of view are introduced, highlighting their risks and benefits and focusing on their quantitative formulation. Then, a description of the GA-Integrated Neutral tool, a MATLAB set of functions based on genetic algorithms for active portfolio management, is given. The algorithm specializes in the parameter optimization of trading signals for an integrated market neutral strategy. The chapter concludes showing an application of the tool as a support to decisions in the Absolute Return Interest Rate Strategies sub-fund of Generali Investments.

Parte I

Background

2

A gentle introduction to evolutionary computation

This chapter introduces the basic concepts and major areas of evolutionary computation. It presents a brief history of three major types of evolutionary algorithms, i.e. evolution strategies, evolutionary programming and genetic algorithms, and points out similarities and differences among them. Moreover it gives an overview of genetic algorithms and describes classical and genetic multi-objective optimization techniques.

2.1 Introduction

Operations research (OR) and management science (MS) are disciplines that attempt to aid managerial decision making by developing mathematical models that describe the essence of a problem and then by applying mathematical procedures to solve these models. In recent years evolutionary-based techniques have attracted increasing attentions for solving complex optimization problems, and in particular for many real world OR/MS problems, because these techniques are more robust than traditional methods based on formal logics or mathematical programming. Roughly speaking, evolutionary computation is the study of computational systems which use ideas and get inspirations from natural evolution and adaptation. It aims at understanding such computational systems and developing more robust and efficient ones. The problems dealt with such computational systems are usually highly nonlinear and contain inaccurate and noisy data.

2.2 Traditional versus non-traditional search

Before solving an optimization model, it is important to consider the form and mathematical properties of the objective function, constraints, and de-

cision variables. For example, the objective function might be linear or nonlinear, differentiable or non-differentiable, concave or convex. The decision variables might be continuous or discrete. The feasible region might be convex or non-convex. Since these differences impact on how the model can be solved, optimization models are classified according to them. Search and optimization techniques may be broadly classified into three categories:

- Enumerative schemes
  They are perhaps the simplest search strategy. Within some defined finite, or discretized infinite, search space each possible solution is evaluated. An example of this search is dynamic programming. However, enumerative technique is inefficient or even infeasible as search spaces become large.

- Deterministic methods
  As many real-world problems are computationally intensive, some means of limiting the search space must be implemented to find "acceptable" solutions in "acceptable" time. Deterministic algorithms attempt this by incorporating problem domain knowledge. Greedy and calculus-based methods, branch and bound tree/graph search schemes are all deterministic methods successfully used in solving a wide variety of problems (the interested reader my refer to Brassard and Bratley (1988) and Neapolitan and Naimipour (1996)). More precisely, greedy algorithms make locally optimal choices, assuming optimal sub-solutions are always part of the globally optimal solution. Thus, these algorithms fail when this assumption results false. Calculus-based methods, also called numerical methods, use a set of necessary and sufficient conditions that must be satisfied by the optimization problem for an optimal value to be found. These methods can be further subdivided into two categories: direct and indirect strategies. Direct search methods perform a hill-climbing on the search space by moving in a direction related to the local gradient. In indirect methods, the solution is sought by solving a set of equations resulting from setting the gradient of the objective function to zero. Since they assume the existence of derivatives and are local in scope, calculus-based algorithms work best on unimodal functions, but the presence of local optima, plateaus, or ridges in the fitness/search landscape reduce their effectiveness, as highlighted in Russell and Norvig (1995). Branch and bound search techniques need problem specific heuristics and decision algorithms to limit the search space. They compute some bound at a given node which determines whether the node is "promising"; several

nodes' bounds are then compared and the algorithm branches to the "most promising" node. For a more detailed description of the procedure one may refers to Neapolitan and Naimipour (1996).
Mathematical programming is further classified according to the properties of constraints and/or the objective function. Linear programming is designed to solve problems in which the objective function and all constraint relations are linear. Conversely, nonlinear programming techniques solve problems not meeting those restrictions but require convex constraint functions. It is noted here that many problem domain assumptions must be satisfied when using linear programming, and that many real-world scientific and engineering problems may only be modeled by nonlinear functions. Finally, stochastic programming is used when random-valued parameters and objective functions subject to statistical perturbations are part of the problem formulation. Depending on the type of variables used in the problem, several variants of these methods exist (i.e. discrete, integer, binary, and mixed-integer programming, following the classification made by Schwefel (1995)).

- GUIDED RANDOM SEARCH TECHNIQUES
  All the aforementioned search methods are successfully used in solving a wide variety of problems. When problems are high dimensional, or discontinuous or multimodal, or NP-complete, these techniques are unsuitable because they require problem domain knowledge (heuristics) to direct or limit search in these exceptionally large search spaces (see for example Michalewicz and Fogel (2004)). Guided random search and optimization approaches such as simulated annealing (see Kirkpatrick et al. (1983)), Monte Carlo methods (see dagpunar (2007)), tabu search (see Glover and Laguna (1997)) and evolutionary computations (see Goldberg (1989), Michalewitz (1994) and Bäck (1996)), also called evolutionary algorithms, have been developed as alternative techniques for solving these problems. Stochastic techniques are based on enumerative methods, but they use additional information about the search space to guide the search to potential better regions. They require a function assigning fitness values to possible (or partial) solutions and an encode/decode mapping mechanism between the problem and the algorithm domains. Although some stochastic techniques are shown to "eventually" find an optimum, most cannot guarantee the optimal solution. They usually provide good/near-optimal solutions to a wide spectrum of optimization problems which traditional deterministic techniques find difficult to

solve.

A random search is the simplest stochastic search strategy, as it simply evaluates a given number of randomly selected solutions. A random walk is very similar, except that the next solution evaluated is randomly selected using the last evaluated solution as a starting point. Like enumeration, though, these strategies are not efficient for many problems because of their failure to incorporate problem domain knowledge. Random searches can generally expect to do no better than enumerative ones, as shown in Goldberg (1989).

Simulated annealing is an algorithm explicitly modeled on an annealing analogy, where for example a liquid is heated and then gradually cooled until it freezes. Where hill-climbing chooses the best move from some initial point, simulated annealing picks a random one. If the move improves the current optimum it is always executed, else it is made with some probability $p < 1$. This probability exponentially decreases either with time or with the amount by which the current optimum is worsened (see Russell and Norvig (1995)). If water's temperature is lowered slowly enough it attains the lowest-energy configuration; the analogy for simulated annealing is that if the "move" probability decreases slowly enough, the global optimum is found.

In general, Monte Carlo methods involve simulations dealing with stochastic events; they employ a pure random search where any selected trial solution is fully independent of any previous choice and its outcome. The current "best" solution and associated decision variables are stored as a comparator.

Tabu search is a meta-strategy developed to avoid getting "stuck" on local optima. It keeps a record of both visited solutions and the "paths" which reached them in different "memories". This information restricts the choice of solutions to evaluate next. Tabu search is often integrated with other optimization methods, as described in Glover and Laguna (1997) and Schwefel (1995).

Evolutionary computation (EC) is a generic term for several stochastic search methods which computationally simulate the natural evolutionary process. EC embodies the techniques of genetic algorithms (GAs), evolution strategies (ESs), and evolutionary programming (EP). These techniques are loosely based on natural evolution and the Darwinian concept of survival of the fittest. Common between them are the reproduction, random variation, competition, and selection of contending individuals within some population. In general, an EA consists of a population of encoded solutions, called individuals, manipulated

by a set of operators and evaluated by some fitness function. Each solution's associated fitness determines which survive into the next generation.

2.3 Evolutionary algorithms terminology

The mimicked search process of natural evolution can yield very robust, direct computer algorithms, although these imitations are crude simplifications of biological reality. The resulting evolutionary algorithms (EAs) encode the decision variables of a search problem into finite-length strings of alphabets of certain cardinality. The string which are candidate solutions are referred to as chromosomes, the alphabets are referred to as genes and the values of genes are called alleles (see Table 2.1 for a complete nomenclature). EAs are based on the collective learning process within a population of individuals, which is arbitrarily initialized and evolves towards better and better regions of the search space by means of randomized processes of selection, mutation, and recombination. The environment delivers a quality information of the search points, represented by the value of a given fitness function, and the selection process favours those individuals of higher fitness to reproduce more often than worse individuals. The recombination mechanism allows the mixing of parental information while passing it to their descendants and mutation introduces innovation into the population. To formally define an EA, its general structure has to be described in mathematical terms, allowing for exact specification of various EA instantiations. Having discussed the relevant background terminology, an EA can be then identified by the following definition and template which are inspired from Coello et al. (2007).

Definition 1 (Evolutionary algorithm)**.** *Let I be the space of coded individuals, an algorithm structured to search for the best solution to a given problem, represented by an objective function $\Phi : I \rightarrow \mathbb{R}$, that uses the following features:*

1. *$\{\mu^{(i)}\}_{i \in \mathbb{N}}$ and $\{\mu'^{(i)}\}_{i \in \mathbb{N}}$, which are two sequences in $\mathbb{Z}^+$ representing the parent and the offspring population sizes respectively,*

2. *$\mathcal{T} : I \rightarrow I$, which is a population transformation,*

3. *$\iota : \bigcup_{i=1}^{\infty} I^{\mu^{(i)}} \rightarrow \{true, false\}$, which is the termination criterion,*

EA term	Significance
Genotype	the code, devised to represent the parameters of the problem in the form of a string
Chromosome	one encoded string of parameters (binary, Gray, floating point number)
Individual	one of more chromosomes with an associated fitness value
Gene	the encoded version of a parameter of the problem being solve
Allele	value which a gene can assume (binary, integer)
Locus	the position that the gene occupies in the chromosome
Phenotype	Problem version of the genotype (algorithm version), suited for being evaluated
Fitness	real value indicating the quality of an individual as a solution to the problem
Environment	the problem and is represented as a function indication the suitability of phenotypes
Population	a set of individuals with their associated statistics (fitness average, Hamming distances)
Selection	policy for selecting one individual from the population (selection of the fittest)
Crossover	operation that merges the genotypes of two selected parents to yield two new children.
Mutation	operation than spontaneously changes one or more alleles of the genotype

Tabella 2.1: Nomenclature in EA

4. $\{r^{(i)}\}_{i\in\mathbb{N}}$, which is a sequence of recombination operators, such that

$$r^{(i)} : \Theta_{r^{(i)}} \times \Omega_{r^{(i)}} \to \mathcal{T}\left(I^{\mu^{(i)}}\right)$$

where $\mathcal{T}\left(I^{\mu^{(i)}}\right) \subseteq I^{\mu'^{(i)}}$, $\Theta_{r^{(i)}}$ indicates the parameter space and $\Omega_{r^{(i)}}$ is the subset of $I^{\mu^{(i)}}$ to which the operator $r^{(i)}$ is applied,

5. $\{m^{(i)}\}_{i\in\mathbb{N}}$, which is a sequence of mutation operators, such that

$$m^{(i)} : \Theta_{m^{(i)}} \times \Omega_{m^{(i)}} \to I^{\mu'^{(i)}}$$

where $\Theta_{m^{(i)}}$ indicates the parameter space and $\Omega_{m^{(i)}}$ is the subset of $I^{\mu'^{(i)}}$ to which the operator $m^{(i)}$ is applied,

6. $\{s^{(i)}\}_{i\in\mathbb{N}}$, which is a sequence of selection operators, such that

$$s^{(i)} : \Theta_{s^{(i)}} \times \Omega_{s^{(i)}} \to I^{\mu^{(i+1)}}$$

where $\Theta_{s^{(i)}}$ indicates the parameter space and $\Omega_{s^{(i)}}$ is the subset of $I^{\mu^{(i)}} \cup I^{\mu'^{(i)}}$ to which the operator $s^{(i)}$ is applied,

and that may be designed as showed in Figure 2.1 is called an evolutionary algorithm.

In particular, Definition 1 takes into account for EA variants where the resulting populations have a size different to their predecessors, introducing the population transformation function $\mathcal{T}$. Moreover this definition admits sequences $\{r^{(i)}\}_{i\in\mathbb{N}}$ of different types of recombination operators that combine alleles of solutions from $I^{\mu^{(i)}}$ in a "crossover" operation, inspired by the crossover of DNA strands that occurs in reproduction of biological organisms, to generate the offsprings in $I^{\mu^{(i)}}$. The sequences of mutation operators, $\{m^{(i)}\}_{i\in\mathbb{N}}$, are applied to a subset of the descendants, $I^{\mu'^{(i)}}$, in order to make random changes or mutations in one or more individuals, yielding new candidate solutions (which may be better or worse than existing population members). Instead, selection operators consider both the population of parents and that of offsprings, $I^{\mu^{(i)}} \cup I^{\mu'^{(i)}}$, to determine the next population, $I^{\mu^{(i+1)}}$.

Evolutionary Algorithm

$t := 0;$
Initialize
    $P(0) = \{\mathbf{a_1}(0), \ldots, \mathbf{a}_\mu(0)\} \in I^{\mu(0)};$
While $\iota(\{P(0), \ldots, P(t)\}) \neq$ true **do**

- recombine: $P'(t) := r^{(t)}_{\theta^{(t)}_r}(P(t))$ %   where $\theta^{(t)}_r \in \Theta^{(t)}_r$ %

- mutate: $P''(t) := m^{(t)}_{\theta^{(t)}_m}(P'(t))$ %   where $\theta^{(t)}_m \in \Theta^{(t)}_m$ %

- select:
    If χ is true
        then $P(t+1) = s^{(t)}_{\theta^{(t)}_s}(P''(t))$ %   where $\theta^{(t)}_s \in \Theta^{(t)}_s$ %
        else $P(t+1) = s^{(t)}_{\theta^{(t)}_s}(P''(t) \cup P(t))$
    end

- update: $t = t + 1;$

end

Figura 2.1: The fundamental structure of an evolutionary algorithm

2.4 Evolutionary algorithms foundations

Three main streams of instances of the algorithm in Figure 2.1, developed independently of each other, can nowadays be identified:

- *evolutionary programming* (EP), originally developed by Fogel et al. (1966) in the U.S. and recently refined by Fogel (1991);

- *evolution strategies* (ESs), developed in Germany by Rechenberg (1973) and Schwefel (1981);

- *genetic algorithms* (GAs) by Holland (1992) developed in the U.S. with refinements by DeJong (1975), Grefenstette (1983) and Goldberg (1989).

All evolutionary algorithms have two prominent features which distinguish themselves from other search algorithms. First, they are all population-based. Second, there is communications and information exchange among individuals in a population. Such communications and information exchange are the result of selection and/or recombination. From a philosophical perspective, theese algorithms differ mainly in the level at which they simulate

EA type	Representation	Evolutionary operators
EP	real values	mutation and $(\mu + \lambda)$ selection alone
ES	real values	mutation, recombination and $(\mu + \lambda)$ or (μ, λ) selection
GA	binary and real values	mutation, recombination and selection

Tabella 2.2: Key EA implementation differences

evolution. From a computational point of view, they differ mainly in their representations of potential solutions and their operators used to modify the solutions. Table 2.2 reports the major differences of these main stream algorithms in terms of two key issues: representation and search. In particular, for evolutionary programming and evoution strategies, there are two major deterministic selection schemes, i.e. $(\mu+\lambda)$-scheme and (μ, λ)-scheme, where μ is the population size (which is the same as the number of parents) and λ the number of offspring generated from all parents. In $(\mu + \lambda)$-schemes the μ fittest individuals from $\mu + \lambda$ candidates will be selected to form the next generation. In (μ, λ)-schemes, the μ fittest individuals from λ offspring only will be selected to form the next generation. As a result, $\lambda \geq \mu$ is required. For an in-depth analysis of general evolutionary operators and EA components, interested readers are directed to Bäck et al. (1997).

In literature, there are numerous variants of classical evolution strategies, evolutionary programming and genetic algorithms described above. Some of the evolutionary algorithms can hardly be classified into any of these three categories. Evolutionary computation includes much more than just three kinds of algorithms. It also covers topics such as artificial immune systems, artificial ecological systems, co-evolutionary systems, evolvable hardware, self-adaptive systems. All the above EAs have clearly demonstrated their capability to yield good approximate solutions even in case of complicated multimodal, discontinuous, non-differentiable, and even noisy or moving response surfaces of optimization problems. Within each research community, parameter optimization has been a common and highly successful theme of applications. Evolutionary programming and especially genetic algorithms, however, were designed with a very much broader range of possible applications in mind and confirmed their wide applicability by a variety of important examples in fields like machine learning, control, automatic programming and planning.

This thesis focuses on the applicability of GAs in the financial field for parameter optimization purposes as well as for time series pattern recognition and econometric model-building. In this sense, the following sections describe more deeply features and characteristics of this type of EAs.

2.5 Genetic algorithms

GAs are probably the best known EAs, receiving remarkable attention all over the world. John Holland's research interests in the sixties were devoted to the study of general adaptive processes, concentrating on the idea of a system receiving sensory inputs from the environment by binary detectors (see Holland (1992)). Structures in the search space were progressively modified in this model by operators selected by an adaptive plan, judging about the quality of previous trials by means of an evaluation measure. Holland (1992) points out how to interpret the so-called reproductive plans in terms of genetics, economics, game-playing, pattern recognition, and parameter optimization. His genetic plans, or GAs, were applied to parameter optimization for the first time by DeJong (1975), who laid the foundations of this application technique. Nowadays, numerous modifications of the original GA, usually referred to as the canonical GA, are applied to all (and more) fields Holland had indicated. However, many of these applications show enormous differences to the canonical GA and the boundary to the other algorithms discussed above becomes blurred. Important examples of non-canonical GAs include the GENITOR system developed by Whitley (1989), the SAMUEL system proposed by Gordon and Grefenstette (1990), and the GAs in Davis (1992). The last few decades have witnessed great strides toward the development of the so-called competent GAs, which solve hard problems quickly, reliably and accurately, basing on a design decomposition methodology and mechanistic versions of certain modes of human innovation (see Goldberg (2002)).

2.5.1 Design guidelines

There are six issues leading to practical GA design. These are described below.

1. REPRESENTATION
   This issue is primarily related to the encoding scheme. Individuals are represented by binary codes, real-valued (i.e. floatingpoint) codes, and program code. Moreover, the length of individuals may be constant

or variable. In general, it is hard to find an encoding method that transforms a problem so as to reduce or preserve the difficulty of the problem itself. Hence, the encoding method that has identical genotype and phenotype of the decision variables is advisable. Although fixed-length individuals are generally desirable, their variability is not a critical factor provided their design is easy.

2. INITIALIZATION

In general, there are two issues to be considered for population initialization of GAs: the initial population size and the procedure to initialize the population. At first, the initial population size connected to the supply of raw building-blocks, or shortly BBs (see the next section for a comprehensive explanation) is crucial for efficiency of GAs in terms of both optimality and complexity. Secondly, there are two ways to generate the initial population: random and heuristic initialization. If no prior information on the problem is available, random initialization is the natural choice; otherwise, heuristic initialization is favored. Although the mean fitness of the heuristic initialization is already high so that it may help the GAs to find solutions faster, it may just explore a small part of the solution space and never find global optimal solutions because of lack of diversity in the population (see Beyer and Schwefel (2002)). In the heuristic case, thus, a portion of the population can still be generated randomly to ensure some diversity in the population. It is noted that the random initialization is generally desirable for stability and simplicity of GAs even when a valuable piece of information is available.

3. FITNESS FUNCTION

The fitness function interprets the individual in terms of physical representation and evaluates its fitness based on desired traits (in the solution). But, the fitness function must accurately measure the quality of the individuals in the population. The definition of the fitness function, therefore, is very crucial. It is suggested that the fitness function fully reflect the physical objective of the problem.

4. GENETIC OPERATORS

The genetic operators must be carefully designed as they directly affect the performance of GAs. They are usually divided into three cathegories:

i) *Selection operators*

Selection focuses on the exploration of promising regions in the solution space. As proportionate selection is very sensitive to the selection pressure, a scaling function is employed for redistributing the fitness range of the population. The selection pressure of the ordinal selection is independent of the fitness distribution, and is based solely based on the relative ranking of the population although it may also suffer from high selection pressure. In general, the ordinal selection is preferable. Among the selection schemes (in the ordinal selection), tournament selection without replacement is perceived to be effective in achieving low (selection) noise. Recall that tournament selection without replacement works by means of choosing nonoverlapping random sets of s individuals (i.e. tournament size of s) from the population and then selecting the best individual from each set to serve as a parent for the next generation. Typically, the tournament size s is 2 (viz., pairwise tournament), and it would adjust the selection pressure: the selection pressure increases as the tournament size s becomes larger. In this regard, pairwise tournament selection without replacement is advisable.

ii) *Recombination operators*

Recombination, also known as crossover, is the primary operator that increases the exploratory power of GAs. In order to successfully achieve the crossfertilizing type of innovation, crossover operator must ideally intermix good subsolutions without any disruption of the partitions, i.e. BBs. For example, uniform crossover is very promising in the absence of any inter-gene linkage while building-block crossover is better otherwise. Here, building-block crossover uniformly shuffles the genes on the basis of entire partitions, i.e. subsolutions. In practice, uniform crossover is pessimistic as most of real-world problems have the decision variables that are closely interacted each other. Moreover, building-block crossover may also be undesirable because the capability of learning linkage is an essential prerequisite of the operator. Instead of pursuing the maximum BB-wise mixing in the population, it can be also efficient to increase the population size and employ a simple crossover that has a low probability of disrupting the BBs found so far. Therefore, it is recommended that building-block crossover is suitable if the evaluation of

fitness function requires a high computational cost; otherwise, one- or two-point crossover is desirable. Naturally, the crossover probability must be relatively high.

iii) *Mutation operators*

Mutation is the secondary operator of GAs to explore a solution space. In other words, a local search is performed in the case of altering nonsalient genes or getting away from local optima is possible when the salient genes are changed. To carry out the continual improvement type of innovation, as in nature, the probability of applying mutation must be very low. Hence, the suggestion with respect of mutation is that any type of mutation designed is applicable as long as its probability is quite small. Moreover, it is possible to get rid of mutation when the design of mutation operator is complicated.

5. TREATING INFEASEABLE INDIVIDUALS

When a problem has some constraints, crossover or mutation may often generate infeasible individuals that violate the constraints. There are two strategies to deal with infeasible individuals: one is to impose a penalty and the other is to repair them (see Beyer and Schwefel (2002) for a detailed description). A classical method employs penalty functions. It must be noted that the penalty function is critical to ensure quick convergence and high quality of solution. But it is not easy to come up with an appropriate penalty function. Moreover, this technique may sacrifice some feasible individuals as well because the infeasible individuals might continue to be reproduced. On the other hand, the repair method is applied extensively. But it is not always simple to cure infeasible individuals. Hence, the repair strategy is always advisable unless developing a repair function is an arduous task or the designed function is computationally too expensive by far.

6. POPULATION SIZE

A problem that arises with GAs is to properly estimate the values of parameters. Most of the parameters can be determined by the transcendental cognition of practitioners so as to attain good performance. However, it is not easy to estimate the population size that guarantees an optimal solution quickly enough. Thus, the population size has generally been perceived as the most important factor. A recent study has developed a refined population-sizing model by integrating the requirements of the BB supply and decision making. It provides an

accurate bound on determining an adequate population size that guarantees a solution with desired quality for GAs. However, it requires stochastic information such as the variance of fitness (i.e. noise) and the expected difference value of fitness (i.e. signal) between the best and second-best BBs, which may not be available in many practical problems.

2.5.2 Why do GAs work?

Exactly how and why GAs work is still hotly debated. There are various schools of thought, and none can be said to provide a definitive answer. One of the difficulties in analyzing GAs is that there is not a single generic GA, the behaviour of which will characterize the class of algorithms that it represents. In practice, there is a vast number of ways to implement a GA and what works in one case may not work in another. Some researchers have therefore tried to look for ways of predicting algorithm performance for particular problem classes. A comprehensive survey is available in Reeves and Rowe (2001). Meanwhile, the following is a brief guide to the main concepts that have been used.

The traditional view

Holland's explanation of why it is advantageous to search the space of coded chromosomes rather than the search space hinges on three main ideas. Central to this understanding is the concept of schema. A schema is a subset of the space of coded chromosomes in which all the strings share a particular set of defined values. This can be represented by using the alphabet $\mathcal{A} \cup *$; in the binary case, $1 * * 1$, for example, represents the subset of the 4-dimensional hypercube $\{0, 1\}^4$ in which both the first and last genes take the value 1, i.e. the strings $\{1\ 0\ 0\ 1,\ \ 1\ 0\ 1\ 1,\ \ 1\ 1\ 0\ 1,\ \ 1\ 1\ 1\ 1, \}$.

The first of Holland's ideas is the intrinsic, or implicit, parallelism in accordance to which the information on many schemata can be processed in parallel. Under certain conditions that depend on population size and schema characteristics, Holland estimated that a population of size N contains information on $\mathcal{O}(N^3)$ schemata. However, these schemata cannot actually be processed in parallel, because independent estimates of their fitness cannot be obtained in general.

The second concept is expressed by the so-called Schema Theorem, in which Holland showed that if there are $N(S, t)$ instances of a given schema S in the population at time t, then at the next generation, or time step, the expected

number of instances in the new population can be bounded from below as
follows:

$$E[N(S, t+1)] \geq \frac{F(S,t)}{\overline{F}(t)} N(S,t)[1 - \epsilon(S,t)],$$

where $F(S,t)$ is the fitness of schema S, $\overline{F}(t)$ is the average fitness of the
population, and $\epsilon(S,t)$ is a term which reflects the potential for genetic op-
erators to destroy instances of schema S.

Some conclusions have been drawn from this theorem, expressed in the fre-
quently made statement that good schemata will receive exponentially in-
creasing numbers of trials in subsequent generations. However, as recent
literature highlights, it is clear that the Schema Theorem is a result in ex-
pectation only for one generation. Thus, any attempt to extrapolate this
result for more than one generation is doomed to failure because the terms
are then no longer independent of what is happening in the rest of the pop-
ulation. Also, given a finite population, it is clear that any exponential
increase cannot last very long.

Holland also attempted to model schema processing (or hyperplane compe-
titions) by means of an analogy to stochastic two-armed bandit problems.
This is a well known statistical problem: there are two "levers" which if
pulled give payoff values according to different probability distributions.
The problem is to use the results of previous pulls in order to maximize the
overall future expected payoff. In Holland (1992) it is argued that a GA ap-
proximates an optimal strategy which allocates an exponentially increasing
number of trials to the observed better lever; this is then used to contend
for the supposed efficiency of a GA in distinguishing between competing
schemata or hyperplanes.

Following these ideas, early studies on GAs suggested quite strongly that in
a GA it had thus been discovered an algorithm that used the best available
search strategy to solve not merely one, but many hyperplane competitions
at once. Recently, however, "No-Free-Lunch" Theorem (NFLT) developed
by Wolpert and Macready (1997) has rather destroyed such dreams. In fact,
intrinsic parallelism turns out to be of strictly limited application; it merely
describes the number of schemata that are likely to be present in some num-
bers given certain assumptions about string length, population size and,
most importantly, the way in which the population has been generated.
Even then, only in very unusual circumstances could the hyperplane compe-
titions actually be processed in parallel; normally, the competitions are not
independent. The two-armed bandit analogy also fails in at least two ways:
first, Macready and Wolpert (1996) have recently argued that there is no

reason to believe that the strategy described by Holland as approximated by a GA is an optimal one.

This is not to say that the Schema Theorem in particular, or the idea of a schema in general, is useless, but that what it says is of limited and mainly short-term value, i.e. certain schemata are likely to increase their presence in the next population and will be on the average fitter, and less resistant to destruction by crossover and mutation, than those that do not.

This brings to the third assumption implicit in the implementation of a GA – that the recombination of small pieces of the genotype (good schemata) into bigger pieces is indeed a sensible method of finding optimal solutions. Goldberg (1989) calls this the building-block hypothesis (BBH). There is certainly some negative evidence, in that problems constructed to contain misleading building-blocks may indeed be hard for a GA to solve. The failure of the BBH is often invoked as an explanation when a GA fails to solve particular complex problems.

However, the properties of these problems are not usually such that they are uniquely difficult for GAs. Holland himself, with two other co-workers, looked for positive evidence in favour of the BBH (see Mitchell et al. (1994)) and found the results rather problematical: functions constructed precisely to provide a "royal road" made up of building blocks of increasing size and fitness turned out to be more efficiently solved by "non-genetic" methods.

Other approaches

By writing his theorem in the form of a lower bound, Holland was able to make a statement about schema S that is independent of what happens to the other schemata. However, in practice what happens to schema S will influence the survival of other schemata, and what happens to other schemata will affect what happens to S as is made plain by the exact models of Vose (1993) and Whitley (1993).

Markov chain theory has been applied to GAs to gain better understanding of the GA as a whole. However, while the results are fascinating in illuminating some nuances of GA behaviour, the computational requirements are formidable for all but the smallest of the problems, as shown by K. A. DeJong and Gordon (1995) and Rees and Koehler (1999).

Shapiro et al. (1994) first examined GAs from a statistical perspective and Peck and Dhawan (1995) have linked GAs to global randomized search methods. Studies such as Jones (1995) have also pointed out connections between GA and neighbourhood search methods. But one of the difficulties in analyzing GAs is that there is not a single generic GA. Reeves and Wright

(1999) explain a perspective based on relating GAs to statistical methods of experimental design, which draws upon biological concept of epistasis. This represents the idea that the expression of a chromosome is not merely a sum of the effects of its individual alleles, but that the alleles located in some genes influence the expressions of the alleles in others. From a mathematical viewpoint, epistasis is equivalent to the existence of interactions in the fitness function. If one knew the extent of this non-linearities, he/she might be able to choose an appropriate algorithm. Unfortunately, as explained in Reeves (1999), it is unlikely that this approach will be successful, although the literature surrounding the question of epistasis has produced some useful insights into GAs.

2.6 Multi-objective genetic algorithms

In many real-world situations there may be several objectives that must be optimized simultaneously in order to solve a certain problem. This is in contrast to the problems tackled by conventional GAs, which involve optimization of just a single criterion. The main difficulty in considering multi-objective optimization is that there is no accepted definition of optimum in this case, and therefore it is difficult to compare one solution with another one. An approach for solving such multi-objective problems is to optimize each criterion separately and combine the solutions thus obtained. However, this method is seldom likely to provide a solution where each criterion is optimally balanced. In fact, it may happen that optimizing one objective may lead to unacceptably low performance of another objective. For example, consider the case of planning a trip from place X to place Y. The objectives here are to minimize the time as well as the cost of the trip. It is evident that the two objectives are conflicting in nature, i.e. if time is minimized, the cost goes up and vice versa. Therefore, there cannot be a single optimum in this case. Thus, for solving multi-objective problems all the objectives need to be treated together. In general, these problems admit multiple solutions, each of which is considered acceptable and equivalent when the relative importance of the objectives is unknown. The best solution is subjective and depends on the need of the designer or decision maker.

2.6.1 Multi-objective optimization terminology

In this section I refer to the minimization problem for the definition of a multi-objective optimization problem. It may be observed that the maxi-

mization problem may be expressed in terms of a minimization one, simply by considering the opposite of the objective functions.

In this manner, the multi-objective optimization problem can be stated as follows

$$\min[f_1(\mathbf{x})\ f_2(\mathbf{x})\ \ldots\ f_k(\mathbf{x})] \tag{2.1}$$

subject to the m inequality constraints

$$g_i(\mathbf{x}) \leq 0 \qquad i = 1, 2, \ldots, m \tag{2.2}$$

and the p equality constraints

$$h_i(\mathbf{x}) = 0 \qquad i = 1, 2, \ldots, p \tag{2.3}$$

where k is the number of objective functions $f_i : \mathbb{R}^n \to \mathbb{R}$. The vector $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]^T$ is called vector of decision variables. The decision maker wish to determine from among the set $\mathcal{F}$ of all vectors which satisfy Equations 2.2 and 2.3 the particular set of values $x_1^*\ x_2^*\ \ldots\ x_n^*$ which yield the optimum values of all the objective functions.

2.6.2   Pareto optimality

It is rarely the case that there is a single point that simultaneously optimizes all the objective functions of a multi-objective optimization problem. Therefore, one normally looks for "trade-offs", rather than single solutions when dealing with multi-objective optimization problems. The notion of "optimality" is therefore different in this case. The most commonly adopted notion of optimality is the Paretian.

Definition 2 (Pareto optimality). *A vector of decision variables $\mathbf{x}^* \in \mathcal{F}$ is Pareto optimal if there does not exist another $\mathbf{x} \in \mathcal{F}$ such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*)$ for all $i = 1, \ldots, k$ and $f_j(\mathbf{x}) < f_j(\mathbf{x}^*)$ for at least one j.*

In words, this definition says that $\mathbf{x}$ is Pareto optimal if there exists no feasible vector of decision variables $\mathbf{x} \in \mathcal{F}$ which would decrease some criterion without causing a simultaneous increase in at least one other criterion. Unfortunately, this concept almost always gives not a single solution, but rather a set of solutions called the Pareto-optimal set. The vectors $\mathbf{x}^*$ correspoding to the solutions included in the Pareto-optimal set are called non-dominated. The image of the Pareto-optimal set under the objective functions is called Pareto front.

2.6.3 Multi-objective algorithms

The multi-objective optimization includes many different techniques. Thus, it is hard to summarize all of them, however, to give an idea about the most common methods used in the literature, a short overview follows. for a more deep insight the interested reader may consult to Coello et al. (2007) and the references therein included.

- AGGREGATING TECHNIQUES:

    - *Weighted sum approach*, where the different objectives are combined using weighting coefficients w_i, $i = 1, 2, \ldots, k$. The objective to minimize become $\sum_{i=1}^{k} w_i f_i(\mathbf{x})$.

    - *Goal programming-based approach*, where the user is required to assign targets, or goals, T_i, $i = 1, 2, \ldots, k$, for each objective. The aim then becomes the minimization of the deviation from the targets to the objectives, or $\sum_{i=1}^{k} |f_i(\mathbf{x}) - T_i|$.

    - *Goal attainment-based approach*, where the user is required to provide, in addition to the vector of goals, a vector of weights w_i, $i = 1, 2, \ldots, k$, linking the relative under- or over-attainment of the desired goals.

    - ϵ-*constraint approach*, where the primary objective function is minimized and the other objectives are treated as constraints bound by some allowable levels ϵ_i.

- POPULATION-BASED NON-PARETO TECHNIQUES:

    - *Vector evaluated genetic algorithm* (VEGA) that incorporates a special selection operator in which a number of sub-populations were generated by applying proportional selection according to each objective function in turn.

    - *Lexicographic ordering*, where the objectives are ranked in order of importance by the user, and optimization is carried out on these objectives according to this order.

    - Use of *game theory*, where it is assumed that a player is associated with each objective.

    - *Gender-based schemes* for identifying the objectives, where panmitic reproduction, in which several parents combine to produce a single child, is allowed.

- *Contact theorem based procedures* for detecting Pareto-optimal solutions, where the fitness of an individual is set according to its relative distance with respect to the Pareto set.

- *Non-generational GAs*, where multi-objective problem is transformed into a single-objective one through a set of appropriate transformations and the fitness of an individual is calculated incrementally. Genetic operations are utilized to produce a single individual that replaces the worst individual in the population.

- PARETO-BASED NON-ELITIST STRATEGIES:

  - *Multiple objective GA* (MOGA) in which an individual is assigned a rank corresponding to the number of individuals in the current population by which it is dominated increased by 1. All non-dominated individuals are ranked 1. Fitness of individuals with the same rank are averaged so that all of them are sampled at the same rate. A niche formation method is used to distribute the population over the Pareto-optimal region.

  - *Niched Pareto GA*, where a Pareto dominance-based tournament selection with a sample of the population is used to determine the winner between two candidate solutions. Around ten individuals are used to determine dominance, and the non-dominated individual is selected. If both the individuals are either dominated or non-dominated, then the result of the tournament is decided through fitness sharing.

  - *Non-dominated sorting GA* (NSGA), where all non-dominated individuals are classified into one category, with a dummy fitness value that is proportional to the population size. Then this group is removed and the process repeated on the remaining individuals iteratively until all the individuals are classified. Stochastic-remainder-proportionate selection is used in this technique. With this scheme, any number of objective functions, and both maximization and minimization problems can be solved.

- PARETO-BASED ELITIST STRATEGIES:

  - *Strength Pareto Evolutionary Algorithm* (SPEA) introduces the elitism explicitly by maintaining an external population called an archive. At any generation t, two populations co-exist:

    i) population P_t of size N;

ii) external population, also called archive, P'_t of maximum size N'.

All the non-dominated solutions of P_t are stored in P'_t. Fitness is assigned to all the individuals in population P_t and archive P'_t as follows: individuals in the archive P'_t are assigned strength S_i using the equation

$$S_i = \frac{n}{N+1}$$

where n is the number of population members dominated by individual i of P'_t in the population and N is the total number of the individuals in the population. The fitness of the members in the archive is taken to be equal to their strength values. The fitness of the individual i in P_t is calculated as:

$$f_i = 1 + \sum_{j \in P'_t,\ j \geq i} S_j$$

where $j \geq i$ indicates that member j of the archive dominates member i of the population P_t. Fitness is determined relative to the individuals stored in the archive, irrespective of the relative dominance between the members of P_t. Binary tournament selection with assigned fitness followed by crossover and mutation operators creates the new population P_{t+1} of size N. The non-dominated members of P_{t+1} are copied into the archive, which is updated by deleting any dominated solutions if present. In case the size of the archive exceeds N' , clustering-based niching is used to reduce it.

– *Strength Pareto Evolutionary Algorithm II (SPEA2)*: two potential weaknesses of SPEA are

 i) Fitness assignment is determined entirely on the basis of the strength of archive members. This results in individuals having the same fitness value in P_t, if the corresponding set of dominating members in the archive is the same. In the worst case, if the archive contains one member, then all the members of the population will have the same rank.

 ii) The clustering technique used for ensuring diversity may lose the outer solutions, which should be kept in the archive to obtain good spread of the non-dominated solutions.

SPEA2 was developed to avoid the situation where individuals dominated by the same archive members have the same fitness

values. SPEA2 considers both the population and the archive to determine the fitness. Strength S_i of each individual i, belonging to either P_t or P'_t, is set equal to the number of individuals it dominates. A raw fitness R_i is then assigned to the individual i as

$$R_i = \sum_{j \in \{P_t \cup P'_t\}, \ j \geq i} S_j$$

Here $R_i = 0$ corresponds to the non-dominated members. The final fitness of i is computed as

$$F_i = R_i + D_i$$

where D_i is the density of individual i, computed based on its distance to the k-th nearest neighbor in the objective space. A different scheme for updating the archive that prevents the loss of the boundary solutions is adopted.

– *Elitist Nondominated Sorting Genetic Algorithm* (NSGA-II) was proposed to resolve the weaknesses of NSGA, specially its non-elitist nature. Here the chromosomes in a population are first sorted based on their domination status using the non-dominated sorting, which results in all the chromosomes being assigned a rank. The selection that follows uses the crowded tournament strategy. Crossover and mutation are then performed to generate a new child population. The parent and the child population are combined, and elitism is applied to generate the next population.

Some other related and promising techniques in this regard are based on other methods such as evolutionary strategies (Knowles and Corne (2000)), tabu search (Baykasoglu (2001)), particle swarm optimization (Baumgart-ner et al. (2004)), integration of tabu search and evolutionary algorithm (Tan et al. (2003)), and ant colony optimization (Garcia-Martinez et al. (2004)).Multi-objective GAs have evolved as a viable alternative for solving problems where the purpose is to optimize several objectives simultaneously. An important aspect of such techniques is that the decision maker is provided with a set of possible alternative solutions, as well as an intermediate solution, which the decision maker may subsequently refine.

A study establishing the effectiveness of some multi-objective optimization techniques over their single-objective counterparts for the multi-objective set covering problem with varying degrees of difficulty has been carried out in Jaszkiewicz (2003). Theoretical analysis characterizing multi-objective optimization algorithms is made by Laumanns et al. (2002). The possibility of

generating a unified framework for these algorithms, where individual multi-objective evolutionary algorithms could be generated by varying some parameter have been taken up in recent studies, such as Bosman and Thierens (2003).

Bibliografia

Ashlock, D.: 2006, *Evolutionary Computation for Modeling and Optimization*, Springer.

Bandyopadhyay, S. and Pal, S. K.: 2007, *Classification and Learning Using Genetic Algorithms*, Springer.

Baumgartner, U., Magele, C. and Renhart, W.: 2004, Pareto optimality and particle swarm optimization, *IEEE Transactions on Magnetics* **40**(2), 1172–1175.

Baykasoglu, A.: 2001, Goal programming using multiple objective tabu search, *Journal of the Operational Research Society* **52**(12), 1359–1369.

Bäck, T.: 1996, *Evolutionary Algorithms in Theory and Practice*, Oxford University Press.

Bäck, T., Fogel, D. and Michalewicz, Z.: 1997, *Handbook of Evolutionary Computation*, Vol. 1, IOP Publishing Ltd. and Oxford University Press.

Beyer, H.-G. and Schwefel, H.-P.: 2002, Evolution strategies – a comprehensive introduction, *Natural Computing: An International Journal* **1**(3), 3–52.

Bosman, P. A. N. and Thierens, D.: 2003, The balance between proximity and diversity in multiobjective evolutionary algorithms, *IEEE Transactions Evolutionary Computation* **7**(2), 174–188.

Brassard, G. and Bratley, P.: 1988, *Algorithmics: Theory and Practice*, Prentice-Hall.

Coello, C. A. C., Lamont, G. B. and Veldhuizen, D. A. V.: 2007, *Evolutionary Algorithms for Solving Multi-Objective Problem*, Springer.

Coello, C. C. C.: 2003, *Evolutionary Optimization*, Kluwer, chapter Evolutionary Multiobjective Optimization, pp. 117–146.

dagpunar, J. S.: 2007, *Simulation and Monte Carlo*, John Wiley & Sons.

Davis, L.: 1992, *Handbook of genetic algorithms*, Van Nostrand Reinhold.

Deb, K.: 2005, *Search Methodologies*, Springer, chapter Multi-objective Optimization, pp. 273–316.

Deb, K., Pratap, S. and Meyarivan, T.: 2002, A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Transactions of Evolutionary Compuation* **6**, 182–197.

DeJong, K.: 1975, *An analysis of the behaviour of a class of genetic adaptive systems*, PhD thesis, University of Michigan, Ann Arbor, MI.

Fogel, L. J.: 1991, *System identification through simulated evolution: A Machine Learning Approach to Modeling*, Ginn Press.

Fogel, L. J., Owens, A. J. and Walsh, M. J.: 1966, *Artificial Intelligence through Simulated Evolution*, Wiley.

Garcia-Martinez, C., Cordon, O. and Herrera, F.: 2004, An empirical analysis of multiobjective ant colony optimization algorithms for the bi-criteria tsp, *in* B. Heidelberg (ed.), *Proceedings of Ant Colony Optimization and Swarm Intelligence, Lecture Notes in Computer Science*, Vol. 3172, Springer, New York, pp. 61–72.

Glover, F. and Laguna, M.: 1997, *Tabu Search*, Kluwer Academic Publishers.

Goldberg, D. E.: 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, New York.

Goldberg, D. E.: 2002, *Design of Innovation: Lessons From and For Competent Genetic Algorithms*, Kluwer.

Gordon, D. A. and Grefenstette, J. J.: 1990, Explanations of empirically derived reactive plans, *Proceedings of the seventh international conference on machine learning*, pp. 198–203.

Grefenstette, J. J.: 1983, Optimization of control parameters for genetic algorithms, *IEEE Transactions on Systems, Man and Cybernetics SMC* **16**(1), 122–128.

Holland, J. H.: 1992, *Adaptation in Natural and Artificial Systems*, second edn, University of Michigan Press.

Jaszkiewicz, A.: 2003, Do multiple-objective metaheuristics deliver on their promises? a computational experiment on the set-covering problem, *IEEE Transactions Evolutionary Computation* **7**(2), 133–143.

Jones, T. C.: 1995, *Genetic Algorithms*, Doctoral dissertation, University of New Mexico, Albuquerque, NM.

K. A. DeJong, W. M. S. and Gordon, D. F.: 1995, *Foundations of Genetic Algorithms 3*, Morgan Kaufmann, San Mateo, CA, chapter Using Markov chains to analyze GAFOs, pp. 115–137.

Kirkpatrick, S., Gellatt, C. and Vecchi, M.: 1983, Optimization by Simulated Annealing, *Science* **220**(4598), 671–680.

Knowles, J. D. and Corne, D.: 2000, Approximating the nondominated front using the pareto archived evolution strategy, *Evolutionary Computation* **8**(2), 149–172.

Laumanns, M., Thiele, L., Deb, K. and Zitzler, E.: 2002, Combining convergence and diversity in evolutionary multi-objective optimization, *Evolutionary Computation* **10**(3), 263–282.

Macready, W. G. and Wolpert, D. H.: 1996, On 2-armed gaussian bandits and optimization, *Technical report SFI-TR-96-03-009*, Santa Fe Institute, New Mexico.

Michalewicz, Z. and Fogel, D. B.: 2004, *How to Solve It: Modern Heuristics*, Springer.

Michalewitz, Z.: 1994, *Genetic Algorithms + Data Structure = Evolution Program*, Springer-Verlang, New York.

Mitchell, M., Holland, J. H. and Forrest, S.: 1994, When will a genetic algorithm outperform hill climbing?, *in* J. D. Cowan, G. Tesauro and J. Alspector (eds), *Advances in Neural Information Processing Systems*, number 6, Morgan Kaufmann, San Mateo, CA.

Neapolitan, R. and Naimipour, K.: 1996, *Foundations of Algorithms*, D. C. Heath and Company.

Peck, C. C. and Dhawan, A. P.: 1995, Genetic algorithms as global random search methods: An alternative perpective, *Evolutionary Computation* **3**, 39–80.

Rechenberg, I.: 1973, *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Frommann Holzboog.

Rees, J. and Koehler, G. J.: 1999, *Evolutionary Algorithms*, Vol. 111 of *IMA Volumes in Mathematics and its Applications*, Springer-Verlang, New York, chapter An investigation of GA performance results for different cardinality alphabets, pp. 191–206.

Reeves, C.: 2003, *Handbook of Metaheuristics*, Kluwer Academic Publishers, chapter Genetic Algorithms, pp. 55–82.

Reeves, C. R.: 1999, Predictive measures for problem difficulty, *Proceedings of 1999 Congress on Evolutionary Computation*, IEEE Press, pp. 736–743.

Reeves, C. R. and Rowe, J. E.: 2001, *Genetic Algorithms: Principles and Perspectives*, Kluwer.

Reeves, C. R. and Wright, C. C.: 1999, *Evolutionary Algorithms: IMA Volumes in Mathematics and its Applications*, Vol. 111, Springer-Verlang, New York, chapter Genetic algorithms and the design of experiments, pp. 207–226.

Russell, S. and Norvig, P.: 1995, *Artificial Intelligence: A Modern Approach*, Prentice-Hall.

Sastry, K., Goldberg, D. and Kendall, G.: 2005, *Search Methodologies*, Springer, chapter Genetic Algorithms, pp. 97–125.

Schwefel, H.-P.: 1981, *Numerical optimization of computer models*, Wiley.

Schwefel, H.-P.: 1995, *Evolution and Optimum Seeking*, John Wiley & Sons.

Shapiro, J. L., Prügel-Bennett, A. and Rattray, M.: 1994, *Lecture Notes in Computer Science*, Vol. 865, Springer-Verlang, chapter A statistical mechanics formulation of the dynamics of genetic algorithms, pp. 17–27.

Tan, K. C., Khor, E. F., Lee, T. H. and Yang, Y. J.: 2003, A tabu-based exploratory evolutionary algorithm for multiobjective optimization, *Artificial Intelligence Review* **19**(3), 231–260.

Vose, M. D.: 1993, *Foundations of Genetic Algorithms 2*, Morgan Kaufmann, San Mateo, CA, chapter Modeling simple genetic algorithms, pp. 63–73.

Whitley, D.: 1989, The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best, *in* J. D. Schaffer (ed.), *Proceedings of the third international conference on genetic algorithms and their applications*, Morgan Kaufmann, San Mateo, CA, pp. 116–121.

Whitley, L. D.: 1993, *Foundations of Genetic Algorithms 2*, Morgan Kaufmann, San Mateo, CA, chapter An executable model of a simple genetic algorithm, pp. 45–62.

Wolpert, D. H. and Macready, W. G.: 1997, No free lunch theorems for optimization, *IEEE Transactions of Evolutionary Computation* **1**, 67–82.

Yao, X.: 2003, *Evolutionary Optimization*, Kluwer, chapter Evolutionary Computation, pp. 27–53.

3

Efficiency and predictability of financial markets

This chapter first presents an overview of the literature on the predictability of financial time series. In particular, the extent to which the efficiency paradigm is affected by the introduction of new theories, such as behavioral finance, is described in order to justify the market forecasting methodologies developed by practitioners and academics in the last decades. Then, a description of the econometric and financial techniques that will be used in conjunction with evolutionary algorithms in the next chapters is provided. Special attention is paid to the economic significance, in order to highlight merits and shortcomings from a practitioner perspective.

3.1 The behavior of prices

Economists have long been fascinated by the sources of variations in stock markets. By the early 1970s a consensus had emerged among financial economists suggesting that stock prices could be well approximated by a random walk model and that changes in stock returns were basically unpredictable. The cornerstone of these early studies is the concept of market efficiency. In the following I describe the ideas behind the efficient market hypothesis and its improvements that have lead to the justification and understanding of the predictability of stock returns from a theoretical point of view.

3.1.1 Efficient market hypothesis

Financial markets are difficult to predict because predictions influence markets themselves. Indeed, forecasts are potential sources of profits or losses and thus they produce market movements that in turn cause immediate

changes in prices, thereby invalidating the predictions themselves. This consideration leads to the following

Definition 3 (Market efficiency). *A market is efficient if all new information about the future behavior of prices is immediately reflected in the prices themselves.*

This definition is also known as the efficient market hypothesis (EMH henceforth) and has long been a dominant paradigm in describing the behavior of prices in speculative markets. For example, Working (1949) provided an early version of the hypothesis:

> *"If it is possible under any given combination of circumstances to predict future price changes and have the predictions fulfilled, it follows that the market expectations must have been defective; ideal market expectations would have taken full account of the information which permitted successful prediction of the price changes."*

Successively, in Working (1962), he revised his definition of a perfect futures market to

> *"one in which the market price would constitute at all times the best estimate that could be made, from currently available information, of what the price would be at the delivery date of the futures contracts."*

This definition of a perfect futures market is in essence identical to the definition of an efficient market given by Fama (1970), who stated

> *"A market in which prices always fully reflect available information is called efficient."*

This is exactly Definition 3. A more practical definition is given by Jensen (1978):

Definition 4. *A market is efficient with respect to the information set Ω_t if it is impossible to make economic profits by trading on the basis of Ω_t.*

Jensen (1978) groups the various versions of the EMH into the following three testable forms based on the definition of the information set Ω_t:

1. *the weak form of the EMH*, in which the information Ω_t only comprises past and current prices as well as possibly dividends and variables such as trading volume;

2. *the semi-strong form of the EHM*, in which Ω_t expands to include all publicly available information;

3. *the strong form of the EMH*, in which Ω_t includes all public and private information.

Timmermann and Granger (2004) extends Jensen's definition by specifying how the information variables in Ω_t are used in actual forecasting

Definition 5. *A market is efficient with respect to the information set Ω_t, search technologies S_t and forecasting models M_t if it is impossible to make economic profits by trading on the basis of signals produced from a forecasting model M_t defined over predictor variables in the information set Ω_t and selected using a search technology in S_t.*

This definition emphasizes three points:

i) the importance of the information set Ω_t adopted in the test;

ii) the ability to exploit this information in a trading strategy;

iii) that the yardstick for testing if the EMH holds is measured in economic profits (i.e. risk-adjusted and net of transaction costs).

The concept of informational or market efficiency must be sharply distinguished from the notion of operational efficiency with which it is often confused:

Definition 6 (Operational efficiency). *A market will be called operationally efficient if trades are executed at the lowest possible cost, i.e. if transaction costs are minimal.*

Transaction costs can be measured as the difference between the total cost of an item to the buyer (ask price) and the net proceeds from that same item to the seller (bid price). If there is active competition both in the stock market and among brokers, these transaction costs are presumably minimal and the market is operationally efficient. What matters, actually, is not the size of transaction costs in the stock market compared with those in other markets but what effect they have on the frequency of trading. Clearly, individual investors tend to keep their shares for long periods of time because of the high costs of commission and bid-ask spread they have to pay in order to sale and purchase a modest number of shares. It is not surprising, therefore, that institutional investors, with their lower transaction costs, turn

47

over their portfolios more rapidly, while floor traders and specialists, who have the lowest transaction costs of all, are in and out of the market many times every day.

Traders with low transaction costs can respond quickly and easily to any item of information relevant to share prices that reach them. Depending on whether the news is favorable or unfavorable, they will buy or sell immediately without having to worry unduly about transaction costs. Provided there are enough traders with low transaction costs, all information available to these traders will have resulted in virtually instantaneous sales and purchases and will be reflected in current prices. The market will then be informationally efficient. It follows that operational efficiency is a prerequisite for informational efficiency. In a market where transaction costs are relatively large, such as the housing market, high informational efficiency should not be expected. Neither the stock market should be expected to be informationally efficient if all traders had transaction costs as high as those presently affecting small investors using "full-service" brokers. Even though most important news items are widely disseminated, traders with substantial transaction costs may not find them worth enough to modify their investments. Whether the stock market is operationally efficient, i.e. whether there is a sufficient number of traders with negligible transaction costs, is essentially an empirical question on which a great deal of research has been done in recent years.

In the academic literature, it is often tacitly assumed that transaction costs can be ignored and, according to the EMH yet formulated, the stock market is informationally efficient.

3.1.2 Other market hypothesis

Over the last two decades the efficient market paradigm has been increasingly challenged by a growing number of alternative theories such as noisy rational expectations models, feedback models, disequilibrium models, herding models, agent-based models and chaos theory. All these models postulate that prices adjust sluggishly to new information due to noise, market frictions, market power, investors' sentiments, herding behavior or chaos. In these models there exist profitable trading opportunities that are not being exploited.

- NOISY RATIONAL EXPECTATIONS MODELS
  The efficient markets model implies instantaneous adjustment of prices to new information by assuming that the current equilibrium prices fully impound all available information. It implicitly assumes that

market participants are rational and they have homogeneous beliefs about information. In contrast, noisy rational expectations equilibrium models assume that current prices do not fully reveal all available information because of noise (unobserved current supply of a risky asset or information quality) in current equilibrium prices. Thus, prices show a pattern of systematic slow adjustment to new information and this implies the existence of profitable trading opportunities. Noisy rational expectations equilibrium models were developed on the basis of asymmetric information among market participants on one hand, and of the speed and the efficiency with which a speculative market responds to new information on the other hand (see, for example, Hellwig (1982) and Brown and Jennings (1989)).

- NOISY TRADERS AND FEEDBACK MODELS
  In the early 1990s, several financial economists developed the field of behavioral finance, which is, roughly specking, the finance from a broader social science perspective including psychology and sociology (the interested reader may consult Shiller (2003) for an introduction to the subject). In the behavioral finance model, there are two types of investors: arbitrageurs (also called sophisticated investors or smart money) and noise traders (feedback traders or liquidity traders). Arbitrageurs are defined as investors who form fully rational expectations about security returns, while noise traders are investors who irrationally trade on noise as if it were information (Black (1986)). Noise traders may obtain their pseudo-signals from technical analysts, brokers, or economic consultants and irrationally believe that these signals impound information. The behavioralists' approach, also known as feedback model, is then based on two assumptions. First, noise traders' demand for risky assets is affected by their irrational beliefs or sentiments that are not fully justified by news or fundamental factors. Second, since arbitrageurs are likely to be risk averse, arbitrage, defined as trading by fully rational investors not subject to such sentiment, is risky and therefore limited (Shleifer and Sumers (1990)).
  In feedback models, noise traders buy when prices rise and sell when prices fall, like trend chasers. For example, when noise traders follow positive feedback strategies (buy when prices rise), this increases aggregate demand for an asset they purchased and thus results in a further price increase. Arbitrageurs having short horizons may think that the asset is mispriced above its fundamental value, and sell it short. However, their arbitrage is limited because it is always pos-

sible that the market will perform very well (fundamental risk) and that the asset will be even more overpriced by noise traders in the near future because they can be even more optimistic ("noise trader risk", DeLong et al. (1990a)). As long as there exists risk created by the unpredictability of noise traders' opinions, sophisticated investors' arbitrage will be reduced even in the absence of fundamental risk and thus they do not fully counter the effects of the noise traders. Rather, it may be optimal for arbitrageurs to jump on the "bandwagon" themselves. Arbitrageurs optimally buy the asset that noise traders have purchased and sell it out much later when the asset price rises high enough. Therefore, although ultimately arbitrageurs make prices return to their fundamental levels, in the short run they amplify the effect of noise traders (DeLong et al. (1990b)). On the other hand, when noise traders are pessimistic and thus follow negative feedback strategies, downward price movement drives further price decreases and over time this process eventually creates a negative bubble. In the feedback models, since noise traders may be more aggressive than arbitrageurs due to their overoptimistic/overpessimistic or overconfident views on markets, they bear more risk with higher expected returns. As long as risk-return tradeoffs exist, noise traders may earn higher returns than arbitrageurs. DeLong et al. (1991) further showed that even in the long run noise traders as a group survive and dominate the market in terms of wealth despite their excessive risk taking and excessive consumption. Hence, the feedback models suggest that technical trading profits may be available even in the long run if technical trading strategies (buy when prices rise and sell when prices fall) are based on noise or "popular models" and not on information such as news or fundamental factors (Shleifer and Sumers (1990)).

- DISEQUILIBRIUM MODELS
  Beja and Goldman (1980) introduced a simple disequilibrium model that explained the dynamic behavior of prices in the short run. The rationale behind their model was

  > "When price movements are forced by supply and demand imbalances which may take time to clear, a nonstationary economy must experience at least some transient moments of disequilibrium. Observed prices will then depend not only on the state of the environment, but also on the state of the market."

The state of the economic environment represents agents' endowments, preferences and information generally changing with time. In the disequilibrium model, therefore, the investor's excess demand function for a security includes two components:

  i) fundamental demand which is the aggregate demand that the auctioneer would face if at time t one were to conduct a Walrasian auction in the economy;

  ii) the difference between actual excess demand and corresponding fundamental demand.

With non-equilibrium trading, the demand should reflect the potential for direct speculation on price changes, including the price's adjustment towards equilibrium. In general, this is a function of both speculators' average assessment of the current trend in the security's price and the opportunity growth rate of alternative investments in non-equilibrium trading with comparable securities. The process of trend estimation is adaptive because price changes include some randomness. Beja and Goldman showed that when trend followers have some market power, an increase in fundamental demand might generate oscillations, although the economy, dominated by fundamental demand, is stable and non-oscillatory. Furthermore, increasing the market impact of trend followers causes oscillations and makes the system unstable. These situations imply poor signaling quality of prices. On the other hand, they also demonstrated that moderate speculation might improve the quality of price signal and thus accelerate the convergence to equilibrium. This happens when the speculators' response to changes in price movements is relatively faster than the impact of fundamental demand on price adjustment.

- HERDING MODELS
  Froot et al. (1992) demonstrated that herding behavior of short horizon traders can lead to informational inefficiency. Their model showed that an informed trader who wants to buy or sell in the near future could benefit from their information only if it is subsequently embedded into the price by the trades of similarly informed speculators. Thus, short horizon traders would make profits when they can coordinate their research efforts on the same information. This kind of positive informational spillover can be so powerful that herding traders may even analyze information that is not closely related to the asset's long run value. In their model, such an equilibrium is possible even in

the condition in which prices follow a random walk and hence publicly available information has no value in forecasting future price changes.

- AGENT-BASED MODELS
An agent-based model is a computational model for simulating the simultaneous operations of multiple agents, in an attempt to re-create and predict the actions of a very large group of heterogeneous artificial traders. The process wants to describe the relations between prices and market information from the lower level (micro-structure) of the market to the higher level (macro-structure) of the prediction of its movements. The individual agents are presumed to be acting in what they perceive as their own interests, such as economic benefit or social status, and their knowledge is limited (see LeBaron (1998) for an introduction).

3.2 Market predictability

Predictability is related to the possibility of generating excess returns by using past information. The sources of the predictability of stock returns are well documented. Indeed, there are two competing points of view. The first considers that predictability is attributed to market inefficiencies and the second argues that predictability is the result of variation in the expected returns driven by economic fundamentals. The rational expectations theory has as a consequence that expected stock returns should be predictable if they are related to predetermined variables, which predict the variation over time. Bekaert (2001) argues that predictability may also reflect irrational behavior of part of market participants, or should be the result of poor statistical inference. Asset pricing models make a relation between expected returns and their sensitivity to changes in economic factors, or factor loadings (the betas coefficients). Hence, the predictable variation of stock returns can be attributed to the changes in the betas, in the risk premium or in both of them. Recently, a large number of studies in the finance literature have confirmed the evidence of the predictability of stock returns by means of interest rates, dividend yields and a variety of macroeconomic variables reflecting business cycle variations. Ferson and Harvey (1991) explain the evidence of the predictability by standard risk factors in a multiple beta model. They identify the prespecified economic factors used by Chen et al. (1986) and find that risk premiums vary over the time and are higher during recessionary periods. Summers (1986) finds that the logarithm of stock price index can be described by a component of random walk and a com-

ponent of stationary mean reversion. Poterba and Summers (1986) suggest a similar model, with an alternative hypothesis which states that investors are irrationals. However, Lo and McKinley (1988) reject the postulate of random walk and mean reversion using data from US market. The most significant variables used in studies by American scholars are past returns of stock market, the dividend rate of a market index, the earning to price ratio and variables of the term structure. Fama (1986) find that stock returns are negatively correlated to the expected inflation and to the level of short-term interest rates. Keim and Stambaugh (1986) develop predetermined variables that are able to predict expected stock and bond returns. Fama and French (1988) show that past returns can predict 40% of future stock returns at long horizons. Fama and French (1989) suggest that the predictable variation in expected returns is rational and largely common across security classes (stocks and bonds), thus the predictability may reflect changing in business conditions. Jagannathan and Wang (1996) argue that during a recession period the financial leverage of firms in relative poor shape may increase compared to other firms. It follows that their systematic risk (stock betas) should increase. Kothari and Shanken (1997) find that the book to market ratio have a strong ability to predict future returns. Finally, several studies, which are based on daily and weekly data, make a weak evidence of predictability using t-student statistics, R^2 and p-value.

The concept of stock returns predictability is often related to market efficiency and investor rationality. Balvers et al. (1990) argue that predictable movements in economy are consistent with efficient markets. Pesaran (2003) shows that stock market returns will be not predictable only if market efficiency is combined with risk neutrality. In particular, excess returns can be predictable at an efficient stock market if investors are risk averse. The extent to which excess returns can be predicted will depend on the presence of a stable relation between the risk premium and predetermined variables. In the same direction, Rey (2004) argues that stock market predictability on its own would not imply stock market inefficiency and irrational behavior. Indeed, investors' risk aversion should also be considered.

From a practitioner point of view, due to risk aversion, investors require a small positive expected return in risky markets. Thus, in markets in which long positions prevail, like stock markets, this implies a positive upward drift. In symmetric markets, where traders are as likely to be long as they are short, like futures and foreign exchange markets, the implication is that one would expect the price to be predictable to some degree. Furthermore, government intervention in foreign exchange markets may even make them more predictable. So, for theoretical reasons, one may expect that foreign

exchange markets should be the most predictable, futures markets intermediate and stock markets the least predictable. The empirical evidence found in Park and Irwin (2004) and James (2006) confirms this theory. However, a buy-and-hold strategy in the stock market should make money because stock markets are a positive sum games, whilst the same can not be said for futures or FX markets.

3.3 Empirical tools

In practice, model selection involves some interplay between business requirements and economic intuition. Business needs can be of various kinds, including the need to build or modify a suite of models and the need to model specific markets, market segments, or market regimes. Given a broad business need, the starting point of the modeler will be economic intuition; a purely data-mining approach in which the modeler probes data automatically in a search for patterns is not feasible. For example, intuition will suggest whether to use an explanatory model, which is based on exogenous factors, or a recursive predictive model, which makes predictions on the basis of its own past. The process of model selection consists of four steps:

1. formulating the econometric hypothesis;

2. building the model;

3. estimating the model;

4. testing the model.

An important lesson in the theory of learning is that a key virtue for models is simplicity. Complex models require huge amounts of data for estimation and testing. Trade-offs always have to be made among model complexity, explanatory power and the size of available data sets. Despite their apparent superabundance, economic and financial data are actually scarce relative to what is needed to estimate many kinds of models. For example, the 125000 individual possible pairwise correlations in the S&P 500 need to be reduced. Only a tiny fraction of the potential correlation structure is revealing; the rest is noise. If two models have roughly the same explanatory power, the simpler model, i.e. the one with a smaller number of parameters to estimate, is preferable.

There has been a great effort in the development of models able to capture and predict the structure of financial time series. For a detailed survey on

these techniques I refer to Fabozzi et al. (2006). In this thesis, in particular, I focus on the applicability of econometric models and technical analysis to forecast returns in the European and US markets. The choice of these methods is motivated by their widely use in practice, their flexibility and ease of interpretation.

3.3.1 Predictive return models

A number of econometric models are currently being used in equity portfolio management to model risk and returns in a predictive environment. In particular, a predictive return model makes conditional forecasts of expected returns that are dependent on the present information set. Three of the major families of this type of econometric models are regressive models, linear autoregressive models and dynamic factor models.

- REGRESSIVE MODELS

  They are generally based on linear regressions on "factors" (which are also referred to as "predictors"). Linear regressions are simple yet powerful statistical models. Conceptually, regressive models may be categorized into two fundamental kinds:

  i) *Static regressive models*, that do not make predictions about the future. These models regress present returns on present factors. The best known example of a static regressive model of returns is the capital asset pricing model (CAPM). Suppose the risk-free return is r_f and the return of the market portfolio is r_M. The CAPM states that each stock return r_i is characterized by a constant β_i such that the expected excess return of that stock (i.e. the difference between the return of that stock and the risk-free return) is proportional to the expected market excess return. The proportionality constant β_i is the covariance between the stock and the market portfolio divided by the variance of the market portfolio and is a measure of the stock's systematic risk.

  The CAPM can also be expressed as a static linear regression in which each stock's excess return is regressed on the market excess return plus a noise term. Static multifactor models can also be expressed as linear regressions. For example, the arbitrage pricing theory model (APT) is a linear regression of each stock's return on a small number of factors. In general, these factors can be interpreted as portfolios.

  These regressions are not predictive because there is no time lag

between the return and the factor. For example, in the CAPM, the conditional expectation of each stock's return at time t is proportional to the excess return of the market portfolio, which is not known at time t. Predictions would be possible only if one could predict the excess return of the market portfolio. If one wants to use the CAPM or APT to build a portfolio or to compute portfolio risk measures such as value at risk, some assumptions about how to forecast the factor(s) is needed. The usual assumption is that the factors, and thus the returns, are sequences of independent and identically distributed random variables.

Note that nothing in the CAPM regression precludes from assuming that the market portfolio return can be predicted. However, should the market portfolio return be predictable, the theoretical static CAPM relationship would have to be replaced by a dynamic model because prices would be explosive if betas did not change. Theoretical dynamic asset-pricing models have been developed, but they have limited practical applicability because understanding if and how changes of the model parameters actually occur requires long time series and great experience to understand if model coefficients are noisy and do not carry genuine information. In practice, for portfolio management, one needs simple models that can be estimated from the limited amount of empirical data. Note, however, that dynamic models, in which both expected returns and risk are predictable, do not contradict the basic principles of absence of arbitrage and market efficiency.

ii) *Predictive regressive models*, that regress future returns on present and past factors to make predictions.

They have been developed in the quest for models that predict returns. Consider a stock return r_t and a number of predictors, for example a number of company financial ratios, F_{it}. A predictive linear regressive model assumes that the stock return at any given time t is a weighted average of its predictors at an earlier time plus a constant and some error.

Predictive regressions can also be defined by regressing returns on factors at different lags. Models of this type are called distributed lag models (DL). The advantage of these models is their ability to capture the eventual dependence of returns not only on factors but also on the rate of change of factors. To appreciate the economic importance of DL models, suppose one wants

to create a predictive model based on, among other factors, the market sentiment, which is typically measured as a weighted average of analysts' forecasts. A reasonable assumption is that stock returns will be sensitive to the value of market sentiment but will be even more sensitive to changes in market sentiment. Hence, DL models will be useful in this setting.

- LINEAR AUTOREGRESSIVE MODELS
  In a linear autoregressive model, a variable is regressed on its own lagged values, that is on its own past. If the model involves only one variable, it is called autoregressive model (AR). If more than one variable is regressed contemporaneously in the model, it is called vector autoregressive model (VAR) because the model variables are now vectors, i.e. arrays of variables.
  An AR model prescribes the value of a variable at time t as a weighted average of the values of the same variable at times $t-1$, $t-2$ and so on, depending on number of lags, plus an error term. The weighting coefficients are the model parameters. If the model includes p lags, then p parameters must be estimated. Similarly, if a VAR model includes n variables and p lags, each equation includes (up to) $n \times p$ lagged values. Because the model has n variables (and thus n equations), it has (up to) $n^2 \times p$ parameters.
  A VAR model is "richer" than an AR model, in the sense that it is able to capture cross-autocorrelations among a set of different variables, i.e. how a variable at time t is linked to another variable at some other time. An important question is whether these links are causal or simply correlations, but this is beyond the scope of this introduction (see Lütkepohl (2006)). These considerations make clear that a VAR model can model only a small number of series. For example, the return processes for the individual securities making up such aggregates as the S&P 500 Index would result in a huge number of parameters to estimate. For example, if one wanted to model the daily returns of the S&P 500 with a VAR model that included two lags, the number of parameters to estimate would be $500 \times 500 \times 2 = 500000$ parameters. To have at least as many data points as parameters, one would need at least four years of data, or 1000 trading days, for each stock return process, which is $1000 \times 500 = 500000$ data points. Under these conditions, estimates would be extremely noisy and the estimated model, meaningless.

- DYNAMIC FACTOR MODELS

  In general, a model in which factors follow a VAR model and returns (or prices) are linearly regressed on these factors is a dynamic factor model. This type of econometric models are often cointegrated models in which factors are the common trends.

  Recall that for a model to be useful, the number of parameters to be estimated needs to be small. A dynamic factor model fulfills this requirement by concentrating the dynamics of an aggregate (such as the S&P 500) into a small number of dynamic factors. For example, a modeler might identify in the S&P 500 three dynamic factors modeled by a VAR with, say, four lags, which results in only 12 factors rather than the initial 500.

  The salient characteristic of dynamic models of stock returns and stock prices is their ability to predict expected returns on the basis of present and past values of the same returns plus other variables.

3.3.2 Technical analysis

Technical analysis is a forecasting method of price movements using past prices, volumes and open interests. Pring (2002), a leading technical analyst, provides a more specific definition:

Definition 7 (Technical analysis). *"The technical approach to investment is essentially a reflection of the idea that prices move in trends that are determined by the changing attitudes of investors toward a variety of economic, monetary, political, and psychological forces. The art of technical analysis, for it is an art, is to identify a trend reversal at a relatively early stage and ride on that trend until the weight of the evidence shows or proves that the trend has reversed."*

As explained previously, technical analysis has no value if the weak, and therefore the semi-strong and strong form of the EMH hold, or if the market price follows a Markov process. However, this methodology has been extensively used among market participants, such as brokers, dealers, fund managers, speculators and investors in the financial industry. In contrast to the views of many practitioners, most academics are skeptical about technical analysis, basing on the efficiency paradigm. Nevertheless, in recent decades rigorous explanations for the widespread use of technical analysis have been developed and a growing interest in its applicability to predict finacial time series has come out (see for example Pruitt and White (1988), Brock et al. (1992) and Covel (2005)).

There are some psychological explanations of why a large number of people have a strong belief in technical analysis:

i) *representativeness*, that is the attitude of people to often predict future uncertain events by taking a short history of data and asking what broader picture this history is representative of (see, in particular, Tversky and Kahneman (1974));

ii) *communal reinforcement*, that is a social construction in which a strong belief is formed when a claim is repeatedly asserted by members of a community, rather than due to the existence of empirical evidence for the validity of the claim;

iii) *selective thinking*, that is the process by which one focuses on favourable evidence in order to justify a belief, ignoring unfavourable evidence;

iv) *confirmation bias*, that is a cognitive bias whereby one tends to notice and look for information that confirms one's existing beliefs, whilst ignoring anything that contradicts those beliefs. It is a type of selective thinking;

v) *self-deception*, that is the process of misleading ourselves to accept as true or valid what we believe to be false or invalid by ignoring evidence of the contrary position.

Technical analysis includes a variety of forecasting techniques such as chart analysis, pattern recognition analysis, seasonality and cycle analysis and computerized technical trading systems. However, academic research on technical analysis is generally limited to techniques that can be expressed in mathematical forms, namely technical trading systems, although some recent studies attempt to test visual chart patterns using pattern recognition algorithms. A technical trading system consists of a set of trading rules that result from parameterizations. Each trading rule generates trading signals (long, short, or out of market) according to their parameter values. Several popular technical trading systems are moving averages, channels and momentum oscillators.

Since Donchian (1960), numerous empirical studies have tested the profitability of technical trading rules in a variety of markets for the purpose of either uncovering profitable trading rules or testing market efficiency, or both. Most studies have concentrated on stock markets, both in the US and outside the US, and foreign exchange markets, while a smaller number of studies have analyzed futures markets (the interested reader may refers to

Park and Irwin (2004)).

The literature gives rise to the following conclusions about technical analysis applicability:

1. there is evidence in support of the usefulness of moving averages, momentum, support and resistance and some patterns;

2. technical analysis works best on currency markets, intermediate on futures markets, and worst on stock markets

3. chart patterns work better on stock markets than currency markets;

4. non-linear methods work best overall, this is not at all surprising in light of the non-linearities found in the markets (see for example Ammermanna and Patterson (2003)).

5. technical analysis does not work as well as it used to: as transaction costs decrease, available computing power increases and the number of market participants increases, one would expect markets to become increasingly efficient and thus it is not surprising that the efficacy of technical analysis should diminish.

Bibliografia

Ammermanna, P. A. and Patterson, M.: 2003, The Cross-Sectional and Cross-Temporal Universality of Nonlinear Serial Dependencies: Evidence from World Stock Indices and the Taiwan Stock Exchange, *Pacific-Basin Finance Journal* **11**(2), 175–195.

Balvers, R., Cosimano, T. and McDonald, B.: 1990, Predicting stock returns in an efficient market, *Journal of Finance* **45**(4), 1109–1128.

Beja, A. and Goldman, M. B.: 1980, On the Dynamic Behavior of Prices in Disequilibrium, *Journal of Finance* **35**, 235–248.

Bekaert, G.: 2001, Editor's Foreword to the special issue: On the predictability of asset returns, *Journal of Empirical Finance* **8**, 451–457.

Black, F.: 1986, Noise, *Journal of Finance* **41**, 529–543.

Brock, W., Lakonishok, J. and LeBaron, B.: 1992, Simple technical trading rules and the stochastic properties of stock returns, *The Journal of Finance* **XLVII**(5), 1731–1764.

Brown, D. P. and Jennings, R. H.: 1989, On Technical Analysis, *Review of Financial Studies* **2**, 527–551.

Chen, N., Roll, R. and Ross, S.: 1986, Economic forces and the stock market, *Journal of Business* **59**, 383–403.

Covel, M.: 2005, *Trend Following: How Great Traders Make Millions in Up or Down Markets*, Financial Times Prentice Hall.

DeLong, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J.: 1990a, Noise Trader Risk in Financial Markets, *Journal of Political Economy* **98**, 703–738.

DeLong, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J.: 1990b, Positive Feedback Investment Strategies and Destabilizing Rational Speculation, *Journal of Finance* **45**, 379–395.

DeLong, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J.: 1991, The Survival of Noise Traders in Financial Markets, *Journal of Business* **64**, 1–19.

Donchian, R. D.: 1960, High Finance in Copper, *Financial Analysts Journal* pp. 133–142.

Fabozzi, F. J., Focardi, S. M. and Kolm, P. N.: 2006, *Financial Modeling of the Equity Market*, John Wiley & Sons, Inc.

Fama, E.: 1986, Stock returns, real activity, inflation and money, *American Economic Review* **71**, 545–565.

Fama, E. F.: 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* (25), 383–417.

Fama, E. and French, K.: 1988, Dividend yield and expected stock returns, *Journal of Financial Economics* **23**, 3–25.

Fama, E. and French, K.: 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* **25**, 23–49.

Ferson, W. and Harvey, C.: 1991, Fundamental determinants of national equity market returns: a perspective on conditional asset pricing, *Journal of Banking and Finance* **21**, 1626–1965.

Froot, K. A., Scharfstein, D. S. and Stein, J. C.: 1992, Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation, *Journal of Finance* **47**, 1461–1484.

Hellwig, M.: 1982, Rational Expectations Equilibrium with Conditioning on Past Prices: A Mean-Variance Example, *Journal of Economic Theory* **26**, 279–312.

Jagannathan, R. and Wang, Z.: 1996, The conditional CAPM and cross-section of stock returns, *Journal of Finance* **51**, 3–53.

James, J.: 2006, Fx markets the most inefficient, *The Technical Analyst* **19**(4).

Jensen, M.: 1978, Some anomalous evidence regarding market efficiency, *Journal of Financial Economics* (6), 95–101.

Keim, D. and Stambaugh, R.: 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* **17**, 357–390.

Kothari and Shanken: 1997, Book to market, dividend yield and expected market retuns: a time series analysis, *Journal of Financial Economics* **44**, 169–203.

LeBaron, B.: 1998, Agent Based Computational Finance: Suggested Readings and Early Reasearch. graduate School of International Economics and Finance.

Lo, A. and McKinley, A.: 1988, Stock market prices do not follow random walks: evidence from a simple specification test, *Review of Financial Studies* **1**, 41–66.

Lütkepohl, H.: 2006, *New Introduction to Multiple Time Series Analysis*, Springer.

Park, C.-H. and Irwin, S. H.: 2004, The Profitability of Technical Analysis: A Review, *AgMAS Project Research Report 2004-04*, University of Illinois, Urbana-Champaign, Urbana.

Pesaran, H.: 2003, Market efficiency and stock market predictability. Mphil Subject 301.

Poterba, J. and Summers, L.: 1986, The persistence of volatility and stock market fluctuations, *American Economic Review* **76**, 1142–1151.

Pring, M. J.: 2002, *Technical Analysis Explained*, McGraw-Hill, New York, NY.

Pruitt, S. and White, R.: 1988, The CRISMA trading system: who says technical analysis can't beat the market?, *Journal of Portfolio Management* (1), 55–58.

Rey, D.: 2004, Stock market predictability: is it there? A critical review. Working paper n° 12/03.

Shiller, R. J.: 2003, From Efficient Markets Theory to Behavioral Finance, *Journal of Economic Perspectives* **17**, 83–104.

Shleifer, A. and Sumers, L. H.: 1990, The Noise Trader Approach to Finance, *Journal of Economic Perspectives* **4**, 19–33.

Summers, L.: 1986, Does the stock market rationally reflect fundamental values?, *Journal of Finance* **41**, 591–601.

Timmermann, A. and Granger, C. W. J.: 2004, Efficient market hypothesis and forecasting, *International Journal of Forecasting* (20), 15–17.

Tversky, A. and Kahneman, D.: 1974, Judgment under uncertainty: Heuristics and biases, *Science* **185**(4157), 1124–1131.

Working, H.: 1949, The Investigation of Economic Expectations, *American Economic Review* **39**, 150–166.

Working, H.: 1962, New Concepts Concerning Futures Markets Prices, *American Economic Review* **52**, 431–459.

Parte II

Trading Systems

4

Predicting EU Energy Industry Excess Returns on EU Market Index via a Constrained Genetic Algorithm

This chapter introduces an automated procedure to simultaneously select variables and detect outliers in a dynamic linear model using information criteria as objective functions and diagnostic tests as constraints for the distributional properties of errors. A robust scaling method is considered to take into account the sensitiveness of estimates to abnormal data. A genetic algorithm is developed to these purposes. Two examples are presented where models are designed to produce short-term forecasts for the excess returns of the MSCI Europe Energy sector on the MSCI Europe index and a recursive estimation-window is used to shed light on their predictability performances. In the first application the data-set is obtained by a reduction procedure from a very large number of leading macro indicators and financial variables stacked at various lags, while in the second the complete set of 1-month lagged variables is considered. Results show a promising capability to predict excess sector returns through the selection, using the proposed methodology, of most valuable predictors.

4.1 Introduction

Studies in variable selection have mainly followed two directions: the development of stepwise algorithms (see Miller (1990)) and Monte Carlo-based Bayesian procedures (George and McCulloch (1993, 1997)), but these techniques tend to become impractical with more than 25 variables. Up to about ten years ago, in academic works on macroeconomic modeling and economic forecasting, this was usually not a serious limitation, because variables on which researchers focused were in a handful number, e.g. the fifteen ma-

jor macroeconomic aggregates measures. Recently there has been growing attention to the problem of managing financial data-sets with thousands of potentially relevant time series to meet the need of practitioners that use such a number of variables when making their economic valuations. Computationally straightforward solutions have been proposed in literature. A possibility is to consider forecast pooling that entails the combination of two or more models with a different set of predictors to produce a single combined forecast (the interested reader may see Stock and Watson (2006) for a detailed review). An empirical comparison of different methods from the extant literature for combining the forecasts generated by the individual autoregressive distributed lag models to predict monthly US employment growth in the presence of many potentially relevant predictors is conducted by Rapach and Strauss (2008). In the same direction, an algorithm-based procedure to increase the efficiency of forecasting combining methods is provided by Costantini and Pappalardo (2008), where the Harvey et al. (1998) encompassing test is used to compare all forecasting models and eliminate those that are encompassed by others. The robustness of this methodology is assessed through an empirical application to Italian monthly industrial production using ISAE short-term forecasting models. From another perspective, Grenouilleau (2004, 2006) introduces the sorted leading indicators dynamic factor model, an approximate method based on the extraction of principal components from a data-set of thousands of series.

Hendry and his colleagues at the London School of Economics have instead developed the econometric methodology of general-to-specific modeling, where the modeler simplifies an initially general equation that adequately characterizes the empirical evidence within his or her theoretical framework (the interested reader may refer to Gilbert (1986) and Campos et al. (2005a) for a detailed description). Empirical applications of this approach including numerous countries and different sectors of the economy can be found, for example, in Hendry (1983), Ahumada (1985) MacDonald and Taylor (1992) and Campos and Ericsson (1999). Hendry (1983) exposits the main aspects of general-to-specific modelling, illustrating throughout with the modeling of U.K. consumers' expenditure. Ahumada (1985) illustrates various features of the evolving general-to-specific methodology, such as empirical implementation of the theories of reduction and encompassing as a tool for assessing alternative models of Argentina's balance of trade. MacDonald and Taylor (1992) model U.S. real money demand with an ADL, where cointegration is determined in a VAR scheme by the Johansen procedure. Campos and Ericsson (1999) illustrate constructive data mining by modelling Venezuela consumers' expenditure. For a more

comprehensive overview, I refer to Campos et al. (2005b).

A third choice is to define automated model specification searches by genetic algorithms and use information criteria as the objective function to minimize. A genetic algorithm (henceforth GA) is a stochastic procedure that uses the biological paradigm of evolution to solve optimization problems and is usually applied in complex systems with great dimensions and several constraints. Applications in social and economical fields can be found in Koza (1992) and Dorsey and Mayer (1995). Their use in econometric problems of model selection has been recently considered by Balcombe (2005) and Hasheminia and Niaki (2006). In general, the algorithm operates on a set of randomly generated potential solutions, called chromosomes, applying the concept of survival of the fittest to produce better and better approximations of the best solution via cycles of differential replication, recombination and mutation (for a more detailed description of genetic algorithms one can refer to Goldberg (1989) and Michalewitz (1994)). This plan is called simple GA. The literature holds many different versions of GA to constraint handling in evolutionary optimization. A common approach is to apply a penalty function to bias the search toward a feasible solution (for a detailed treatment of the penalty function methods one can see Smith and Coit (1997) and Runarsson and Yao (2002)), another possibility is to treat the constrained optimization problem as a multiobjective where each of the objective function and constraint violations is a separate target to be minimized (see for instance Deb (2001)).

In this chapter I will develop a GA-based procedure capable to take into account the subjective opinions of decision maker on the quality of the solution in a variable section context with the simultaneous and consistent detection of outliers under statistical constraints on errors. Diagnostic tests will be represented as constraints and the bias penalty function method will be used to search the best model in terms of the information criteria. The choice of the penalized method is motivated by the fact that it is able to balance in a more flexible and natural manner the objective function and constraints violations.

Compared to other GA algorithms presented so far, the proposed GA will improve the optimization plan suggested by Balcombe (2005), preserving its simple structure. From the financial point of view, the contribution of this chapter is the employment of the procedure in two examples to predict the excess return of the MSCI Europe Energy sector over the MSCI Europe index. The novelties mainly stay in the choice of the dependent variable on one hand, in fact excess returns are usually defined for stocks and bonds and not for aggregate or industrial indexes, and on the other hand, the use

of a data-set of hundreds of predictors provides the opportunity to exploit a much richer base of information then is conventionally used for financial time series forecasting.

The remainder of the chapter is organized as follow. Section II gives a brief overview of the econometrical and mathematical methodologies used and illustrates the employed algorithm and the proposed automated procedure. Section III describes the data set and Section IV presents the experimental results and discussions. Two different problems associated with variable selection in a large data base are studied to relate the findings to the recent literature on model uncertainty in a predictability context. Finally, Section V concludes the chapter with a summary and some remarks.

4.2 Automated model selection using GAs and diagnostic tests

The problem of variable selection in model construction is often tackled by following a defined plan of action, imposing restrictions sequentially and testing down results. The difficulties to automate this procedure are mainly of two types: the first is considering not rigid levels of significance for all tests and the second is describing a parsimonious and flexible structure for the decision tree, yet able to represent satisfactorily the judgemental criteria. The search may be exhaustive for small model spaces or may involve a stepwise algorithm that combine backward elimination and forward selection for larger spaces (in the order of ten variables). But these procedures become impractical and computationally inefficient when the size of model space increases, because its rate of growth is exponential with the number of regressors (i.e. for m regressors there are 2^m models to evaluate). To overcome this limit, the Bayesian approach uses Monte Carlo Markov Chain algorithms to explore the posterior model probabilities and to average the most likely ones. This methodology has been applied to stock return predictability by Avramov (2002) and Cremers (2002). Another solution is to consider an automated information criterion based search that ranks all meaningful models and to define the solution being the best model in the rank. Balcombe (2005) followed this way and proposed the use of genetic algorithms to detect the solutions as those models with the minimum value of a given information criterion.

4.2.1 Dynamic linear models and outlier detection

In financial literature the commonly used model to predict stock returns is a dynamic linear one because it highlights the leading relations among variables and helps the interpretation of parameter estimates in forecasting (see for example Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999)). I adopt this functional form in this work. Letting m be the number of predictors including the constant term, and n the number of observations, the model in vector form is

$$\mathbf{y} = \mathbf{X}^*\beta + \varepsilon \qquad (4.1)$$

where $\mathbf{y}$ is the variable that is being predicted, $\mathbf{X}^*$ is the $n \times m$ matrix of pre-selected explanatory variables from the complete database $\mathbf{X}$ of size $n \times M$, with $M \gg m$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ is the associated n-vector of residuals and β is the m-vector of regression coefficients. I assume models have to satisfy the following usual assumptions in regression analysis (see Pesaran and Pesaran (1997, p. 71-73)):

1. low degree of collinearity for regressors;

2. homoscedasticity, non-serial correlation and normality of residuals.

The first condition expresses the structure of interrelations among predictors and the last condition specifies the error distributional properties, i.e. $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. A specific problem of variable selection methods is that they are sensitive to outliers. These abnormal observations are typically modeled by either a shift in mean, as in Pynnönen (1992), or via a shift in variance, as in Hoeting et al. (1996). I consider the mean-slippage model to identify outliers in the set of errors and consequently transform ε to $\varepsilon^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)$, where

$$\varepsilon_i^* = \begin{cases} \varepsilon_i, & \text{if } \gamma_{k,i} = 0 \\ \varepsilon_i + \gamma_{k,i}, & \text{if } \gamma_{k,i} \neq 0 \end{cases}$$

and γ_k represents the n-vector of shifts, assuming k outliers, so that $\varepsilon^* \sim \mathcal{N}(\gamma_k, \sigma^2 I_n)$ and equation (4.1) becomes

$$\mathbf{y} = \mathbf{X}^*\beta + \varepsilon^* \qquad (4.2)$$

In multiple outlier detection there are two related problems, known as masking and smearing. Masking means that an outlier prevents another one from being identified, while smearing means that one outlier makes another observation, which is not an outlier at all, to appear as an outlier. These features

emerge mainly in sequential detection procedures as it was pointed out by Hady and Siminoff (1992). The simplest way to identify multiple outliers in a consistent manner would be to consider all possible permutations of the observations into two groups, non-outlier and outlier data, and decide which of them is the best combination based on some criterion. An example of this approach is Pynnönen (1992), who used information criteria. But similar to the case of variable selection, this approach becomes rather impracticable as n has size greater than 30. Since some combinations of data are more likely to form the outlier set and, in the same manner, some combinations of variables fit better the model, Hoeting et al. (1996) proposed a two step Bayesian method to identify outliers and predictors simultaneously. In the first step of the procedure a set of potential outliers is identified through a robust technique and successively a Monte Carlo algorithm considers all possible subsets of models and potential outliers. A procedure based on genetic algorithms to select outliers and more informative variables together in linear regression models is described in Tolvi (2004): the elements of the search space are $m + n$-vectors, where the first m coordinates represent the m predictors and the last n coordinates take into account the presence of outliers among observations. This approach constitutes an appealing alternative to Bayesian methodologies when data-set is in the order of 40–50 observations and 10–12 predictors, but for larger data-sets it is affected by the same computational problems.

Robust scaling

In this chapter scaling is applied to provide information about the location of outliers in the n-vector of residuals. There are two types of scaling: auto and robust scaling. For the sequence of errors $\varepsilon = \{\varepsilon_t\}_{t=1}^n$, the auto scaling produces

$$z_t = \frac{\varepsilon_t - \varepsilon_{\text{mean}}}{\sigma}$$

where $\varepsilon_{\text{mean}}$ is the mean of the errors and σ is their standard deviation. If errors are normally distributed, the probability that $|z_t| > 3$ is about 0.27%. The "3δ edit rule" indicates as an outlier every residual ε_t such that $|z_t| > 3$. As pointed out in Chiang et al. (2003), in the presence of multiple outliers, this rule can perform poorly for its sensitiveness to outliers on the mean and standard deviation. In the context of model selection in particular, Hoeting et al. (1996) used Huber's robust scaling to overcome this problem. The idea is to replace mean with median and standard deviation with median

absolute deviation from median (MAD) as follows

$$\sigma_{\mathrm{MAD}} = K \mathrm{median}\{|\varepsilon_t - \varepsilon_{\mathrm{median}}|\}$$

where $\varepsilon_{\mathrm{median}}$ is the median of $\{\varepsilon_t\}_{t=1}^n$ and the constant $K = 1.4826$ is required to make MAD an unbiased estimate of the standard deviation for normally distributed data. Huber's scaling is resistant to multiple outliers compared to auto scaling but may not be robust enough to detect all abnormal data; moreover, it takes a symmetric view on the variance which may be ineffective for asymmetric distributions. To reduce the effect of outliers on the estimates of mean and standard deviation for normal data, under the hypothesis that errors have considerably less than 50% outliers, I adopt the consistent modified scaling suggested by Chiang et al. (2003). First, the distance between each residual ε_t and the median is computed

$$e_t = |\varepsilon_t - \varepsilon_{\mathrm{median}}|$$

and sorted in ascending order. Then, letting e_{mean} the mean of $\{e_t\}_{t=1}^n$, the fourth order standard deviation

$$\sigma_4(j) = \left(\frac{1}{j + n/2 - 2} \sum_{i=1}^{j+n/2-1} (e_i - e_{\mathrm{mean}})^4 \right)^{1/4}, \qquad j = 1, \ldots, \frac{n}{2} + 1$$

is determined because it is more sensitive to outliers than the standard deviation. When j increases, $\sigma_4(j)$ increases too, in a gradual manner when outliers are not reached and rapidly when they are included in the calculation. The ratio between two consecutive σ_4 is computed

$$r(j) = \frac{\sigma_4(j+1)}{\sigma_4(j)}$$

and the first difference $r(j+1) - r(j)$ is maximized to disclose better normal errors from outliers. Successively the mean and standard deviation of the former data are used to autoscale $\{\varepsilon_t\}_{t=1}^n$ and to apply the 3δ edit rule.

4.2.2 Information Criteria

The most popular selection criteria and evaluation procedures are expressed through a penalized log-likelihood to take a compromise between a measure of in-sample fit versus a measure of model complexity (see for instance

George (2000)). When models are affected by multiple outliers, this unified expression may be written as

$$- 2 \log L_{m,k} + 2g(n, m + k) \qquad \boxed{4.3}$$

where $L_{m,k}$ denotes the maximum likelihood with m parameters and k outliers and $g(n, m + k)$ is a penalty function depending on the sample size n and on $m + k$. I simplify the probabilistic structure of outliers with suitable economic assumptions by considering as outliers singular observations that emerge from different and separated distributions every time. In this way I can write the maximum likelihood for model (4.2) likewise Pynnönen (1992),

$$\log L_{m,k} = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2}\log \widehat{\sigma}_k^2 + \log(n - k)! \qquad \boxed{4.4}$$

where $\widehat{\sigma}_k^2$ is the estimate of the residual variance. The penalty function in the Akaike's information criterion (AIC, Akaike (1973)) becomes $g(n, m + k) = m + k$, and in Schwartz's Bayesian criterion (BIC, Schwarz (1978)) is $g(n, m + k) = (m + k)\log(n)$. Models selected by AIC often result over-parameterized or even misspecified; on the other hand, BIC-based models are consistent only when the true model in included in the search space. This assumption is rarely fulfilled in practice and what one can hope for is a model that provides a useful approximation to the data generation process. In order to select parsimonious proxies, a desirable property of BIC is that 8 or more included parameters should have a higher penalty. However none of these criteria is able to penalize the correlation structure among predictors. Conversely, Bozdogan's information complexity criterion (ICOMP, Bozdogan (1988, 2000)) is specifically designed to be resistant to misspecification and penalizes interdependence between parameter estimates and over-parameterization. For a linear model with m predictors and k outliers it is

$$g(n, m + k) = \frac{m + k}{2} \ln \left(\frac{\mathrm{tr}(\widehat{\Sigma}_{model})}{m + k} \right) - \frac{1}{2} \ln(|\widehat{\Sigma}_{model}|) \qquad \boxed{4.5}$$

where $\widehat{\Sigma}_{model}$ represents the estimated covariance matrix of the parameter vector and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix.
A first purpose of this work is to compare the three criteria in the linear modeling context.

4.2.3 Proposed genetic algorithm

In this chapter I develop an automated procedure based on Balcombe's that is able to rank models via information criteria, detect multiple outliers and

Constraint	Condition	Test
$g_1(\mathbf{x})$	homoskedasticity of residuals	Breush-Pagan test
$g_2(\mathbf{x})$	serial correlation of residuals	Durbin-Watson test

Tabella 4.1: Diagnostic tests adopted as constraints

consider diagnostic restrictions.

The joint problem of variable selection and multiple outlier detection time can be formulated in the context of evolutionary programming. Unlike Tolvi (2004), who parameterized in the same chromosome predictors and outliers, I consider only the combinations of independent variables forming every regression model as chromosome while outliers are designed specifically for each chromosome and are identified by the consistent modified scaling of residuals. Following Balcombe (2005), each chromosome is represented with a binary m-vector, in which 1 at position i means that variable i is in the model, conversely, 0 at the same position means that variable i is out of the model, for $i = 1, \ldots, m$. To compare the different chromosomes AIC, BIC and ICOMP are used and to concern the regularity conditions of residuals in model (4.2) classical diagnostic tests are considered (see for example Kaboudan (2000)) and are displayed in table 4.1. In this way the constrained optimization problem of finding the best model in terms of a given information criterion under a defined underlying distribution of the residuals can be written as

$$\text{minimize} \quad f(\mathbf{x}), \qquad \mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m \qquad (4.6)$$

where $f(\mathbf{x})$ is the objective function, $\mathbf{x} \in \mathcal{S} \cap \mathcal{F}$, $\mathcal{S} \subseteq \{0,1\}^m$ defines the search space which is an m-dimensional space and $\mathcal{F}$ is the feasible region defined by

$$\mathcal{F} = \{\mathbf{x} \in \{0,1\}^m | g_k(\mathbf{x}) \leq 0, k = 1, 2\} \qquad (4.7)$$

where $g_k(\mathbf{x}) \leq 0$, $k = 1, 2$ are the constraints derived from the statistical tests. The constraint associated to Durbin-Watson's test has been reduced to a inequality constraint by imposing $|2 - DW(\mathbf{x})| \leq \delta$, where $DW(\mathbf{x})$ is the Durbin-Watson's test statistic for model $\mathbf{x}$ and δ is a small positive number that indicates the degree of violation.

I have used the penalty function method to transform problem (4.6-4.7) into a non-constrained one. The basis for this method is to define the new objective function

$$\psi(\mathbf{x}) = f(\mathbf{x}) + r_c \, \phi(\mathbf{x}) \qquad (4.8)$$

where $\phi(\mathbf{x})$ is the penalty function built on the constraints and $\{r_c\}_{c=0,...,C}$ is a sequence of penalty coefficients, with the subscript c indicating the generation counter and C is the total number of generations. For minimization problems, the main point is to choose the penalty function to make sure that it is zero for all feasible points and is "high" for all non-feasible points. Then the (unconstrained) minimization of $\psi(\mathbf{x})$ is equivalent to the (constrained) minimization of $f(\mathbf{x})$. To this end, I fix $r_c = 2$ for all generations to equally weight information criteria and penalties, and define

$$\phi(\mathbf{x}) = \lambda_1 \phi_1(\mathbf{x}) + \lambda_2 \phi_2(\mathbf{x}) \qquad (4.9)$$

where λ_k is the relative importance given by the decision maker to the constraint $g_k(\mathbf{x})$ for $k = 1, 2$ and penalties are

$$
\begin{aligned}
\phi_1(\mathbf{x}) &= \mathbb{I}_{\{p_1 \le \alpha_1\}}(\mathbf{x})(1 - p_1) \log L_{m,k}(\mathbf{x}) && (4.10) \\
\phi_2(\mathbf{x}) &= \mathbb{I}_{\{|2-DW|>\alpha_{2,a}\} \cup \{p_2 < \alpha_{2,b}\}}(\mathbf{x})(1 - p_2) \log L_{m,k}(\mathbf{x}) && (4.11)
\end{aligned}
$$

where $\mathbb{I}_A(\mathbf{x})$ is the indicator function, and α's and p's are the significance levels and p-values for tests respectively. From the diagnostic point of view, these expressions are sufficiently flexible to account for the null hypothesis of each test, its p-value p_k and its significance level α_k. From the evolutionary programming perspective, since the space of possible models may be disjoint under constraints, the form of these penalties allow to explore infeasible regions, that acts as bridges connecting the feasible regions and avoid to find solutions of poor quality.

The optimization algorithm consists of the following steps (Fig. 4.1):

Step 1. Select the information criterion to be the fitness function $f(\mathbf{x})$. Input historical data for the financial variables.

Step 2. Initialize the vector of significance levels α, the vector of relative importance of constraints λ and the weight r_c of penalty function with respect to log-likelihood.

Step 3. Define the population size and randomly generate set of binary m-vectors of chromosomes (models).

Step 4. Study statistical and fitness properties of chromosomes:

   a) detect outliers in each chromosome via the consistent modified scaling of data and the "3δ edit rule";

   b) evaluate the penalized fitness function for the consistent models at step 3.

Step 5. Select the solutions giving the minimum value of penalized fitness function. Compare the value of the objective function for this newly selected solutions with the value of the objective function found for the formerly selected solutions (that was determined in the previous iteration). The solutions with the minimum value of objective function is selected to form the next generation.

Step 6 Test the number of the current GA iteration. If the maximum iterations number has not been reached, go to step 7, otherwise go to step 8.

Step 7 Apply GA operators (selection, crossover and mutation) to change the states of the chromosomes. Proceed to step 3.

Step 8 Find the best solution in terms of the information criterion and the constraints. If it is not satisfactory return to step 2 otherwise end the processing.

As suggested in Balcombe (2005), one can run several independent trials of the procedure to obtain a pool of candidates among which operate the choice and to have an indication of the goodness of predictors and successively implement once again the genetic algorithm with the solutions of the trials as initial population to identify the best combinations of predictors.
All the algorithms for this procedure were programmed using MATLAB, version R2007b.

4.3 The data

The set of financial variables has been provided by Generali Investments and consists of 63 explanatory variables measured monthly over the period February 1995 – March 2007 (146 observations). Based on a review of the literature (see Pesaran and Timmermann (1995, 2000) and Bossaerts and Hillion (1999)), data have been classified into five categories: financial conditions, global economy, inflation, revision and technical. In an effort to use stationary variables and to highlight more relations with excess returns over time, I have generated three variations of log-levels for every regressor (changes in 1, 3 and 6 months). Exceptions are data from the revision and technical categories, for which I have adopted a moving average, recently introduced by Leontitsis and Pange (2004), to decrease the impact of historical data revision and have a better trade-off between smoothness and accuracy. I measure the predictability of the 1-month excess returns of the
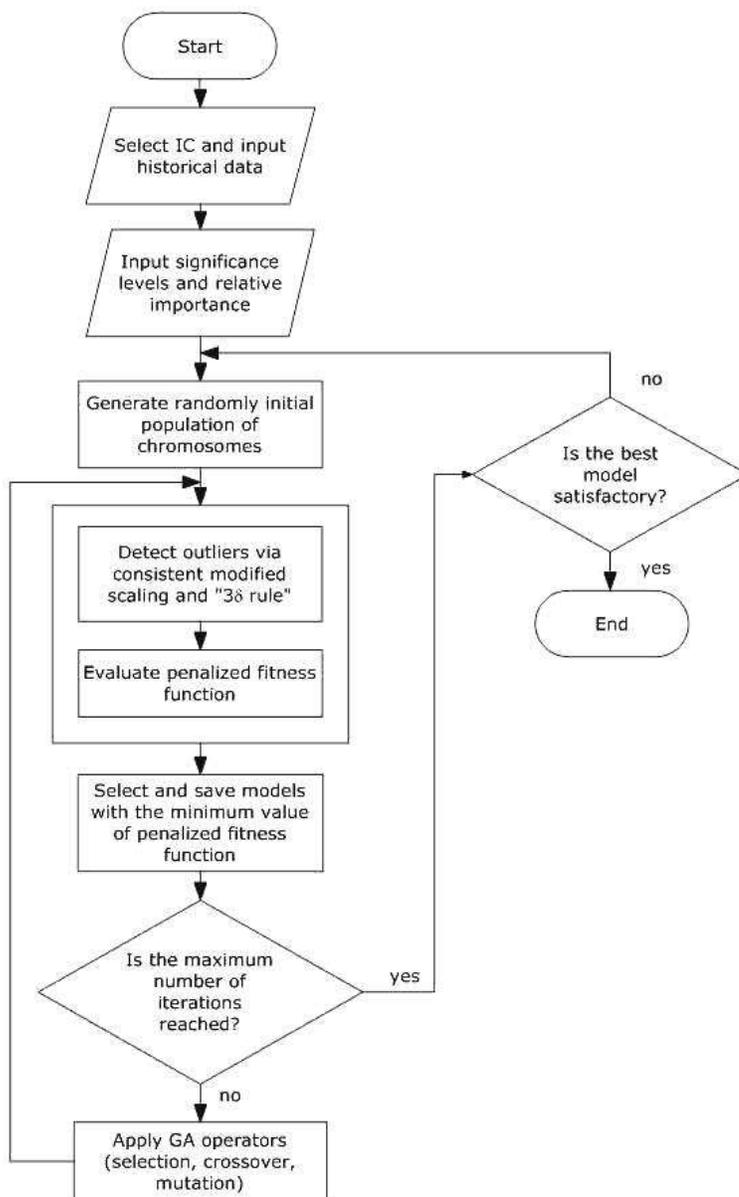
Figura 4.1: Flowchart of the constrained GA-based variable selection and multiple outlier detection algorithm

European Energy sector, defined as $r_t = \log(S_t/S_{t-1}) - \log(P_t/P_{t-1})$, where S_t is the price of MSCI European Energy industry and P_t is the price of MSCI European index with this transformed and augumented data-set of 457 variables. Only data available when a forecast is made are used, and when there is a time lag before a data is announced I use the data from the last obtainable period. I study the effectiveness of the methodology by dividing observations into two subperiods, as in Bossaerts and Hillion (1999): the former provides the estimation sample and corresponds to 80% of the observations, the latter subperiod covers the remaining data and represents the testing sample.

4.3.1 In-sample fitting accuracy measure

The fitting abilities of information criteria to select models in the in-sample tranche are tested with the R^2 adjusted, corrected for the presence of k outliers, that can be presented as below

$$R^2_{adj,k} = 1 - \frac{RSS/(n - (m+k))}{TSS/(n-1)}$$

where RSS is the regression sum of squares (explained deviation) and TSS is total sum of squares (total deviation), both adjusted by their respective degrees of freedom and outliers are handled like independent variables.

4.3.2 Out-of-sample forecasting accuracy measures

I construct out-of-sample forecasts of excess returns using the following recursive approach:

i) the in-sample loadings on the predictive variables suggested by each criterium are used to form the expected excess return forecasts in recursive, step-ahead, out-of-sample periods;

ii) I then expand forward the in-sample end date by one month, reestimate the model and obtain a forecast for the next period. I repeat this process until the end of the out-of-sample period.

The forecasting abilities of models selected by information criteria are then analyzed by five measures commonly used in the forecast performance literature (see Dunis and Williams (2005) and references therein for a complete review): mean error (ME), root mean square error (RMSE), mean absolute error (MAE), Theil's inequality coefficient (U) and correct directional

Performance measure	Description
Mean error	$\text{ME} = \frac{1}{T} \sum_{t=1}^{T} (r_t - \widehat{r}_t)$
Root mean square error	$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (r_t - \widehat{r}_t)^2}$
Mean absolute error	$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |r_t - \widehat{r}_t|$
Theil's inequality coefficient	$\text{U} = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (r_t - \widehat{r}_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (r_t - r_{t-1})^2}}$
Correct directional change	$\text{CDC} = \frac{100}{T} \sum_{t=1}^{T} D_t$
	where $D_t = 1$ if $r_t \widehat{r}_t > 0$ else $D_t = 0$

r_t is the observed excess return at time t
$\widehat{r}_t$ is the forecasted excess return at time t
T is the size of the forecast period

Tabella 4.2: Statistical forecasting accuracy measures

change (CDC), where forecast error is defined as the observed excess return r_t less the forecast $\widehat{r}_t$ at time t. Table 4.2 presents the measures. A positive ME indicates that the forecasts are on average too under-estimated and a negative ME indicates that they are on average over-estimated. The RMSE is computed by taking the square root of the average squared errors and it penalizes large forecast errors relative to smaller ones. The MAE is the average of the forecast errors disregarding their signs. Unlike RMSE and MAE that are scale-dependent measures, U is defined as the ratio between RMSE of the proposed forecasting model and RMSE of a benchmark, in this case the naive model, which assumes that the previous observation is the best predictor of the feature, and results in forecast error estimates independent of the scale of the variables: a value less than one indicates that the forecast is more accurate than the naive alternative. CDC measures the capacity of a model to correctly predict the change of a forecast variable and becomes an important issue in trading strategies that rely on the direction of the forecast rather than its levels.

4.4 Empirical results

In this section I give two examples to demonstrate the practical utility of the proposed methodology in variable selection. In the first application I study the problem of building a model with the best combination of predictors among a preselected pool of informative variables, assuming the true model is in this reduced space. In the second application I analyze the problem of giving a useful approximation to the data generation process from the data-set of all variables lagged 1-month. In addition, the examples are conducted comparing the proposed GA and the simple GA, developed on the plan described in Balcombe (2005).

4.4.1 Example 1

The subset of predictors has been obtained by analyzing correlations of each variable with sector excess returns and selecting the series yielding the highest correlation coefficients at a 5% significance level in the in-sample period with attention to the consistency of their signs with economic theory. The reduced data-set is displayed in Table 4.3 and is constituted by 16 time series from May 1996 to March 2007 for a total of 131 observations, divided in the estimation period that goes from May 1996 to December 2004 (104 observations) and the testing period that goes from January 1997 to March 2007 (27 observations). Transformations of each predictor and their correlations with excess returns are reported in Table 4.4, where it is possible to note that all the correlations are quite low. At the same time a second set of potential predictors has been added including variables that are specific of the Energy sector (i.e. the price of brent, 12-month-forward pro-rata total number of earning per share (EPS) estimates of all companies in Energy sector and weighted 12-month-forward growth in EPS of Energy sector) and US indicators that may condition or lead European market (i.e. US inflation and historical EPS growth of S&P 500 index). As Table 4.5 reveals, there is low correlation among predictors, except between price/book ratios (67%), European CPI with US CPI (45%), European CPI and 12-month-forward of EPS estimates of Energy sector (43%). Another interesting link is between the 3-month EURIBOR and the momentum associated to Europe governments bonds (-45%).

The space of possible linear regression models from a set of 16 predictors and the constant term counts more than 65000 elements, without considering the restrictions in Table 4.1 and the presence of outliers. The proposed GA and the simple GA have been tested by implementing 30 independent

Category	Mnemonic	Variable
financial		
	x_1	EU DS. MARKET DIVIDEND YIELD – BD DISCOUNT RATE / SHORT TERM EURO REPO RATE (MTH. AVG)
	x_2	EU DS. MARKET DIVIDEND YIELD – GERMANY BENCHMARK BOND 10 YR (DS) RED. YIELD
	x_3	EURO RATE 3-month (DS synthetic) offered rate
	x_4	BD BENCHMARK 10 YR DS GOVT INDEX RED. YIELD – BD BENCHMARK 5 YR DS GOVT INDEX RED. YIELD
	x_5	GERMANY BENCHMARK BOND 10 YR (DS) RED. YIELD
	x_6	LEHMAN US CREDIT BOND INDEX RED. YIELD – LEHMAN US AGG. GOVERNMENT RED. YIELD
global economy		
	x_7	EC CONSUMER CONFIDENCE INDICATOR EUROPEAN AGGREGATE (DISC.) SADJ
	x_8	S&P 500 INDEX HIST. EPS GROWTH
	x_9	GSCI PRECIOUS METAL SPOT PRICE INDEX
	x_{10}	Crude oil-brent dated FOB U$/BBL
inflation		
	x_{11}	EM CPI ALL ITEMS (HARMONISED) NADJ
	x_{12}	US CPI ALL ITEMS LESS FOOD & ENERGY (CORE) NADJ
revision and technical		
	x_{13}	MSCI EURP OIL & GAS 12 MTH FWD YOY GROWTH
	x_{14}	MSCI Eurp Oil & Gas REVISION
	x_{15}	MSCI EMU PRICE / BOOK RATIO
	x_{16}	MSCI EUROPE OIL, GAS & C. FUEL PRICE / BOOK RATIO

Tabella 4.3: Data sorted by category for Example 1

Mnemonic	Transform	Lag	Correlation	p-value
x_1	Δ_1	1	0.19	0.06
x_2	Δ_1	3	-0.16	0.10
x_3	Δ_1	3	-0.23	0.02
x_4	Δ_3	3	0.20	0.04
x_5	Δ_3	6	-0.19	0.05
x_6	Δ_6	6	0.23	0.02
x_7	Δ_1	1	0.16	0.10
x_8	Δ_1	6	0.20	0.04
x_9	Δ_3	6	0.22	0.02
x_{10}	Δ_1	1	0.12	$0.22^{\dagger}$
x_{11}	Δ_6	1	0.23	0.02
x_{12}	Δ_6	3	0.24	0.01
x_{13}	$\mathrm{WSMA}(3, -0.4, 2)$	6	0.15	$0.13^{\dagger}$
x_{14}	$\mathrm{WSMA}(3, -0.4, 2)$	6	0.06	$0.54^{\dagger}$
x_{15}	Δ_6	6	-0.16	0.10
x_{16}	Δ_6	6	-0.18	0.08

Tabella 4.4: Correlation between transformed predictors and excess returns in the in-sample period for Example 1

Note. The second column reports the transformation associated with each variable, for the predictor x_i, Δ_k is the difference $x_{i,t} - x_{i,t-k}$ and WSMA(3,-0.4,2) is the Weighted and Simple Moving Average with 3, -0.4 and 2 parameters. Predictors lags are displayed in the third column. In the fourth and fifth columns are reported the correlation of each predictor with excess returns and its significance in the in-sample period respectively. A † superscript denotes variables that have low statistical significance in the estimation period but that are considered of deep economic interest for the Energy sector in the forecasting period.

	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
x_1	-0.14	-0.03	0.12	-0.02	0.02	0.02	-0.06	0.06	0.01	-0.01	-0.08	-0.15	-0.16	0.10	-0.01
x_2		0.18	-0.07	-0.09	-0.13	-0.03	0.09	0.08	0.08	-0.01	-0.18	-0.04	-0.01	-0.05	-0.06
x_3			-0.45	0.33	-0.14	-0.19	0.19	0.08	-0.09	0.05	0.08	0.28	0.21	0.34	0.36
x_4				-0.13	0.18	0.12	-0.08	0.02	0.05	0	0.01	-0.22	-0.17	-0.46	-0.22
x_5					-0.19	-0.23	0.08	0.14	-0.08	-0.12	0	-0.02	0.19	0.22	0.32
x_6						0.01	-0.12	0.14	0.04	0.05	0.16	-0.14	0.09	-0.43	-0.33
x_7							-0.07	0.18	-0.03	0.12	0.09	-0.06	-0.10	-0.11	-0.26
x_8								0.03	0.23	0.04	0.19	0.22	0.13	0.22	0.23
x_9									0.02	0.24	0.12	0.24	-0.15	-0.07	-0.20
x_{10}										0.18	-0.01	0.01	0	0.01	0.01
x_{11}											0.45	0.43	0.08	-0.24	-0.24
x_{12}												0.27	0.24	0.03	-0.05
x_{13}													0.07	0	0.06
x_{14}														0.26	0.30
x_{15}															0.67

Tabella 4.5: Correlation structure of predictors in the estimation period for data considered in Example 1

trials with a population of 100 individuals and 200 generations for each criterium. The suitable forecasting models selected by information criteria for the optimization schemes are depicted in Table 4.6. The findings of implementations are reported in Table 4.7. In general, for the in-sample period, models produced by the proposed GA have a better performance in terms of adjusted R^2 and diagnostic tests, whereas in the out-of-sample period, they behave in a manner similar to those suggested by the simple GA, unless for CDC, for which the simple GA offers more accurate valuations. Analyzing more accurately the results of the proposed GA, all selected models have an outlier corresponding to January 1999; the BIC-based model results to be the smallest one, with only four independent variables, the biggest one is the ICOMP solution, with 10 predictors and AIC suggest a model with 7 indicators. As noted from Table 4.8, their errors pass the constraints on the distributional form, they are normally distributed at a significance level of 1% for the Jarque-Bera's test, so the associated penalties are zero. Comparing in-sample results, BIC and ICOMP behave in a similar manner, with $R^2_{adj,1} = 0.41$, AIC performs quite better, with $R^2_{adj,1} = 0.46$. In the out-of-sample period, the suggested models have a negative ME, indicating a general over-estimate of forecasts. In particular, by considering also RMSE and MAE, for which the optima reach the same values, 0.04 and 0.03 respectively, the best choice is that proposed by AIC, because has the minimum ME (-0.0019). Conversely, Theil's U measure indicates BIC-based model as the best choice with respect to the naive model. From CDC point of view, ICOMP-based model reaches the best performance with 67% of correctly predicted direction changes against only the 48% of BIC and 44% of AIC.

4.4.2 Example 2

The data-set, formed by all variables and their transformations lagged 1-month is considered now, for a total of 150 possible predictors, divided as follow: 72 are financial indicators, 54 describe global economy, 5 are inflation indexes, 1 represents revision for the Energy sector, defined as the ratio of the difference between 12-month-forward number of EPS estimates up with 12-month-forward number of EPS estimates down and the total number of 12-month-forward EPS estimates, and 18 measures are aggregate earnings for European market and sector. Each predictor has 136 observations, from December 1996 to March 2007, divided into two groups: from December 1996 to November 2004 (108 observations), that represents the estimation period and from December 1996 to March 2007 (26 observations), that consitues the testing period.

	simple GA			Proposed GA		
	AIC	BIC	ICOMP	AIC	BIC	ICOMP
x_1	1	0	0	1	1	0
x_2	0	0	0	0	0	1
x_3	1	1	1	1	1	1
x_4	0	0	0	0	0	0
x_5	0	0	1	0	0	1
x_6	1	1	1	1	0	1
x_7	0	0	0	0	0	0
x_8	1	1	1	1	1	1
x_9	1	1	1	1	0	1
x_{10}	0	0	0	0	0	0
x_{11}	0	0	1	1	1	1
x_{12}	1	0	1	0	0	1
x_{13}	1	0	1	0	0	1
x_{14}	0	0	0	1	0	0
x_{15}	0	0	0	0	0	1
x_{16}	0	0	0	0	0	1

Tabella 4.6: Model choices over the estimation period with respect to the three selection criteria for the simple and the proposed GA in Example 1

			Selection Criteria		
Criterium	variables	dummies	AIC	BIC	ICOMP
simple GA					
AIC	7	–	$-\mathbf{345.4}$	-324.2	-351.4
BIC	4	–	-340.1	$-\mathbf{326.9}$	-345.3
ICOMP	8	–	-338.1	-314.3	$-\mathbf{360.3}$
proposed GA					
AIC	7	1	$-\mathbf{1118.6}$	-1092.2	-1119.8
BIC	4	1	-1113.8	$-\mathbf{1095.3}$	-1118.1
ICOMP	10	1	-1105.5	-1068.5	$-\mathbf{1135.1}$

Tabella 4.7: Characteristics of the solutions and values attained by the information criteria, applying the simple and proposed GA respectively, in Example 1

| Criterium | Diagnostic tests | | | | | |
| | Breush-Pagan | | Durbin-Watson | | Jarque-Bera | |
	statistic	p-value	statistic	p-value	statistic	p-value
simple GA						
AIC	11.1375	0.9158	2.2832	0.3127	35.3647	< 0.001
BIC	8.9344	0.9698	2.3225	0.1643	18.5842	0.0040
ICOMP	32.9518	1	2.3883	0.1592	22.1262	0.0026
proposed GA						
AIC	12.9561	0.9268	2.0630	0.8987	3.2037	0.1299
BIC	6.8379	0.8553	1.9407	0.6031	3.1944	0.1306
ICOMP	23.3492	0.9842	2.0479	0.5941	3.1572	0.1334

Tabella 4.8: Diagnostic tests for models selected by information criteria with the simple and the proposed GA in Example 1

| Criterium | Fit measure | Forecast measures | | | | |
	$R^2_{adj,k}$	ME	RMSE	MAE	U	CDC
simple GA						
AIC	0.2491	-0.0150	0.0393	0.0296	0.7849	70%
BIC	0.1883	-0.0082	0.0376	0.0283	0.7499	48%
ICOMP	0.2013	-0.0127	0.0394	0.0299	0.7862	70%
proposed GA						
AIC	0.4576	-0.0019	0.0379	0.0295	0.7568	44%
BIC	0.4124	-0.0047	0.0354	0.0288	0.7072	48%
ICOMP	0.4108	-0.0136	0.0383	0.0288	0.7639	67%

Tabella 4.9: In-sample fitting and out-of-sample forecasting results of the Energy excess returns obtained by the simple and the proposed GA in Example 1

I undertake the problem of giving an approximation to the excess return process with one variable per category, for a total of five predictors. This choice may not be the best in prediction terms, but it allows to identify the determinants of movements of excess returns in each economic category. Because the resultant space is very large (there are about 350.000 possible configurations), with multiple optima and disjointed regions, a reasonable way to tackle this search is to seek a local optimum via more chromosomes and generations than the previous example. Following this approach, I defined 100 simulations of the proposed GA and the simple GA respectively, with 50 chromosomes for 200 generations, then a new implementation of each procedure, where the starting population is constituted by the previous 100 solutions, is done. Final models are reported in Table 4.10: comparing the results relative to the two plans, one can conclude that models suggested by the proposed algorithm are better even if forecast accuracies are similar because they satisfy more appropriately fit measure and diagnostic tests. In particular, focusing on the choices made by the proposed GA, the first two criteria have suggested the same model whereas the ICOMP-based choice differs from the previous one only for the financial component by considering German bond business expectations in place of the momentum variable (the difference between the dividend yield of European market and European BDP rate). All the solutions have an outlier in correspondence of January 1999, pass the diagnostic tests and present similar values for AIC and BIC (see Tables 4.11 and 4.12). The only real difference is in ICOMP criterium: the first model reaches -1161.4 and the second -1171.3, suggesting a more complex correlation structure among predictors in the first regression. This is much more evident if one considers the complexity measure in (4.5), for which the values are 5.14 and -0.48 respectively. In the estimation and in the testing periods they behave in a similar manner as displayed in Table 4.13 with a major capability to capture the forecast movements of excess returns for the ICOMP-based choice, equal to 61% with respect to the 54% of the other model.

4.5 Conclusions

In this chapter, I developed a new automated model search based on information criteria and genetic algorithms to produce linear dynamic models satisfying usual diagnostic constraints on errors. The penalized function approach was used to represent statistical tests and model performances like a constrained evolutionary optimization problem. At the same time a multiple

Criterium	Variable	Transform
simple GA		
AIC & BIC		
	EU DS. MARKET DIVIDEND YIELD – BD DISCOUNT RATE / SHORT TERM	Δ_1
	GERMAN MARK TO US $ (GTIS) - EXCHANGE RATE	Δ_6
	EM CPI ALL ITEMS (HARMONISED) NADJ	Δ_6
	MSCI Eurp Oil & Gas REVISION	$WSMA(3, -0.4, 2)$
	MSCI EUROPE - CAL FY0 WTD EPS	Δ_1
ICOMP		
	BD INDUSTRIAL PRODUCTION INCLUDING CONSTRUCTION VOLA	Δ_1
	GSCI Industrial Metals Spot - PRICE INDEX	Δ_1
	EM CPI ALL ITEMS (HARMONISED) NADJ	Δ_6
	MSCI Eurp Oil & Gas REVISION	$WSMA(3, -0.4, 2)$
	MSCI EUROPE - CAL FY0 WTD EPS	Δ_3
proposed GA		
AIC & BIC		
	EU DS. MARKET DIVIDEND YIELD – BD DISCOUNT RATE / SHORT TERM	Δ_1
	S&P 500 COMPOSITE PRICE INDEX / MSCI EUROPE PRICE INDEX	Δ_1
	EM CPI ALL ITEMS (HARMONISED) NADJ	Δ_6
	MSCI Eurp Oil & Gas REVISION	$WSMA(3, -0.4, 2)$
	MSCI EUROPE - CAL FY0 WTD EPS	Δ_3
ICOMP		
	BD BUSINESS EXPECTATIONS (PAN GERMANY) SADJ	Δ_1
	S&P 500 COMPOSITE PRICE INDEX / MSCI EUROPE PRICE INDEX	Δ_1
	EM CPI ALL ITEMS (HARMONISED) NADJ	Δ_6
	MSCI Eurp Oil & Gas REVISION	$WSMA(3, -0.4, 2)$
	MSCI EUROPE - CAL FY0 WTD EPS	Δ_3

Tabella 4.10: Variables selected by information criteria for the simple and proposed GA in Example 2

	Dummies	Selection Criteria		
Criterium	k	AIC	BIC	ICOMP
simple GA				
AIC & BIC	–	**−344.1**	**−328.0**	-345.1
ICOMP	–	-343.9	-327.9	**−356.1**
proposed GA				
AIC & BIC	1	**−1155.7**	**−1134.3**	-1161.4
ICOMP	1	-1154.4	-1132.9	**−1171.3**

Tabella 4.11: Characteristics of the solutions and values attained by the information criteria, applying the simple and proposed GA respectively, in Example 2

	Diagnostic tests					
Criterium	Breush-Pagan		Durbin-Watson		Jarque-Bera	
	statistic	p-value	statistic	p-value	statistic	p-value
simple GA						
AIC & BIC	14.6335	0.9945	2.2227	0.4224	36.1626	< 0.001
ICOMP	11.6267	0.9796	2.1430	0.6547	28.3599	0.0013
proposed GA						
AIC & BIC	8.1971	0.8543	1.7928	0.1589	1.8707	0.3096
ICOMP	9.4480	0.9075	1.8602	0.2828	0.3949	> 0.5

Tabella 4.12: Diagnostic tests for models selected by information criteria with the simple and the proposed GA in Example 2

	Fit measure	Forecast measures				
Criterium	$R^2_{adj,k}$	ME	RMSE	MAE	U	CDC
simple GA						
AIC & BIC	0.0962	-0.0274	0.4615	0.2485	9.2359	57%
ICOMP	0.0951	-0.0041	0.0348	0.0274	0.6955	61%
proposed GA						
AIC & BIC	0.3460	-0.0088	0.0356	0.0293	0.7134	54%
ICOMP	0.3379	-0.0098	0.0387	0.0296	0.7362	61%

Tabella 4.13: In-sample fitting and out-of-sample forecasting results of the Energy excess returns obtained by the simple and the proposed GA in Example 2

outlier detection method, presented by Chiang et al. (2003), was included to improve predictive accuracy.

Two examples of the applicability of the algorithm to real problems are described and its performance is compared to the GA-plan developed by Balcombe (2005), following the purpose of producing short-term forecasts for the excess returns of the MSCI Europe Energy sector on the MSCI Europe index. In the first application, a reduced pool of potential informative predictors was selected on the basis of the correlation between dependent and independent variables and the undertaken problem was to detect the model with the best dimension and its constituents. In the second application, I considered all predictors lagged 1-month and the objective was to give an approximation to the excess return process with a linear model of fixed dimension. Furthermore, I investigated predictability performances of the different versions of the best model suggested by information criteria with a recursive estimation-window approach. Results show that the compared procedures behave in a similar manner in the out-of-sample period, but, differently from the simple GA, the proposed plan is able to detect models satisfying decision maker assumptions and outliers in the in-sample period, improving consistency and versatility of the econometric modeling. Moreover, a remarkable capability to predict excess returns of the Energy sector may be seen.

This work presents promising initial efforts on searching efficiently models with several constraints in complex solution spaces with high dimensions in financial problems of portfolio selection and econometric modeling.

Bibliografia

Ahumada, H. A.: 1985, An Encompassing Test of Two Models of the Balance of Trade for Argentina, *Oxford Bulletin of Economics and Statistics* **47**(1), 51–70.

Akaike, H.: 1973, Information theory and an extension of the maximum likelihood principle, *in* P. B. N. and C. F. (eds), *Second international symposium on information theory*, Academial Kiado, Budapest, pp. 267–281.

Avramov, D.: 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* **64**, 423–458.

Azadeh, A., Ghaderi, S. F., Tarverdian, S. and Saberi, M.: 2007, Integration of artificial neural networks and genetic algorithm to predict electrical

energy consuption, *Applied Mathematics and Computation* **186**, 1731–1741.

Balcombe, K. G.: 2005, Model Selection Using Information Criteria and Genetic Algorithms, *Computational Economics* **25**, 207–228.

Bossaerts, P. and Hillion, P.: 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, *The Review of Financial Studies* **12**(2), 405–428.

Bozdogan, H.: 1988, *Classification and Related Methods in Data Analysis*, North-Holland, Amsterdam, chapter ICOMP: A New Model Selection Criterion, pp. 599–608.

Bozdogan, H.: 2000, Akaike's Information Criterion and Recent Developments in Information Complexity, *Journal of Mathematical Psychology* **44**, 62–91.

Campos, J. and Ericsson, N. R.: 1999, Constructive Data Mining: Modeling Consumers' Expenditure in Venezuela, *Econometrics Journal* **2**(2), 226–240.

Campos, J., Ericsson, N. R. and Hendry, D. F.: 2005a, *General-to-Specific Modeling*, Edward Elgar, Cheltenham.

Campos, J., Ericsson, N. R. and Hendry, D. F.: 2005b, General-to-specific Modeling: An Overview and Selected Bibliography, *International Finance Discussion Papers 838*, Board of Governors of the Federal Reserve System.

Chan, L. K. C., Jegadeesh, N. and Lakonishok, J.: 1996, Momentum Strategies, *The Journal of Finance* **51**(5), 1681–1713.

Chiang, L. H., Pell, R. J. and Seasholtz, M. B.: 2003, Exploring process data with the use of robust outlier detection algorithms, *Journal of Process Control* **13**, 437–449.

Costantini, M. and Pappalardo, C.: 2008, Combination of Forecast Methods Using An Algorithm-Based Procedure, *Economic Series 228*, Institute for Advanced Studies, Vienna.

Cremers, K. J. M.: 2002, Stock Return Predictability: A Bayesian Model Selection Perspective, *The Review of Financial Studies* **15**(4), 1223–1249.

Deb, K.: 2001, *Multi-Objective Optimization Using Evolutionary Algorithms*, Wiley, New York.

Dorsey, R. E. and Mayer, W.: 1995, Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features, *Journal of Business and Economic Statistics* **13**, 53–66.

Dunis, C. L. and Williams, M.: 2005, *Applied Quantitative Methods for Trading and Investment*, Wiley Finance Series, Wiley, chapter Applications of Advanced Regression Analysis for Trading and Investment, pp. 1–40.

George, E. I.: 2000, The Variable Selection Problem. University of Texas at Austin.

George, E. I. and McCulloch, R. E.: 1993, Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association* **88**(423), 881–889.

George, E. I. and McCulloch, R. E.: 1997, Approaches for bayesian variable selection, *Statistica sinica* **7**, 339–373.

Gilbert, C. L.: 1986, Professor Hendry's Econometric Methodology, *Oxford Bulletin of Economics and Statistics* **48**(3), 283–307.

Goldberg, D. E.: 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, New York.

Grenouilleau, D.: 2004, A sorted leading indicators dynamic (SLID) factor model for short-run euro-area GDP forecasting, *Economic Paper 219*, Directorate-General for Economic and Financial Affairs, European Economy.

Grenouilleau, D.: 2006, The Stacked Leading Indicators Dynamic Factor Model: A Sensitivity Analysis of Forecast Accuracy using Bootstrapping, *Economic Paper 249*, Directorate-General for Economic and Financial Affairs, European Economy.

Hady, A. and Siminoff, J.: 1992, Comments on Paul and Fung (1991), *Technometrics* **34**(3), 373–374.

Harvey, D. I., Leybourne, S. and Newbold, P.: 1998, Tests for forecast encompassing, *Journal of Business and Economic Statistics* **26**, 254–259.

Hasheminia, H. and Niaki, S. T. A.: 2006, A genetic algorithm approach to find the best regression/econometric model among the candidates, *Applied Mathematics and Computation* **183**, 337–349.

Hendry, D. F.: 1983, Econometric Modelling: The 'Consumption Function' in Retrospect, *Scottish Journal of Political Economy* **30**(3), 193–220.

Hendry, D. F.: 1995, *Dynamic Econometrics*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Hendry, D. and P.Clements, M.: 2002, Pooling of Forecasts, *Econometrics Journal* **5**, 1–26.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T.: 1998, Bayesian Model Averaging, *Technical report 9814*, Department of Statistics, Colorado State University.

Hoeting, J. A., Raftery, A. E. and Madigan, D.: 1996, A method for simultaneous variable selection and outlier identification in linear regression, *Journal of Computational Statistics* **22**, 251–271.

Holland, J. H.: 1992, *Adaptation in Natural and Artificial Systems*, second edn, University of Michigan Press.

Jegadeesh, N.: 1990, Evidence of Predictive Behavior of Security Returns, *The Journal of Finance* **45**(3), 881–898.

Kaboudan, M. A.: 2000, Genetic Programming Prediction of Stock Prices, *Computational Economics* **16**, 207–236.

Koza, J.: 1992, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge, MA.

Leontitsis, A. and Pange, J.: 2004, WSMA: in between weighted and simple average, 17th Annual Pan-Hellenic Conference on statistics, Leukada, Greece, pp. 519–526.

MacDonald, R. and Taylor, M. P.: 1992, A stable us money demand function, 1874–1975, *Economics Letters* **39**(2), 191–198.

Michalewitz, Z.: 1994, *Genetic Algorithms + Data Structure = Evolution Program*, Springer-Verlang, New York.

Miller, A. J.: 1990, *Subset Selection in Regression*, Chapman and Hall, New York.

Pesaran, M. H. and A.Timmermann: 2003, How Costly is it to Ignore Breaks when Forecasting the Direction of a Time Series? DAE Working Paper No. 0306.

Pesaran, M. H. and Pesaran, B.: 1997, *Working with Microfit 4.0*, Oxford University Press.

Pesaran, M. H. and Timmermann, A.: 1995, Predictability of Stock Returns: Robustness and Economic Significance, *The Journal of Finance* **50**(4), 1201–1228.

Pesaran, M. H. and Timmermann, A.: 2000, A Recursive Modelling Approach to Predicting UK Stock Returns, *The Economic Journal* **110**, 159–191.

Pesaran, M. H. and Timmermann, A.: 2004, Real Time Econometrics, *Discussion Paper 1108*, IZA.

Pynnönen, S.: 1992, Detection of Outliers in Regression Analysis by Information Criteria, *Discussion Paper 146*, University of Vaasa.

Rapach, D. E. and Strauss, J. K.: 2008, Forecasting US Employment Growth Using Forecast Combining Methods, *Journal of Forecasting* (27), 75–93.

Runarsson, T. P. and Yao, X.: 2002, *Evolutionary Optimization*, Vol. 48, Kluwer Academic Publisher, USA, chapter Constrained Evolutionary Optimization - the penalty function approach, pp. 87–113.

Schwarz, G.: 1978, Estimating the dimension of a model, *Annals of Statistics* **6**, 461–464.

Smith, A. E. and Coit, D. W.: 1997, *Handbook of Evolutionary Computation*, Oxford University Press, chapter Penalty Functions, pp. C5.2:1–6.

Stock, J. H. and Watson, M. W.: 2006, *Handbook of Economic Forecasting*, Vol. 1, North-Holland, chapter Forecasting with many predictors, pp. 515–554.

Tolvi, J.: 2004, Genetic algorithms for outlier detection and variable selection in linear regression models, *Soft Computing* **8**, 527–533.

<div style="text-align: right; font-size: 4em; color: gray;">**5**</div>

Investment using evolutionary learning methods and technical rules

In this chapter, I propose a genetic learning approach to generate technical trading systems for stock timing. The most informative technical indicators are selected from a set of almost 5000 signals by a multi-objective genetic algorithm with variable string length. Successively, these signals are combined into a unique trading signal by a learning method. I test the expert weighting solution obtained by the plurality voting committee, the Bayesian model averaging and Boosting procedures with data from the the S&P 500 Composite Index, in three market phases, up-trend, down-trend and sideways-movements, covering the period 2000–2006. Computational results indicate that the near-optimal set of rules varies among market phases but presents stable results and is able to reduce or eliminate losses in down-trend periods.

5.1 Introduction

A major concern in investment decisions is to exploit short-term market opportunities such as trend spotting and momentum. A method that is widely used by practitioners to this end is technical analysis, or more generally speaking, technical trading strategies, which attempts to forecast future movements by analyzing past prices and volumes, focusing on particular charts and patterns that are supposed to have predictive value and dismissing exogenous factors, such as political events, fiscal policies or economic environments. For example, Covel (2005) and Faber (2007) deal with the application of trend following methodologies to financial markets, whereas Jegadeesh (1990) and Chan et al. (1996) take into account price momentum, finding that the profitability of strategies based on momentum is not due to their systematic risk or to delayed stock price reactions to common factors. Trend following and momentum trading systems work because markets ex-

hibit momentum due to a cognitive bias, i.e. undereaction and overeaction at different timescales, as described in Kahneman and Tversky (1979): humans have an irrational tendency to be less willing to gamble with profits than with losses, thus investors tend to sell their winners too early and hold on to losers too long.

Since technical analysis has a visual nature and refers mainly to human expectations and attitudes, it has hardly been accepted in the academic community. The first results by Alexander (1964) during the 1960s and successively by Fama and Blume (1966) and Fama (1970) during the 1970s supported the efficient market hypothesis, which states that all relevant information is contained in current prices, i.e. prices follow Markov processes. Conversely, some recent results give empirical evidence on the potential of this methodology: for example, Neftci (1991) has showed that the moving average rule can detect useful information in nonlinear time series and in particular in financial markets, whereas Brock et al. (1992) have developed simple technical trading rules based on moving averages and support and resistance that may be applied to the Dow Jones Industrial Average to predict stock price changes.

From these seminal works a considerable amount of academic literature have been developed in the last 20 years to improve the existent technical trading systems or create new ones. Two main approaches have been followed. On one hand, by using some combinations of simple indicators, such as trend indicators and moving averages, to dispatch trading signals and, on the other hand, by considering artificial intelligence techniques to identify optimal trading rules.

To the first class belong the CRISMA trading system developed by Pruitt and White (1988) and the trading rules described by Gençay (1999). The CRISMA trading system is based on three components: a moving average indicator, a relative strength indicator and the cumulative volume of transactions. This system showed positive returns over a 10-year period, from 1976 to 1985, accounting for 2% of transaction costs. The Gençay trading rule are based on non-parametric models which maximize the total return of an investor strategy. These rules provided a significant forecast profitability over the random walk model for the period from 1973 to 1992 using daily spot rates for the British pound, Deutsche mark, French franc, Japanese yen and Swiss franc.

To the second class belong automated trading systems that embed genetic algorithms, neural networks and expert systems (the interested reader may refer to Trippi and Turban (1990) for a detailed review of these methodologies). In particular, the problem of parameter optimization in technical

trading rules in order to avoid data-snooping and survivorship bias has been solved with a genetic algorithm in Fernandez-Rodríguez et al. (2005). These authors predict the General Index of Madrid Stock Market over an out-of-sample test period of 2188 daily observations, from 16 December 1988 to 15 November 1997. This approach showed a remarkable profitability on the classical buy-and-hold strategy.

The problem of automatically generate new trading rules is tackled for the first time in Allen and Karjalainen (1999), where a genetic programming approach is applied to the S&P 500 with daily prices from 1928 to 1995. However, these rules were not consistently better than the buy-and-hold strategy in the out-of-sample test periods. A similar approach, based on genetic programming, has been considered in Potvin et al. (2004) to analyze 14 Canadian companies listed on the Toronto Stock Exchange Market, including daily stock prices and transaction volumes from 30 June 1992 to 30 June 2000. The results suggests that trading rules generated by genetic programming are useful when the market falls or is quite stable and are not beneficial on the buy-and-hold strategy when the market is rising.

Expert systems generated by a genetic algorithm are proposed in Korczak and Roger (2002) to analyze 24 of the most important stocks of the CAC 40 Index. These authors used three years of daily data, from 1997 to 1999, with transaction costs of 0.25%. On each stock 10 experiments with random beginning dates were performed, where the training period was 261 days long and the test period 7-days long. The results suggest that genetic algorithms are capable of efficiently extract the relevant indicators and function parameters in a short-term trading perspective.

The system proposed in Dourra and Siy (2002) uses technical indicators with fuzzy logic to create a fuzzy indicator that recommends the position on the market. The method examined various companies from 1995 to 1999, considering different rules in terms of risks and trends of the stock price, and proved to be effective.

Kwon and Moon (2007) presents a hybrid neurogenetic system with a context based ensemble method of neural networks which dynamically changes relative to the test day's context. The procedure was tested on 36 companies in NYSE and NASDAQ from 1992 to 2004, showing on average a notable improving over the buy-and-hold strategy.

The applicability of machine learning methods to formulate trading strategies using technical indicators has grown considerably in the last years. A non-parametric kernel regression for technical pattern recognition is developed in Lo et al. (2000) in order to forecast a large number of U.S. stocks for the period 1962–1996. Results show that technical indicators provide

incremental information comparing the unconditional empirical distribution of daily stocks returns to the conditional distribution on specific technical indicators. In Andrada-Félix and Fernandez-Rodríguez (2008) learning algorithms are used to increase the accuracy of the prediction by averaging the decision of a set of variable moving averages: boosting, Bayesian model averaging and committee methods are compared for predicting the NYSE Composite Index from January 1993 to December 2002 and have supplied better out-of-sample performance than any single moving average.

In this chapter, I predict the market price using an evolutionary learning ensemble approach, where a wide range of different types of technical indicators is considered as input variables, from which a genetic algorithm selects the more profitable signals and a learning machine generate the trading expert system.

Inspired by Andrada-Félix and Fernandez-Rodríguez (2008), the capabilities of three learning methods are compared (boosting, Bayesian model averaging and committee) in different phases of the market: up-trend, downtrend and sideways-movements.

With respect to the aforementioned literature, where the stock timing is predicted by considering a single economic or statistical objective, this chapter introduces a learning-based multi-objective genetic algorithm that takes into account financial and statistical criteria at the same time. Since the algorithm generates linear combinations of classical technical indicators, the proposed solutions are easier to be interpreted than those suggested by other artificial intelligence techniques, such as neural networks and genetic programming. Moreover, in comparison to the solutions generated by classical pooling forecasting methods, where hundreds of technical indicators are combined, admitting a variable length chromosome representation makes possible a reduction in the complexity of solutions, since the algorithm is able to detect the more informative technical indicators. The results in terms of predicability are in line with the classical literature.

The remainder of the chapter is organized as follow. Section II gives a brief overview of the types of technical indicators adopted and illustrates the trading rules they generate. Section III describes the employed methodologies, section IV presents the experimental results and discussions. Finally, Section V concludes the chapter with a summary and some remarks.

5.2 Trading rules

5.2.1 Technical indicators

The inputs to the algorithm are some of the most common technical indicators considered in academic studies and technical analysis literature (see, for example, Murphy (1998) and Sullivan et al. (1999)).

- RATE OF CHANGE MOMENTUM INDICATOR
  The rate of change (ROC) indicator represents the speed at which a variable changes over a specific period of time. In this study, it is calculated as the ratio between the current closing price P_t and the closing price n days in the past P_{t-n-1}, i.e.

$$\text{ROC}(n)_t = \frac{P_t}{P_{t-n-1}}.$$

  The trading rule at time t associated with this indicator is

$$\begin{cases} Buy, & \text{if } \text{ROC}(n)_{t-1} \leq 1 \text{ and } \text{ROC}(n)_t > 1; \\ Sell, & \text{if } \text{ROC}(n)_{t-1} \geq 1 \text{ and } \text{ROC}(n)_t < 1; \\ Hold, & \text{otherwise.} \end{cases}$$

- MOVING AVERAGE INDICATORS
  A moving average is a mean value calculated over a previous rolling period of fixed length n. I use two types of moving averages for a rolling window of length n at time t: the simple moving average (SMA), defined by

$$\text{SMA}(n)_t = \sum_{i=0}^{n-1} \frac{1}{n} P_{t-i}$$

  and the weighted moving average (WMA), calculated as

$$\text{WMA}(n)_t = \sum_{i=0}^{n-1} \frac{n-i}{\widehat{n}} P_{t-i}.$$

  where $\widehat{n} = \sum_{j=1}^{n-1} j$. Further, I consider the exponential moving average (EMA), expressed by

$$\text{EMA}(n)_t = \frac{1}{n} P_t + \left(1 - \frac{1}{n}\right) \text{EMA}(n)_{t-1}$$

$\text{EMA}_0(n) = P_0$ and 0 is a reference date.

Generally speaking, a moving average indicator uses the down crossing of a shorter moving average with respect to a longer moving average as a buy signal and the crossing in the opposite direction as a sell signal. I focus on three variations of this definition: the variable-length moving average (VMA) rule, applied for example in Brock et al. (1992), a variation of the weighted and simple moving average (WSMA) rule, proposed in Leontitsis and Pange (2004), and the moving average convergence divergence (MACD) rule, described in detail in Fusai and Roncoroni (2008).

The VMA at time t is the difference between a n_1 days SMA (shorter) and a n_2 days SMA (longer):

$$\text{VMA}(n_1, n_2)_t = \text{SMA}(n_1)_t - \text{SMA}(n_2)_t$$

with $n_1 < n_2$.

The WSMA is a triple smoothed linear combination of the difference of a n_1 WMA and a n_2 SMA and is expressed by

$$GD(n_1, n_2)_t \quad = \quad (1+v)\,\text{WMA}(n_1)_t - v\,\text{SMA}(n_2)_t$$

with $n_1 < n_2$ and $v = 0.4$. The WSMA is obtained by applying twice the procedure used to compute GD (see Leontitsis and Pange (2004) for the exact definition). I consider a generalized WSMA (gWSMA) that can be obtained by allowing different number of linear combinations and different choices of v.

The MACD indicator combines two EMA of past prices:

$$\text{MACD}(n_1, n_2)_t = \text{EMA}(n_1)_t - \text{EMA}(n_2)_t$$

with $n_1 < n_2$. It is a trend follower procedure that performs better during strong trending periods and, conversely, tends to lose money during periods of choppy trading. In connection with the MACD, a trigger signal SL, expressed as a k period EMA of the MACD, is also used to obtain the MACD histogram (MACDH) indicator, defined by

$$\text{MACDH}_t = \text{MACD}_t - \text{SL}_t,$$

which highlights variations in the spread between fast and slow signals. Denoting with MA the generic moving average indicator, the trading

rule is

$$\begin{cases} Buy, & \text{if } \text{MA}(n)_{t-1} \leq 0 \text{ and } \text{MA}(n)_t > 0; \\ Sell, & \text{if } \text{MA}(n)_{t-1} \geq 0 \text{ and } \text{MA}(n)_t < 0; \\ Hold, & \text{otherwise.} \end{cases}$$

- DYNAMIC SUPPORT/RESISTANCE INDICATOR

  The support/resistance (S&R) indicator represents key market junc-
  tures where supply and demand meet. In particular, support is the
  price level at which buyers are expected to enter the market and the
  price stops going down, whereas resistance is the price level at which
  sellers are expected to enter the market and the price stops going up.
  The dynamic S&R (dS&R) takes into account price levels and their
  volatility, expressed through the standard deviation σ on the last n
  days. It is computed by taking 2σ above and 2σ below the n days
  price average. The dynamic support and resistance at time t relative
  to the last n observations are respectively

$$\begin{aligned} \text{Sup}(n)_t &= \text{SMA}(n)_t - 2\sigma(n)_t \\ \text{Res}(n)_t &= \text{SMA}(n)_t + 2\sigma(n)_t \end{aligned}$$

  where $\text{SMA}(n)_t$ is the n days moving average at time t and σ_t is its
  standard deviation, i.e.

$$\sigma(n)_t = \sqrt{\frac{\sum_{i=t-n}^{t}(P_i - \text{SMA}(n)_t)^2}{n}}$$

  The trading rule developed on the dS&R indicator is

$$\begin{cases} Buy, & \text{if } P_{t-1} \leq \text{Sup}(n)_{t-1} \text{ and } P_t > \text{Sup}(n)_t; \\ Sell, & \text{if } P_{t-1} \geq \text{Res}(n)_{t-1} \text{ and } P_t < \text{Res}(n)_t; \\ Hold, & \text{otherwise.} \end{cases}$$

- STOCHASTIC MOMENTUM INDICATOR

  The stochastic (K&D) indicator is a momentum oscillator intended to
  help determine the strength of price trends and to highlight potential
  short term market overbought and oversold levels. It has two main
  variables, %K and %D, defined at each time t as:

$$\%\text{K}(n)_t = 100\frac{P_t - P_t^{min}(n)}{P_t^{max}(n) - P_t^{min}(n)},$$

where $P_t^{min}(n)$ and $P_t^{max}(n)$ are the minimum and maximum closing prices over a n days period, respectively, and

$$\%\text{D}_t = \sum_{i=1}^{3} \frac{\%\text{K}(n)_{t-i}}{3}, \, n \leq 3$$

is the 3 days moving average of %K. In this case the trading signal becomes

$$\begin{cases} Buy, & \text{if } \%\text{K}(n)_{t-1} \leq \%\text{D}_{t-1} \text{ and } \%\text{K}(n)_t > \%\text{D}_t; \\ Sell, & \text{if } \%\text{K}(n)_{t-1} \geq \%\text{D}_{t-1} \text{ and } \%\text{K}(n)_t < \%\text{D}_t; \\ Hold, & \text{otherwise.} \end{cases}$$

- ON BALANCE VOLUME INDICATOR
  The on balance volume (OBV) represents the flow of volume in a stock and is calculated as a running cumulative total of the daily volume transactions, adding the amount of daily volume when the closing price increases, and subtracting the daily volume when the closing price decreases:

$$\text{OBV}_t = \begin{cases} \text{OBV}_{t-1} - \text{Vol}_t, & \text{if } P_t < P_{t-1} \\ \text{OBV}_{t-1}, & \text{if } P_t = P_{t-1} \\ \text{OBV}_{t-1} + \text{Vol}_t, & \text{if } P_t > P_{t-1} \end{cases}$$

where Vol_t is the volume at day t.
The rule employed for trading is obtained by comparing the OBV level with the simple moving average on the last n days of the OBV itself:

$$\begin{cases} Buy, & \text{if } \text{OBV}_{t-1} \leq \text{SMA}(n)_{t-1} \text{ and } \text{OBV}_t > \text{SMA}(n)_t; \\ Sell, & \text{if } \text{OBV}_{t-1} \geq \text{SMA}(n)_{t-1} \text{ and } \text{OBV}_t < \text{SMA}(n)_t; \\ Hold, & \text{otherwise.} \end{cases}$$

- EASE OF MOVEMENT VALUE
  The ease of movement value (EMV) is a momentum indicator which attempts to identify the amount of volume required to move prices and is calculated as

$$\text{EMV}_t = \frac{\text{Midpoint}_t - \text{Midpoint}_{t-1}}{\text{BoxRatio}_t}$$

where

$$\text{Midpoint}_t = \frac{P_t^{high} + P_t^{low}}{2}$$

$$\text{BoxRatio}_t = \frac{\text{Vol}_t / 10000}{P_t^{high} - P_t^{low}},$$

Indicator	Parameter
ROC	$n = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60$
VMA	$n_1 = 1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 50,$ $75, 100, 125, 150$ $n_2 = 25, 10, 15, 20, 25, 30, 35, 40, 50, 75,$ $100, 125, 150,$ with $n_2 \geq n_1$
gWSMA	$n_1 = 5, 10, 15$ $n_2 = 30, 45, 60$ $v = -0.4, -0.2, 0, 0.2, 0.4$ $T = 1, 3$
MACDH	$n_1 = 8, 9, 10, 11, 12$ $n_2 = 17, 20, 23, 26, 29$ $k = 8, 9, 10$
dS&R	$n = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60$
K&D	$n = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60$

Tabella 5.1: Parameterizations of the technical indicators

with P_t^{high} and P_t^{low} indicating the highest and lowest price in day t. In defining BoxRatio$_t$, I follow the custom in literature to scale the volume by dividing by 10000 to overcome flat signals. The investment signal is:

$$\begin{cases} Buy, & \text{if EMV}_t > 0; \\ Sell, & \text{if EMV}_t < 0; \\ Hold, & \text{otherwise.} \end{cases}$$

The parameterizations I use for the six groups of technical indicators are reported in Table 5.1.

5.2.2 Filters

In order to avoid deceptive signals and to reduce transactions costs, I include three classical types of filters applied to each investment signal:

- the *fixed percentage band* filter, that requires the buy or sell signal to exceed a fixed amount;

- the *time delay* filter, that requires the buy or sell signal to remain valid for a predefined number of days before action is taken;

Filter	Parameter
fixed percentage band	$p = 0, 0.1\%, 0.5\%, 1\%, 1.5\%,$ $2\%, 3\%, 4\%, 5\%$
time delay	$d = 2, 3, 4, 5$
holding time	$h = 5, 10, 15$

Tabella 5.2: Parameterizations of the three filter employed

- the *holding time* filter, that consists in holding a long or short position for a prespecified number of days, ignoring all other signals generated during that time.

Table 5.2 reports the parameterizations of the filters.
The complete universe of trading indicators after filtering is equal to 5286.

5.3 Evolutionary learning

Evolutionary learning may be thought as the evolutionary computation approach applied to machine learning (for an introductory tutorial, see Yao and Liu (2005)). From this point of view, the problem of building a promising investment signal for future trading by pooling technical indicators translates into the problem of finding the best combination of finite-state machines from a population of solutions enhanced by a genetic algorithm (GA).
The developed procedure is shown in Figure 5.1: it is constituted by two interconnecting modules, the learning module, which is built up on the boosting and statistical learning methods, and the evolutionary module, which interconnects with the previous and where a non-dominated sorting GA, the NSGA-II procedure by Deb et al. (2002), finds multiple Pareto-optimal solutions. A detailed description of these steps follows.

5.3.1 Learning module

The filtered signals are transformed in binary-classifier machines dividing price movements into two classes, upward and downward. In this manner the resulting label space Y, representing the foregoing movements of the market, is also binary, taking a value of -1 for a sell signal and a value of $+1$ for a buy signal. The holding signal is obtained as a confirmation of the previous sell or buy indication. Given a trading rule h, its value at t for a time series x is denoted by $h(x)_t \in Y$.
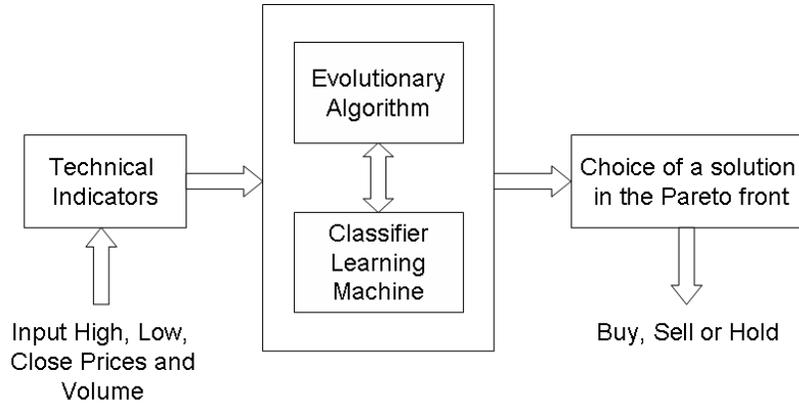
Figura 5.1: Outline of the evolutionary learning algorithm for trading analysis

Learning algorithms are used at this stage to increase the accuracy of technical predictions by averaging the decision of an ensemble of these classifier machines. To this end, three different methods are considered: the Plurality Voting Committee (PVC), the Bayesian Model Averaging (BMA) and the Boosting method (BOOST). All the algorithms take as input a training set belonging to the time series x and produce a set of trading rules that are collected to form a new trading signal, $H(x)$, called expert system.

Plurality Voting Committee

The committee classifier uses the plurality voting rule to decide the output of the committee. For the market timing it becomes a simple average of the predictions from each trading rule. The PVC signal is given by

$$H_{\mathrm{PVC}}(x) = \mathrm{sign}\left(\frac{1}{n}\sum_{i=1}^{n} h_i(x)_t\right) \qquad (5.1)$$

where the average is made on the n rules $\{h_i\}_{i=1,\dots,n}$ selected by the GA and $\mathrm{sign}(x) = 1$ if $x \geq 0$ and -1 otherwise.

Bayesian model averaging

Given a set of candidate classifiers, the BMA approach consists of taking a linear combination of the predictive distribution of each classifier, weighted

by its posterior distribution (see Hastie et al. (2001) for a detailed explanation).

As shown in Andrada-Félix and Fernandez-Rodríguez (2008), this definition translates for the market timing problem in a combination of trading rules weighted by the success rate on the training set X. Defining the rate of success of the i-th trading rule h_i, $i = 1, ..., n$ as

$$\alpha_{\text{BMA},i} = \frac{s_i}{\sum_{j=1}^{n} s_j}$$

where $s_j = \sum_{t=1}^{T} I_{\{y_t = h_j(x)_t\}}$ represents the number of correct predictions and I_A is the indicator function of the event A, the BMA signal is then

$$H_{\text{BMA}}(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_{\text{BMA},i} h_i(x)\right). \qquad \boxed{5.2}$$

In this manner, weights assigned by this classifier to each trading signal depend on how well this indicator fits.

Boosting

Boosting finds a highly accurate classification rule by combining many classifiers, each of which has a performance at least slightly better than random guessing, and for this reason they are called weak hypothesis.

The algorithm I consider is the Discrete AdaBoost for the two class problem due to Freund and Shapire (the interested reader may consult Freud and Shapire (1997, 1999)) and also known as AdaBoost.M1 algorithm. Its pseudo code is described in Figure 5.2. Once a weak hypothesis h_i has been received, AdaBoost chooses a parameter $\alpha_{\text{BOOST},i}$ that measures the importance assigned to h_i. Note that $\alpha_{\text{BOOST},i} \geq 0$ if the error rate ϵ_i satisfies $\epsilon_i \leq 1/2$ and $\alpha_{\text{BOOST},i}$ gets larger as ϵ_i gets smaller. Next, the distribution of weights $w_i(t)$ is updated modifying the individual weights, increasing the weight of the examples misclassified by h_i and decreasing the weight of correctly classified examples. Thereby each successive classifier is forced to focus on the observations that are missed by the previous classifiers in the sequence.

5.3.2   Evolutionary module

The investment timing translates into a multi-objective evolutionary optimization problem where candidate solutions are detected among the expert trading systems generated by learning methods.

Algorithm AdaBoost.M1

Input:
- sequence of T examples $\{(x_1, y_1), ..., (x_T, y_T)\}$ with labels $\quad y_t \in Y = \{-1, +1\}$
- integer n specifying number of hypothesis

Initialize the weights vector: $w_1(t) = \frac{1}{T}$ for $t = 1, ..., T$

Do for $i = 1, ..., n$

1. Fit the classifier $h_i(x)$ to the training data using weights $w_i(t)$

2. Compute the error rate of $h_i(x)$:
$$\epsilon_i = \frac{\sum_{t=1}^{T} w_i(t) I_{\{y_t \neq h_i(x)_t\}}}{\sum_{t=1}^{T} w_i(t)}$$

3. Set
$$\alpha_{\text{BOOST},i} = \frac{1}{2} \log\left(\frac{1 - \epsilon_i}{\epsilon_i}\right) \qquad \boxed{5.3}$$

4. Set the new weights vector to be
$$w_{i+1}(t) = w_i(t) \exp(-\alpha_{\text{BOOST},i} y_t h_i(x)_t)$$

5. Normalize $w_{i+1}(t)$

Output: final hypothesis

$$H_{\text{BOOST}}(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_{\text{BOOST},i} h_i(x)\right) \qquad \boxed{5.4}$$

Figura 5.2: The discrete adaptive boosting algorithm

NSGA-II

The evolutionary procedure adopted is the NSGA-II algorithm and its pseudocode is outlined in Figure 5.3. This algorithm builds a population of competing individuals and sorts them with a non-dominated sorting procedure, which results in all the chromosomes being assigned a rank. The selection that follows uses the crowded tournament strategy. Crossover and mutation are then applied to create a new pool of offsprings, which are combined with parents, before partitioning the enlarged pool into fronts. Elitism is then conducted by adding a crowding distance to each member to generate the next population. This ensures diversity in population and helps the algorithm to explore the fitness landscape (see Coello et al. (2007) for a detailed explanation of the procedure). The chromosomes in the first front of the final population constitute the set of solutions. A validation function is defined to select a chromosome in the solution set which is best suited for the investment timing problem. To this end, the data set is divided into three parts: the training set that is defined over the period $[1, T_{tr}]$, the validation set, defined over $[T_{tr} + 1, T_{val}]$ and the testing set defined over $[T_{val} + 1, T_{tst}]$.

Chromosome representation

Each individual is represented by a subset $\mathcal{S}$ of technical trading rules from the entire universe $\mathcal{S}_{all}$. Coding solutions as fixed length strings has several limitations, such as overfitting of the training data and presence of redundant signals in the final classifier. To overcome these drawbacks, the chromosomes that represent each solution have variable length and are coded on a discrete alphabet χ of cardinality $|\mathcal{S}_{all}| + 1$:

$$\chi = \{0, 1, \ldots, |\mathcal{S}_{all}|\},$$

where each non-zero number corresponds to a technical signal in $\mathcal{S}_{all}$ and the zero index has been added in order to utilize the existing evolutionary operators as much as possible and corresponds to the "no signal" input. After having fixed a maximum length l_{max}, strings are filled with 0's until they have l_{max} alleles, in order to make them comparable. Subsequently, each string is rearranged so that all 0's are pushed at the end. For example, chromosomes

$$string_1 = 10\ 34\ 50\ 0\ 0\ 0\ 0\ 0$$
$$string_2 = 1\ 27\ 45\ 120\ 157\ 0\ 0\ 0$$

have length $l_{max} = 8$ but $string_1$ represents 3 signals whereas $string_2$ has 5 signals.

An individual is valid if it has at least two non-zero alleles and all the included signals are different. Moreover, relative to PVC and BMA methods, chromosomes are sorted in an increasing way to guarantee more diversity among population.

Evolutionary operators

Crossover. Uniform crossover is used to avoid the positional and distributional bias that may prevent the production of good solutions, see Reeves (2003). A control on the composition of each offspring is included to guarantee its admissibility.

Mutation. I use an alternating mutation probability $\mu_m(g)$ that depends on the generation $g \in \{1, \ldots, G\}$, throughout a triangle wave relation, to provide a better balance between exploration and exploitation of the search space. An example of its graph is given in Figure 5.4. I consider the mutation operator developed in Bandyopadhyay and Pal (2007): for each position in a string, it is determined whether conventional mutation can be applied or not with probability μ_m. Otherwise, the position is set to 0 with probability μ_{m_1} and each "no signal" is set to a signal according to another mutation probability μ_{m_2}. The string is then reordered to have the admissible form previously described.

Objectives

The problem of expert trading system selection is formulated by using statistical and economic criteria at the same time, because the outcomes suggested by one criterium alone may be very different from the outcomes recommended by the other, in the sense that standard forecasting criteria are not necessarily well suited for assessing the economic significance of the predictions and vice versa, as highlighted in Sullivan et al. (1999). The first objective considers both the profitability and the risk profile by finding a subset of technical trading rules $\mathcal{S}$ from $\mathcal{S}_{all}$ that are able to capture market directional changes that influence the net profit in a simple trading strategy, where total funds are invested in either a stock market or in a risk free security. Following Allen and Karjalainen (1999), the forecast of the learning machine is used to classify trading days into periods in (earning the market return) or out of the market (earning the risk free return). The return

obtained following the forecasts of $\mathcal{S}$ at time t is given by

$$R(\mathcal{S}, t) = r_t I_b(t) + r_t^f I_s(t) \tag{5.5}$$

where r_t is the daily continuously compounded return constructed on the closing price on day t; r_t^f is the daily risk free rate; $I_b(t)$ and $I_s(t)$ are indicator variables equal to one if the classifiers signal buy and sell, respectively, and zero otherwise, satisfying the relation $I_b(t) \cdot I_s(t) = 0$, for all $t \in [1, T_{tr}]$. The expert system net return at time t is defined as

$$R_{\mathrm{net}}(\mathcal{S}, t) = R(\mathcal{S}, t) + C(t) \tag{5.6}$$

where $C(t)$ is equal to $\log \frac{1-c}{1+c}$ if a transaction took place at time t and 0 otherwise, with c denoting the transaction costs (expressed as a fraction of the price and supposed constant over time). The risk measure is expressed by the standard deviation of the daily net profit over a given time period $[t_a, t_b]$, denoted by $\sigma(\mathcal{S}, t_a, t_b)$. Thus the performance of the expert system $\mathcal{S}$ may be evaluated with the modified version of the Sharpe ratio introduced in Andrada-Félix and Fernandez-Rodríguez (2008):

$$f_{\mathrm{econ}}(\mathcal{S}, 1, T_{tr}) = \frac{\mu(\mathcal{S}, 1, T_{tr})}{\sigma(\mathcal{S}, 1, T_{tr})}. \tag{5.7}$$

where $\mu(\mathcal{S}, 1, T_{tr})$ is the mean net return over the training period $[1, T_{tr}]$. On the other hand, to pursue high classification power, the second objective is designed to select expert systems by minimizing the number of missclassified observations in the training period and by attempting to reduce the number of trading rules involved. This objective function is defined by

$$f_{\mathrm{stat}}(\mathcal{S}, 1, T_{tr}) = \alpha_1 \frac{\mathrm{miss}(\mathcal{S}, 1, T_{tr})}{T_{tr}} + \alpha_2 \frac{|\mathcal{S}|}{l_{max}} \tag{5.8}$$

where $\mathrm{miss}(\mathcal{S}, 1, T_{tr})$ is the number of missclassifications for $\mathcal{S}$ on $[1, T_{tr}]$ and the coefficients α_1 and α_2 reflect the relative importance of the two statistics. The resulting bi-criteria optimization problem can be stated as

$$\begin{cases} \max f_{\mathrm{econ}}(\mathcal{S}, 1, T_{tr}) \\ \min f_{\mathrm{stat}}(\mathcal{S}, 1, T_{tr}) \end{cases}, \text{ subject to } \mathcal{S} \subseteq \mathcal{S}_{all}. \tag{5.9}$$

When the data set used is clear from the context, I will suppress the dependence on the time period in the functions introduced so far.

Validation and testing

The validation function, defined on the interval $[T_{tr} + 1, T_{val}]$, is used to detect the best strategy for investments and is a weighted mean of the two objectives:

$$f_{\text{val}}(\mathcal{S}) = w_{\text{econ}} f_{\text{econ}}(\mathcal{S}) + w_{\text{stat}} f_{\text{stat}}(\mathcal{S}), \qquad (5.10)$$

where w_{econ} and w_{stat} are the relative importance of economic and statistical criteria respectively. On the testing set the selected expert system is tested to assess its capability of prediction and its profitability.

5.4 Computational results

5.4.1 Parameter values

The best parameter setting for the GA is obtained from preliminary experiments and is as follows:

- $l_{max} = 20$;

- population size $N = 500$;

- maximum number of generations $G = 500$;

- crossover probability $\mu_c = 0.85$;

- mutation probabilities $\mu_m \in [0.002, 0.45]$ and $\mu_{m1} = \mu_{m2} = 0.95$;

- $\alpha_1 = 0.7$ and $\alpha_2 = 0.3$

- $w_{\text{econ}} = 0.6$ and $w_{\text{stat}} = 0.4$

5.4.2 Data

I use daily data for the Standard & Poor's Composite Index (S&P500) from 3 January 2000 to 29 December 2006. The data series include the high, low and closing prices and the volume of transactions. No data on dividends declared by the firms is used in the learning process. For the corresponding period, following the custom in literature (see, for example, Allen and Karjalainen (1999)), the risk-free rate of return I adopt is the 3-month Treasury Bill rate[1], and I set the one-way transaction costs at 0.25%. The time series of

---

[1]The quotations for the S&P500 Index are taken from http://www.finance.yahoo.com and for the 3-month Treasury Bill from http://www.federalreserve.gov/releases/h15/data.htm

Experiment	Start of Training	Start of Validation	Start of Testing
down-trend	03/01/2000	02/01/2002	01/04/2002
no-trend	02/01/2002	02/01/2004	01/04/2004
up-trend	02/01/2004	03/01/2006	03/04/2006

Tabella 5.3: Dates of the experiments

Learning method	Ensemble constituents	filter
PVC	VMA(25,125)	p (10%)
	gWSMA(5,45,0.4,3)	p (3%)
	MACDH(9,26,10)	h (10)
	MACDH(12,29,9)	d (3)
BMA	ROC(30)	p (1.5%)
	VMA(40,150)	p (10%)
BOOST	gWSMA(5,45,-0.4,3)	p (3%)
	VMA(2,250)	p (1.5%)
	ROC(5)	p (10%)

Tabella 5.4: Comparisons of ensemble indicators for the period 3 January 2000 to 31 December 2002 (down-trend period)

the S&P500 closing prices in log-form and the discount rate of the Treasury Bill are displayed in Figure 5.5.

5.4.3 Experiments

Three sets of experiments are conducted to study the applicability of the developed evolutionary learning methods in topical market phases, i.e. when prices follow a downward trend (bear market), when they display sideways-movements and when they have an upward trend (bull market), for short-term investments. To this end, the data for each experiment, are divided in three subsamples, the training duration, for fitting the technical signals that is fixed to 24 months, the validation period, set to 3 months, and the testing period, of 9 months. The starting dates of each subsample for the experiments are reported in Table 5.3.   The solutions suggested by the analyzed learning methods are listed in Table 5.4, 5.5 and 5.6. Each table depicts an experiment. In the second column the ensemble constituents are reported with their parameterizations and in the third the corresponding

Learning method	Ensemble constituents	filter
PVC	VMA(50,200)	d (4)
	MACDH(8,26,100)	h (30)
BMA	VMA(40,250)	p (0.1%)
	gWSMA(5,60,0.4,1)	h (30)
	MACDH(10,20,10)	h (15)
BOOST	ROC(10)	h (30)
	VMA(50,200)	d (5)
	gWSMA(15,30,-0.2,1)	h (30)

Tabella 5.5: Comparisons of ensemble indicators for the period 2 January 2002 to 31 December 2004 (sideways-movement period)

Learning method	Ensemble constituents	filter
PVC	gWSMA(10,90,0,3)	p (1.5%)
	MACD(8,23,8)	h (10)
BMA	VMA(40,150)	d (2)
	VMA(200,250)	p (10%)
	gWSMA(15,45,-0.2,3)	h (30)
	MACDH(10,17,8)	h (30)
BOOST	dS&R(30)	p (10%)
	gWSMA(15,45,0,3)	h (30)

Tabella 5.6: Comparisons of ensemble solutions for the period 2 January 2004 to 29 December 2006 (up-trend period)

filter with its parameter (p = fixed percentage band, d = time delay, h = holding time). Comparing the solutions, it can be noted that the number of considered signals varies from two to four, making easy the interpretations of the compounded indicator. In all the solutions the final signal is governed mainly by the momentum indicators. The most important difference among the experiments is the way technical information is filtered in the final trading indicator:

   i) during down-trend periods, almost all signals are passed under the fixed percentage band filter;

   ii) during sideways-movement periods, the most informative trading signals emerge from the holding time rule;

   iii) in the up-trend experiment, the constituents information are filtered by the fixed percentage band and the holding time rule.

The time delay filter holds only a marginal role and is seldom considered in the solutions. Table 5.7, 5.8 and 5.9 display the economic and statistic performances of the solutions divided for experiments and for training, validation and testing sets. To facilitate the comparisons among the economic results, the net returns and the Sharpe ratio refer to annualized data.
Relative to the percentage of forecasting direction success, the PVC method shows the best results, with a mean value of about 77%, against the 55% of BMA and BOOST. From the profitability perspective, BMA and BOOST have net returns similar or quite better than PVC. In particular, during the first two experiments, these three learning procedures are able to reduce or eliminate the potential losses that characterize the period from 2000 to 2004. The Sharpe ratio confirms these findings, highlighting the great stability of the results and the flexibility of the evolutionary learning approach.
Results are in line with the findings of Andrada-Félix and Fernandez-Rodríguez (2008), however, the evolutionary learning approach developed in this chapter generates solutions less complex, and thus more easier to interpret, making possible to focus on those signals that show to be more informative for the future behavior of market prices. This selection is conducted on a wide range of technical indicators, extending previous literature that solely analyzed simple moving average rules (see for example Brock et al. (1992), Kwon and Kish (2002)).

Learning method	Subset	% forecasting direction success	Transactions	Annualized net return	Sharpe ratio
PVC	train	77.15	19	0.0717	0.1478
	validation	77.97	0	-0.0252	-0.1464
	test	74.87	9	-0.2049	-1.1712
BMA	train	68.14	6	-0.0540	-0.1581
	validation	52.54	0	0.0112	0
	test	56.54	0	0.0111	0
BOOST	train	54.11	15	0.0953	1.6644
	validation	57.63	4	0.1864	2.9417
	test	53.93	1	-0.0795	-1.2483

Tabella 5.7: Statistic and economic performances for the period 3 January 2000 to 31 December 2002 (down-trend period)

Learning method	Subset	% forecasting direction success	Transactions	Annualized net return	Sharpe ratio
PVC	train	75.94	3	0.2228	0.6801
	validation	85.25	0	0.0636	0.6160
	test	75.13	4	0.0124	0
BMA	train	56.66	3	0.2601	1.4640
	validation	50.82	0	0.0636	0.6160
	test	49.74	4	0.0833	0.9882
BOOST	train	58.25	7	0.3504	1.9083
	validation	50.82	0	0.0636	0.6160
	test	55.56	2	0.0776	0.8624

Tabella 5.8: Statistic and economic performances for the period 2 January 2002 to 31 December 2004 (sideways-movement period)

Learning method	Subset	% forecasting direction success	Transactions	Annualized net return	Sharpe ratio
PVC	train	83.70	0	0.1188	0.4720
	validation	73.77	0	0.0812	0.8470
	test	72.87	0	0.1184	1.2163
BMA	train	57.46	6	0.1795	0.9730
	validation	54.10	0	0.0812	0.8470
	test	52.32	2	0.0535	0.5270
BOOST	train	54.27	7	0.1384	0.7776
	validation	50.82	1	0.0732	3.1623
	test	59.57	2	0.1004	1.5058

Tabella 5.9: Statistic and economic performances for the period 2 January 2004 to 29 December 2006 (up-trend period)

5.5   Conclusions

In this chapter, I propose a method based on the evolutionary learning paradigm for the daily stock index trading, by taking into account statistical and economic objectives at the same time. A variable number of technical indicators as input is selected by the algorithm to recommend buy/sell of S&P 500 Index when certain price formation exist. The performance of three different learning procedures is tested from 2000 to 2006, by dividing the entire period in three trend-following subperiods. The analysis used MATLAB program to generate the outputs.
Results indicate the stability of the methods analyzed over time both from a statistical point of view and from an economic perspective. In particular, in sideways-movement periods and especially in down-trend periods, they predict correctly the major losses, a desirable characteristic for an automatic trading system.

Bibliografia

Aksela, M.: 2003, Comparison of classifiers selection methods for improving committee performance, *Lecture Notes in Compure Science* **2709**, 84–93.

Aksela, M. and Laaksonen, J.: 2006, Using diversity of errors for selecting members of a committee classifier, *Pattern Recognition* **39**(4), 608–623.

Alexander, S. S.: 1964, *The random character of stock market prices*, Vol. 2, MIT Press, Cambridge, MA, chapter Price movements in speculative markets: trends or random walks, pp. 338–372.

Allen, F. and Karjalainen, R.: 1999, Using genetic algorithms to find technical trading rules, *Journal of Financial Economics* (51), 245–271.

Andrada-Félix, J. and Fernandez-Rodríguez, F.: 2008, Improving Moving Average Trading Rules with Boosting and Statistical Learning Methods, *Journal of Forecasting* (27), 433–449.

Bandyopadhyay, S. and Pal, S. K.: 2007, *Classification and Learning Using Genetic Algorithms*, Springer.

Brock, W., Lakonishok, J. and LeBaron, B.: 1992, Simple technical trading rules and the stochastic properties of stock returns, *The Journal of Finance* **XLVII**(5), 1731–1764.

Chan, L. K. C., Jegadeesh, N. and Lakonishok, J.: 1996, Momentum Strategies, *The Journal of Finance* **51**(5), 1681–1713.

Coello, C. A. C., Lamont, G. B. and Veldhuizen, D. A. V.: 2007, *Evolutionary Algorithms for Solving Multi-Objective Problem*, Springer.

Colby, W. and Meyers, T.: 1990, *The Encyclopedia of Technical Market Indicators*, Dow Jones Irwin.

Covel, M.: 2005, *Trend Following: How Great Traders Make Millions in Up or Down Markets*, Financial Times Prentice Hall.

Deb, K., Pratap, S. and Meyarivan, T.: 2002, A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Transactions of Evolutionary Compuation* **6**, 182–197.

Dourra, H. and Siy, P.: 2002, Investment using technical analysis and fuzzy logic, *Fuzzy sets and Systems* (127), 221–240.

Faber, M. T.: 2007, A quantitative approach to tactical asset allocation, *The Journal of Wealth Management* (1), 1–11.

Fama, E. F.: 1970, Efficient capital markets: a review of theory and empirical work, *Journal of Finance* **25**, 383–417.

Fama, E. F. and Blume, M. E.: 1966, Filter rules and stock market trading, *Journal of Business* **39**, 226–241.

Fernandez-Rodríguez, F., González-Martel, C. and Sosvilla-Rivero, S.: 2005, Optimizaton of technical rules by genetic algorithms: evidence from the madrid stock market, *Applied Financial Economics* (15), 773–775.

Freud, Y. and Shapire, R. E.: 1997, A decision-theoretic generation of on-line learning and application to boosting, *Journal of Computer and System Sciences* **55**(1), 119–139.

Freud, Y. and Shapire, R. E.: 1999, A short introduction to boosting, *Journal of the Japanese Society for Artificial Intelligence* **14**(5), 771–780.

Fusai, G. and Roncoroni, A.: 2008, *Implementing Models in Quantitatative Finance: Methods and Cases*, Springer.

Gençay, R.: 1999, Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules, *Journal of International Economics* **47**, 91–107.

Goldberg, D. E.: 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, New York.

Gupta, B. and Chatiras, M.: 2003, The interdependence of managed futures risk measures, *Technical report*, Center for International Securities and Derivatives Markets.

Hastie, T., Tibshirani, R. and Friedman, J.: 2001, *The Ellemnts of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York.

Hendry, D. and P.Clements, M.: 2002, Pooling of Forecasts, *Econometrics Journal* **5**, 1–26.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T.: 1999, Bayesian Model Averaging: A Tutorial, *Statistical Science* **14**(4), 382–417.

Jegadeesh, N.: 1990, Evidence of Predictive Behavior of Security Returns, *The Journal of Finance* **45**(3), 881–898.

Johnston, F. R., Boyland, J. E., Meadows, M. and Shale, E.: 1999, Some properties of a simple moving average when applied to forecasting a time series, *Journal of the Operational Research Society* (50), 1267–1271.

Kaboudan, M. A.: 2000, Genetic Programming Prediction of Stock Prices, *Computational Economics* (16), 207–236.

Kahneman, D. and Tversky, A.: 1979, Prospect theory: an analysis of decision under risk, *Econometrica* **47**(2), 263–292.

Korczak, J. and Roger, P.: 2002, Stoch timing using genetic algorithms, *Applied Stochastic Models in Business and Industry* (18), 121–134.

Kwon, K. Y. and Kish, R.: 2002, A comparative study of technical trading strategies and return predictability: an extension of Brock, Lakonishok, and LeBaron (1992) using NYSE and NASDAQ indices, *The Quarterly Review of Economics and Finance* (42), 611–631.

Kwon, Y. K. and Moon, B. R.: 2007, A hybrid neurogenetic approach for stock forecasting, *IEEE Transactions on Neural Networks* **18**(3), 851–864.

Leontitsis, A. and Pange, J.: 2004, WSMA: in between weighted and simple average, 17th Annual Pan-Hellenic Conference on statistics, Leukada, Greece, pp. 519–526.

Lo, A. W., Mamaysky, H. and Wang, J.: 2000, Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation, *Journal of Finance* **LV**(4), 1705–1765.

Michalewitz, Z.: 1994, *Genetic Algorithms + Data Structure = Evolution Program*, Springer-Verlang, New York.

Murphy, J.: 1998, *Technical Analysis of the Financial Markets*, New York Institute of Finance, New York.

Neftci, S.: 1991, Naïve trading rules in financial markets and Wiener-Kolmogorov prediction theory:a study of technical analysis, *Journal of Business* (64), 549–571.

Papadamou, S. and Stephanides, G.: 2007, Improving technical trading systems by using a new MATLAB-based genetic algorithm procedure, *Mathematical and Computer Modelling* (46), 189–197.

Potvin, J. Y., Soriano, P. and Vallée, M.: 2004, Generating trading rules on the stock markets with genetic programming, *Computers & Operations Research* **31**, 1033–1047.

Pruitt, S. and White, R.: 1988, The CRISMA trading system: who says technical analysis can't beat the market?, *Journal of Portfolio Management* (1), 55–58.

Ready, M. J.: 2002, Profits from technical trading rules, *Financial Management* (3), 43–61.

Reeves, C.: 2003, *Handbook of Metaheuristics*, Kluwer Academic Publishers, chapter Genetich Algoritms, pp. 55–82.

Reveiz, A. and León, C.: 2008, Efficient Portfolio Optimization in the Wealth Creation and Maximum Drawdown Space, *Working Paper 520*, Borradores de Economia, Banco de la República, Colombia.

Sullivan, R., Timmermann, A. and White, H.: 1999, Data-Snooping, Technical Trading Rule Performance, and the Boostrap, *The Journal of Finance* **LIV**(5), 1647–1691.

Trippi, R. and Turban, E.: 1990, *Investment management: decision support and expert systems*, Van Nostrand Reinhold, New York.

Yao, X. and Liu, Y.: 2005, *Search Methodologies*, Springer, chapter Machine Learning, pp. 341–373.

Algorithm NSGA-II

Input:
- population size N
- number of generations G

Initialize random population P_1

Evaluate objective values

Generate offspring population Q_1 from P_1 by

   i) selection

  ii) crossover and mutation operators

Evaluate objective values

Do for $g = 1, \ldots, G$

  1. Combine parent and child populations and create $R_g = P_g \cup Q_g$. Assign rank to chromosomes in R_g based on Pareto dominance and identify different fronts $\mathcal{F}_i$, $i = 1, 2, \ldots$

  2. Set next population $P_{g+1} = \emptyset$, set a counter $i = 1$.
     Do until $|P_{g+1}| + |\mathcal{F}_i| < N$

      i) perform $P_{g+1} = P_{g+1} \cup \mathcal{F}_i$

     ii) $i = i + 1$

  3. Niching: calculate crowding distances between points on each front and use them to include the most widely spread $N - |P_{g+1}|$ solutions

  4. Create offspring population Q_{g+1} from P_{g+1} by

     i) crowded tournament selection

    ii) crossover and mutation operators

Output: limited number of different solutions
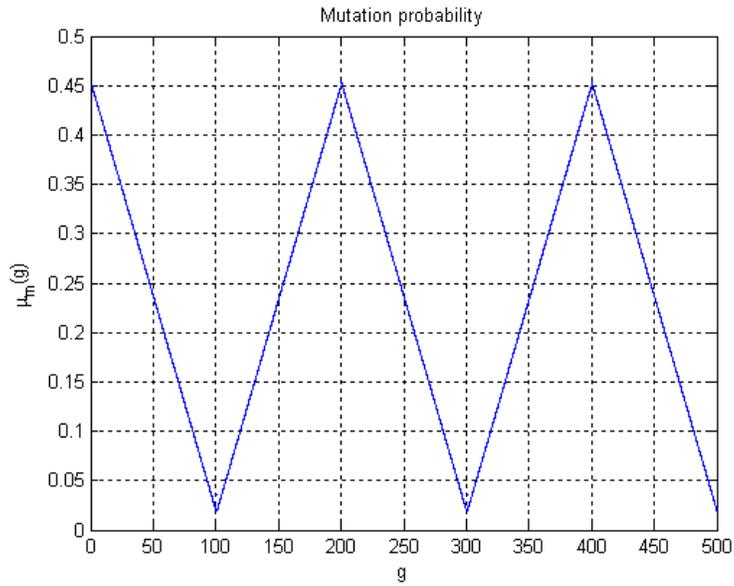
Figura 5.3: Pseudocode of the NSGA-II algorithm

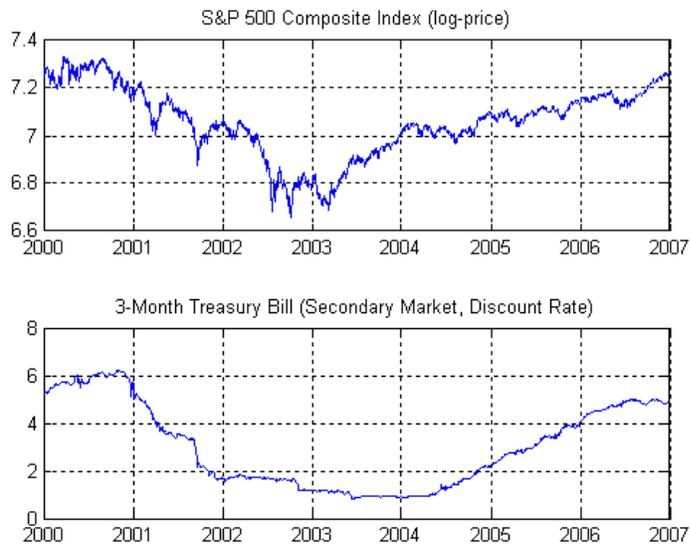Figura 5.4: Alternating mutation probability graph

Figura 5.5: S&P500 log-prices and 3-month Treasury Bill discount rates for the period from 03/01/2000 to 29/12/2006

125

Parte III

Portfolio Selection

6

GA-based market neutral strategies

In this chapter, a MATLAB tool based on genetic algorithms for active portfolio management is described. The tool I have developed deals with parameter optimization of trading signals for an integrated market neutral strategy. In this chapter I will only provide a qualitative description of the ideas and principles underlying the procedure, refraining from exploring technical details. The chapter is organized as follows. First, an introduction to market neutral strategies from an economic point of view is provided, highlighting risks and benefits associated to these rules and focusing on their quantitative formulation of these rules. In particular, the concepts of long-plus-short and integrated neutral strategies are derived. A short introduction to the evolutionary approach with expert trading systems for portfolio optimization serves as bridge to explain the GA-Integrated Neutral tool. The chapter concludes showing an application of the tool as a support to decisions in the Absolute Return Interest Rate Strategies sub-fund of Generali Investments.

6.1 Market neutral strategies

In general, market neutral strategies are not market timing strategies. Rather than seeking to profit from correctly forecasting underlying market movements, they seek to profit from detecting perceived mispricings in individual securities and constructing portfolios that deliver the excess return and/or risk associated with those securities, regardless of underlying market changes. This is accomplished by holding balanced long and short positions in various securities and/or by holding these securities in conjunction with long or short positions in derivative securities so that the overall portfolio's exposure to primary risk factors, such as equity market and interest rate risks, is neutralized. These strategies may use in-depth fundamental analysis, technical approaches and/or quantitative valuation and portfolio

construction techniques.

Market neutral tactics have the same basic aim as more conventional strategies, i.e. to "buy low and sell high". In more traditional active approaches, however, the buying and selling are sequential events, whereas in market neutral investing they are more often concurrent. A market neutral investor buys underpriced securities and simultaneously sells an offsetting amount of fundamentally related overpriced securities. Examples of this type of investments are

- MARKET NEUTRAL EQUITY STRATEGY
  It is a strategy that seeks to exploit investment opportunities unique to some specific group of stocks by being long and short while maintaining a neutral exposure to broad groups of stocks defined for example by sector, industry, market capitalization, country or region.

- CONVERTIBLE BOND HEDGING
  This form of hedging typically involves purchasing a convertible security and shorting the stock into which it is convertible. Shorting reduces the investor's exposure to changes in the stock price, because price movements in the convertible are at least partially offset by the price movements of the short stock position. More sophisticated variants include hedging so that the net expected position is fully hedged with respect to changes in the stock price, or hedging so that the net expected position is also fully hedged with respect to changes in interest rates and/or credit spreads.

- FIXED INCOME STRATEGIES
  These type of neutral strategies try to exploit perceived mispricing among one or more fixed income instruments. Some of the most widely used are:

  i) *Swap-spread arbitrage*, that is a bet on the direction of swap rates, Libor, treasury coupon rates and repo rates. A typical swap-spread arbitrage trade would consist of a fixed receiver swap and a short position in a Treasury bond of the same maturity. The proceeds of the sale of the Treasury bond would be invested in a margin account earning the repo rate. This trade is a simple bet that the difference between the swap rate and coupon rate will be hugher than the difference between Libor and the repo rate.

  ii) *Yield curve arbitrage strategies*, that are designed to profit from shifts in the steepness of, or kinks in, the Treasury yield curve by

taking long and short positions on various maturities. This could take the form of a butterfly trade, where, for example, the investor goes long five-year bonds and short two and ten-year bonds, or it may take the form of a spread trade, where the investor goes short the front end of the curve and long the back end of the curve. The strategy requires the investor to identify some points along the yield curve that are "rich" or "cheap".

iii) *Government bond arbitrage* refers to the difference between the amount of interest gained on funds which have been borrowed at a tax-free rate, and the interest on funds which have been invested at a taxable rate rendering a greater yield.

iv) In their simplest form, *volatility arbitrage strategies* profit from the tendency of implied volatilities to exceed subsequent realized volatilities. This is done by selling options of fixed income instruments and then delta-hedging the exposure to the underlying asset.

v) *Capital structure arbitrage strategies*, that exploit the lack of coordination between various claims on a company, like its debt and stock. The strategy involves buying one instrument of a company's capital structure and hedging that exposure by selling another. For example, a trader who believes that the debt of a company is overpriced relative to its equity would short the company's debt and buy its stock. Capital structure arbitrage trades may also trade junior vs. senior debt or even convertible bonds vs. stock.

- MERGER ARBITRAGE
  This is also called risk arbitrage. Typically merger arbitrage represents situations in which one invests simultaneously long and short in the companies involved in a merger or acquisition. Risk arbitrageurs are typically long the stock of the company being acquired and short the stock of the acquirer. By shorting the stock of the acquirer, the manager hedges out market risk, and isolates his/her exposure to the outcome of the announced deal.

6.1.1   Risks related to neutral strategies

Because market neutral strategies are designed to eliminate systematic risk factors, such as stock market or interest rate risk, they are often perceived to have low risk but in general this is not true. In fact, risk levels may vary

across different types of market neutral strategies and across portfolios in a given strategy. The risk of any given strategy will depend upon multiple factors, including the volatility of the underlying securities, the sources of uncertainty impacting those securities, the models and methods used in the investment process and the degree of leverage employed. At least in the short term, equities are inherently more volatile than fixed-income instruments, so that one may expect market neutral equity strategies to be inherently more volatile than fixed-income strategies. Furthermore, the instruments underlying some bond-based strategies, including mortgage and convertible arbitrage, may be subject to extreme bounces of volatility because they include option-like elements that can cause them to behave in nonlinear ways. As any investment strategy, market neutral strategies are subject to uncertainty beyond anticipated volatility. Unexpected events can cause actual and expected portfolio performance to diverge. Sources of uncertainty can be introduced by unanticipated changes at company-specific level, by developments in the broader economy and by regulatory and legal events. In particular, credit risk, i.e. the risk that a counterparty to a trade will default, may be a major issue for market neutral strategies than for more conventional investment approaches, to the extent that the former rely more heavily on over-the-counter derivatives. Traders using organized exchanges are largely protected against counterparty default by the guarantees provided by exchange clearinghouses. For market neutral strategies that require over-the-counter derivatives such as options and interest rate swaps, due diligence must be conducted to ensure that counterparties are creditworthy. The primary line of defense against uncertainty is diversification. This is true for market neutral as well as for conventional investment strategies. For example, diversification across different securities protects against company-specific risks. Diversification across counterparties may provide some protection against credit risk.

However, other problems, such as issues concerning portfolio constructions, valuation process or risk measurement, may be subtle and difficult to detect. As a partial solution, quantitative investment approaches may have the advantage over more judgmental ones when it comes to detecting and correcting these sources of error. They can thus provide a transparent audit trail of cause and effect that can be used to detect and remedy potential trouble spots.

In general, market neutral strategies are more dependent on leverage than conventional investing. Leverage can take many forms, among them outright borrowing, repo arrangements, purchase of securities on margin, and the short sale of borrowed securities. By increasing the number and size

of positions a strategy can take, leverage can increase the return to that strategy, but also the risk. If the strategy performs as expected, leverage will multiply the profits. But it will also multiply the losses if the strategy goes awry. In this sense, leverage magnifies all the risks discussed here. A leveraged market neutral strategy (or any leveraged investment strategy) in effect invests more capital than it has. When things go wrong, losses can exceed the invested money, and as a result the fund can lose more than it started with. With short selling, for example, the owner of shares sold short may demand them back; in certain instances, the short seller may have to liquidate positions in order to meet this demand, regardless of the impact on the portfolio, having disastrous results if it cannot be met via a liquidity reserve, the sale of assets or an infusion of new capital. In such cases, lenders and other counterparties may liquidate the portfolio at large losses to investors.

6.1.2 Advantages of neutral strategies

Because of their ability to deliver returns that are independent of the performance of the underlying market, market neutral strategies have often been thought as "hedges" against market downturns. For this reason, market neutral strategies are often used as a tool for diversification. When added to an institution's existing investment in bonds and stocks, market neutral portfolios may be able to increase overall return and/or reduce risk. However, market neutral strategies have much to offer beyond diversification. For example, to the extent that they neutralize underlying market risk, these strategies can be used to exploit profit opportunities in markets that might otherwise be considered too risky for suitable investment. Neutral structures can also allow investors to fine-tune portfolio risk exposures and offer advantages in terms of return enhancement. Most obviously, the ability to sell securities short enables the investor to seek out opportunities in overvalued securities, as well as in undervalued ones.

Perhaps the major advantage of market neutral construction is that it allows the investor to extract the return available from selecting securities in one asset class and, by using derivatives, to "transport" that return to an entirely different asset class. For example, when fixed-income futures or swaps are added to a market neutral equity strategy, any excess return available from the market neutral equity portfolio can be used to enhance a bond market return. This affords a great deal of flexibility in overall fund management and, most importantly, it allows the investor to reap the rewards of both individual security selection and asset class selection.

6.2 Long-plus-short and integrated neutral portfolios

The combination of a long and a short market portfolio generates the so-called long-plus-short portfolio. In particular, the excess return on this combined portfolio equals the excess return due to the short portfolio and the interest earned on the proceeds from the short sales, increased by the excess return obtained from the long portfolio. In this sense, the ability to short, by increasing the investor's freedom to act on his/her insights, has the potential to enhance returns from active security selection. This potential may be especially appealing if short-sale candidates are less efficiently priced than purchase candidates. Even if this is not the case, however, market neutral construction can improve upon the results of long-only portfolio management.

The improvement offered by market neutral investing depends critically on the way in which the portfolio is constructed. There are two possible approaches that now I describe.

- LONG-PLUS-SHORT MARKET NEUTRAL PORTFOLIO
  Many investors construct market neutral portfolios by combining a long-only portfolio, perhaps a preexisting one, with a short-only portfolio, following a two-portfolio strategy. The long side of this portfolio is identical to a long-only portfolio, hence it offers no benefits in terms of incremental return or reduced risk. Furthermore, by assuming no greater inefficiencies on the short side, the short side of this portfolio is then statistically equivalent to the long one and hence equivalent to the long-only portfolio. These relations reflect the fact that all these portfolios, the long-only portfolio and the long and short components of the long-plus-short portfolio, are constructed relative to a benchmark index. Each portfolio is active in pursuing excess return relative to the underlying benchmark only insofar as it holds securities in weights that depart from their benchmark weights. However, departures from benchmark weights introduce residual risk. Controlling portfolio risk thus involves balancing expected excess (to benchmark) returns against the added risk they introduce. In this balancing act, investors face the probability of having to forgo some increment of expected return in order to reduce portfolio residual risk. Thus, this portfolio construction is benchmark-constrained. In long-plus-short portfolios, the advantage offered by the flexibility to short is moreover curtailed by the need to control risk by holding or shorting securities in benchmark-like weights.

The performance of the equity portion of the long-plus-short portfolio is measurable as the long and the short performances in excess of the underlying benchmark and can be compared to the performance of the long-only portfolio in order to highlight the differences. To this end, as a measure for portfolio performances one may consider the information ratio (IR), i.e. the ratio of excess return to residual risk. More precisely, denoting with α_{LO}, α_L and α_S the excess returns of the equity portions of the long-only portfolio, long portfolio and short portfolio respectively, and letting σ_{LO}, σ_L and σ_S be the associated residual risks, the performance of the equity portion of the long-only portfolio is given by

$$\text{IR}_{LO} = \frac{\alpha_{LO}}{\sigma_{LO}} \qquad (6.1)$$

and, similarly, the information ratio for the equity portion of the long-plus-short portfolio is

$$\text{IR}_{L+O} = \frac{\alpha_{L+S}}{\sigma_{L+S}} \qquad (6.2)$$

where α_{L+S} represents the excess return of the combined portfolio and σ_{L+S} is the residual risk. Since this portfolio can be viewed as an equally weighted two-assets portfolio, its excess return is given by

$$\alpha_{L+S} = \frac{1}{2}\alpha_L + \frac{1}{2}\alpha_S. \qquad (6.3)$$

This expression, according to the considerations previously stated, may be simplified to

$$\alpha_{L+S} = \alpha_{LO} \qquad (6.4)$$

since $\alpha_L = \alpha_S = \alpha_{LO}$. With similar arguments, one can observe that $\sigma_{LO} = \sigma_L = \sigma_S$. Hence, after simple computations (see, for example, Cvitanic and Zapatero (2004) for a detailed derivation), the residual risk of the equity portion of the long-plus-short portfolio is given by

$$\sigma_{L+S} = \sqrt{\frac{1}{2}\sigma_{LO}^2(1 + \rho_{L+S})}, \qquad (6.5)$$

where ρ_{L+S} is the correlation between the excess return of the long and short sides of the long-plus-short portfolio.

Combining relations 6.4 and 6.5 with 6.1, the relative performance of the two portfolios may be computed and the result is

$$\frac{\text{IR}_{L+S}}{\text{IR}_{LO}} = \sqrt{\frac{2}{1 + \rho_{L+S}}}, \qquad (6.6)$$

It may be observed that if ρ_{L+S} is less than one, the long-plus-short portfolio will enjoy greater diversification and reduced risk relative to a long-only portfolio, for an improvement in IR. However, a long-only portfolio can derive a similar benefit by adding a less than fully correlated asset with comparable risk and return, so this is not a benefit unique to long-short. The long-only portfolio can also engage in leverage, just like the long-plus-short portfolio (however, a long-only portfolio would have to borrow funds to achieve leverage, and this can have tax consequences for otherwise tax-exempt investors; borrowing shares to sell short does not result in unrelated business taxable income). Furthermore, derivatives such as index futures contracts can be used to make the long-only portfolio market neutral, just like the long-short portfolio. Thus neither market neutrality, nor leverage, nor even shorting constitutes an inherent advantage over long-only portfolio construction.

- Integrated market neutral portfolio
  This strategy represents an integrated optimization that considers both long and short positions simultaneously according to a one portfolio strategy that takes into account the expected returns of the individual securities, the standard deviations of those returns and the correlations between them as well as the investor's tolerance for risk. This approach frees the investor from the nonnegativity constraint imposed on long-only portfolios and, at the same time, frees the market neutral portfolio itself from the restrictions imposed by securities' benchmark weights. In fact, once an underlying benchmark has been used to determine the systematic risks of the candidate securities, its role in market neutral construction is effectively depleted. The offsetting market sensitivities of the aggregate long and aggregate short positions eliminate market sensitivity and the need to consider benchmark weights in establishing security positions. The investor is not constrained to moving away from or toward benchmark weights in order to pursue return or control risk. Rather, capital can be allocated without regard to securities' weights in the underlying benchmark, as offsetting long and short positions are used to control portfolio risk. For example, to establish a 1% overweight or a 1% underweight, the investor merely has to allocate 1% of capital long or allocate 1% of capital short.
  Rather than being measurable as long and short performance in excess of an underlying benchmark, the performance of the equity portion of

the integrated portfolio is measurable as the overall return on the long and short positions, or the spread between the long and short returns, relative to their risk. Compared with the excess return/residual risk of long-only management, this performance should be enhanced by the elimination of benchmark constraints, which allows the market neutral portfolio increased flexibility to implement investment insights, both long and short.

6.3 Evolutionary algorithms for portfolio optimization

In the previous chapters I have described models and procedures through which I exploit the capabilities of evolutionary algorithms to select/extract information for predicting financial time series, such as stocks, sectors and indices. In these pages I want to explain another attracting characteristic of evolutionary computations: their usefulness to seek an optimal parameter configuration for a model in an optimization problem. In the literature, for example, Fernandez-Rodríguez et al. (2005) investigate the profitability of the generalized moving average trading rule for the General Index of Madrid Stock Market by optimizing parameter values with a genetic algorithm. They concludes that the optimized trading rules are superior to a risk-adjusted buy-and-hold strategy if the transaction costs are reasonable. Similarly, Papadamou and Stephanides (2007) present the GATradeTool, a parameter optimization tool based on genetic algorithms for technical trading rules. In the description of this software, they compare it with other commonly used, non-adaptive tools in terms of stability of the returns and computational costs. Results of the tests on the historical data of a UBS fund shows that GATradeTool outperforms the other tools.
As a natural consequence of these studies, evolutionary algorithms may constitute a promising tool also for portfolio optimization, where the main goal is the optimal allocation of funds among various financial assets. An interesting application of evolutionary computations in this field is the approach developed by Korczak and Lipinski (2001) that leads to the optimization of portfolio structures by making use of artificial trading experts, previously discovered by a genetic algorithm (see Korczak and Roger (2002)), and evolutionary strategies. The approach has been tested using data from the Paris Stock Exchange. The profits obtained by this algorithm are higher than those of the buy-and-hold strategy.

6.4 The GA-Integrated Neutral tool

In the last year, I have developed an evolutionary approach to active portfolio management, on the basis of practical requirements and suggestions of financial analysts of Generali Investments and, in particular, the ideas of Fabrizio Barbini.
The tool builds a portfolio of six assets, according to an integrated neutral strategy, where the estimates of the future assets' movements pass through the judgement of a set of artificial trading experts generated by a genetic algorithm, hence the name GA-Integrated Neutral tool (shortly GAIN)[1]. As Figure 6.1 shows, the structure of the GAIN procedure has several steps:

1. Given the assets, the investor/analyst appoints four financial variables that are used to predict each asset[2]. According to this choice, he/she has to select the lag and the type of transformation these leading/trading indicators are subject to.

2. Once data are collected together, GAIN starts with the generation of the artificial trading experts that optimize the parameters in each transformation. To this end, one "factor" at a time is considered and GAIN is applied to search the parameter values that produce the highest performance of the integrated neutral strategy, in terms of the percentage of positive returns achieved by the portfolio over the given period.
   This step may be iterated several times, say N, to obtain a set of N sub-optimal configurations. Successively, GAIN can be applied to this optimized population in order to generate a better solution. Finally, the investor/analyst may control the economic significance of the parameters and, if they are meaningless, he/she may repeat the procedure to find other values.

3. The experts are now collected into a weighted classifier, where the weights represents the relative importance of each "factor". Weights are optimized with a local search procedure, such as exhaustive, simulated annealing or threshold acceptance, that maximize the absolute return of the neutral portfolio on the basis of the indications of the artificial experts.

---

[1]In this context, for integrated neutral strategy I mean that the investor purchases the two most promising assets, sells the two worst and is neutral on the rest of the portfolio.
[2]The four indicators for each asset may be both bottom-up asset-specific fundamentals and top-down economic fundamentals
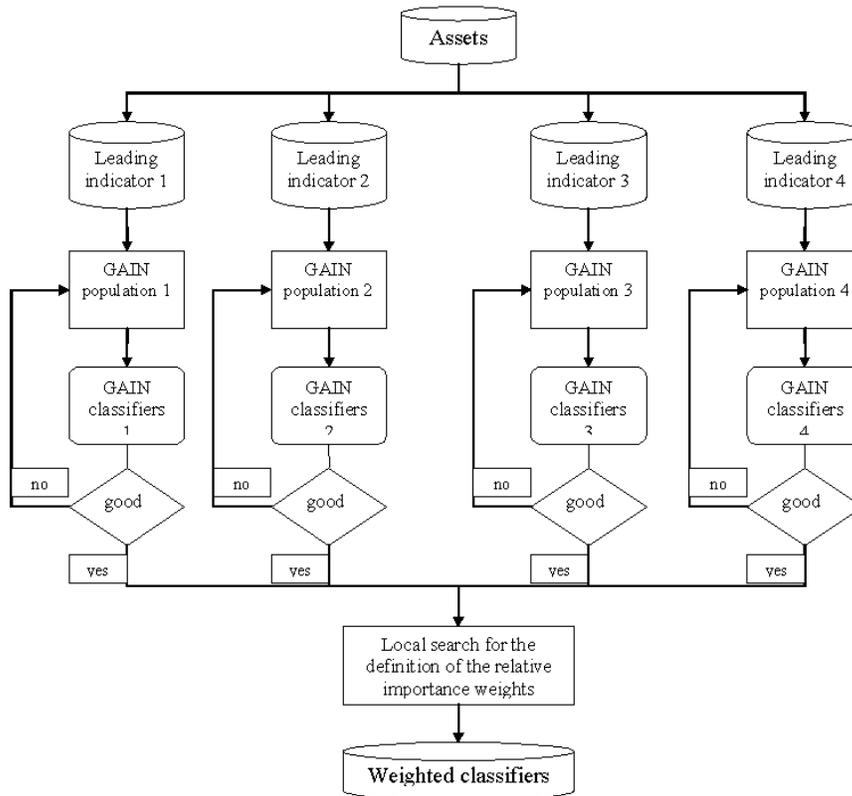
Figura 6.1: GAIN flowchart

The GAIN tool may be very useful as a support for asset allocation in active portfolio management mainly because it benefits from the interaction between practitioners and evolutionary computations during its implementation and, least but not last, it is a general tool that can be applied to stocks, bonds and interest rates.

6.5 GAIN tool in a real fund

The GAIN tool is already implemented as a support to management decisions in the Absolute Return Interest Rate Strategies sub-fund of Generali

Investments. The characteristics of this fund are derived from the web site `www.generali-investments.com`:

> *The Sub-fund will invest mostly in Euro-denominated bonds issued by governments, agencies and supranational entities. Furthermore, the Sub-fund will use interest rate and currency derivatives, both on regulated markets and over the counter, for investment purposes. The sub-fund is an absolute return fund implying a low tolerance to risk, a low correlation to markets and to the main traditional indexes, with the objective to realize an absolute net performance. The Sub-fund seeks to generate an absolute return in excess of the EONIA rate by an active asset allocation into a diversified portfolio of debt related securities and currency derivatives, taking into account the Investment Manager's views on interest rates and yield curves in the main OECD markets. At any time at least two third of the Sub-fund's total assets are invested in fixed and floating rate government/supranational bonds, including derivatives on such instruments. The average duration of the Sub-fund's portfolio will range from minus 5 years to plus 7 years.*
>
> *None of the assets of the Sub-fund shall be invested in convertible bonds, shares and other participation rights in corporate bonds. The maximum unhedged exposure to currencies other than Euro will not exceed 30% of the Sub-fund's net assets.*
> *The benchmark of the portfolio is the EONIA rate.*

Figure 6.2 displays the performance of the fund, relative to its benchmark for the period from 1 July 2008 to 13 January 2009. It can be shown that from the beginning of July the fund has yielded more than the EONIA rate (+285 bps), with a total return of 4.83% (9.18% as annualized return) and a risk-profile (1.50%) less than that of the 2-year bund (2.40%). These results, in general, demonstrate that the application of both qualitative and quantitative tools may considerably improve the portfolio management.

Bibliografia

Cvitanic, J. and Zapatero, F.: 2004, *Introduction to the Economics and Mathematics of Financial Markets*, The MIT Press.
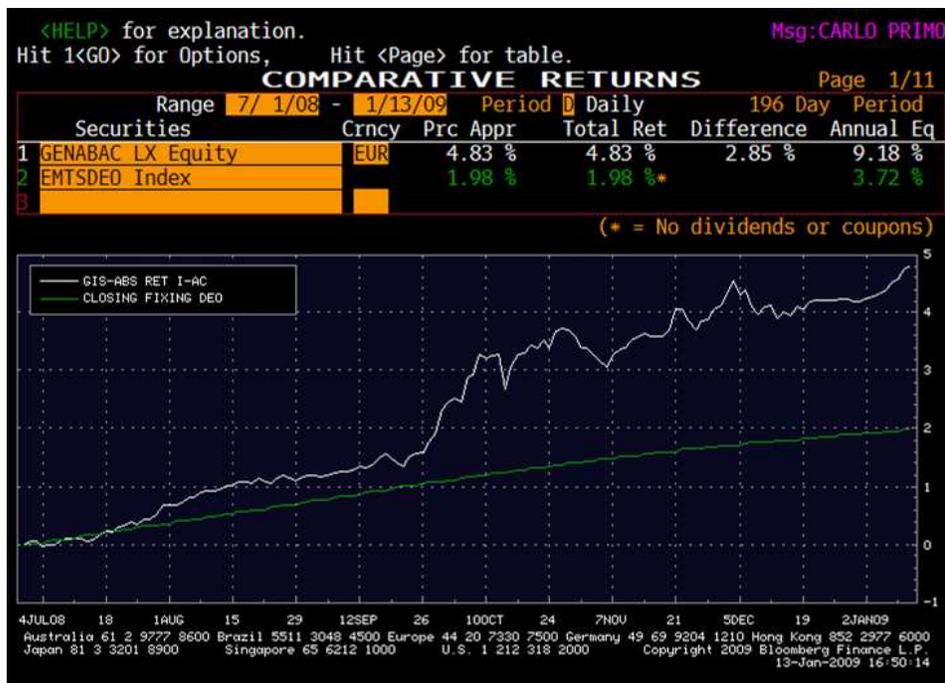
Figura 6.2: Performance of the fund

Fernandez-Rodríguez, F., González-Martel, C. and Sosvilla-Rivero, S.: 2005, Optimization of technical rules by genetic algorithms: evidence from the Madrid stock market, *Applied Financial Economics* **15**, 773–775.

Jacobs, B. I. and Levy, K. N.: 2005, *Market Neutral Strategies*, John Wiley & Sons, Inc.

Korczak, J. and Lipinski, P.: 2001, Evolutionary approach to portfolio optimization, *Proceedings of Workshop on Artificial Intelligence for Financial Time Series Analysis*, Porto.

Korczak, J. and Roger, P.: 2002, Stock timing using genetic algorithms, *Applied Stochastic Models in Business and Industry* **18**, 121–134.

Papadamou, S. and Stephanides, G.: 2007, Improving technical trading systems by using a new MATLAB-based genetic algorithm procedure, *Mathematical and Computer Modelling* **46**, 189–197.

Pole, A.: 2007, *Optimal Trading Strategies: Quantitative Approaches for Managing Market Impact and Trading Risk*, John Wiley & Sons, Inc.

Weir, D. J.: 2006, *Timing the market : how to profit in the stock market using the yield curve, technical analysis, and cultural indicators*, John Wiley & Sons, Inc.

Whistler, M.: 2004, *Trading Pairs*, John Wiley & Sons, Inc.