Evolving Investment Models using Genetic Network Programming and Genetic Relation Algorithm

Victor Parque

The Graduate School of Information, Production and Systems

Waseda University

A thesis submitted for the degree of

Doctor of Engineering

2011/06

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract

As global financial innovation opens innumerable risks and opportunities, the economic markets are evolving from highly localized trading places toward global platforms for risk sharing. In this context, devising global investment strategies that tackle the increased market complexity and boost the sustainable economic growth is becoming a key issue in public and corporate agendas. And to a large degree, the recent financial crisis has also enforced the need of enhancing the resiliency of the global investment systems to mitigate the risk exposure to reasonable levels.

Traditional Finance has created global investment strategies that maximize return and minimize risk by focusing on interdependent risks and distributional assumptions of *Modern Portfolio Theory*(MPT). Over the last decade, recent studies have also integrated investor's behaviors and risk preferences in order to get more accurate models and reasonable economic performances.

Developing global investment models also involves the systematic quest of the margin of safety, or a favorable difference between the price and the intrinsic value. Although this variable might not be quantified with exact precision, it may be approximated through the underlying relationships in financial markets and the real economy. In this context, key factors such as multiple risks, global asset classes, and intrinsic value creation play crucial roles.

This thesis aims at incorporating these variables while complementing upon *Genetic Network Programming*(GNP) and *Value Investing*(VI) principles to build global and diversified portfolios that tackle risk misspricing and behavioral bias involved in speculative investing. The principles of VI uses the factors related with the margin of safety to create wealth over the long term; and GNP is a robust search architecture that enables the improved exploration ability of the factors that determine the margin of safety. The

juncture of both constructs implies designing an improved heuristic to assess the interdependent risk in the context of traditional finance. This field has received limited attention compared with portfolio trading and optimization, and should contribute to building the value investing models of the next generation that are commensurate with the new realities of global risk interdependence.

Therefore, the objective of this thesis is to develop algorithms for building global investment models, which are able to diversify the risks while allocating the scarce economic resources in multiple asset classes, spread in developed financial markets, by using the principles of evolution of GNP and pricing of VI. The following chapters are organized as follows:

Chapter 1 presents the motivations, the aims and the structure of this thesis.

Chapter 2 proposes a methodology to build optimal asset selection models using *Genetic Network Programming* (GNP), which builds network oriented risk pricing models embedded with intrinsic and extrinsic risk factors. The basic idea of GNP is to build flexible decision making networks to assess, through risk factors, how valuable and attractive the assets in financial markets are. The number and the type of risk factors relevant to build the risk pricing model are decided by the evolutionary principles of GNP. The proposed methodology is compared to relevant benchmarks used in financial practice, such as the widely known *value*, *growth* and *capitalization* indexing strategies.

Chapter 3 introduces a methodology to build robust asset selection models by using *Robust Genetic Programming*(r-GNP). The basic idea of r-GNP is inspired by how *evolvabilty* and *robustness* are complementary properties in the development of biological organizations, that is, individuals have better chances to survive and have better generalization ability if they acquire and accumulate crucible experiences when they are exposed to a relevant set of environments/experiences. Simulations show that the generalization ability of r-GNP has benefits over the standard GNP approach and the benchmarks used in financial practice.

Chapter 4 introduces a methodology to build adaptive asset selection models by using *Genetic Network Programming with Changing Structures*

(GNP-cs). The basic idea of this system comes from biologically adaptable systems which incorporate *control* functions in their organization to monitor and guide the self-adaptation to the changing environments. The unique point of GNP-cs is to introduce a guiding control mechanism to self-change the structure for the asset selection depending on the fluctuations in the real economy. GNP-cs is compared to the standard GNP approach and benchmarks used in financial practice.

Chapter 5 proposes a methodology to build optimal asset allocation models by using *Genetic Relation Algorithm* (GRA). The basic idea of GRA is inspired by how compact networks survive by focusing on partial and relevant relationships. Thus, GRA models the asset portfolios though undirected network structures where each node in the network represents financial assets, such as *stocks*, *bonds* and *currencies*, and each relationship focuses on the systematic risk, in terms *of portfolio beta*. Which nodes and which connections are relevant are decided by the evolutionary structure of GRA, which is compared to relevant benchmarks in the asset allocation context.

Chapter 6 introduces a methodology to build optimal portfolio diversification models by using *Genetic Relation Algorithm with Variable Size* (GRA-vs). The basic idea of GRA-vs comes from biological organizations that expand/shrink their internal structure during the period of evolution to systematically enhance their survivorship ability. vs-GRA has the role of building flexible portfolio structures considering variable size structures during the evolution process, in which the expansion/shrinkage is guided probabilistically using diversity metrics. vs-GRA is compared with the standard GRA in the portfolio allocation problem.

Chapter 7 concludes the thesis by highlighting the remarks of each chapter.

CHAPTER

1

Introduction

1.1 Background

Over the last decade, much work has done in terms of Modern Portfolio Theory to tackle risk management problems in global investment strategies. Studies are mainly based on mean-variance analysis and market-investor rationality assumptions, which are firmly integrated into a discipline of econometric models of uncertainty. Recent studies have also integrated investor's behaviors and risk preferences in order to get more accurate models and reasonable economic performances.

However, a crucial insight behind risk management is that it needs to assure a primary reason in investment: to create value in the long term. In this context, key factors such as multiple risks, asset classes, and intrinsic value creation play crucial roles. This thesis aims at incorporating these variables while complementing upon *Genetic Network Programming(GNP)* and *Value Investing(VI)* principles to get well diversified portfolios and to tackle *risk-misspricing* and *behavioral bias* involved in speculative investing. This field received limited attention in comparison with portfolio trading and optimization, and contributes to building the investing models in the next generation, which commensurates with the new realities of increasing financial complexity and global risk interdependence.

1.2 Contents

1.2.1 Objective

This thesis develops algorithms for making investment models, which are able to diversify the risks while allocating the scarce economic resources in multiple asset classes spread in developed financial markets, by using the principles of evolution of *GNP* and pricing of *VI*.

1.2.2 Research topics

- Chapter 2 proposes a methodology to build *asset selection models* using *Genetic Network Programming(GNP)*, which builds the risk pricing models based on *intrinsic* and *extrinsic* risk factors by using the evolutionary structures of *GNP*. The number and type of factors needed to build the risk pricing model is decided by evolution. The proposed methodology is compared with relevant benchmarks used in financial practice. It is found from simulations that GNP-based asset selection outperforms the benchmarks in terms of wealth accumulation over the long term, implying an improved ability to identify undervalued opportunities.

- Chapter 3 introduces a methodology to build *robust asset selection models* by using *Robust Genetic Network Programming r-GNP)*. The basic idea of *r-GNP* is inspired by how *evolvabilty* and *robustness* are complementary properties in the development of biological organizations, that is, individuals have better chances to survive and better generalization ability if they acquire and accumulate different meaningful experiences when they are exposed to a relevant set of environments. Simulations show that the generalization ability of *r-GNP* brings benefits on return, risk and liquidity over the standard *GNP* approach and the benchmarks used in financial practice.

- Chapter 4 introduces a methodology to build *adaptive asset selection models* by using *Genetic Network Programming with Changing Structures*. The basic idea comes from biologically adaptable systems which incorporate control functions in their organization to monitor and guide the self-adaptation to the changing environments. The unique point of *GNP-cs* is to introduce a guiding control mechanism to self-change the structure for the *asset selection* depending on the fluctuations in the real economy. Simulations show that the adaptive mechanism of GNP-cs brings benefits in wealth accumulation over the long term over the standard GNP and benchmarks used in financial practice.

- Chapter 5 proposes a methodology to build *asset allocation models* by using *Genetic Relation Algorithm(GRA)*. The basic idea of GRA is to model a portfolio in an undirected graph structure that focuses on risk relationships to capture the *systematic risk*. GRA is compared to relevant benchmarks for in the asset allocation context. It is found from simulations that evolving asset allocation structures through GRA has benefits in profit accumulation over a global market index (DJ Gloabl 1800) and standard portfolio optimization techniques in Traditional Finance and Computer Science literature.

- Chapter 6 presents a methodology to build *portfolio diversification* models by using *Genetic Relation Algorithm with Variable Size(GRA-vs)*. The basic idea of GRA-vs comes from biological organizations that expand/shrink their structure during the evolution process and systematically enhance their survivorship ability. vs-GRA has the role of building flexible portfolio structures considering the variable size during the evolution, which are guided probabilistically using diversity metrics. vs-GRA is compared with the standard GRA approach, which shows that the flexibility of GRA has benefits to decide on the optimal spread over asset classes, sectors and countries, implying an improved return performances over the long term.

CHAPTER

2

Asset Selection with Genetic Network Programming

2.1 Aims of the Proposed Method

This chapter:

- Introduces a methodology to build *asset selection* models using *Genetic Network Programming(GNP)*.
  - The methodology uses *evolutionary computing* and *value investing* principles to find the *optimal asset selection* models.
  - *Judgment* and *processing* nodes in the network structure of GNP use the *intrinsic* and the *extrinsic* risk factors to decide on the *asset selection* decision.
  - The *fitness function* is designed to asses *return*, *risk* and *liquidity* as main objectives.
- Compares the proposed scheme with benchmarks used in financial practice through simulations.

- Simulations use the assets listed in the *Russell Developed Index*.
- Simulations are executed through *sliding time periods* between Jan 2000 and Dec 2010.
- Benchmarks include widely known *indexing strategies* such as *Value*, *Capitalization* and *Growth*.

2.2 Background

Financial markets are continuously evolving from highly localized trading places toward sophisticated and intertwined global platforms. This fact has brought significant challenges for global risk management and, particularly, for asset selection practices. The attractive and non-toxic assets are to be identified so that the health of the financial markets, the smooth return premium for institutional investors and the competitive edge for businesses are safeguarded.

Apart from portfolio optimization, *asset selection* is primarily concerned with the task of identifying prospective assets from financial markets for investment purposes. Depending on the asset composition, formal practices of asset selection include active and passive strategies.

- *Active* strategies assume that the underlying value of each asset(or a group of them) can be estimated though explicit risk *pricing* models, which incorporate either *fundamentals* and/or *stock market* elements to separate the undervalued assets from the overvalued financial assets. Remarkable models include the *Value Investing* approach(1), the *Mean Variance* model(2), the *Capital Asset Pricing Model* (3), the *Option Valuation* model(43) and the *Multiple Valuation(MV)* approach(5). The advantage of this strategy is that companies that have potential to grow in the future can be identified through the individually developed pricing models. On the other hand, although the pricing models developed independently may lead to less correlative issues since individual thinking may not be correlated with the market behavior, and human intervention may also incite behavioral bias and thus make models miss-price risk.

- *Passive* strategies assume that the prospective assets are grouped in leading financial *market indexes*. All forms of *indexing*, including the frequently used *long-only 130/30*(6), are examples under this category. For example, a simple passive strategy would replicate the S&P 500 *index* in an investable portfolio of 500 assets with weights proportional to their market capitalization. The advantage of this approach lies in the conservatism to protect individual investors from making mistakes in stock markets. However, since strategies tend to be correlated in terms that thinking/trading is correlated, systemic risk is an issue.

The models mentioned above shed light on important building blocks in the finance; however, the active and passive strategies may endure some limitations such as:

- *Pricing issues.* Conventional strategies require fixed a-priori measure definitions to price the risk exposure or the expected return of assets. Widely known measures include variance(2), value at risk, price to book(1) and the book to market ratio(7). Nevertheless, complex factors such as financial innovation and bounded rationality of investors call the presence of dynamic measures to avoid risk mispricing issues(8, 9, 10).
- *Structural issues.* In the last two decades, AI techniques have emerged to aid the asset selection task. However, structural issues such as the trees's bloating problem of GP, the strings' inability to express underlying relationships of GA and the black box issue of Neural Networks undermine their efficiency or legibility to deal with risk pricing issues.

In order to tackle these issues, this chapter introduces an asset selection model based on Genetic Network Programming(GNP)(11), which is one of the first models that tackle the asset selection problem by using evolutionary networks.

2.3 Definitions

An asset selection strategy has the role of identifying the prospective assets from the financial markets, which can be generally characterized by:

$$Prospective\ Assets_t = f(IC_t, M_t, m_t), \tag{2.1}$$

$$m_t = f(History_{\leq t}, Modeling\ Theory_t, Expectations_t) \tag{2.2}$$

The notation of $f()$ means "*is a function of*". The above suggests that the prospective assets at time period t depends on the investor's characteristics IC_t at time period

t, the investment universe in the financial market M_t at time period t, and the asset selection solution m_t at time period t; which in turn depends on the $History_{\leq t}$ of the underlying risk factors of every asset $i \in M_t$ for the periods up to time period t, the $Modeling\ Theory_t$ to represent and build m_t, and the $Expectations_t$ of the future prospects of every asset $i \in M_t$.

Assuming that,

- IC_t represents the features of a risk averse investor,
- M_t is the market index M at time period t,
- $History_{\leq t}$ is given by fundamentals and market related factors of every asset $i \in M$ for the periods up to time period t as shown in Table 2.1, and
- $Expectations_t$ is given by the Market Values MV_t(prices) of every asset $i \in M$ at time period t,

Thus, in order to find the $Prospective\ Assets_t$ at time period t, we need explicit $Modeling\ Theory_t$; that is, how to represent the asset selection solution m_t, and how to build the optimal solution m_t^*. These questions will be discussed next.

2.4 Genetic Network Programming on Asset Selection

2.4.1 Basic Concept

The basic idea of this chapter is to use evolutionary networks to build solutions (m_t) as risk pricing models to evaluate the assets in a financial market M by using the *value investing principles*: which is to consider the asset *value* and *attractiveness* into the GNP in the context of asset selection.

2.4.2 Main Features

In the context of asset selection, the proposed approach contributes to the following features:

- The compact network of reusable nodes in the GNP structure balances the aspects of the asset value and attractiveness in the selection mechanism. To put it more bluntly, it means introducing the concepts of evolution when building an asset selection model through an active indexing strategy; which implies designing an improved heuristic to asses risk in the context of traditional finance.

- GNP enables a parallel search mechanism of evaluation measures which may be unimaginable for financial experts, overcoming local optima or bias issues in formal asset selection approaches. Designing a compact network through GNP has implications for building a compact asset selection model, which could also minimize financial behavioral anomalies such as *conservatism* and *representativeness* in expert based approaches(12).

- Compared to other evolutionary algorithms, the GNP's network structure avoids the bloating and black box issues, making the asset selection process efficient and legible.

The different points from the conventional financial engineering methods are the following:

- The asset selection using GNP is an extension of the passive strategy in the sense that it considers the market index M as an investable universe(6). However, instead of relying on the full set, we aim at identifying a basket of prospective assets from M by using the asset selection models built upon the evolvability concepts of GNP. To put it more bluntly, not only the optimal combination of metrics, but also the set of prospective assets for investment is decided by evolution of *GNP*. This scheme has implications on building not only the enhanced but also the robust *indexing strategies*.

- Second, The asset selection using GNP is an active strategy in the sense that it builds explicit models for risk pricing. Instead of using ratios(1), statistical equations(2), compounded indexes(3), pricing rules(13), *trees* structures (14), and more recently, syntaxes(15), we propose using the networks that incorporate not only the *intrinsic*, but also the *extrinsic* risk factors embedded in *judgment* and *processing* nodes. This scheme implies an exhaustive tool for the building *risk pricing models*.

2.4.3 Structure of GNP for Asset Selection

Every GNP individual is a solution m_t, and is expressed in a graph structure which contains nodes connected by directed edges, as shown in Fig. 2.1. Concretely, there are four elements embedded in every GNP individual:

- A single *start node* indicates the first node to be executed.

- The *judgement nodes* evaluate the value and attractiveness of assets.
- The *processing nodes* makes the asset selection decision based on evaluation results of judgment nodes, i.e. whether to add assets or not in the *Asset pool*.

Gene structure

Each node in the GNP individual is encoded in a gene structure, whose complete set shapes the asset selection model. To illustrate this mechanism, Fig. 2.2 shows the encoding scheme of a node r, whose elements are defined by the following:

- *Node type*(NT_r) which defines the type of node r, where $NT_r = 0$ implies that node r is a starting node, $NT_r = 1$ implies that node r is a judgement node and $NT_r = 2$ implies that node r is a processing node.
- *Intrinsic factor component*(IF_r) which stores the identification number of the financial metric quantifying *intrinsic factors* and their thresholds to decide how valuable an asset is. For example, if $NT_r = 1$ and $IF_r = 2$, then node r refers to I_2 in *Metric Library*.
- *Extrinsic factor component*(EF_r) which stores the identification number of the financial metric quantifying *extrinsic factors* and their thresholds to decide how attractive an asset is. For example, if $NT_r = 1$ and $EF_r = 3$, then node r refers to E_3 in *Metric Library*.
- *Connections*(C_{rs}) which defines the node connected from node r using the s_{th} branch according to the arguments of node r in the graph structure.
- *Delay time*(d_r and d_{rs}) which represent the delay times in node r and connection C_{rs}, respectively. The delay times limit the number of judgment nodes to be executed, preventing from the loops in the route transitions.

Judgment node

To give a balanced view about the risks that assets may be exposed to, the judgment nodes are explicitly designed to evaluate both the *intrinsic* and *extrinsic factors* that determine the asset's expected return.

- *Intrinsic factors* quantify the asset value and growth as fundamental financial metrics(16, 17, 18). For instance, the management quality and the firm value

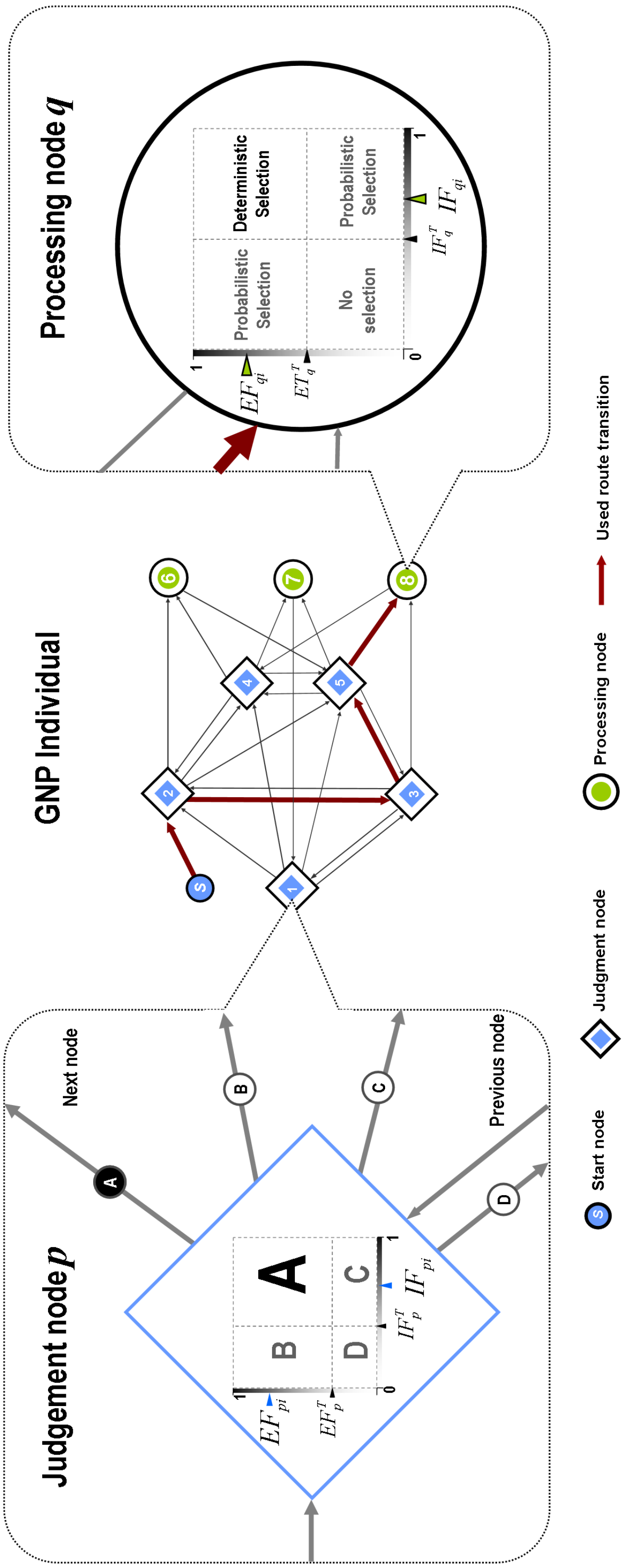


Figure 2.1: GNP Individual, judgment and processing node

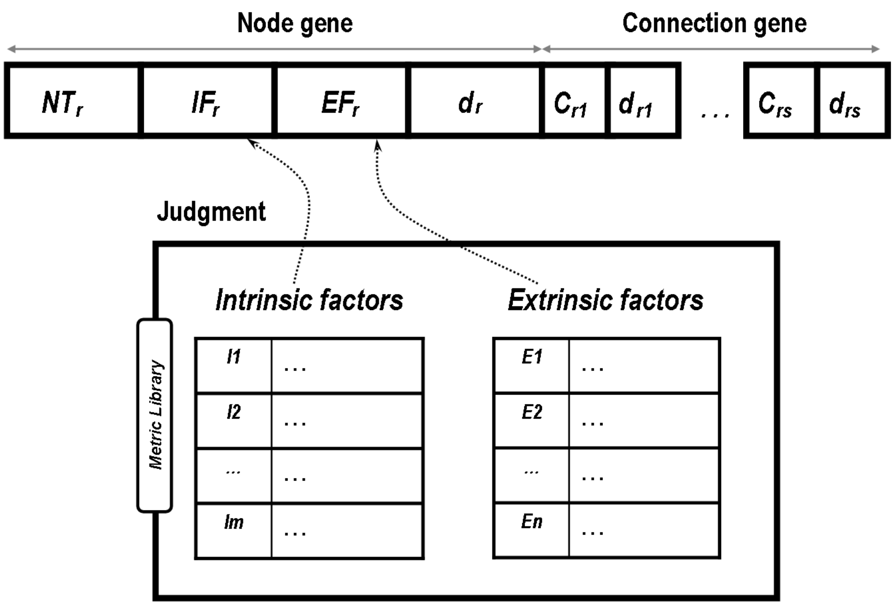


Figure 2.2: Gene structure of GNP for Asset Selection

chain efficiency are implicitly measured by these metrics. We use relevant metrics in finance literature and practice, as shown in Table 2.1.

To evaluate the *intrinsic factors*, each judgment node has two components: a *metric* and a *threshold*. While the *metric* component picks up financial information using one of the metrics in Table 2.1, the *threshold* divides the metric into high and low levels. For instance, the left side of Fig. 2.1 shows the judgement mechanism of asset i by judgment node p. In our example, IF_{pi} is the metric and IF_p^T is its threshold. Each metric is normalized in the range of 0 and 1 to allow the comparison and aggregation among different metrics.

- *Extrinsic factors* quantify the asset attractiveness as market driven factors. Metrics under this category focus on the interactions of agents in specific financial markets that significantly influence the asset expected return. Concretely speaking, we use components of return, volatility and liquidity, as shown in Table 2.1.

  To evaluate the *extrinsic factors*, each judgment node has a normalized metric EF_{pi} and its threshold EF_p^T, as described in Fig. 2.1.

The judgment nodes combine the normalized metrics and the thresholds of the *intrinsic* and *extrinsic factors* into if-then type decision functions, in which four areas are determined, as shown in Fig. 2.1

Table 2.1: Metric Library in GNP

Id.	Description
	Intrinsic factors
I_1	Dividend to price
I_2	Earnings to price
I_3	Cash flow to price
I_4	Book value to price
I_5	Sales to price
I_6	Short term change in earnings to price(3 months)
I_7	Long term change in earnings to price(2 years)
I_8	Short term change in cash flow to price(3 months)
I_9	Earnings surprise
I_{10}	Profit margin(Net operating income to sales)
I_{11}	Return on assets(Net operating income to total assets)
	Extrinsic factors
E_1	Excess return to S&P500
E_2	Excess return to risk free asset
E_3	Rate of return
E_4	Sharpe Ratio
E_5	Beta
E_6	Volatility of return rate
E_7	Volatility of beta
E_8	Market price per share
E_9	Trading volume
E_{10}	Turn over ratio
E_{11}	Bid ask ratio

Processing node

After the evaluations of the judgment node, the processing nodes perform the decision making on the asset selection in two phases. Fig. 2.1 shows the decision making mechanism of processing node q to deal with asset i:

- First, compute how valuable and attractive asset i is, i.e., IF_{qi} and EF_{qi} by using the judgement nodes in transition TN, which is defined as the judgment nodes from the previous processing node to the current processing node q.

$$IF_{qi} = \frac{1}{|TN|} \sum_{p \in TN} IF_{pi}, \tag{2.3}$$

$$EF_{qi} = \frac{1}{|TN|} \sum_{p \in TN} EF_{pi}, \tag{2.4}$$

where TN is the set of suffixes of judgement nodes in the transition; IF_{pi} and EF_{pi} are the *intrinsic and extrinsic factors* in judgement node $p \in TN$ that drive the expected return and risk exposure for asset i; and IF_{qi} and EF_{qi} are the *intrinsic and extrinsic factors* to perform the decision of selecting asset i by processing node q.

- Second, perform the decision on selecting asset i:

  - [Deterministic selection]
    Select asset i into *asset pool*.
    If $IF_{qi} \geq IF_q^T$ and $EF_{qi} \geq EF_q^T$.
  - [Probabilistic selection]
    Select asset i in *asset pool* with the probability of ϵ_1.
    If $IF_{qi} \geq IF_q^T$ and $EF_{qi} < EF_q^T$.
  - [Probabilistic selection]
    Select asset i in *asset pool* with the probability of ϵ_2.
    If $IF_{qi} < IF_q^T$ and $EF_{qi} \geq EF_q^T$.
  - [No selection]
    Discard asset i.
    If $IF_{qi} < IF_q^T$ and $EF_{qi} < EF_q^T$.

The deterministic selection concentrates on identifying the assets which are highly valuable and attractive, while the probabilistic selection is designed to enhance the exploration ability of GNP to find undervalued or overvalued assets in financial markets, i.e., opportunities of mispriced positions as a consequence of financial market inefficiency factors(17, 18).

2.4.4 Fitness function of GNP

By using the information in the judgment and processing nodes, every *GNP individual* selects a group of assets from the market universe M into the *Asset pool* A_m. Thus, evaluating the fitness of the *GNP individual* implies measuring the economic performance of the *Asset pool* in a specified time period, i.e., buying and holding the selected assets of the *Asset pool* throughout time period t. Thus, to guide the evolution mechanism in the training period, the fitness of a *GNP individual* measures the performance as follows:

$$F^m = \frac{\sigma_{A_m}.\beta_{A_m}}{(R_{A_m} - R_F).L_{A_m}}, \tag{2.5}$$

$$R_{A_m} = \frac{1}{|A_m|} \sum_{i \in A_m} \frac{(P_i^f - P_i^o + div_i)}{P_i^o}, \tag{2.6}$$

$$L_{A_m} = \frac{1}{|A_m|} \sum_{i \in A_m} MC_i, \tag{2.7}$$

where,

A_m : set of suffixes of assets in *Asset Pool* A_m selected by model m at time period t.
R_{A_m} : average return of A_m at time period t.
P_i^o : the opening price of asset i at the beginning of time period t.
P_i^f : the closing price of asset i at the end of time period t.
div_i : the dividends paid by asset i at time period t.
σ_{A_m} : the standard deviation of the returns of the assets in A_m at time period t.
β_{A_m} : the average beta coefficient of A_m relative to market M at time period t.
R_F : average risk free rate defined by 3-month U.S. Treasury Bill at time period t.
L_{A_m} : average liquidity level of A_m at time period t.
MC_{it} : the normalized market capitalization of asset i in time period t.

The lower fitness values are preferred over the larger ones. The advantages of using F^m as the fitness function are as follows:

- It measures the market and volatility risk exposure. While standard deviation σ_{A_m} conventionally measures the volatility of returns as the risk inherent to each asset, beta β_{A_m} captures the systematic risk, which is the component associated with aggregated returns in dependant financial markets(19).

- It maximizes liquidity per unit of risk, avoiding the asset exposure to liquidity risk issues when financial markets turn volatile or distressed. The works of Amihud and Mendelson(20), and Hill (9) show additional discussions on liquidity risk.

2.4.5 Genetic Operators of GNP

To make evolution successful toward the optimal solutions m_t^*, which means exhaustive and effective asset selection, the graph structure and parameters in the GNP individuals are evolved through generations. Basically, the selection, crossover and mutation operators perform this task.

Selection

Tournament selection is carried out. The elite individual, i.e., the one with the best fitness function, is moved to the next generation. Tournament selection is used because selection pressure can be easily adjusted by the tournament size.

Crossover

Crossover generates two offspring by exchanging the information in two parent individuals. As shown in Fig. 2.3, the following procedure is carried out:

- Select two GNP parent individuals by tournament selection.
- Select the nodes in GNP parent individuals with the probability of P_c.
- Exchange the selected nodes in parent individuals.
- New individuals consist of the new population in the next generation.

Mutation

Mutation generates a new individual by changing parameters in a parent individual. As shown in Fig. 2.4, the following procedure is executed:

- Select a GNP parent individual with tournament selection.
- Perform mutation operation to node r of the individual:
  - *Node Type.* NT_r is selected with the probability of P_m and changed to a different one.
  - *Node Measure.*
    - *Intrinsic factor measure.* IF_r is selected with the probability of P_m and changed to other variable in the *metric library.*

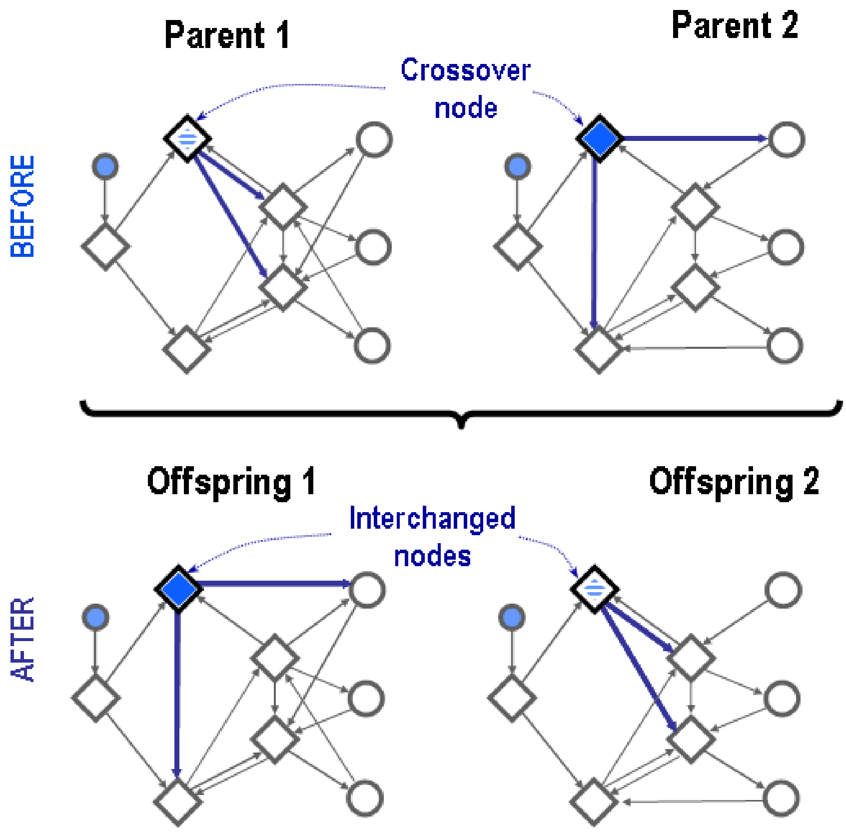


Figure 2.3: Crossover operation in GNP for Asset Selection

  * *Extrinsic factor measure.* EF_r is selected with the probability of P_m and changed to other variable in the *metric library*

- *Node Threshold.*
  * *Intrinsic measure threshold.* IF_r^T is changed to other value with the probability of P_m.
  * *Extrinsic measure threshold.* EF_r^T is changed to other value with the probability of P_m.
- *Node connection.* C_{rs} is selected with the probability of P_m and reconnected to a different node.

• New individuals consist of the new population in the next generation.

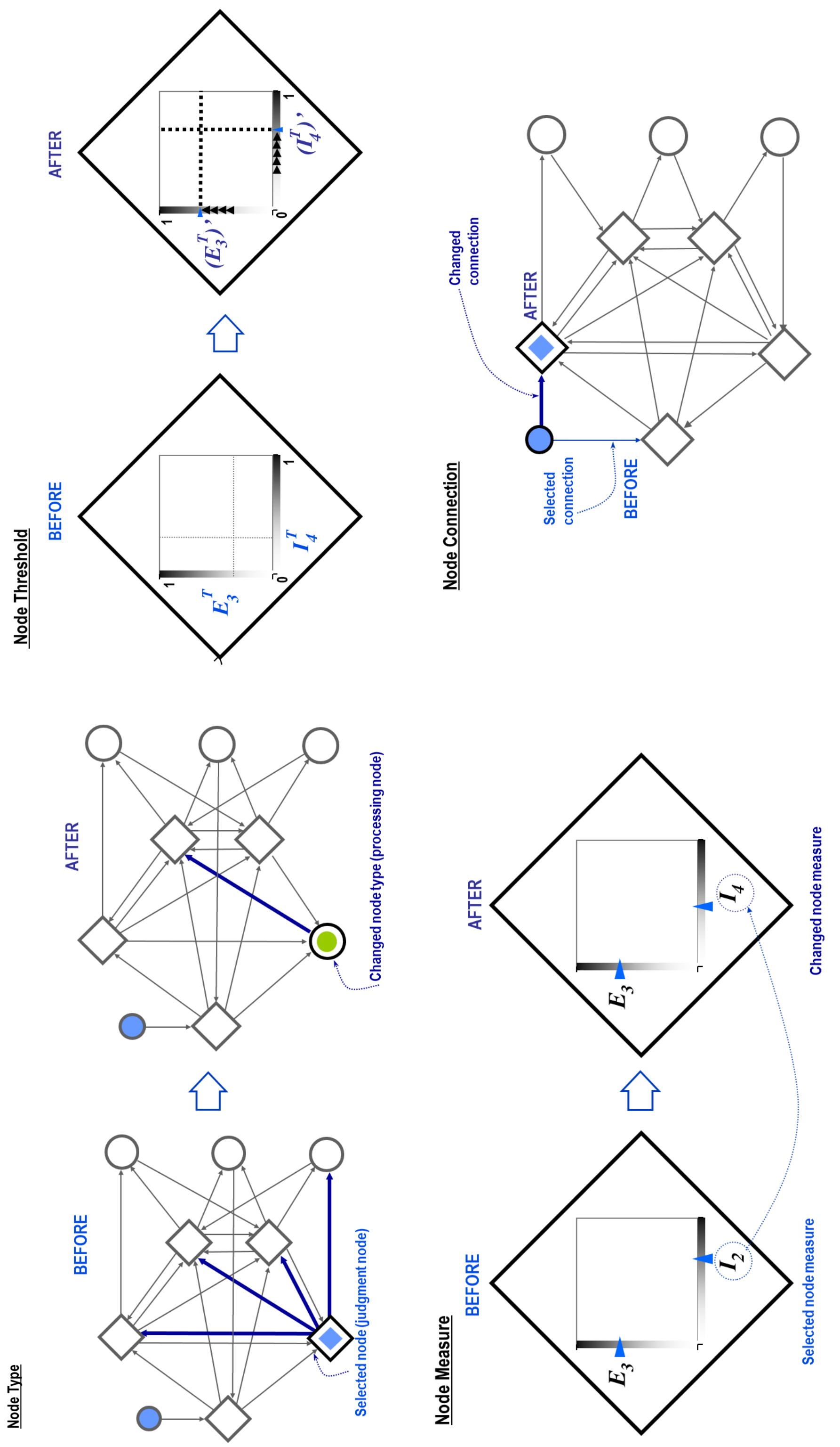


Figure 2.4: Mutation operation in GNP for Asset Selection

2.5 Simulations

2.5.1 Problem definition

An optimal subset A_m of common stocks is to be picked up from an investable market universe, which is defined as a stock market index M. The resulting subset is hypothetically invested using a buy and hold strategy over 1 month. Investment performance of GNP is compared with other common styles used in financial practice.

2.5.2 Investment Universe

The *market index M* consists of 2372 assets listed in the *Russell Developed Index*, which is a result of a pre-filtering process to avoid the assets with the following features: (1) with less than 3 years of data history, (2) lacking of market prices during the selection period, (3) with limited market capitalization and/or limited economic relevance(e.g. micro-cap stocks), and (3) with high correlation and reduced heterogeneity in the investment universe.

We choose the *Russell Developed Index* as the investment universe due to the fact that the representative assets from developed financial markets in U.S., Europe and Asia are identified in a single index, allowing better scope for diversification when performing the asset selection in a global scale.

The benchmarks considered for this paper include: the standard and widely used indexing strategies based on *value* and *growth* strategies(16, 17, 18, 21). All methods, including the proposed one, focus on long-only strategies.

2.5.3 Time Span

The total *time span* T for simulation is from Jan. 3rd of 2000 to Dec. 31st of 2009, which is divided in time periods $t \in T$, each consisting of $|TR|$ time units for *training*, followed by $|TE|$ units for *testing*. Each period time period $t \in T$ lags $|TE|$ time units one another, as the example that Fig. 2.5 shows. During the *training* phase of every time period t, GNP obtains the optimal asset selection model m_t^*; and during the *testing* phase, GNP obtains m_t^* to select subset A_m of prospective assets from market index M, which in turn is traded using a buy and hold strategy over $|TE|$ time units.

This *time span* is used because relevant financial collapses occurred in this period of time, providing better opportunities to compare the proposed method with other benchmarks in financial practice.

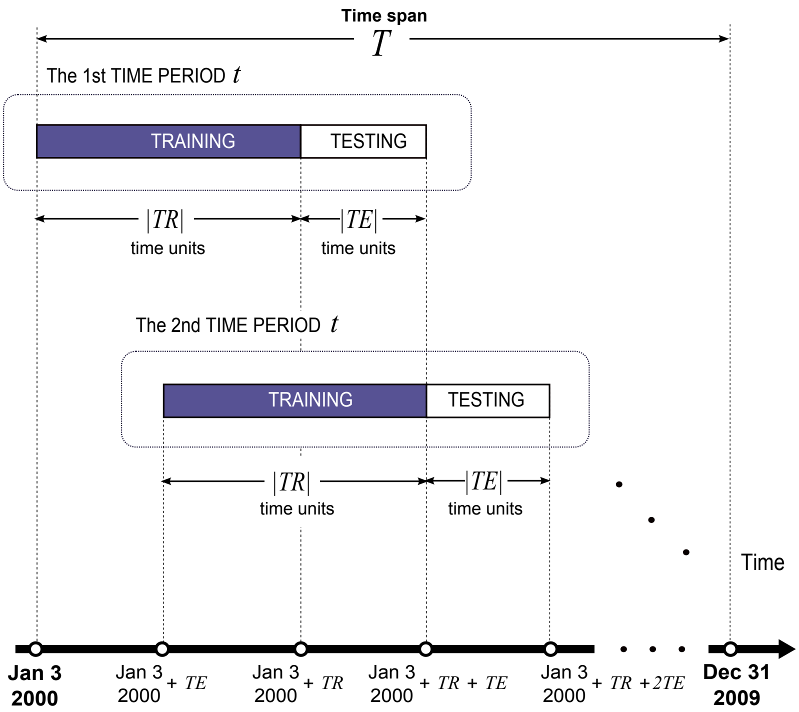


Figure 2.5: The *time span* used for simulations

2.5.4 Parameters

Table 2.2 shows the parameters for simulations, in which 30 independent runs are executed. The number of GNP individuals is 200, where 79 individuals are generated by crossover, 120 are generated by mutation and one is the elite individual. The probability of crossover and mutation are 0.1 and 0.01, respectively. The initial capital is $10, 000. The base currency for the fitness evaluation is the dollar.

2.5.5 Performance

Fig. 2.6 shows the average accumulated returns during the testing periods. *Wealth* means initial capital plus cumulative monthly returns: the initial *wealth* is 1(initial capital is 100%) and subsequent losses/gains are added/deducted in every testing period. The variables labeled as *Value*, *Growth* and *Cap* refers to the cumulative *wealth* of the assets selected by the *growth*, *value* and *capitalization* strategies. All methods include dividends, and assume the policy for reinvesting profits and no frictional expenses. We can see that GNP has better performance in terms of the return accumulation during

Table 2.2: Simulation Parameters

Item	Description	Value
GE	the number of generations for GNP evolution	200
I	the number of individuals in GNP	200
IC	the number of individuals by crossover	79
IM	the number of individuals by mutation	120
IE	the number of elite individuals	1
Pm	probability of mutation	0.01
P_c	probability of crossover	0.1
JN	the number of judgement nodes	25
PN	the number of processing nodes	12
SN	start nodes	1
$\|TR\|$	time span for training	2 years
$\|TE\|$	time span for testing	1 month

the testing period.

All methods show competitive and smooth cash flow generation, showing their effectiveness for the asset selection purpose. In the same line as Chan(21) and Brush(17), value based selection beats the growth based selection, showing its effectiveness for searching areas with mispriced positions. GNP-based asset selection outperforms smoothly the value based selection approach, implying a better ability to search undervalued positions.

2.6 Summary

In this chapter, an evolutionary based approach for the asset selection using GNP has been proposed. GNP evolves the value and attractiveness as selection measures. How many and what kind of measures are needed is decided by evolution. It is clarified from simulations that the proposed approach selects the prospective and valuable assets from the developed financial markets effectively.

However, since our approach is trained using historical financial information, it lacks of generalization ability, which means the overfitting of the GNP-based asset selection model to historical data. In finance, it means overreliance that the past economic performance can be extrapolated into the future. To tackle this issue, the next chapter

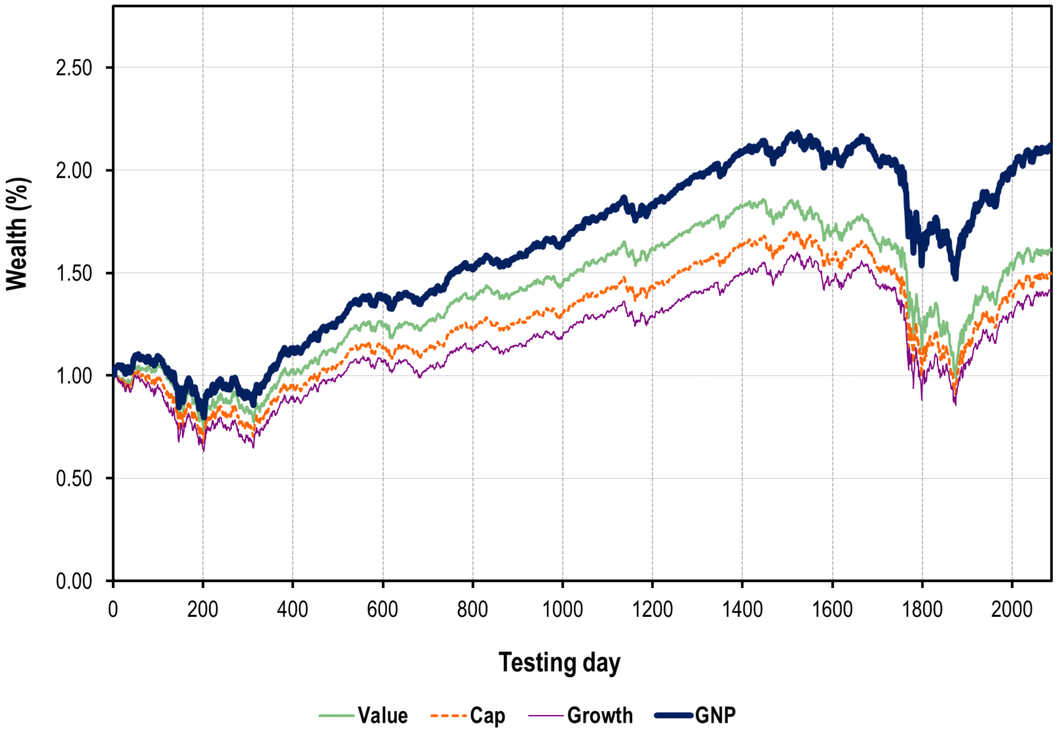


Figure 2.6: Accumulative wealth by GNP and the benchmarks

proposes a method to improve the generalization ability in the proposed asset selection model.

CHAPTER

3

Robust Asset Selection with Genetic Network Programming

3.1 Aims of the Proposed Method

This chapter:

- Introduces a methodology to enhance the *generalization ability* of the *asset selection* models based on *Robust Genetic Network Programming(r-GNP).*
  - The methodology introduces *robustness* principles into the *evolvability* concept of the standard GNP, where *individuals* use multiple and divergent(noisy) *environments* to enhance systematically the *survivorship ability* during the evolution process.
  - The *robust fitness function* evaluates not only the main performance function, but also the variability of performance through the *environments* in the *evolution* process of the standard GNP.
  - To track the dynamic optimum, r-GNP stores recent past optimal solutions through an *explicit memory mechanism* in time periods.

  - The aim of r-GNP is to stress-test the GNP-based *asset selection* models to avoid the *overfitting* to historical risk factors.

- Compares to the standard GNP approach and benchmarks used in financial practice.

3.2 Introduction

An asset selection model based on Genetic Network Programming(GNP) was proposed in the previous chapter. Although this is a population based scheme, implying not only the ability to avoid failing in local optimal due to its parallel-like search mechanism, but also the ability to handle uncertainty in optimization relatively well(22), it conventionally assumes that single and deterministic environments are relevant when searching for the optimum. This assumption leads to *overfitting* issues, implying the limited generalization ability, and lack of robustness to unseen cases in highly uncertain environments, which is the case of global financial markets.

To tackle this issue, we need to consider the fact that the *robustness* is a complementary property in evolution of complex organizations and living systems. Research on the *robustness* over different fields have shown certain commonalities. For instance, studies in Biology(23), Operations Research(24), Evolutionary Computation(25) and Finance(26) consider that an individual is robust if it is insensitive to small changes caused by internal and/or external variables.

Our interest in this chapter is to enhance the *robustness* of the *asset selection* task of our previous work(27), for which Robust Genetic Network Programming(r-GNP) is proposed to build generalized models and avoid *overfitting* issues in unseen uncertain environments. The basic idea of r-GNP comes from *Robust Universal Learning Networks*(28), which uses the second order derivatives of the main evaluation function with respect to inputs to improve the generalization ability of connected networks.

3.3 Robust Genetic Network Programming on Asset Selection

3.3.1 Basic Concept

r-GNP is essentially an algorithm inspired by nature, in which a population of candidate *solutions* are exposed to heterogeneous environments and evolved through darwinian principles of selection, reproduction and survival. Thus, a *solution* is *optimal* if it is able to survive in multiple environments continuously. More specifically, r-GNP incorporates the concept of the *robustness* into the *evolvability* property of the conventional GNP for the asset selection problem in the following forms.

- **Training more.** By adding multiple and divergent environments during evolution,
- **Adding perturbation.** By adding noises to the newly generated environments,
- **Tracking the dynamic optimum.** By using accumulative strategies of *individuals* through sliding time frames.

The first two approaches aim at improving the generalization ability of the individuals for unexperienced environments, and thus avoiding the overfitting issues of conventional evolutionary algorithms when trained in a single environment. In place of using multiple environments, other strategies exist as well. We could also enlarge the population size(24) or use adaptive genetic operators(14). However, for these approaches, the overfitting to a single training environment is still a potential issue since the individuals are not exposed to divergent training cases, which is essential for improving the generalization ability, thus improving the performance in unseen testing cases(29, 30).

The third approach aims at adapting to the continuous changes in the environments, which is important in applications where the optimum changes as time goes on.

3.3.2 Main Features

- r-GNP is a new scheme to enhance the robustness of the asset selection models; in which not only an strategy for tracking the dynamic optimum, but also a technique for creating the non-parametric and perturbed scenarios are integrated into a population based optimization algorithm. r-GNP evaluates the asset selection models against changing and divergent scenarios extensively, which improves the model's generalization ability. In Finance, it implies departing from the exclusive use of distributional assumptions for *stress testing*(31) and evading the extrapolation of the past performance into the future(*behavioral issues*).

- r-GNP uses accumulative strategies through time frames to track the optimum in dynamic environments. Instead of re-starting the evolution when the market environment changes, useful information from the recent past time frames is used for further evaluation. This feature enables adjusting the risk exposure of the asset selection models to the changing market conditions.

3.3.3 Basic Algorithm

Outline

r-GNP is designed to find solutions in a lifelong optimization context, where a time-variant problem over a period of time T is given. Algorithm 1 shows its basic procedures, and Fig. 3.1 shows its main components.

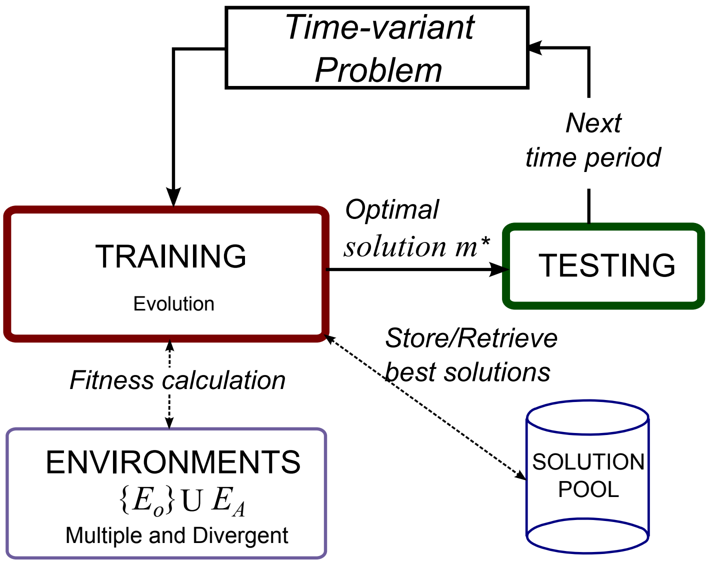


Figure 3.1: Main components of r-GNP

The algorithm of r-GNP follows the standard *training-testing* mechanism:

- During the *TRAINING* phase, the optimal solution m_t^* at time period t is evolved and stored/retrieved into/from the *Solution Pool* continuously by using not only the original(given) training environment E_o, but also the set E_A of artificially generated environments. Thus, the robust fitness F^m of the candidate solution m depends not only its the fitness function F_o^m in the original environment E_o,

Algorithm 1: Basic algorithm for r-GNP

input : A time-variant problem over a time span T

Solution Pool $\leftarrow \emptyset$;

for *each time period t in T* **do**

  `/* Training */`

  $E_o \leftarrow$ *Training environment(t)* ;

  $E_A \leftarrow$ *Create artificial environments(E_o)*, where $E_A = \{E_1, E_2, ..., E_g, ..., E_{|G|}\}$ and E_g is the $g - th$ artificially generated environment for time period t, and G is the suffixes of environments;

  Create a *Population* P of $|I|$ solutions by retrieving up to the $|I_{SP}|$ best solutions from the *Solution Pool*, and generating the rest randomly, where m_t is a solution candidate in P in time period t.

  while *termination condition is not met* **do**

    Evaluate the fitness F_o^m and F_g^m of each solution m_t in the environment E_o and each artificial environment E_g, respectively;

    Calculate the robust fitness $F^m = F_o^m + \lambda S_A^m$ of each solution m_t, where λ is a coefficient of user's aversion to volatility performance; and $S_A^m = \sqrt{\frac{1}{|G|}\sum_{g \in G}(F_o^m - F_g^m)^2}$, is the fitness deviation of each solution $m \in P$ in the set of environments E_A;

    Store the best solution m_t into the *Solution Pool*;

    Evolve the *Population* P;

  `/* Testing */`

  $m_t^* \leftarrow$ Pick the best solution m_t out of the current population P;

  $E_{test} \leftarrow$ *Testing environment(t)* ;

  Test(m_t^*,E_{test});

but also depends on its derived fitness function S_A^m in the set of artificial environments E_A. Whereas, the function F_o^m is the main objective evaluation function related to the problem itself, the function S_A^m is related to the notions of stability, smoothness and robustness of the candidate solution m. Evolving the population P of candidate solutions implies following the conventional selection, crossover and mutation of GNP in the previous chapter.

- During the *TESTING* phase, the optimal solution m_t^* is validated by using the given testing environment E_{test}.

Asset Selection with r-GNP

In order to find the optimal model m_t^*, Algorithm 1 evolves a population P of candidate models m_t for every time period t. However, in order to deal with the asset selection problem using r-GNP, we need to use the following elements that are peculiar in this chapter:

Environment E_o

The environment E_o is the original(given) training environment defined by the time series:

$$E_o = \{X_1, X_2, ..., X_h, ..., X_{(|TR|)}\} \tag{3.1}$$

where, X_h is the h-th data point of $History_{\leq t}$ and the $Expectations_t$ during time period t, and $|TR|$ is the length of the time series at time period t in the training phase. Fig. 3.2 shows an example of the environment E_o at time period t.

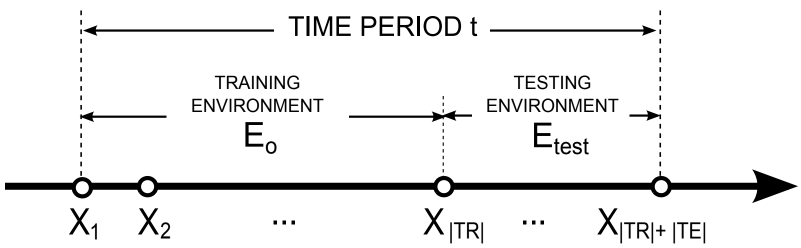


Figure 3.2: Explanation of time period t

Environment E_{test}

The environment E_{test} is the original testing environment defined by the time series:

$$E_{test} = \{X_{|TR|+1}, X_{|TR|+2}, ..., X_{(|TR|+|TE|)}\} \tag{3.2}$$

where, $|TE|$ is the length of the time series at time period t in the testing phase. Fig. 3.2 shows an example of the environment E_{test} at time period t.

Artificial Environments E_A

The set E_A of artificial environments is created by Algorithm 2, whose input is environment E_o and output is set E_A of artificially generated environments. The basic idea of this scheme is based on the *block bootstrapping* technique, which creates artificial scenarios for stress-testing by re-sampling consecutive *blocks*(sequential points in the time series) randomly(32). To estimate the *block* size l correctly, we refer to Hall et al.(33).

Generally speaking, conventional methods create multiple artificial environments either by using a longer period of time which is usually divided into *training, validation and testing*(*machine learning perspective*); or by using distributional assumptions of the time series, such as Monte Carlo or generalized ARCH(31)(*parametric perspective*). However, Algorithm 2 has the advantage of creating not only multiple, but also divergent environments, acting as unexperienced events that might be unimaginable for the *machine learning* or the *parametric* perspective. Divergent environments are useful to evaluate the internal structure of m_t and avoid its *overfitting* to historical data.

Algorithm 2: *Create artificial environments*(E_A)

input : A training environment $E_o = \{X_1, X_2, ..., X_h, ..., X_{(|TR|)}\}$
output: $E_A = \{E_1, E_2, ..., E_g, ..., E_{|G|}\}$.

1 Divide E_o into b disjoint blocks, where $E_o = \{B_1, B_2, ..., B_k, ..., B_b\}$, where $B_k = \{X_{(k-1).l+1}, X_{(k-1).l+2}, ..., X_{k.l}\}$ refers to the k-th block, and l refers to the size of each block;

2 $E_A \leftarrow \emptyset$;

3 **for** $g \leftarrow 1$ **to** $|G|$ **do**

4 $\quad E_g \leftarrow \emptyset$;

5 $\quad$ Choose $E_g = \{B_1^*, B_2^*, ..., B_k^*, ..., B_b^*\}$ by re-sampling $\{B_1, B_2, ..., B_k, ..., B_b\}$ randomly, where $B_k^* = \{X^*_{(k-1).l+1}, X^*_{(k-1).l+2}, ..., X^*_{k.l}\}$;

6 $\quad$ Add gaussian noise η to every data point of E_g;

7 $\quad E_A \leftarrow E_A \cup \{E_g\}$;

8 **return** E_A

Fitness F_o^m

The fitness F_o^m is the performance of the subset that solution m_t is able to select, thus is measured by the conventional fitness function of GNP in Eq. (2.5).

Fitness F_g^m

The procedure to calculate fitness F_g^m of solution m_t in environment E_g is the same as the procedure to calculate fitness F_o^m, in which environment E_g is used instead of E_o.

Evolve *Population* P

Evolving population P implies following the conventional operators of GNP, which is selection, crossover, mutation in Section 2.4.5.

3.4 Simulations

3.4.1 Problem

The same as chapter 1, indicated in Section 2.5.1, that is, an optimal subset A_m of common stocks is to be picked up from an investable market universe, which is defined as a stock market index M. The resulting subset is hypothetically invested using a buy and hold strategy over a period of $|TE|$ units of time. Investment performance of r-GNP is compared with the standard GNP approach and other common styles used in financial practice.

3.4.2 Investment Universe

The same as chapter 1, indicated in Section 2.5.2. In this chapter, more financial metrics are used, which are grouped into *intrinsic* and *extrinsic* factors as shown in Fig. 3.3. These variables represent a relevant set of four broad investment areas: value, growth, profitability, and momentum.

3.4.3 Time Span

The same as Chapter 1, indicated in Section 2.5.3, that is the total *time span* T for simulation is from Jan. 3rd of 2000 to Dec. 31st of 2009, which is divided in time periods $t \in T$, each consisting of $|TR|$ time units for *training*, followed by $|TE|$ units for *testing*. Each period time period $t \in T$ lags $|TE|$ time units one another, as the example that Fig. 2.5 shows. During the *training* phase of every time period t, r-GNP obtains the optimal asset selection model m_t^*; and during the *testing* phase, r-GNP obtains m_t^* to select subset A_m of prospective assets from market index M, which in turn is traded using a buy and hold strategy over $|TE|$ time units.

Intrinsic Measures	Extrinsic Measures
IF_1 Dividend to price	EF_1 Excess return to S&P500
IF_2 Earnings to price	EF_2 Excess return to risk free asset
IF_3 Cash flow to price	EF_3 Rate of return
IF_4 Book value to price	EF_4 Sharpe Ratio
IF_5 Sales to price	EF_5 Stirling Ratio
IF_6 Earnings surprise	EF_6 Volatility of return rate
IF_7 Return on Equity share	EF_7 Beta of return
IF_8 Return on Assets	EF_8 Volatility of beta
IF_9 Capital turnover	EF_9 Conditional value at risk
IF_{10} Profit margin	EF_{10} Market capitalization
IF_{11} Short term change in IF_1(1 month)	EF_{11} Trading volume
IF_{12} Long term change in IF_2(1 year)	EF_{12} Turn over ratio
IF_{13} Short term change in IF_3(1 month)	EF_{13} Bid ask ratio
IF_{14} Long term change in IF_3(1 year)	EF_{14} Price per share
IF_{15} Short term change in IF_4(1 month)	EF_{15} Short term volatility of IF_3(1 month)
IF_{16} Long term change in IF_4(1 year)	EF_{16} Long term volatility of IF_3(1 year)
IF_{17} Short term change in IF_5(1 month)	EF_{17} Short term volatility of IF_4(1 month)
IF_{18} Long term change in IF_5(1 year)	EF_{18} Long term volatility of IF_4(1 year)
IF_{19} Long term change in IF_7(1 year)	EF_{19} Short term change of EF_1 (1 month)
IF_{20} Long term change in IF_8(1 year)	EF_{20} Long term change of EF_1 (1 year)
IF_{21} Long term change in IF_9(1 year)	EF_{21} Short term price momentum (1 month)
	EF_{22} Long term price momentum (1 year)
	EF_{23} 14-Day MACD
	EF_{24} 26-Day MACD
	EF_{25} 14-Day RSI
	EF_{26} 20-Day Lane's Stochastic Indicator

Figure 3.3: Metric Library

3.4.4 Parameters

Each simulation has 30 independent runs, and it is executed for every time period t. The parameter settings for both r-GNP and GNP are shown in Table 3.1. The terminal condition for Algorithm 1 is 500 generations, the population size is 301, where 100 individuals are generated by crossover, 200 are generated by mutation and 1 is the elite individual. The number of judgment and processing nodes are 48 and 24, respectively; whose delay times are 1 and 8 units. The maximum delay time is set at 8, which means that the transition through nodes ends when at least 1 processing node or 8 judgment nodes are used. Individuals with internal loops cannot execute a processing node, thus their uncompetitive fitness values will automatically exclude them from the population during the evolution of Algorithm 1. Other parameters exclusive in r-GNP include the number of solutions retrieved from the *Solution Pool*, the gaussian noise and the number of artificial environments which are set considering the reasonable balance of exploration and exploitation to converge to the optimal fitness values.

The parameters for the asset selection algorithm include the coefficient λ of user's aversion to volatility performance, which is set at 1 considering risk averse investors. The length of the training period is set at 2 years, and the testing period at 1 month, with policy for reinvesting and no frictional expenses. The initial capital K is $\$10,000$, in which the base currency for the fitness evaluation is the dollar.

Table 3.1: Parameters for r-GNP and GNP

Item	Description	Value
GE	the terminal condition	500 generations
$\|I\|$	the number of individuals	301
IC_{cross}	the number of individuals by crossover	100
IM_{mut}	the number of individuals by mutation	200
IE_{elite}	the number of elite individuals	1
P_m	the probability of mutation	0.01
P_c	the probability of crossover	0.1
JN	the number of judgement nodes	48 nodes
PN	the number of processing nodes	24 nodes
SN	the number of start nodes	1 node
d_j	the time delay of each judgement node	1 unit
d_p	the time delay of each processing node	8 units
d_c	the time delay of each branch	0 unit
$Max\ delay$	the maximum time delay	8 units
$\|I_{SP}\|$	the number of retrieved individuals from the *Solution Pool*	10
λ	user's aversion to volatility performance (r-GNP only)	1
η	the gaussian noise in artificial environments (r-GNP only)	0.025
$\|G\|$	the number of artificial environments (r-GNP only)	45
$\|TR\|$	the length for training	2 years
$\|TE\|$	the length for testing	1 month

3.4.5 Performance on the Training Phase

To show the influence of the probability of ϵ_1 and ϵ_2 on the fitness performance of r-GNP and GNP, Fig. 3.4 shows the average best fitness over the training phases of time span T. The parameters of ϵ_1 and ϵ_2 control the proportion of the selected *valuable* and *attractive* assets. A low/high value of ϵ_1 suggests that few/many *valuable* assets are selected; similarly, a low/high value of ϵ_2 suggests that few/many *attractive* assets are selected.

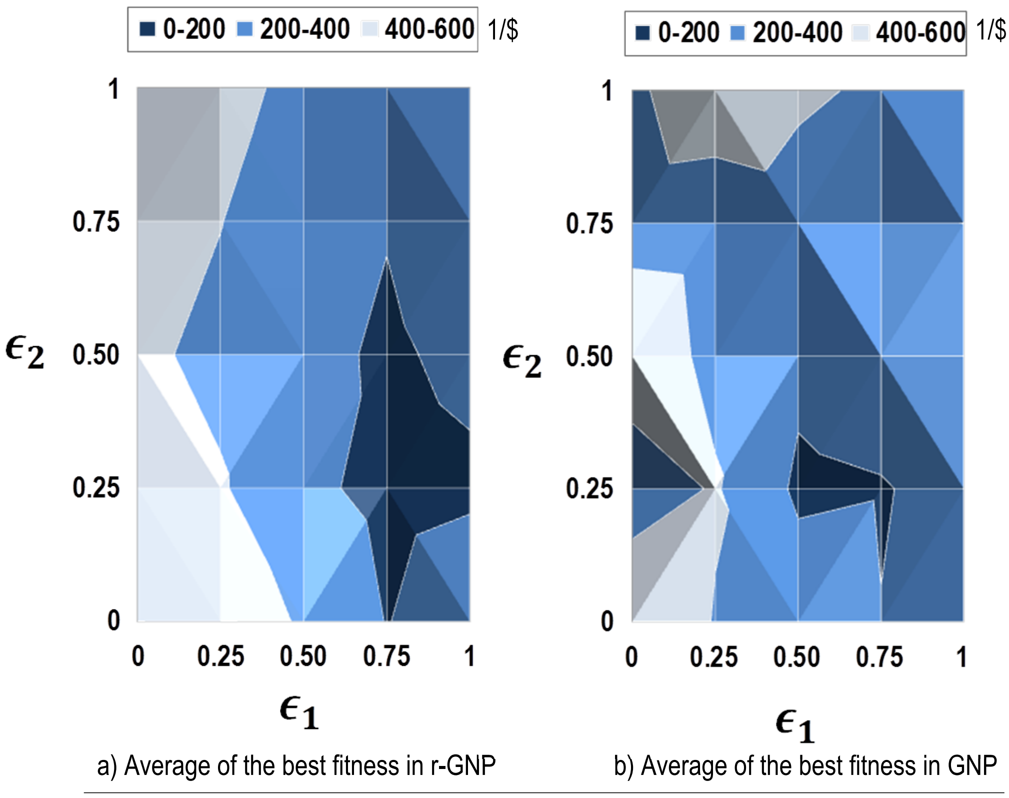
Fig. 3.4 suggests that areas with better fitness performance relate to higher values of ϵ_1 and lower values of ϵ_2, which implies that the prospective assets are more exposed to the *intrinsic* risk factors than exposed to the *extrinsic* risk factors. In finance, it has implications on developing risk pricing models that focus more on the underlying intrinsic value of the assets, rather than on the extrinsic growth or price momentum. In addition, Fig. 3.4 suggests that, on average, r-GNP has better chances to find the solutions with improved fitness performance. This is because r-GNP not only penalizes the volatility of the performance in multiple and divergent training environments, but also retrieves the partial solutions from the recent past; so the optimal solutions with small performance deviation are obtained in the last generation. Since the standard GNP neither penalizes the volatility of performance nor uses multiple environments for training, GNP tends to be trapped in local optima; which is a direct consequence of the overfitting to a single training environment. To compare both methods in further simulations, the probability of ϵ_1 and ϵ_2 are set at 0.75 and 0.25, respectively.

3.4.6 Performance on the Testing Phase

Performance in monthly holding periods

To compare the performance of the monthly testing periods, Fig. 3.5 - Fig. 3.7 summarizes the return performance and the number of assets held during the monthly testing periods($|TE| = 1$). Fig. 3.5 (a) shows that in the long term, the *wealth* behaves like the standard indexing strategies because of using the same investment universe. *Wealth* means initial capital plus cumulative monthly returns: the initial *wealth* is 1(initial capital is 100%) and subsequent losses/gains are added/deducted in every testing period. Fig. 3.5 (a) suggests that, on average, the generalization ability of r-GNP does yield *wealth* benefits over the standard GNP, the *value* and *growth* strategies. From the financial view point, since r-GNP penalizes volatility over the divergent environments, which are generated using a *nonparametric* scheme to avoid the form of the probability distribution of the time series, a type of stress-testing mechanism is realized, which is useful to avoid the extrapolation of historic performance into future horizons.

A year by year comparison is more readily displayed in Fig. 3.5 (b), which shows the annual accumulated returns in the proposed method and the benchmarks. Fig. 3.5 (b) exhibits gains for return performance of r-GNP over the standard GNP, *value*

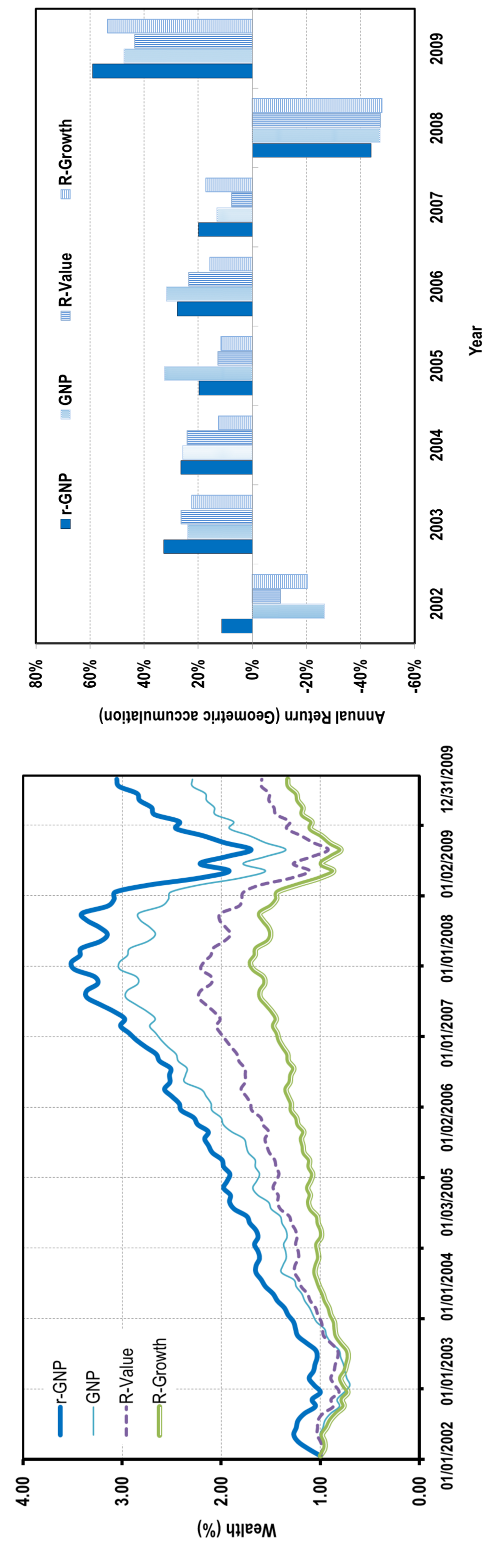


Note: Averages are over all the training periods of time span T, the larger the better.

Figure 3.4: Influence of ϵ_1 and ϵ_1 on F_o

and *growth* strategies. However there are periods such as 2005 and 2006 where r-GNP can underperform the standard GNP. A reason linked to this fact is that GNP uses its overfitting to historical data as an advantage during periods of time with high growth, such as 2005 and 2006. From a financial viewpoint, it implies that GNP is prone to select the assets that are likely to gain economic momentum in the short term because of relying on the recent past growth performance to identify future economic factors that drive the asset's expected return.

A month by month return comparison is shown in Fig. 3.6, which shows the monthly returns of r-GNP, standard GNP, *value* and *growth* strategies. In most of the cases, r-GNP shows increased gains or decreased losses, and the closer view shows that the average monthly return of r-GNP is 1.33% with a standard deviation of 5.64%; the average monthly return of GNP is 1.03% with standard deviation of 5.68%; the average



a) Accumulated wealth by r-GNP and the benchmarks

b) Yearly accumulated return rates by r-GNP and the benchmarks

Figure 3.5: Comparison of the return performance when $|TE| = 1$ month.

					r-GNP Monthly Returns								Annual
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	(Geom)
2002	12.30	9.08	3.58	-1.99	-1.42	-5.29	-9.46	3.14	-7.74	6.12	4.94	-3.88	11.40
2003	-1.55	-2.02	1.52	8.68	7.73	1.87	1.66	4.36	2.77	5.20	3.04	5.11	32.68
2004	2.84	3.30	-0.13	-2.06	0.91	2.21	-2.12	1.20	3.76	1.87	7.12	2.59	26.37
2005	-0.41	3.55	-1.81	-1.38	3.22	0.92	4.78	1.90	1.83	-1.66	4.97	1.88	19.77
2006	5.43	0.93	3.05	2.95	-2.12	0.20	-0.45	4.36	1.35	3.91	3.46	2.65	27.67
2007	2.89	-1.41	3.44	4.55	4.25	0.12	-3.29	1.02	6.06	1.00	-2.51	-0.35	19.93
2008	-5.65	-2.22	2.55	3.98	1.33	-7.50	-2.12	-0.38	-11.10	-18.73	-13.38	15.29	-43.93
2009	-11.62	-13.18	15.70	11.20	12.36	-1.30	10.32	1.08	4.45	0.98	6.34	0.73	59.03
Mean	0.53	-0.24	3.49	3.24	3.28	-1.10	-0.08	2.08	0.17	-0.16	1.75	3.00	
SD	7.22	6.45	5.28	4.96	4.84	3.50	5.88	1.71	6.17	7.92	6.78	5.62	

					GNP Monthly Returns								Annual
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	(Geom)
2002	-3.40	1.08	2.85	-3.00	-2.04	-5.89	-9.94	2.82	-8.32	-7.79	4.40	2.52	-26.61
2003	2.15	3.61	2.02	9.38	5.55	2.18	8.41	3.58	3.10	4.41	2.55	4.67	23.82
2004	2.06	9.74	-0.98	-2.40	0.39	1.48	-2.45	0.67	3.06	1.31	6.30	2.01	25.79
2005	7.07	3.04	-2.34	-1.71	2.62	0.14	4.06	1.30	1.31	7.07	4.32	1.22	32.44
2006	4.48	0.48	2.29	2.40	7.47	-0.24	-0.67	3.72	0.94	3.28	2.75	2.14	31.68
2007	2.24	-2.04	2.88	4.14	3.68	-0.43	-3.69	0.58	5.69	0.31	-2.70	-0.84	13.08
2008	-6.16	-2.83	2.10	3.51	0.57	-8.05	-2.78	-0.87	-11.41	-18.98	-13.70	14.49	-47.25
2009	-12.30	-13.57	15.22	10.41	11.66	-2.02	9.75	0.43	3.79	0.62	5.96	0.02	47.35
Mean	-0.48	-0.06	3.00	2.84	3.74	-1.60	0.34	1.53	-0.23	-1.22	1.23	3.28	
SD	6.36	6.71	5.29	5.14	4.41	3.59	6.63	1.66	6.18	8.38	6.65	4.82	

					R-Value Monthly Returns								Annual
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	(Geom)
2002	-2.25	2.20	3.51	-0.49	-0.39	-4.80	-10.89	2.72	-8.97	5.01	4.90	-3.60	-10.45
2003	-1.61	-2.82	0.28	8.15	8.08	1.08	1.30	3.58	2.23	5.10	2.56	5.53	26.29
2004	2.09	3.45	-0.29	-2.76	0.08	2.21	-1.14	1.19	3.45	1.61	6.76	2.00	24.11
2005	-0.33	3.54	-1.96	-2.08	2.08	0.99	3.57	1.74	1.17	-2.35	3.94	1.34	12.72
2006	5.50	0.79	2.52	3.03	-2.32	-0.07	0.31	3.33	1.07	3.19	2.87	2.80	23.55
2007	2.04	-1.91	3.02	3.98	3.40	-0.90	-4.78	-0.05	4.88	-1.13	-3.34	-1.00	7.75
2008	-5.37	-3.44	2.65	3.10	-0.59	-9.08	-1.60	-1.06	-9.40	-19.66	-14.06	14.75	-47.30
2009	-14.41	-15.50	16.17	11.84	11.99	-2.33	10.49	1.56	3.47	-0.32	5.31	-0.27	43.38
Mean	-1.79	-1.71	3.24	3.10	2.79	-1.61	-0.34	1.63	-0.26	-1.07	1.12	2.69	
SD	6.06	6.21	5.56	5.03	4.91	3.74	6.20	1.61	5.65	8.00	6.83	5.58	

					R-Growth Monthly Returns								Annual
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	(Geom)
2002	-4.29	-0.06	2.56	-4.68	-3.75	-6.81	-9.07	2.42	-7.48	6.01	3.91	-5.18	-20.40
2003	-2.61	-2.20	1.68	8.01	6.24	1.58	0.86	4.04	2.18	4.09	2.49	3.49	22.46
2004	2.47	2.12	-1.12	-2.43	0.68	1.08	-4.25	0.06	2.91	1.05	6.29	2.00	12.59
2005	-1.60	2.47	-2.83	-1.68	3.31	-0.32	4.94	0.86	1.38	-2.00	4.87	1.30	11.59
2006	4.18	0.05	2.39	1.81	-3.07	-0.65	-2.28	4.22	0.56	3.50	2.92	1.40	15.64
2007	2.57	-1.90	2.73	4.02	3.91	0.13	-2.82	0.94	6.18	1.98	-2.75	-0.75	17.19
2008	-6.98	-2.03	1.39	3.73	2.09	-6.94	-3.85	-0.73	-13.99	-18.67	-13.55	14.50	-47.84
2009	-9.53	-11.65	13.98	9.32	11.54	-1.33	8.84	-0.56	4.34	1.25	6.31	0.67	53.58
Mean	-1.97	-1.65	2.60	2.26	2.62	-1.66	-0.95	1.41	-0.49	-0.35	1.31	2.18	
SD	4.87	4.43	5.00	4.99	4.95	3.35	5.66	1.95	6.79	7.78	6.66	5.61	

Figure 3.6: Monthly return rates in r-GNP and the benchmarks when $|TE| = 1$ month.

monthly return of the *value* strategy is 0.65% with standard deviation of 5.66%; and the average monthly return of the *growth* strategy is 0.44% with standard deviation of 5.34%. It is well-known in Modern Portfolio Theory that a portfolio with a large number of assets has improved diversification benefits, and thus lower standard deviation of returns. One might attribute the relatively low standard deviation of r-GNP due to the large number of assets that it selects. However, r-GNP uses fewer assets than GNP, as shown in Fig. 3.7. On average, r-GNP has 290 assets with standard deviation of 38, and GNP has 381 assets with standard deviation of 43. In this context, r-GNP resembles an enhanced long-only indexing strategy for which the model based on composite risk factors is more flexible than the conventional fixed indexing strategies. The implications of these results in finance lie in the possibility to find miss-pricing opportunities in developing robust risk pricing models that keep the simplicity of the buy and hold indexes without loosening the long-only constraint.

r-GNP													
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean
2002	239	312	298	331	234	286	265	225	346	245	317	295	283
2003	254	227	314	291	263	226	325	261	263	316	230	296	272
2004	252	226	258	338	297	301	304	337	261	345	250	343	293
2005	281	290	328	300	242	337	242	287	276	294	339	250	289
2006	315	276	251	279	346	343	340	264	299	290	263	311	298
2007	337	325	340	317	236	284	254	317	225	244	321	348	296
2008	299	342	314	340	286	314	294	326	337	284	334	325	316
2009	271	249	227	341	259	269	298	248	343	292	285	242	277
Mean	281	281	291	317	270	295	290	283	294	289	292	301	
SD	34	44	41	24	38	38	34	40	45	34	41	39	

GNP													
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean
2002	305	371	363	446	423	422	423	440	357	439	433	374	400
2003	440	399	319	304	387	353	418	448	362	431	438	344	387
2004	303	406	362	434	362	345	431	349	395	388	410	316	375
2005	434	377	428	393	352	416	442	412	362	430	419	410	406
2006	374	303	357	383	318	335	379	426	397	368	449	305	366
2007	383	354	380	429	313	387	340	399	302	410	338	377	368
2008	415	397	426	377	314	386	337	340	423	353	402	372	379
2009	302	379	354	312	385	376	321	378	330	419	416	427	367
Mean	370	373	374	385	357	378	386	399	366	405	413	366	
SD	59	33	37	54	40	32	48	40	39	32	34	42	

Figure 3.7: Monthly number of assets held in r-GNP and GNP when $|TE| = 1$ month.

Performance in different holding periods

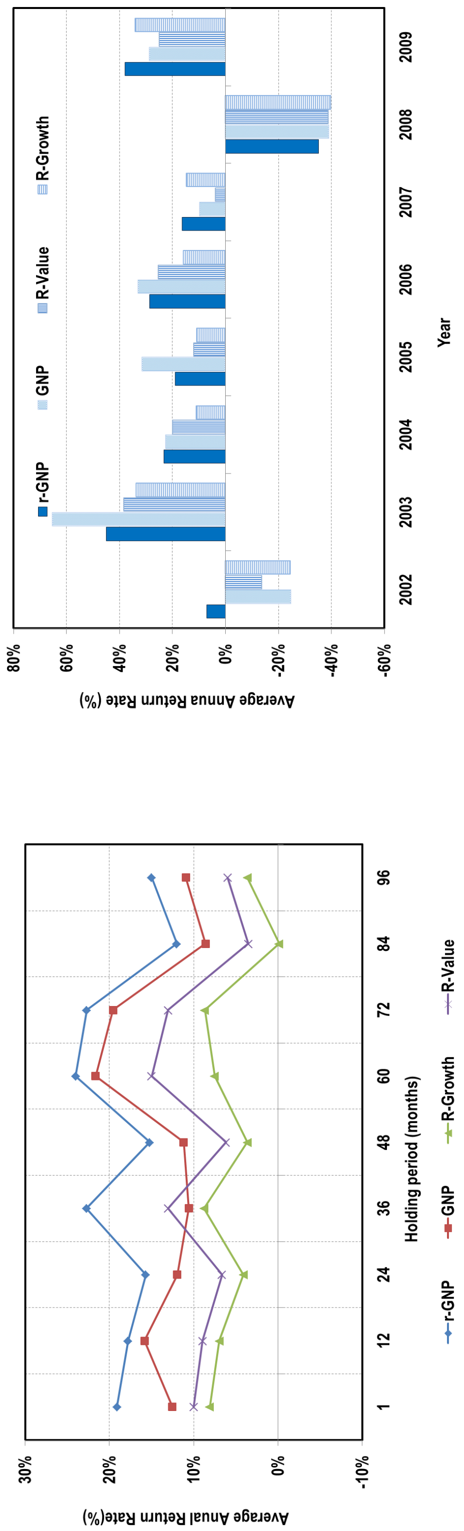
To compare the performance over longer periods of time, that is, relaxing the constraint of the buy and hold over 1 month period, Fig. 3.8 - Fig. 3.10 summarizes the annual return performance of r-GNP, standard GNP, *value* and *growth* strategies in different holding periods, i.e., $|TE|$ is extended to 12, 24, 48, 60, 72, 84 and 96 months.

Fig. 3.8 (a) summarizes the average of the annual returns of r-GNP, GNP and benchmarks through different holding periods, which shows that r-GNP yields better returns on average. A more closer comparison of the average annual returns in different holding cases is provided in Fig. 3.9 and Fig. 3.10, which shows the annual average returns of r-GNP, GNP and benchmarks. A case by case comparison shows that the generalization ability of r-GNP has benefits in increased return performance in most of the testing cases involving longer periods of time. It is because r-GNP avoids the overreliance on the historical performance and focuses more on stressed risk factors that can be useful to estimate and explain the underlying value of the business linked to the assets for a longer period of time.

Apart from return differences, Fig. 3.11 (a) shows the performance in terms of *risk*, and Fig. 3.11 (b) shows the performance in terms of *liquidity*, where *risk* and *liquidity* imply the volatility of annual return rates and the average market capitalization during the testing phase for the different holding periods. Not surprisingly, Fig. 3.11 suggests that the feature of r-GNP to penalize the volatility of the fitness performance over divergent training environments brings benefits not only to minimize the volatility of annual returns over the testing cases, but also to improve the average market capitalization, that is, the price at the end of the holding period multiplied by the number of shares held in each selected asset.

To compare which risk factors r-GNP uses most, Fig. 3.12 shows the average usage ratio, and Fig. 3.13 shows the average threshold values over the holding periods of 1, 12, 24, 48, 60, 72, 84 and 96 months. Fig. 3.13 shows that about 52% of the intrinsic factors, and 35% of the extrinsic factors, respectively, have their threshold values above 0.5. It suggests that r-GNP evaluates the exposure to the intrinsic factors more exhaustively.

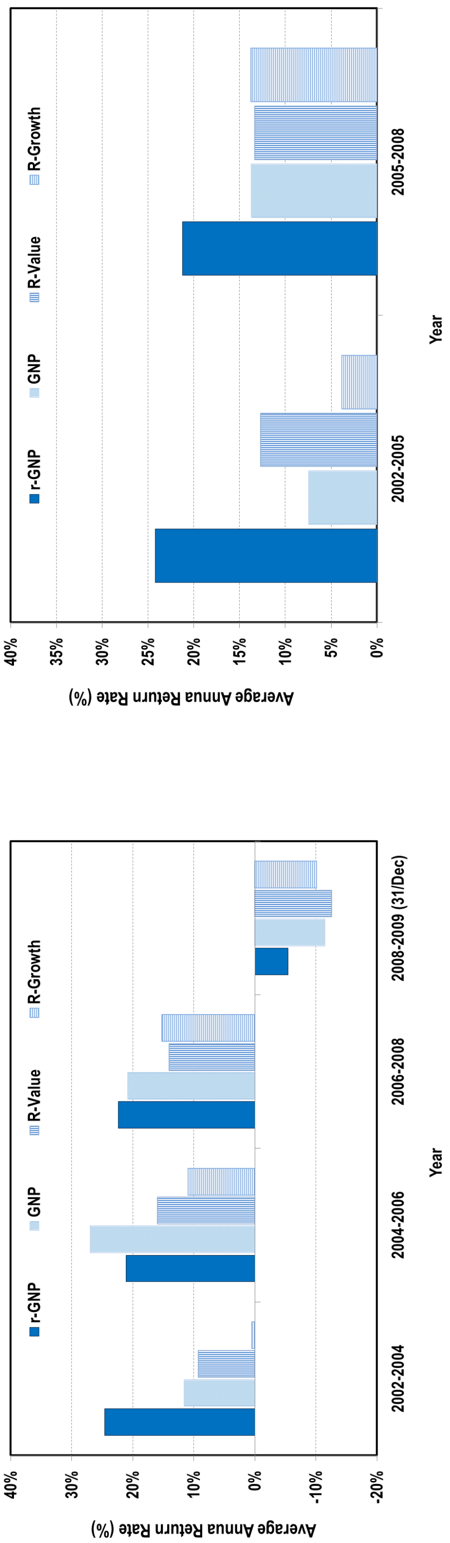
In addition, Fig. 3.12 shows that, on average, the first half of the intrinsic risk factors (IF_1 to IF_{11}) are used for 60.81% of the times, and the second half of the extrinsic risk factors (EF_{13} to EF_{27}) are used for 52.79% of the times. The metrics that measure the long term changes in both the intrinsic and extrinsic factors are used for 25.4% of the times on average; while the metrics that measure the short term changes are used for 6.72% of the times. It means that r-GNP focuses more on the common intrinsic factors relative to long term price changes. In finance, it implies the possibility of finding miss-pricing opportunities by focusing more on the long term view of price changes relative to fundamental factors; and if the pricing of an asset or forecasting of the excess returns(risk premiums) are involved, investors should look at the price change relative to intrinsic factors over more than one year.



a) Annual returns for different holding periods

b) Annual return when holding period TE = 12 months

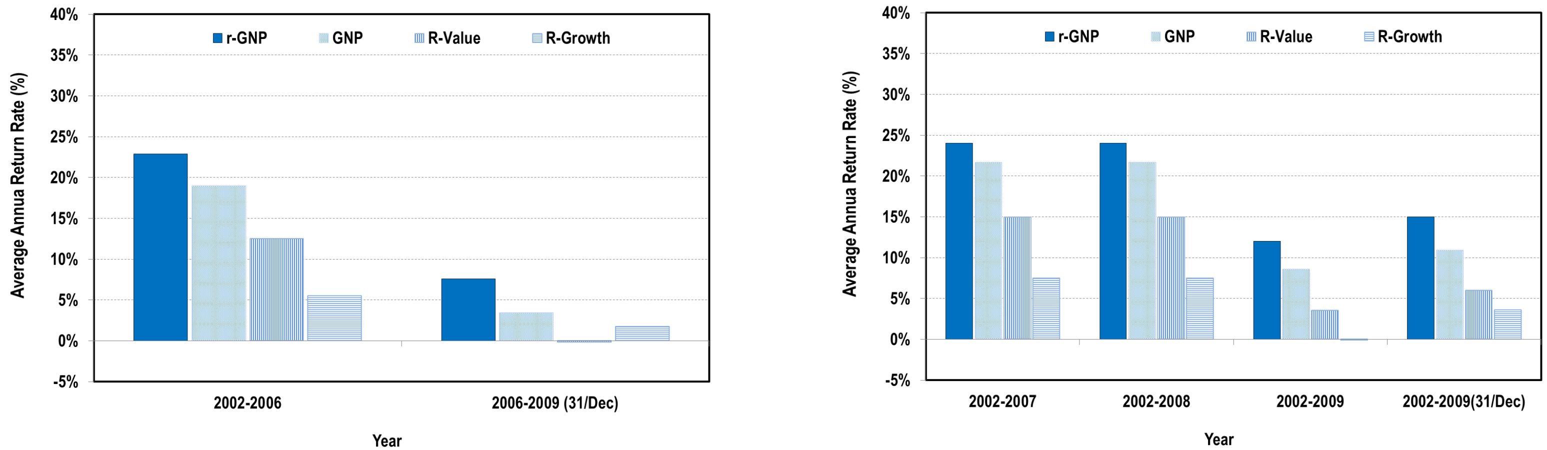
Figure 3.8: Comparison of the return performance in different holding periods and when $|TE| = 12$ months



a) Annual return when holding period TE = 24 months

b) Annual return when holding period TE = 36 months

Figure 3.9: Comparison of the return performance when $|TE| = 24, 36$ months



a) Annual return when holding period TE = 48 months

b) Annual return when holding period TE = 60, 72, 84, 96 months

Figure 3.10: Comparison of the return performance when $|TE| = 48, 60, 72, 84, 96$ months

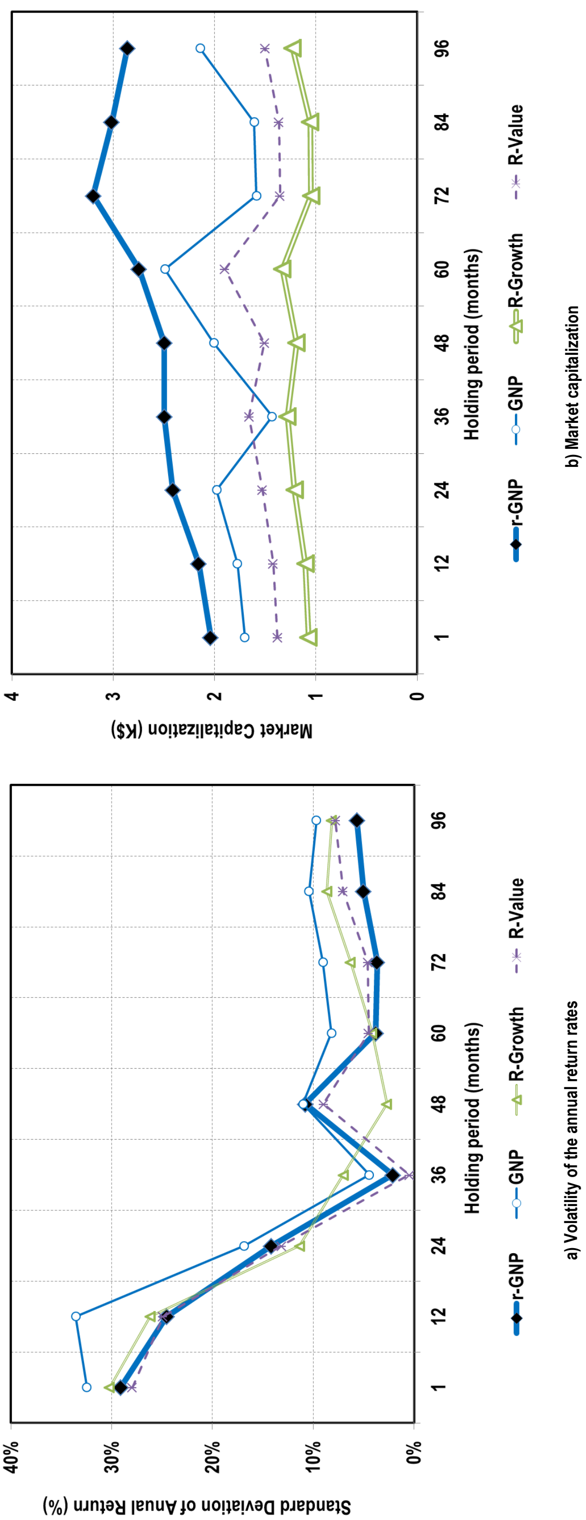


Figure 3.11: Comparison of the volatility and liquidity in different holding periods

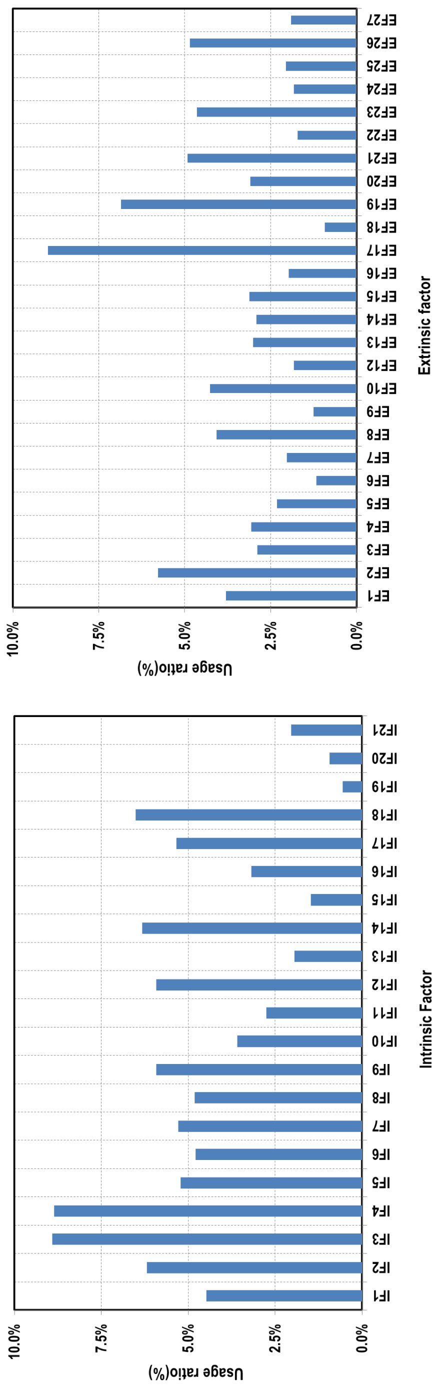


Figure 3.12: Average usage ratio of the *intrinsic* and the *extrinsic* factors over all holding periods

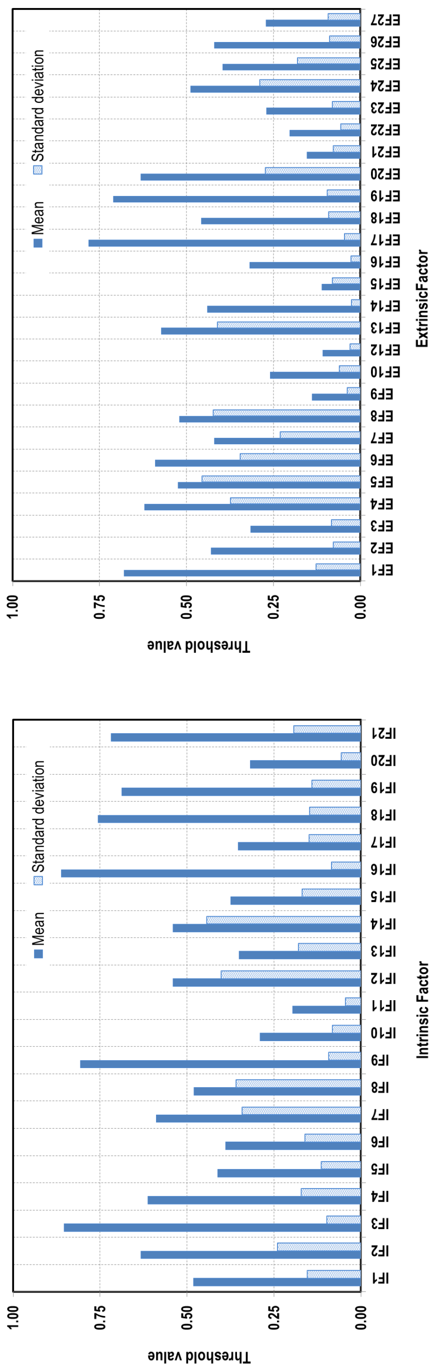


Figure 3.13: Average threshold values of the *intrinsic* and the *extrinsic* factors over all holding periods

3.5 Summary

In this chapter, a robust evolutionary strategy for the asset selection using r-GNP is proposed. One of the key points of r-GNP is that it is inspired by how *evolution* and *robustness* play important roles in the *individual* development. Biological organizations have better chances to survive if they acquire and accumulate different meaningful experiences when they are exposed to a relevant set of environments.

- Instead of using a single environment during evolution(*training* algorithm), r-GNP uses multiple and divergent *environments* which serve as crucible experiences to validate their internal structure of the individuals(*asset selection models*). This schema has a direct effect on avoiding the *overfitting* problem to historical data and improving the generalization ability in the individual's structure.

- In order to adapt to the changing market conditions, r-GNP uses accumulative strategies through time periods to track the dynamic optimum in the *Solution Pool*. Instead of re-starting the evolution when environmental changes occur, useful information from the recent past are used for further evaluation. This schema is useful to track the dynamic optimum when the current environment resembles the recent past history.

Simulations using assets in developed markets show that the generalization ability of r-GNP: (1) enlarges the search space for the optimal asset selection models, (2) outperforms the standard GNP, *value* and *growth* strategies in the long term, and (3) focuses more on the intrinsic risk factors relative to the changing extrinsic factor over the long term, which implies avoiding the overfitting to short term historical data. It brings practical implications in finance to capture *wealth* by focusing more on the long term patterns of prices relative to stressed fundamental factors, without loosening the simplicity of the standard long-only strategies.

Additional improvements on how to design robust and holistic risk management should be addressed. The architecture of r-GNP could be enhanced to consider different sources and levels of risks, such as those coming from the real economic fluctuations. Such system would imply a proactive way to handle risk, and guide the asset selection adaptively. The next chapter proposes a complementary approach to the concept of robustness, which is the case of the adaptive asset selection.

CHAPTER

4

Adaptive Asset Selection with Genetic Network Programming with Changing Structures

4.1 Aims of the Proposed Method

This chapter:

- Introduces a methodology to enhance the adaptability of the asset selection models based on *Genetic Network Programming with Changing Structures (GNP-cs)*.
  - The methodology implements a *control* and *operational* functions to realize the *adaptability* to changing environments; where the *control* function monitors the occurrence of environmental changes, in terms of *economic fluctuations*, and the *operational* function devises strategies, in terms of *asset selection models*, to deal with the detected changes.
  - Both the *control* and *operational* function are built upon jointly evolved functionally distributed systems, where the evaluation function considers

not only the *accuracy* of the prediction of the economic fluctuations(*control* function), but also the *return*, *risk* and *liquidity* of the asset selection strategies(*operational* function).

- Compares the proposed scheme with the standard GNP approach and benchmarks used in financial practice.
  - Simulations use the assets listed in the *Russell 3000 Index*.
  - Simulations are executed through *sliding time periods* between Jan 1995 and Dec 2010.
  - Benchmarks include widely known *indexing strategies* such as *Value*, and *Growth*.

4.2 Background

The main advantage of using nature-inspired methods, such as GNP, to build asset selection models is that they exploit the information extensively and exhaustively, which is supported by the recent developments in financial innovation in the last decade(13, 14, 15, 34, 35, 36, 37, 38, 39).

However, there has been a gap between the practice and real financial world in terms of not only using constant strategies, but also dealing with historical information as permanent factors; when they should be treated as temporal, as suggested by the social nature of markets.

In the previous chapter we demonstrated that models overfitted to historical data may prone to be risky when used in future horizons, specially in periods of financial distress. Considering the fluctuations in the economy is to take different strategies to guide the asset selection in accord with the changing dynamics in financial markets and the real economy. This approach is consistent with biologically adaptable organisms and offers a natural way to handle the risk while distributing the scarce economic resources in financial assets. Thus, an approach that changes and *adapts* to the situations of the markets(whether real or financial) is more likely to bring positive results.

This chapter proposes an evolutionary approach for the adaptive asset selection based on *Genetic Network Programming with changing structures(GNP-cs)*. The unique point is to introduce a guiding control mechanism to self-change the structure for asset selection depending on the fluctuations in the economy. The basic idea comes from biologically adaptable systems which incorporate control functions in their organization that monitor and guide the self-adaptation to the changing environments.

4.3 Genetic Network Programming with Changing Structures on Asset Selection

4.3.1 Basic Concept

GNP-cs is essentially an evolutionary computing algorithm with self adaptive properties that handles the modeling and optimization of a *decision making* system in complex and changing environments. GNP-cs uses *control* and *operational* functions in a collaborative manner to ensure the self-adaptability of the *decision making* system, namely *Control GNP* and *Operational GNP*, respectively, as shown in Fig. 4.1.

- The control function(*Control GNP*) monitors the changes in the environments and issues the relevant signals according to the perceived change. This system answers the question of *what is happening outside in the environment.*
- The operational function(*Operational GNP*) selects and executes the strategy for the *decision making* to deal with the perceived signals. This system answers the question of *what action to do given the current state and recent changes in the environment.*

Both functions are basically supported by GNP systems(11), whose structures are modeled and evolved according to evolutionary computing principles.

4.3.2 Main Features

The distinguishing features of *GNP-cs* from other evolutionary based asset selection schemes are the following points:

- *GNP-cs* incorporates an evolutionary based control mechanism, which is introduced into our previous work(40), and whose advantages involve enlarging the search space over the economic factors that determine the intrinsic value of the assets, and enhancing the adaptability to external changes in the economy.

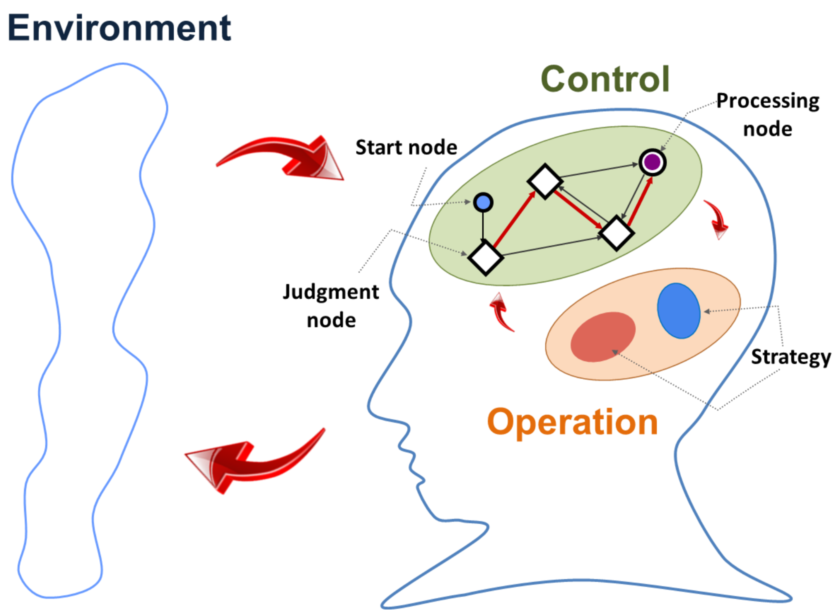


Figure 4.1: Basic idea of *GNP-cs* model

- *GNP-cs* covers flexible solutions over longer periods of time. Instead of using heuristics or statistical-based techniques to tackle the economic fluctuations, we deal with a compact network structure to guide the current asset selection strategies depending on the changes of the environments.

Although the conventional GNP system(11) also aims at handling complex *decision making* problems in dynamic environments, the proposed approach differs in the following points:

- GNP-cs incorporates an implicit guiding mechanism in the form of a control function, whose aim is to self modify the *decision making* structure depending on the changes of the environment. This feature enhances the adaptability and flexibility when building the optimum model based on a large scale GNP.

- GNP-cs is based on jointly evolved functionally distributed systems, implying an improved exploration ability and less internal loops in conventional GNP. The *genotype* in GNP-cs is defined as the concatenation of the *genotypes* of the *Control GNP* and the *Operational GNP* systems.

4.3.3 Basic Algorithm

Outline

An ideal asset selection model would be able not only to monitor the changing *economic cycles*, but also to execute the asset selection strategy considering the *margin of safety*. We use *GNP with changing structure*(GNP-cs) to separate the *control* systematically from the *operational* task. In our context, *control* means continuous monitoring of the state of the *economic cycle*, while *operation* means choosing and executing the adequate strategy for asset selection. Thus, the general scheme of the GNP-cs system for asset selection includes the following components as shown in Fig. 4.2:

- *Control GNP*, which is a component that monitors the changes in the current state of the economy. The inputs for this component include relevant indicators measuring the U.S. real economic activity as shown in Table 4.1. The output of this system is a signal s indicating the state of the economic cycle, i.e., $s = \{Ex, Co\}$, where $\{Ex\}$ refers to *Economic expansion* and $\{Co\}$ refers to *Economic contraction.*

- *Operational GNP*, which is a component that chooses and executes the strategy for asset selection. The inputs for this component include the issued signal s from the *Control GNP* and financial metrics such as *fundamentals* and *market oriented* asset information as shown in Table 2.1. The output of this system is a subset A_m representing the prospective assets stored in the *Asset Pool.*

Table 4.1: Economic variables used by *Control GNP*

Id.	Description	Source
	Leading indicators	
e_1	Dwellings started (number)	Bureau of the Census
e_2	Net new orders for durable goods (\$)	Bureau of the Census
e_3	Share prices: NYSE composite	Bureau of the Census
e_4	Consumer sentiment index	University of Michigan
e_5	Weekly manufacturing time (hours)	Bureau of Labor Statistics
e_6	Purchasing managers index(%)	Institute of Supply Management
e_7	Spread of interest rates(%)	Federal Reserve
	Coincident indicators	
e_8	Payroll employment	Bureau of Labor Statistics
e_9	Personal income	Bureau of Economic Analysis
e_{10}	Index of Industrial Production	Federal Reserve

It is important to note that not all variables in Table 4.1 and Table 2.1 are used. The way on how the *Control GNP* issues the signal s, the *Operational GNP* issues the asset set A_m, and how these systems are built up is explained in the next.

Control GNP

In order to issue signal s, the structure of the *Control GNP* incorporates *judgment* and *processing* functions, whose mechanisms are detailed in the following.

- **Judgement nodes**

The *judgment* nodes in the *Control GNP* asses the change in the state of the economy by using leading public U.S. economic indicators as shown in Table 4.1. Each judgment node p is associated with economic variable e_p in Table 4.1, and calculates the following:

$$E_p(t) = \frac{200(e_{pt} - e_{p(t-1)})}{(e_{pt} + e_{p(t-1)})}, \tag{4.1}$$

where,

$E_p(t)$: normalized symmetric variation of variable e_p in judgment node p during time period t.

e_{pt} : value of variable e_p during time period t.

In the case that e_p refers to a variable which is measured in percentage terms (such as e_6 in Table 4.1), the simple arithmetic difference $e_{pt} - e_{p(t-1)}$ is computed to calculate $E_p(t)$.

Each judgement node p has two branches; and the next branch depends on the if-then judgment on the *variable change* $E_p(t)$ against its threshold E_p^T. If $E_p(t)$ is greater or equal to E_p^T then *Branch I* is followed; otherwise *Branch II*. Then, the subsequent judgment node p' is determined, which forms the *node transition* NT as Fig. 4.2 shows. In order to avoid internal loops, we set 5 as the maximum number of judgment nodes in the transition NT.

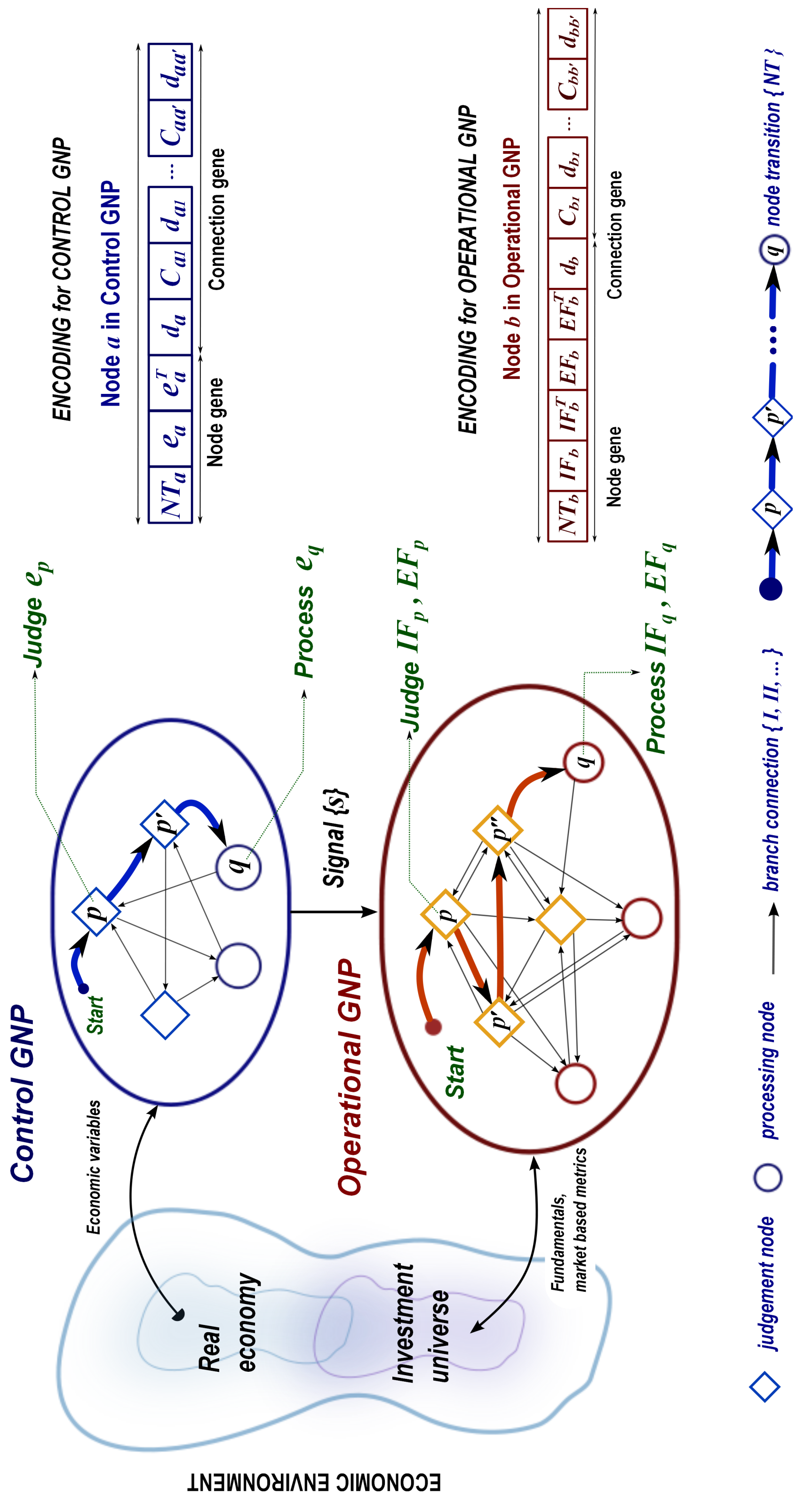


Figure 4.2: Basic structure of *GNP-cs* individual for asset selection

- **Processing nodes**

  The *processing* nodes issue the signal s related to the current state of the economy, i.e., $s = \{Ex, Co\}$. To cope with this task, every *processing* node q in the *Control GNP* uses the information on the *judgment* nodes in the *node transition* NT as follows:

  – First, compute the average normalized *change* in *node transition* NT.

  $$e_q = \frac{1}{|NT|} \sum_{p \in NT} e_p, \tag{4.2}$$

  where, NT is the set of suffixes of *judgement* nodes in node transition NT; and e_q is the average of the *change* used to issue the signal s by processing node q.

  – Second, issue the signal s related with the episode in the economic cycle:
    * *Expansion*. If $e_q \geq e_q^T$ then issue signal $\{Ex\}$.
    * *Contraction*. If $e_q < e_q^T$ then issue signal $\{Co\}$.

  – Third, the issued signal s is used by the *Operational GNP*.

- **Genotype**

  The genotype of node a in a *Control GNP individual* includes the node type NT_a, the *variable* e_a, its threshold e_a^T, time delay d_a and connections with other nodes $C_{aa'}$ as shown in Fig. 4.2. These elements are changed to other values at random by mutation(11).

Operational GNP

The structure of the *Operational GNP* is based on chapter 1.

Asset Selection algorithm

The optimization of the GNP-cs system uses the *training* - *testing* scheme to build and evaluate solutions with in-sample and out-of-sample data, respectively. This scheme consists of the following concepts:

(1) Training

Training implies using in-sample data to evolve the GNP-cs *individuals* until a terminal condition is satisfied. Concretely speaking, during the *training* phase of GNP-cs:

- The *Control GNP* issues signal s indicating the environmental change, which is used by the *Operational GNP* to evolve *sub-strategies* considering the issued signal s.
- The *Operational GNP* evolves two substructures: one for periods in *economic expansion* and other for periods in *economic contraction*, namely *O-GNP(Ex)* and *O-GNP(Co)*, respectively. Each substructure is evolved in different environmental dynamics, which is to say *O-GNP(Ex)* and *O-GNP(Co)* are evolved when the *Control GNP* issues the signal $\{Ex\}$ and $\{Co\}$, respectively.
- The quality of both the signal s and evolved *sub-strategies* determines the fitness value f_{cs} of the GNP-cs *individual*.
- Genetic operators are applied locally, i.e. crossover and mutation are executed only among the *Control GNP genotypes*, or among the *Operational GNP genotypes*.

The advantage of handling substructures in the *Operational GNP* lies in acquiring quite different strategies in order to evaluate the assets independently through *expansion* and *contraction* periods in the real economy, which in turn involves higher rule exploration to better estimate the *underlying value* of common assets in the financial market M.

The quality of *GNP-cs individuals* is measured by the fitness function f_{cs}, which is defined by:

$$f_{cs} = h_c f_c + h_o f_o, \tag{4.3}$$

$$f_c = \sum_{t=1}^{|TR|} IEC_t, \tag{4.4}$$

$$IEC_t = \begin{cases} 0 & \text{if } Y_t = F_t \\ 1 & \text{if } Y_t \neq F_t \end{cases} \tag{4.5}$$

where,

f_{cs}	: fitness of *GNP-cs individual*.
f_c	: fitness of *Control GNP*.
f_o	: fitness of *Operational GNP* defined by Eq. 2.5.
$h_c,\ h_o$	: coefficient for collaborative relationship.
$\lvert TR\rvert$	: number of training months.
IEC_t	: incorrectly estimated economic cycle during time period t.
Y_t	: economic cycle estimated by *Control GNP* during time period t.
F_t	: economic cycle estimated by OECD(41) during time period t.

Smaller fitness values are preferred. The values of h_c and h_o must be positive and greater than zero to consider an implicit collaborative relationship; which are set at 1.0 in this chapter.

(2) Testing

Testing implies evaluating the performance of the best GNP-cs *individuals* in the last generation with out-of-sample data. Concretely speaking, during the *testing*:

- The *Control GNP* issues a signal s, which is used by the *Operational GNP* to select its optimal evolved *sub-strategy*. In the case that the *Control GNP* issues a signal $\{Ex\}$, the *Operational GNP* uses the substructure evolved for *expansion* periods(*O-GNP(Ex)*), otherwise the strategy for *contraction* periods(*O-GNP(Co)*).

- The chosen substructure in the *Operational GNP* evaluates every asset i in M independently by using the *judgment* and *processing* functions explained in section 2.4.3.

4.4 Simulation Results

4.4.1 Problem

The same as chapter 2, indicated in Section 2.5.1.

4.4.2 Investment Universe

In our study, the investment universe M is the set of common assets in the Russell 3000 Index, representing approximately 98% of the investable U.S. security market.

4.4.3 Time Span

The *Time span* performed for simulations is between 1995 and 2010. Each experiment consists of two-year *training* and one-year *testing.* For example, the first experiment to build an optimal asset selection model based on GNP-cs consists of a *training* phase between Jan-1995 and Dec-1996; and a *testing* phase between Jan-1997 and Dec-1997. All subsequent experiments' *training* and *testing* are lagged by one year through the sliding time windows to consider recent arrived data and avoid overfitting issues.

4.4.4 Parameters

Simulation settings for both the standard GNP and GNP-cs are shown in Table 4.2.

Table 4.2: Parameter Settings

Description	**Value**
The number of generations	500
The number of individuals	200
Crossover size	79
Mutation size	120
Elite individuals	1
Mutation probability	0.01
Crossover probability	0.1
The number of judgement nodes	
- *Control GNP*	10
- *Operational GNP*	22
The number of processing nodes	
- *Control GNP*	5
- *Operational GNP*	10

4.4.5 Performance

(1) Training: Fig. 4.3 plots the average of the best fitness values over 30 independent runs in the *training* period, showing the convergence to the optimal asset selection models based on GNP-cs and GNP. We can see that both systems converge to the optimal solutions as the generations go on. Comparing the quality of solutions of GNP-cs and GNP in the last generation, we can observe that GNP-cs converge to more effective solutions than GNP; which is supported by the p-value of 0.00082 of a one side t-test, implying its significant difference. We believe that the improved performance of GNP-cs during the *training* period is mainly due to its higher exploration ability which results from combining functionally distributed systems with an implicit collaborative scheme, such as the *Control GNP* and *Operational GNP*, implying the use of evolutionary based building blocks to tackle the asset selection problem.

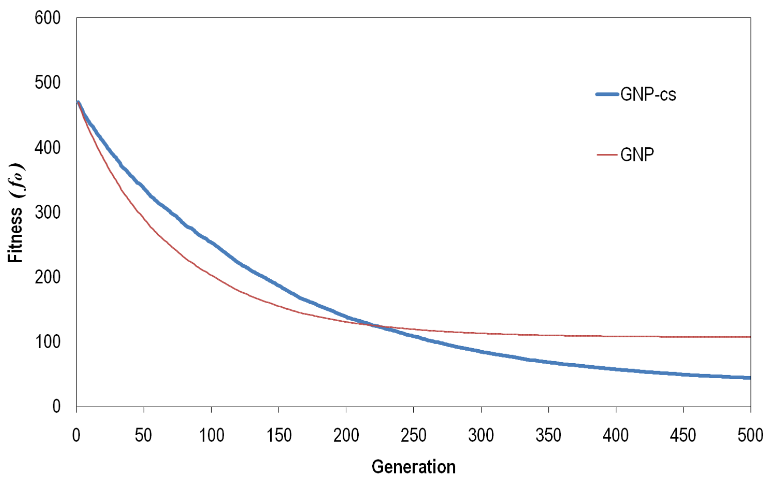


Figure 4.3: Average of best fitness values

(2) Testing: Fig. 4.4 shows the accumulated wealth over all the experiments' *testing* periods, i.e., 168 months between Jan-1997 and Dec-2010. The initial wealth is represented by the totality of the initial funds(100%) in Dec-1996(the initial date of the first testing period); and subsequent monthly returns in the form of gains or losses are accumulated. Fig. 4.5 shows the average yearly return rate and the standard deviation. The results and analysis are summarized in the following:

- The *Value* approach outperforms the *Growth* and the *Broad* indexes in terms of the long term wealth and short term yearly return rate, as Fig. 4.4 and Fig. 4.5 show, respectively. The reason is not because it is a riskier strategy, since its volatility is lower than the *Growth* and *Broad* styles as shown in Fig. 4.5. It is because the undervalued assets generally lie in the areas of high *intrinsic value* per *share*, e.g., high *Earnings per share* ratios. Since the *Value* approach searches in these areas, the undervalued assets are identified reasonably well.

- Both GNP and GNP-cs significantly outperform the *Value* based approach. Mainly, it is because of the following features. First, the optimal combination of factors, whether *intrinsic* or *extrinsic*, is decided by evolutionary computing approach, implying a more robust scheme to build the models for asset selection, in contrast to the fixed compounded index methodology of the *Value* approach. Second, whether an asset is selected or not is decided by the *judgement* and *processing* functions in the evolved network structure of GNP and GNP-cs, implying a more exhaustive risk pricing mechanism, in contrast to the ranking mechanism of the *Value* approach.

- GNP-cs significantly outperforms the selected benchmarks during long periods of time during the *testing* period, mainly due to its enhanced adaptability. Tracking changes in the economic cycles by the *Control GNP* and guiding systematically the strategies for asset selection in the *Operational GNP* implies an enhanced risk management ability, since the changing factors concerned with the real economy are reflected not only in the state of financial markets but also in the investors's return performance as shown in Fig. 4.5.

4.5 Summary

This chapter has introduced a novel approach for the asset selection based on *Genetic Network Programming with changing structures(GNP-cs).*

The distinguishing point from the conventional approaches is the inclusion of an evolutionary based control mechanism to monitor the changing external environments and guide the decision making for the asset selection. This feature brings not only the benefits in return performances as shown by the simulation studies, but also the following implications for building risk pricing models of the next generation:

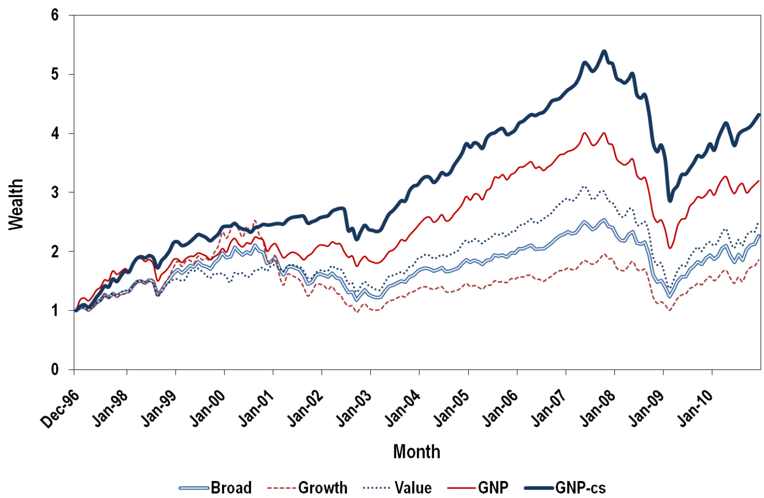


Figure 4.4: Comparison of GNP-cs and GNP in terms of wealth accumulation in the testing period

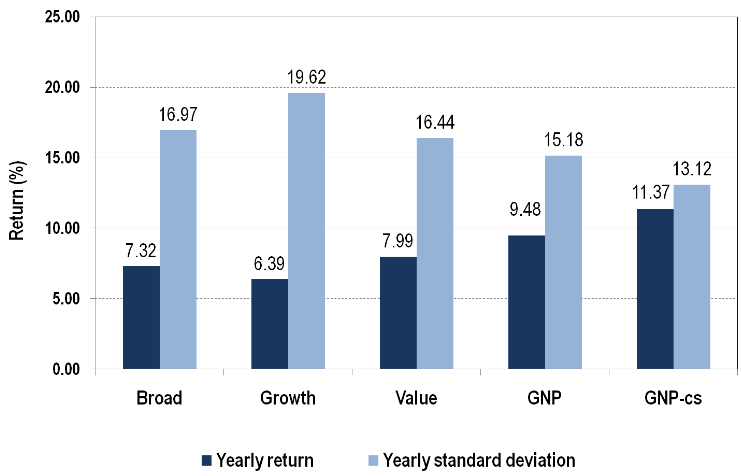


Figure 4.5: Comparison between GNP-cs and GNP in terms of yearly return and volatility rates

- Better adaptability and flexibility of the risk management strategies when financial markets turn uncertain and volatile.

- Exhaustiveness in asset pricing, and thus the avoidance of behavioral issues involved in speculative investing.

Further assessment is being addressed. Although the proposed approach relies on the market factors and fundamentals as sources of risk, other risk sources should be systematically evaluated.

Up to now, Chapters 1 to 3 have discussed how to model and build the risk pricing mechanisms in the form of asset selection models; which have the role of identifying a set of fairly prospective assets to invest with *equal importance*, which means allocating the investor's capital with equal proportion to the selected assets. The next chapter relaxes this assumption and introduces a methodology to build optimal asset allocation models.

CHAPTER

5

Asset Allocation with Genetic Relation Algorithm

5.1 Aims of the Proposed Method

This chapter:

- Introduces a methodology to build asset allocation models using *Genetic Relation Algorithm(GRA)*.
  - The methodology builds optimal asset portfolios by using evolutionary *undirected network structures*.
  - Each *node* in the network models financial *assets*, such as *stocks*, *bonds* and *currencies*, and each *relationship* measures the *systematic risk* between assets.
  - The evolution process uses *accumulative strategies* through *generations* and *time frames* to enhance the search for optimal portfolios by using the elite assets that performed well over the recent past.

- The *fitness function* is designed to asses *return*, *risk* and *liquidity* as main objective functions.

- Compares the proposed approach to relevant benchmarks for the asset allocation context.

  - Simulations use relevant *stocks*, *bonds* and *currencies* in America, Europe and Asia.
  - Simulations are executed through *sliding time periods* between Jan 2000 and Jan 2007.
  - Benchmarks include not only widely known asset allocation techniques, such as the *Black Litterman* model, *Genetic Algorithm-based CAPM* model, *Neural Network-based Markowitz* model, and *Stochastic CAPM* model, but also a global indexing model such as DJ Global 1800.

5.2 Background

Basically, the asset allocation consists of distributing a set of resources into several assets taking account of reasonable balance between the investors' needs and the return performances. The risk and asset allocation has been studied widely. The important contributions in finance include Mean Variance model by Markowitz(2); Indifference Theory by Modigliani(42); Capital Asset Pricing model(3); Options-pricing model(43); Arbitrage Pricing Theory(44) ; Binomial Option Pricing model(45); and a framework for the risk management including hedging(46). Moreover, an important outcome in the banking sector, Basel II Accord(31), encourages developing the robust risk and capital allocation models.

Recent advances in Neural Networks(56), Evolutionary Methods(48), Fuzzy Systems(49) and MonteCarlo Simulation(50) also deal with the issues such as tasks of the risk and asset allocation. Commonly studied cases include index forecasting, automated trading and asset pricing(48, 49).

Despite regular advances, the risk and capital allocation may also endure some limitations such as:

- Market risk. Risk is formally evaluated as variance, loss probability or Value at Risk methods(31). This schema leads to the underestimation issues of the financial risk at the systematic level(8, 9). The systematic risk is the risk component that affects a large number of assets due to the interdependencies in financial markets, and is claimed to be a common factor in financial collapses(8, 9, 10).

- Liquidity risk. Liquidity is often measured with the market capitalization rates. Liquidity becomes an issue when things turn sour in financial markets(10, 31). The investor needs to assure that money will be there when needed. Being locked up with illiquid assets makes significant distinction between attaining acceptable return rates and losing much more at the end. Much of the advances have been done in order to improve the liquidity features for financial markets. However, a few has been done in the context of risk and asset allocation.

In order to deal with the above issues, this chapter proposes an asset allocation model by using Genetic Relation Algorithm(GRA),

5.3 Previous definitions

5.3.1 Portfolio

A portfolio P is defined as a set of n assets, where x_i represents the proportion of the capital invested in asset i.

5.3.2 Return components

The economic dimension of the assets and portfolios are defined by their return components. This subsection describes the return components in detail.

Return at asset level

Stocks reveal the degree of ownership(share) for a company and are often key factors in the daily operations of the stock market. The return rate provided by stock i is defined by the opening and closing prices and dividends paid using the following equation.

$$R_{it}^S = ({P_{it}}^f - P_{it}^o + div_{it})/{P_{it}}^o, \tag{5.1}$$

where, R_{it}^S is the return rate of stock i during the time period t; ${P_{it}}^o$ is the opening price of stock i at the beginning of time period t; ${P_{it}}^f$ is the closing price of stock i at the end of time period t; and div_{it} is the dividend of stock i during time period t.

Bonds show the ownership of a payment contract. The return rate provided by bond i is defined by the coupons, which is the interest of the initial invested money, and face value, or the amount of redemption at maturity of bond using the following equation.

$$R_{it}^{B} = (FV_{it}^{f} - I_{it}^{o} + C_{it})/I_{it}^{o}, \tag{5.2}$$

where, R_{it}^{B} is the return rate of bond i during time period t; I_{it}^{o} is the initial money invested in bond i at the beginning of time period t; FV_{it}^{f} is the face value given by bond i at the end of time period t; and C_{it} is the coupon received by bond i during time period t.

Currencies show the ownership of the fast interchangeability and ensure the real purchasing power. The return rate is measured using the spot value, which is the current exchange rate, and the forward points, which is the value added or deduced from the spot exchange rate by the following equation.

$$R_{it}^{C} = ({S_{it}}^{f} - {S_{it}}^{o} - f_{it})/{S_{it}}^{o}, \tag{5.3}$$

where, R_{it}^{C} is the return rate of currency i during time period t; ${S_{it}}^{o}$ is the opening spot rate of currency i at the beginning of time period t; ${S_{it}}^{f}$ is the ending spot rate of currency i at the end of time period t; and f_{it} is the forward points of currency i during time period t.

Market m is a platform that enables the asset trading and risk transfer. In this chapter, we consider stock, bond and foreign exchange markets. The return of market m is calculated by an index value, which is a passive track measuring return changes.

$$R_{mt} = (Index_{t} - Index_{t-1} + Div_{d})/Index_{t-1}, \tag{5.4}$$

where, R_{mt} is the return of the market index during time period t; $Index_{t}$ is the level of the market index at the end of time period t; $Index_{t-1}$ is the level of the market index at the end of the of time period $t-1$; Div_{d} is the dividend paid by the index during the time period t.

Return at portfolio level

The return component of portfolio P is expressed by the following:

$$R_{Pt} = \sum_{i=1}^{n} x_i . R_{it}, \tag{5.5}$$

where, R_{Pt} is the return performance of P during the period of time t; n is the number of assets in portfolio P; x_i is the proportion invested in asset i; and R_{it} is the return of asset i during the period of time t, in Eqs. (5.1), (5.2) and (5.3).

5.3.3 Systematic risk components

The beta concept is a relevant construct in Modern Portfolio Theory and captures the systematic risk component of the assets and portfolios. Higher asset beta values imply higher levels of volatility and viceversa. For example, the asset with the beta being 0.5 has half of the systematic risk of the market; and the asset with the beta being 2 has twice of the systematic risk of the market.

The beta also exhibits the degree of independence and sensitivity of the asset prices. Positive betas indicate that the asset follows the market behavior. Very positive or very negative values indicate the strong price sensitivity in contrast to the market's behavior. The beta equal to 0 shows the independence from the market.

Betas at asset level

The beta of asset i measures the risk added to portfolio P as the correlation of the asset values with a reference market. The beta is widely used in the asset valuation. In our approach, the beta sheds light on unfamiliar properties as an investment diversifier. The beta coefficient of asset i is defined by:

$$\beta_{it} = \frac{Cov_t(R_{it}, R_{mt})}{Var_t(R_{mt})}, \tag{5.6}$$

where, β_{it} is the beta coefficient of asset i during time period t; $R_{it}$is the return rate of asset i during time period t; R_{mt} is the return rate of market m during time period t; $Cov_t()$ is the covariance during time period t; and $Var_t()$ is the variance during time period t.

In order to capture the risk added to portfolio P by asset i and asset j we use the following measure, which shows the independence and sensitivity by a pair of assets in contrast to Eq. (5.6) :

$$\beta_{(ij)_t} = x_i \beta_{it} + x_j \beta_{jt}, \tag{5.7}$$

where, $\beta_{(ij)_t}$ is the relational *beta* between asset i and asset j during time period t; x_i is the proportion invested in asset i; and β_{it} is the individual *beta* of asset i during the period of time t.

Betas at portfolio level

In order to measure the average correlation between the portfolio performance and movements in the referential markets, the following coefficient is used:

$$\beta_{Pt} = \frac{1}{|A(P)|} \sum_{i \in A(P)} \frac{1}{|A(P_i)|} \sum_{j \in A(P_i)} \beta_{(ij)_t}, \tag{5.8}$$

where, β_{Pt} is the beta coefficient of portfolio P during time period t; $A(P)$ is the set of suffixes of assets in P; $A(P_i)$ is the set of suffixes of assets whose link is defined from asset i in P.

5.3.4 Liquidity components

This subsection describes the liquidity components for the portfolio construction.

Liquidity at asset level

An ideal liquid asset is the one which is traded frequently in large quantities and with little price impact(9). The trading volume ($\overline{V}_{it}$) captures the price impact and trading size factors; the turn over ratio ($\overline{T}_{it}$) measures the price impact and trading time features; the bid ask ratio ($\overline{B}_{it}$) captures the price impact and trading time factors; and the market capitalization ($\overline{M}_{it}$) captures the trading size factors. The following metric is used to measure the liquidity level of asset i:

$$L_{it} = w_V \overline{V}_{it} + w_T \overline{T}_{it} + w_B \overline{B}_{it} + w_M \overline{M}_{it}, \tag{5.9}$$

where,
L_{it}: liquidity index of asset i during time period t.
$\overline{V}_{it}$: trading volume of asset i during time period t.
$\overline{T}_{it}$: turn over ratio of asset i during time period t.
$\overline{B}_{it}$: bid ask ratio of asset i during time period t.
$\overline{M}_{it}$ market capitalization of asset i during time period t.
w_V: importance weight of trading volume.
w_T: importance weight of turn over ratio.
w_B: importance weight of bid ask ratio.

w_M: importance weight of market capitalization.

The weighing values (w_V, w_T, w_B, w_M) consider the investor's interest for trading volume, turn over ratio, bid ask ratio and market capitalization, respectively. In this chapter, equal importance is given to all measures.

Variables such as $\overline{V}_{it}$, $\overline{T}_{it}$, $\overline{B}_{it}$ and $\overline{M}_{it}$ range between 0 and 1, and are normalized with the following equation:

$$\bar{l}_{it} = \frac{l_{it} - min_k l_{kt}}{max_k l_{kt} - min_k l_{kt}}, \tag{5.10}$$

where, $\bar{l}_{it}$ is any of the normalized liquidity metrics defined by Eq. (5.9). Minimum and maximum values are computed considering a set of k assets during time period t.

Liquidity at portfolio level

Having calculated liquidity measures at asset levels, the following measures the liquidity level of portfolio P:

$$L_{Pt} = \sum_{i=1}^{n} x_i . L_{it}, \tag{5.11}$$

where, L_{Pt} is the liquidity level of portfolio P during time period t; n is the number of assets in portfolio P; x_i is the proportion invested in asset i; and L_{it} is the liquidity level of asset i during time period t.

5.4 Genetic Relation Algorithm on Asset Allocation

5.4.1 Basic Concept

GRA is a graph based evolutionary computing algorithm derived from Genetic Network Programming(11), which has been proposed as a rule pruning mechanism in datamining applications(51).

In this chapter, GRA is used as a tool to model and optimize asset portfolios considering the return, risk relationship and liquidity principles among a set of asset classes, such as stocks, currencies and bonds(Fig. 5.1). Unlike strings for solution representation in GA and trees in GP, GRA has the ability to express complex events compactly in directed/undirected node graph structures.

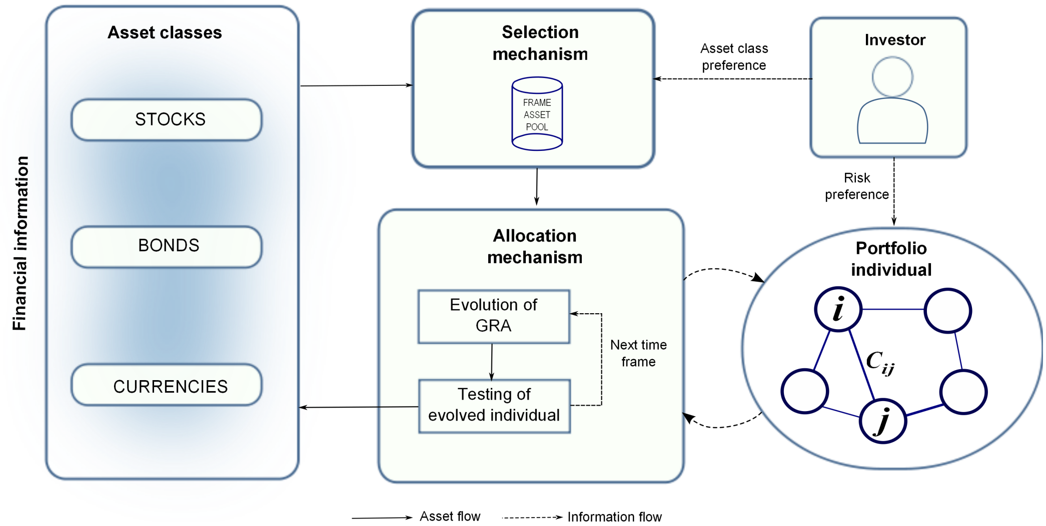


Figure 5.1: Outline of vs-GRA for portfolio diversification

5.4.2 Main Features

The main features of the proposed approach are described as follows:

- The proposed framework finds the competitive asset portfolios not only under risk and return profiles, but also integrating the liquidity aspects. This feature enables the strong basis for mitigating the liquidity risk in the methods such as Mean Variance model by Markowitz and Capital Asset Pricing Model(CAPM).

- The proposed framework evolves the complex and ill portfolio structures toward compact and effective ones through a legible and evolvable graph structure, evading black box issues and exhaustive mathematical properties needed for encoding in other natural inspired algorithms.

- A measure for the systematic risk and relational beta portfolio, enhances the portfolio of risk assessment. The beta is a relevant construct in Modern Portfolio Theory. In the proposed approach, beta portfolio is easily calculated by the GRA structure.

- The GRA-based risk and asset allocation finds the effective asset portfolios and store them in a functional pool. Whereas, in conventional approaches the final result of the optimization is restricted to the optimum portfolio only. This feature makes the flexible portfolio optimization in markets with high volatility.

- Unlike Markowitz or CAPM based models, where a single and representative market is identified a priori, the proposed approach performs an unbiased diversification of investments considering the multiple financial markets in reference.

5.4.3 Structure of GRA

GRA models a *portfolio individual* as an intertwined graph(Fig. 5.2). The *gene information* contains the following information

- The identification ID_i of asset i,
- The function F_i that represents the type of asset i such as stocks, bonds or currencies,
- The node size x_i that represents the proportion of the capital invested to asset i,
- The connections $\beta_{(i1)_t}$, $\beta_{(i2)_t}$, $\beta_{(in)_t}$ of asset i to asset 1, 2,...,n in time period t that represent the relational beta coefficient among assets(19).
- C_{i1}, C_{i2}, ..., C_{in} define the suffixes of assets whose link is defined from asset i.
- The portfolio individual has n assets. In this paper we set n at 30 assets(52).

5.4.4 Fitness function of GRA

The fitness function of the *GRA individual* is defined by the following equation:

$$F_f = \frac{(\beta_f - \beta)^2}{(R_f - RF_f)(L_f)}, \tag{5.12}$$

where,

F_f : fitness during the f_{th} training time frame.
β_f : beta coefficient during the f_{th} training time frame.
β : user defined *beta* value.
R_f : return performance during the f_{th} training time frame.
RF_f : average risk free rate defined by 3-month U.S. Treasury Bill during the f_{th} training time frame.
L_f : liquidity level, i.e., an average of trading volume,

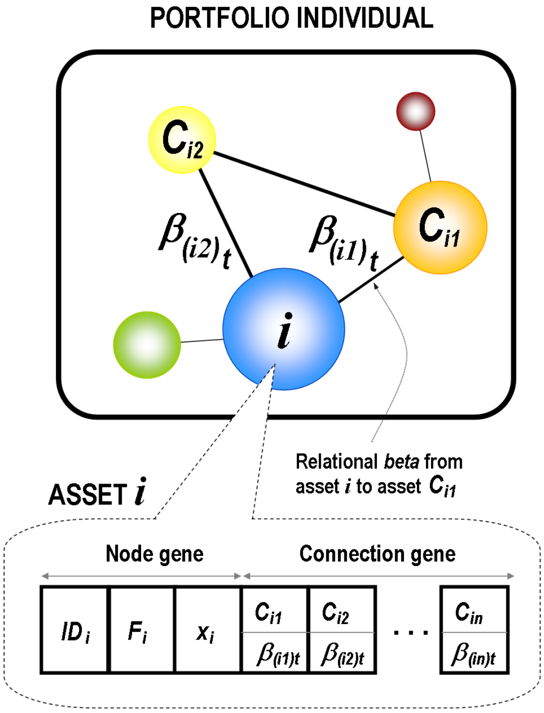


Figure 5.2: Basic structure of GRA portfolio individual

turn over ratio, bid ask ratio and
market capitalization during the f_{th}
training time frame.

Attitudes toward risk can be controlled by parameter β, in which higher values imply strong risk seeking attitudes and low values imply strong risk aversion attitudes.

The fitness function in GRA optimizes not only the excess return for investors; but also liquidity and market risk exposure at a portfolio level(19).

The main reasons for having quite different fitness functions for the GRA-based asset allocation and the GNP based asset selection relate to the context and purpose. Whereas the fitness of a GRA individual depends on the performance of its allocation strategy, which is mainly determined by the type and proportion of each prospective asset of the Asset Pool in the portfolio, the fitness of a GNP individual depends on the performance of the set of assets that is able to select, which is mainly determined by the risk pricing model. Separating the selection task from the allocation task provides a mechanism not only to enhance the transparency of the investment cycle, but also to

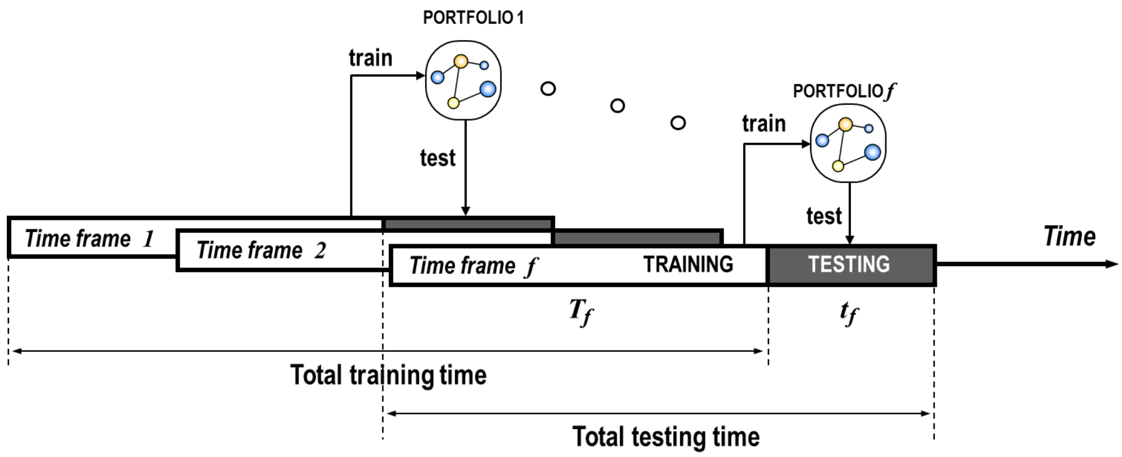


Figure 5.3: Time frames for GRA training and testing

evaluate the financial risk factors systematically.

5.4.5 Genetic Operators in GRA

Selection

We preserve the elite portfolio individuals through the accumulation mechanisms(53); and generate the new portfolio offspring through crossover and mutation in GRA.

Fig. 5.4 shows the role of the *Initial Asset Pool*, which is to store the initial asset candidates for the portfolio optimization. The role of the *Generational Asset Pool (GAP)* is to preserve the elite portfolio individuals during generations with training data; while the *Frame Asset Pool* stores the elite assets after testing with unseen data. As shown in Fig. 5.4, the best individual is obtained by evolution and accumulation mechanisms. In every generation, $a\%$ of the elite portfolio individuals (P_1, P_2, ... , P_I) are stored in *GAP*. This process is repeated until the last $GE - th$ generation. The elite individual from the evolution is validated against newly arrived unseen testing data. $a\%$ of elite assets with good performance are updated in FAP after the testing procedure.

Crossover

The crossover mechanism facilitates the asset rebalancing considering the liquidity features in the portfolio individuals (Fig. 5.5):

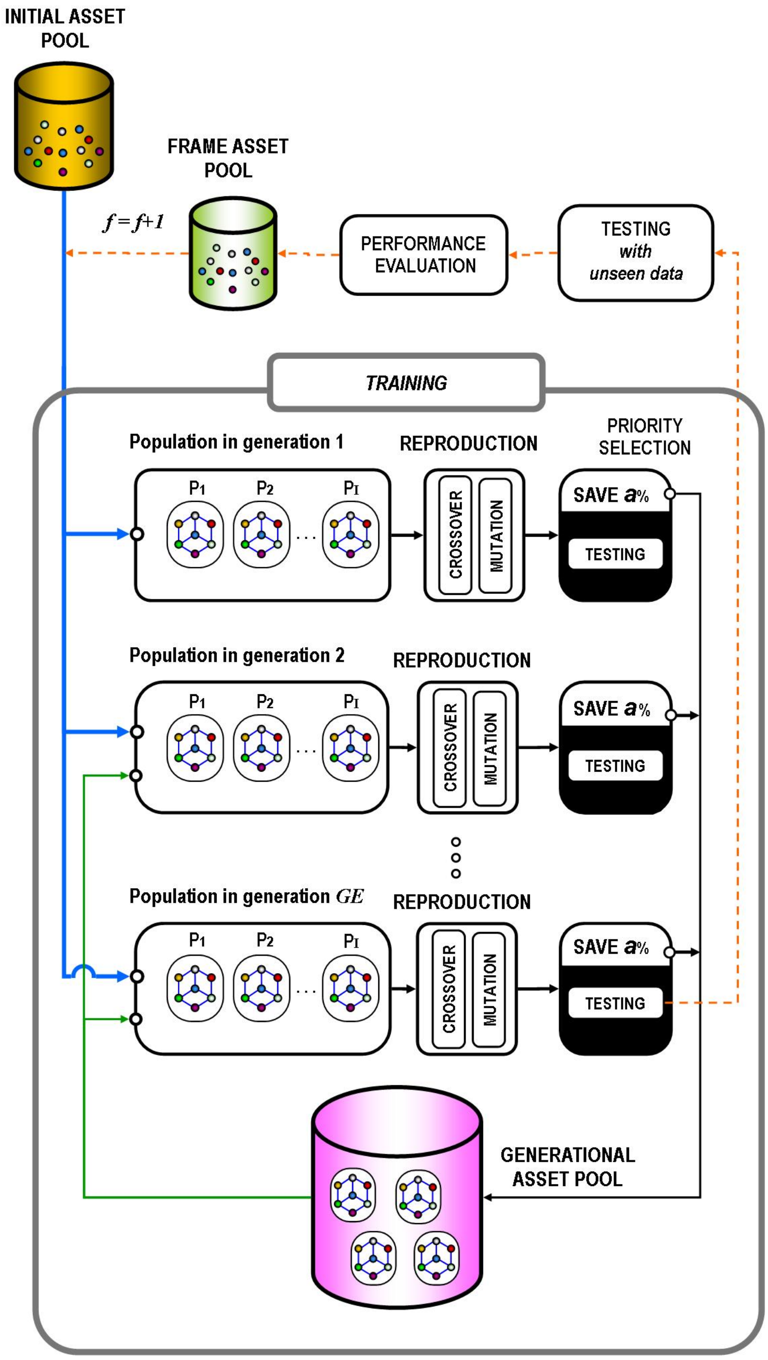


Figure 5.4: Accumulation mechanism through generations

- Using tournament selection, select two parent portfolio individuals (${G_1}^P$ and ${G_2}^P$).
- Crossover nodes are selected with the probability of P_c.
- Gene information is exchanged among the corresponding crossover nodes in parent individuals.
- New offspring individuals, ${G_1}^s$ and ${G_2}^s$, become available for the next generations.

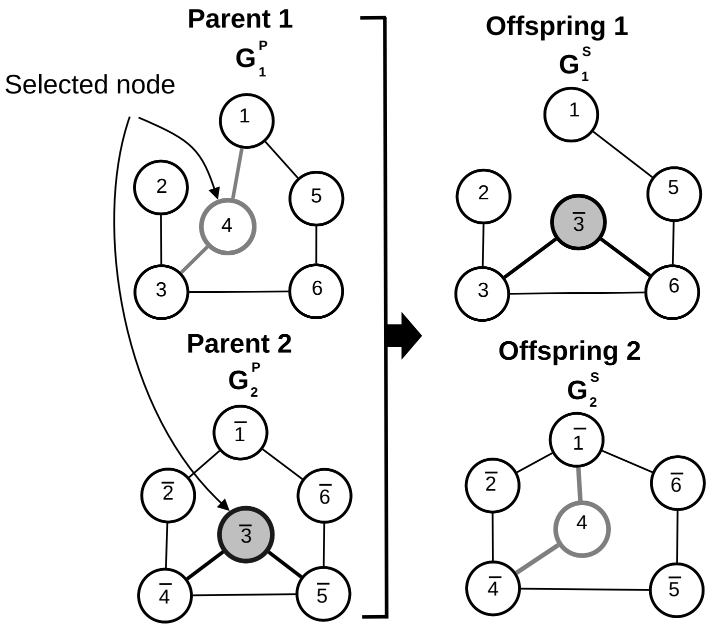


Figure 5.5: Crossover operation in GRA

The following probability is used to generate the offspring according to the liquidity index in each portfolio individual:

$$P_c = 1 - \frac{1}{|A(G)|} \sum_{i \in A(G)} L_{it}, \tag{5.13}$$

where, P_c is the probability of crossover of node i; G denotes GRA individual; $A(G)$ is the set of suffixes of assets in G; and L_{it} is the liquidity index of asset i during time period t.

Since crossover operator tends to break the building blocks of the evolution process, the role of Eq. (5.13) is to preserve the liquidity features in elite portfolio individuals during the crossover operation.

Mutation

Mutation makes the portfolio rebalancing possible. The offspring is generated by changing both the connections and nodes of assets as follows (see Fig. 5.6):

- Using tournament selection, select the parent portfolio individual (G^P).
- Mutation Operation.

  Node connection. Select connections with probability P_m and change them.

  Node function. Select nodes with probability P_m, and changed them.
- New individual G^S becomes available for the next generation.

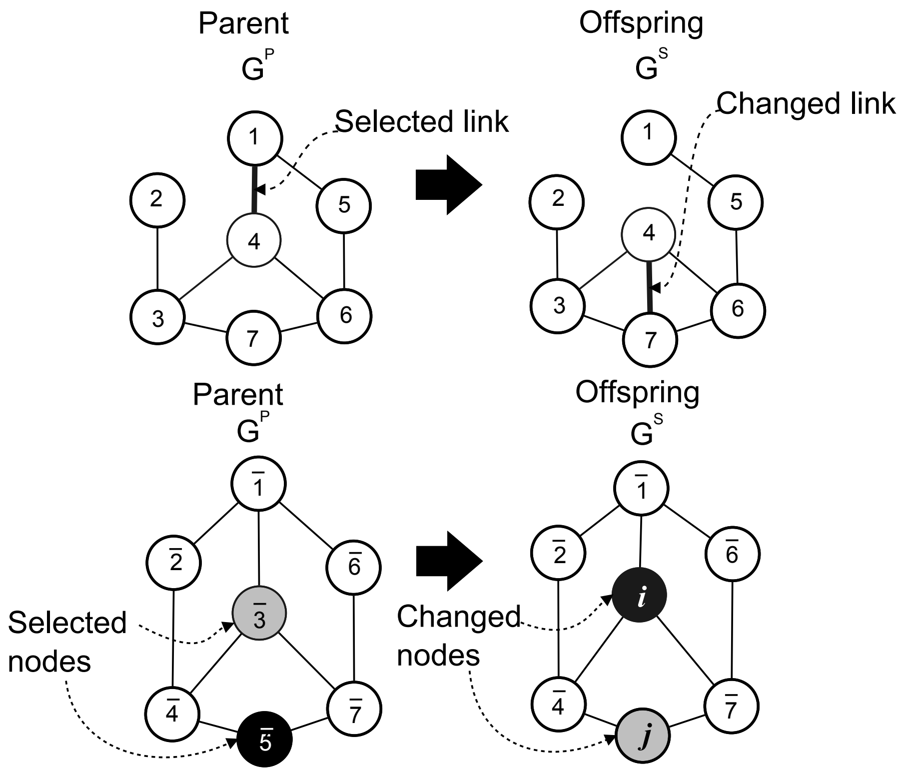


Figure 5.6: Mutation operation in GRA

5.4.6 Time frames mechanism

Time frames (T_f, t_f) are defined for the training and testing in time frame f as shown in Fig. 5.3. The period of time T_f is for the training phase, or portfolio optimization, whereas t_f is for the testing phase, or portfolio validation.

Asset accumulation through time frames is performed consecutively. For example, if asset b performs well in testing phase t_f, then it implies that asset b will be preserved as a candidate in the next training phase T_{f+1}; i.e., asset b will participate in the evolution process of GRA during T_{f+1}. The general schema for the asset accumulation through time frames is depicted in Fig. 5.7.

In this chapter, time frames have the role of including newly arrived non-stationary information into the optimization process of GRA; from another point of view, time frames check the stability and the value of the optimal portfolios, and avoid the over-fitting problem for the portfolio optimization in relatively short time period.

5.4.7 Basic algorithm

Selection phase

This chapter focuses on the *asset allocation* problem in which an *Initial Asset Pool* containing a set of prospective assets is needed. It has been demonstrated from simulations that the methods for *asset selection* proposed in previous chapters identifies the *Initial Asset Pool* effectively. However, in order to avoid influence/bias on the performance of the *asset allocation*, the proposed algorithm and benchmarks(in simulation) use a more simple/conventional *asset selection* mechanism, that is, based on *ranking* with a priority metric.

We use an iterative procedure to pick up a fairly small competitive assets from the financial markets, where, in each asset class A_c and market M_k ($A_c \in M_k$), the following priority metric P_{iS} of asset i is evaluated. Assets with the high P_{iS} are added into Initial Asset Pool (IAP). Then, without asset i, priority values of remaining assets in M_k are evaluated. The procedure is repeated until we reach the size of IAP.

P_{iS} is defined as follows:

$$P_{iS} = \frac{1 + L_{iS}}{(1 + \beta_{iS})(\sigma_{\beta_{iS}})}, \tag{5.14}$$

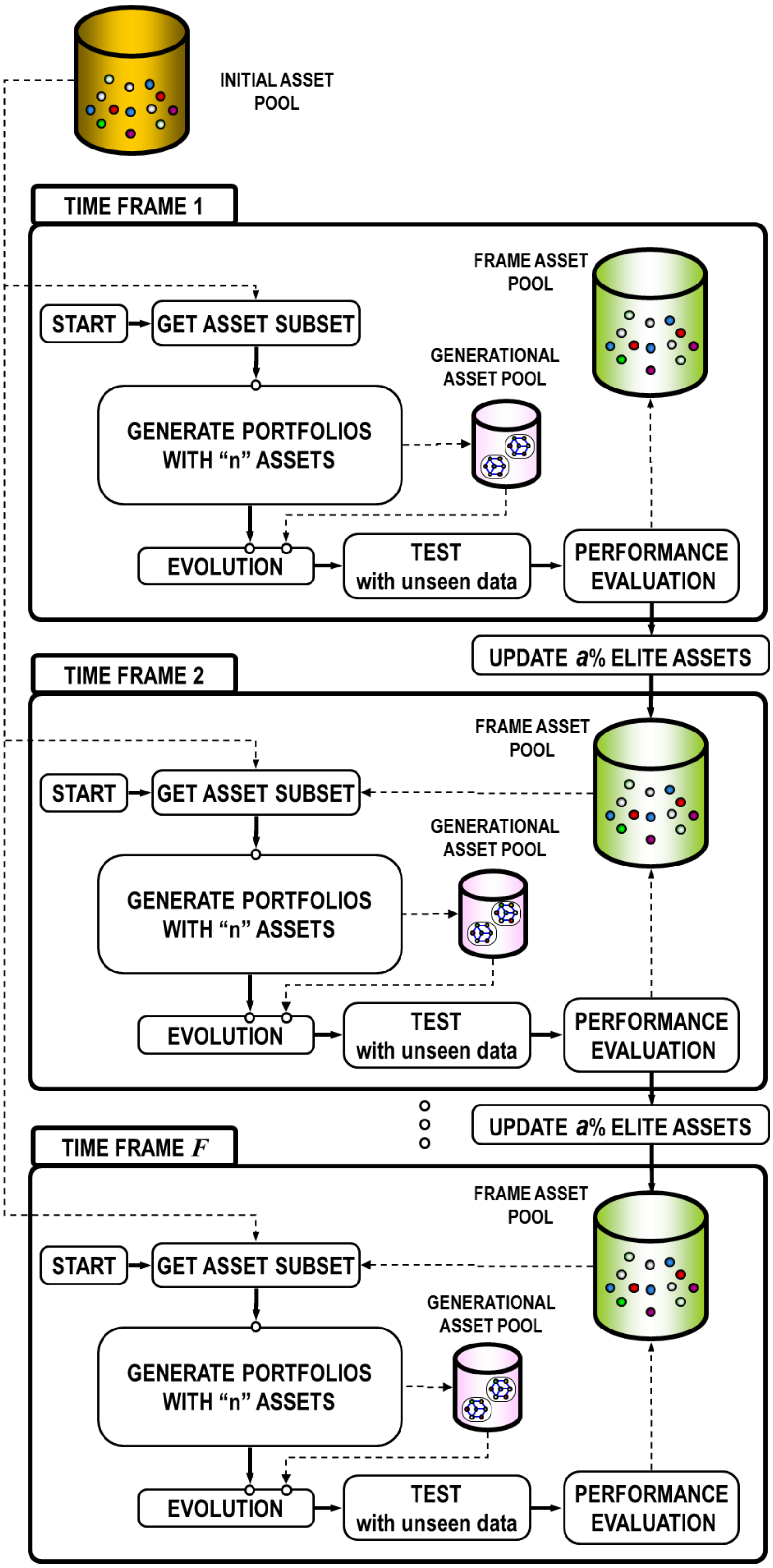


Figure 5.7: Accumulation mechanism through time frames

where, P_{iS} is the priority of asset i during the time frame of S; L_{iS} is the liquidity index of asset i during the time frame of S; β_{iS} is the beta coefficient of asset i during the time frame of S; and $\sigma_{\beta_{iS}}$ is the standard deviation of beta coefficient β_{iS}.

The advantages of using P_{iS} as the selection metric are as follows: (1) P_{iS} captures the systematic risk exposure in long term periods; (2) P_{iS} minimizes the asset exposure to the liquidity risk factors in volatile times; (3) P_{iS} uses the feature that the liquid and less risky assets have good expected return profiles in long term periods; and (4) P_{iS} avoids the noise by averaging the past return rates in volatile times.

Training phase

During the training phase, GRA evolves the optimal portfolios through time frames.

Testing phase

The testing phase has the role of validating the optimal portfolio $\overline{P}$ during the testing phase t_f, where $\overline{P}$ is the best individual in the last generation of T_f. Since the training and testing phase are closely linked by the feedback, an investment performance metric, i.e., Eq. (5.15), is defined to accumulate the $a\%$ of the assets in $\overline{P}$ after t_f, i.e., the assets with outstanding features of return performance, systematic risk and volatility:

$$PM_{it_f} = \frac{(R_{it_f} - R_{ft_f})}{(1 + \beta_{it_f})(\sigma_{R_{it_f}})}, \tag{5.15}$$

where, PM_{it_f} is the performance of asset i in $\overline{P}$ during t_f; R_{it_f} is the return of asset i during t_f; R_{ft_f} is the average risk free rate defined by 3 month U.S. Treasury Bill during t_f; β_{it_f} is the beta of asset i during t_f; and $\sigma_{R_{it_f}}$ is the standard deviation of the return of asset i during t_f.

The benefits of PM_{it_f} as the selection metric are as follows: (1) PM_{it_f} captures the systematic risk exposure, return premium and volatility during the short term period; (2) PM_{it_f} uses the low volatile, less risky and high yielded assets for the next short term periods. See Yakov(20), Hill (9) and Sharpe(3) for additional discussions.

It is important to note that Eq. (5.12), Eq. (5.14) and Eq. (5.15) are used in different contexts and different periods of time. Eq. (5.12) is effective for the portfolio evaluation in the the evolutionary process; Eq. (5.14) is effective for the assets evaluation in the long term periods; and Eq. (5.15) is useful for the assets evaluation in the short term periods.

5.5 Simulation Results

5.5.1 Problem

An *Initial Asset Pool* is to be picked up from an *Investment Universe*, which is defined as leading stock indexes, Treasury Bills and currencies detailed later. The *Initial Asset Pool* is used to build optimal GRA-based portfolios that allocate a capital K into a set of of n assets. The resulting portfolios are hypothetically invested using a buy and hold strategy over 1 week. Investment performance of GRA is compared with other asset allocation algorithms and a global financial index.

5.5.2 Investment Universe

Data used for selection, training and testing belong to listed stocks, bonds and currencies in relevant financial markets in the global scale. Three assets classes are used in the proposed approach, i.e. stocks, bonds and currencies. Daily opening and closing prices of stocks which belong to American, European and Asian indexes are picked up. American indexes include S&P500, DOW, NASDAQ, NYSE, and Rusell 3000. European indexes include S&P EUROPE, S&P 350 and S&P GLOBAL 1500. Nikkei is the index chosen in Asian sector. In the same manner, rates and yields of Treasury Bills from U.S.A., Japan, Germany and France with 3-month and 6-month coupons are used. Finally, initial spot and end prices from foreign exchange rates for dollar, euro and yen are employed.

In addition, during the selection mechanism, assets with the following features are eliminated: (1) with less than 3 years of data history; (2) lacking of market prices during the selection period; (3) with a limited market capitalization and and/or limited economic relevance(e.g. micro-cap stocks); (4) with high correlated and reduced heterogeneity in the investment universe.

5.5.3 Time Span

The proposed approach needs the use of three phases which are selection, training and testing, as denoted in previous sections. The phases are selected according the dates shown in Table 5.1. Selection phase is carried out from 05/01/2000 to 04/01/2003; training phase is performed between 05/01/2003 and 28/12/2006; and testing phase is performed between 08/01/2005 and 04/01/2007. In all time frames, the minimal time unit for data processing is a day and for data testing is a week. In accordance

Table 5.1: Dates for selection, training and testing phases

Phase	Starting date	Ending date	Total days
Selection phase	05/01/2000	04/01/2003	1079
Training phase	05/01/2003	28/12/2006	1433
Testing phase	08/01/2005	04/01/2007	716

Table 5.2: Parameters for simulation

Item	Description	Value
n	the number of assets in portfolio	30
IAP	initial asset pool size	2500
FAP	frame asset pool size	250
$a\%$	ratio of elite assets accumulation	10%
GE	the number of generations for GRA evolution	300
I	the number of individuals in GRA	250
I_C	the number of individuals by crossover	139
I_M	the number of individuals by mutation	110
I_E	the number of elite individuals	1
P_m	Probability of mutation	0.25

with Table 5.1, selection is consisted of 1079 days, training period is 1433 days, and esting period is 716 days or 103 testing weeks. It implies that in one simulation, 103 portfolios($\overline{P}$) are trained and tested, which means the number of time frames $F = 103$.

5.5.4 Parameters

The simulation settings is shown in Table 5.2. For simulations, 30 independent runs are executed and a selection phase followed by various training and testing phases was done per every simulation.

5.5.5 Performance

Training results

The effect of the number of connections in GRA on the fitness behavior is also analyzed. Fig. 5.8 shows the fitness curves of the elite portfolios during the training period. From this figure, we can see that added connections in GRA individuals causes slow fitness convergence. The important reason for this phenomenon is that the GRA structures become complex when the number of connections increases. In all cases, the fitness values converge reasonably well thorough generations. The fitness curves of other individuals have the same tendency.

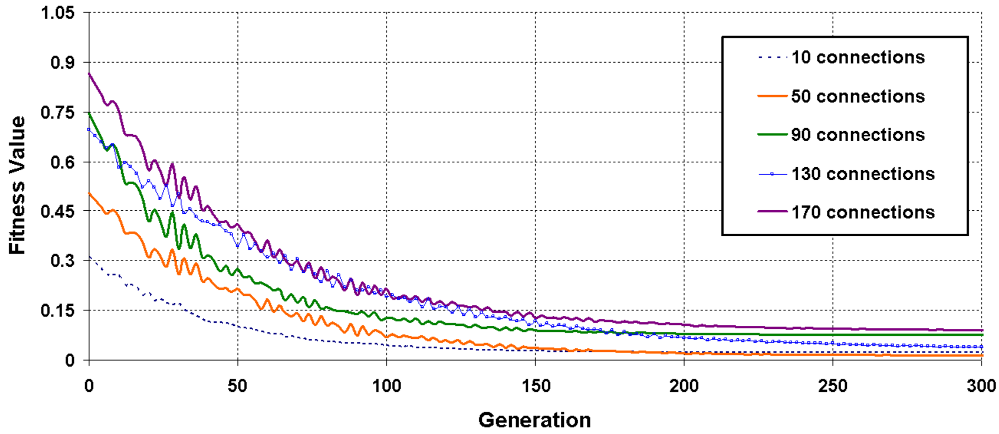


Figure 5.8: Fitness behavior when of GRA with different number of connections

Testing results

In order to show the efficiency of the proposed method, other methods are compared with the proposed method. In the literature of financial research, there is little focus on the optimization of the global asset portfolios considering simultaneously risk, return and liquidity criterion. Benchmarks were selected due to their performance reliability, practical implication and significance in the finance and computing science contexts. The first benchmark is a stochastic approach for international portfolios (Stochastic CAPM)(54). The second benchmark is a CAPM approach with GA(GA-CAPM) (55). The third benchmark is a hybrid heuristic method based on Markowitz and Neural Networks(NN-Markwoitz)(56). The fourth benchmark is a purely financial approach for the global portfolio optimization based on the Markwoitz and CAPM ideas (Black-Litterman)(57). The fifth schema is a passive index of liquid and high yield global assets, which is Dow Jones STOXX Global 1800 Index.

Figure 5.9 shows the average profit accumulation over 30 independent simulations in the proposed algorithm and the above mentioned benchmarks. All benchmarks accumulate profits reasonably well during the testing period. This phenomenon is related to the fact that all benchmarks have an explicit mechanism to include the best return yielding assets during the testing time. Additionally, by the glance of Fig. 5.9, we can see that the proposed method outperforms other benchmarks in the profit accumulation criteria. Jointly with Dow Jones STOXX Global 1800 Index, the

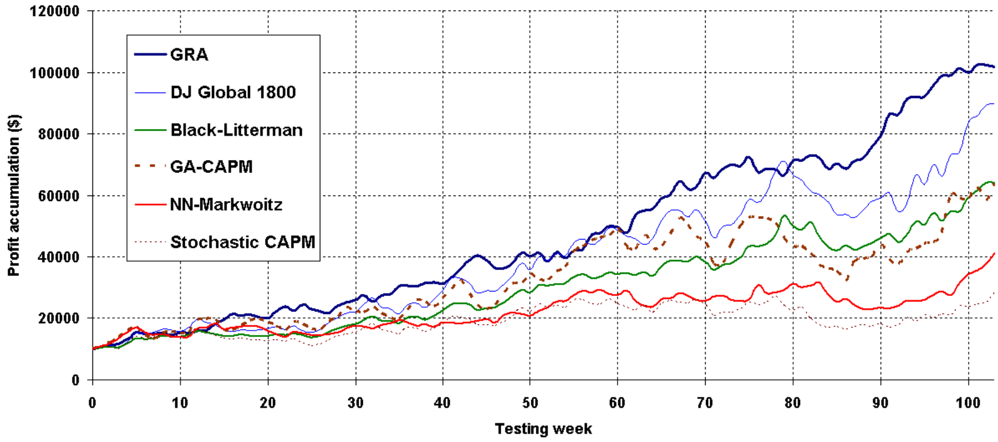


Figure 5.9: Comparison of profit accumulation of GRA and different benchmarks

proposed framework accumulates profits steadily along the testing period. The main reason for this is explained by the fact that the assets selected under these frameworks integrate liquidity measures; while others lack this mechanism.

5.6 Summary

In this chapter, GRA is used as a tool for building asset allocation models. GRA focuses on asset relationships to capture the systematic risk, offers a wider view of liquidity to enable fast, quick and liquid investment transactions. GRA enables dynamic and flexible diversification through time frames, asset accumulative strategies and multiple markets of reference in the global view. Compared to relevant benchmarks in the finance and computer science literature, GRA shows competitive results in terms of return, risk and liquidity at reasonable computational effort.

However, the proposed approach relies on a fixed set of assets in the optimal portfolio structure. Keeping the size fixed has potential issues such as failing at a local optima and over-concentrating the risks in a narrowed and undiversified portfolio of assets. The next chapter introduces a methodology that relaxes the fixed size assumption in portfolio structure and builds flexible portfolios to diversify investments more effectively.

CHAPTER

6

Portfolio Diversification with Variable Size Genetic Relation Algorithm

6.1 Aims of the Proposed Method

This chapter:

- Introduces a methodology to build *portfolio diversification* models using *Genetic Relation Algorithm with Variable Size(vs-GRA)*.
  - The methodology is based on *variable size evolution*, where individuals *shrink/expand* their structure to enhance the survivorship during the *evolution* process even when *environments* turn out distressed, which is the case of financial downturns.
  - The shrinkage/expansion, implemented through variable size crossover and mutation, is guided probabilistically, rather than randomly, to guide toward the contribution of diversification benefits during the evolution process.
  - The objective of vs-GRA is to decide the optimal scope for portfolio diversification, that is, which asset class, which industry and geography to

allocate/spread investor's capital is decided by the evolution of vs-GRA.

- Compares the proposed approach to the standard GRA.
  - Simulations use the assets listed in the *Russell Developed Index*.
  - Simulations are executed through *sliding time periods* between Jan 2005 and Dec 2009.
  - Benchmarks include the *Russell Developed Index* and the standard GRA.

6.2 Introduction

The previous chapter introduced a methodology to build optimal asset allocation models using Genetic Relation Algorithm(GRA). This chapter focuses on building *diversification models.*

Diversification consists of building a portfolio by spreading capitals systematically among diverse asset classes and segmented financial markets, so the robust portfolio can realize the optimal combination of diverse risk sources. Previous research has shown that the proper *diversification* of the risk brings benefits for the robust portfolio management in the sense that:

- *Diversification* mitigates the individual volatility risk, which mainly comes from the asset price fluctuations in changing economic conditions.
- *Diversification* also decreases the uncertainty in financial markets by getting synergies of different risk sources in diverse asset classes.

Since uncertainty and risk are two core issues in finance, portfolio *diversification* has a long-standing history both in financial research and practice. The theoretical base of diversification points some investment practices in the the last century, particularly when investors started to place their money in plural asset classes instead of an individual one, so that their investments were protected against unstable asset price movements of highly changing economic cycles(1). Following this behavior, Markowitz formalized the diversification problem of grouping plural number of assets in a mean-variance portfolio framework, in which a rational investor would maximize the expected return for a given level of risk, or minimize the risk for a given level of expected return(2). Similarly, considering heterogeneous asset classes, other institutional investors included bonds in their portfolios to reduce volatility and improve their liquidity levels(58). An example of this practice is the classical 40/60 fixed rule to mix bonds and stocks. Another example is the *Capital Market Line* of Sharpe(3), who demonstrated that rational investors obtain safer returns compared to the mean-variance efficient frontier by combining bonds and stocks in risk averse portfolios. As financial innovation rendered new investable products such as *derivatives*, *options*, *future contracts* and *swaps*, the idea of diversification turned into *hedging risk*, which rapidly dominated investment practice as a form of *insurance*. Along with these developments and being driven by market integration and deregulation forces, institutional investors headed towards geographically disperse financial markets and different economic sectors in order to gain diversification benefits(8, 10). Nowadays, the current strand on portfolio diversification practice mainly points the practice of of hedge funds and institutional banks in the form of *indexing* , which is a technique used to allocate investments into a large number of assets considering factors that determine the expected return performance, such as *fundamentals* or *market values*.

Despite the fact that portfolio *diversification* was reported as a positive practice in financial risk management, two forces prevent it from achieving consistent results through long term periods.

- First, due to over-concentration and investor home bias issues, dealing with portfolio *diversification* systematically is a scarce practice in finance. It means concentrating on risk in a limited number of assets and being exposed to price fluctuations and behavioral responses of investors' decision making(10, 59). Thus, it is vital to build the robust systems that can handle the portfolio *diversification* systematically not only to protect investors' interest, but also to ensure the health of our economy.

- Second, recent studies have shown that the cost of accessing *diversification* is higher than its potential benefits, making it a selective practice in which only companies with high *market value* in developed financial markets have better access to *diversification*, which means the better control of the risk inherent in their businesses' cash flows(8, 10). Thus, there exists a high potential in the future to spread the practice to small and emerging markets and evolve into a state of more intertwined and robust economy.

Although the idea of portfolio *diversification* is conceptually simple, defining the effective scope is an important issue. The scope for *diversification* defines which asset classes, which sectors and which geography should be included in a portfolio of assets. Generally speaking, institutional investors define the scope for diversification using statistical analysis of historical data and models with strong mathematics(59). For example, in a classical model, the full covariance or correlation matrix is taken into account, where optimizing a portfolio of assets with low mutual correlation, or low covariance, is the main goal(2). However, driven by inherent behavioral biases and restrictive mathematical assumptions of the common practice, *risk misspricing* tested the conventional money allocation schemes during the latest financial meltdown and rendered the diversification practice ineffective(60).

A promising way to define the optimal *diversification* scope is through the use of computational intelligence techniques, which have the ability to handle the information more extensively. Such techniques are inspired by nature dynamics in most cases, and proved to be suitable for complex real-world optimization problems. A representative group of such kind of techniques is under the label of *evolutionary optimization* and include methods such as *Genetic Algorithms(GA)*(61), *Genetic Programming(GP)*(62), *Grammatical Evolution (GE)*, *Evolution Strategies(ES)* and *Genetic Network Programming(GNP)*(11), which are population based evolutionary schemes with enhanced robustness against single point optimization techniques(63).

Since defining the effective scope for *diversification* is a complex search problem, which not only should consider intertwined financial markets, but also the correlated economic sectors, all the techniques from the above group are not always suitable. For instance, *GA* is not only unable to represent the underlying relationships in financial markets, but also its size grows when the complexity of the problem being tackled is higher. Furthermore, *GP* and *GE* suffer from *bloating* issues, which are related to the increase of the solution complexity without performance contribution.

This chapter proposes a unique approach for portfolio *diversification* based on *variable size Genetic Relation Algorithm(vs-GRA)*, which belongs to the class of variable size evolutionary algorithms(vs-EA). The role of *vs-GRA* in this chapter is to model and optimize the scope for portfolio *diversification* considering return, risk relationships and liquidity features for the investment purpose.

6.3 Variable Size Genetic Relation Algorithm(vs-GRA) on Portfolio Diversification

6.3.1 Basic Concept

In general terms, the aim of vs-GRA is to obtain a compact set of diverse *events* from an observed *environment* in a dynamic manner. In the portfolio diversification context, an *event* refers to an *asset*, and *environments* refer to *financial markets*. We use three kind of asset classes, i.e. *stocks*, *bonds* and *currencies*. Additionally, we use the attractive assets in developed financial markets as the market environment. Thus, vs-GRA aims at finding compact diversified portfolios that allocate a given amount of money dynamically.

This chapter is based on previous work on asset allocation, in which compact portfolios are optimized using GRA with accumulative strategies(52). The distinctive point in this chapter is that the optimal scope of portfolio *diversification* is determined by evolutionary principles rather than by users' choices. In (52), the portfolio contains a fixed number of assets and the scope for diversification is not explicitly addressed, which might imply limitations such as over-concentration of risks in a narrowed and undiversified portfolio of assets. To deal with these issues, vs-GRA allows the evolution guided by diversification benefits in sectors and countries, in which variable size *individuals* bring advantages on population diversity, over-fitting avoidance and fitness improvement due to the increased exploration ability.

The components of the proposed approach for asset allocation under diversification principles are the following:

- The *Selection Mechanism* picks up the attractive and valuable assets from financial market indexes into an *Frame Asset Pool*.
- The *Portfolio Diversification* builds a portfolio to diversify the risk into a set of diverse and attractive asset classes.

In this chapter, the *Selection Mechanism* is based on *active indexing* and GNP(Chapter 1), in which assets from defined market indexes are evaluated exhaustively by using *fundamentals* and *market* based financial metrics embedded into GNP structures. Therefore, the optimal subset of attractive assets and financial metrics to use are automatically decided by evolution, assisting investor's decision making. On the other hand, once prospective assets are identified, vs-GRA takes the role of building the *portfolio diversification* models.

6.3.2 Main Features

- The proposed scheme provides a basis for evolving the scope for portfolio diversification, so which asset classes, which sectors and which geography to consider in the portfolio structure is decided by evolution instead of arbitrary choice of users or conventional *indexing* techniques. This feature evades the behavioral bias in portfolio allocation even when financial markets turn out to be highly volatile.
- vs-GRA models and optimizes the portfolios for diversification using the graph structures, avoiding black box issues and exhaustive mathematical assumptions in conventional models, which are based on pure statistics or physics.
- vs-GRA permits handling variable size individuals, allowing the better exploration ability in the search space within the portfolio diversification context. Whereas, in the conventional GRA algorithm, the size is fixed by users' choices. Compared with *GP* and *GE*, the change of the size is systematically controlled during evolution so that benefits on portfolio diversification are ensured.

6.3.3 Evolution of vs-GRA

As any other evolutionary algorithm, vs-GRA also includes genetic operators such as selection, crossover and mutation to evolve a population of initialized *individuals* toward the optimal ones, as shown in Fig. 6.1.

The evolutionary process of Fig. 6.1 is also called *training period*, which is repeated until a terminal condition is met, i.e., a defined number of generations in this chapter. Once evolution is carried out, the best *individual* in the last generation is tested using a subsequent time period, called *testing period.* A *time frame* is a time window composed of a *training and testing period.* Asset accumulation through *time frames* is performed consecutively to track the optimum in the near coming future without the need of re-initializing the optimization process(52, 64).

The initialization of the population considers disperse geographical locations and diverse economic sectors. Every *individual* is initialized by selecting an asset from each geographic location and from each economic sector randomly as Fig. 6.2 shows. A *market niche* is the set M of m geographically disperse market indexes. Each element

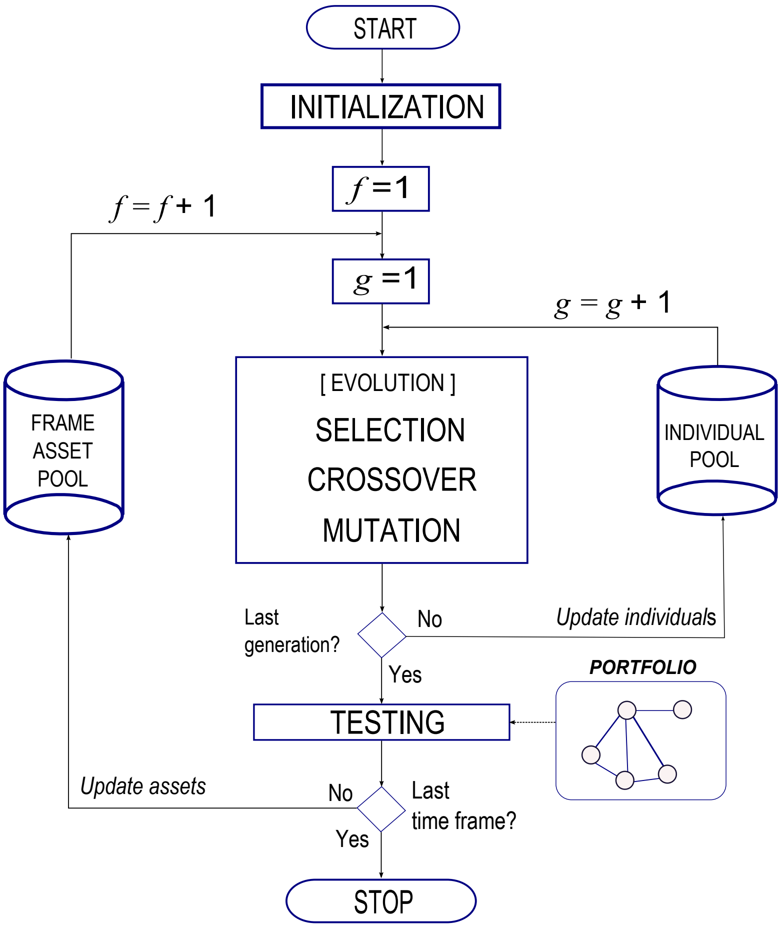


Figure 6.1: Flowchart of vs-GRA

of M is a set S including s economic sectors. In our previous approach, the initialization takes no explicit consideration of locations nor economic sectors. The objective of the diversified initialization is to avoid initial overconcentration in geographical regions or industrial sectors.

After initializing the population of *individuals*, evolution is carried out iteratively by replacing the current population with new *individuals* obtained by selection, crossover and mutation operators, which are described in later subsections.

Although previous studies on variable size evolutionary algorithms (*vs-EA*) take a more realistic view of evolution, i.e., individuals change their genosize and complexity through generations, they implicitly assume that the change of the size occurs randomly. We take a different approach, i.e., individuals change their size guided by the

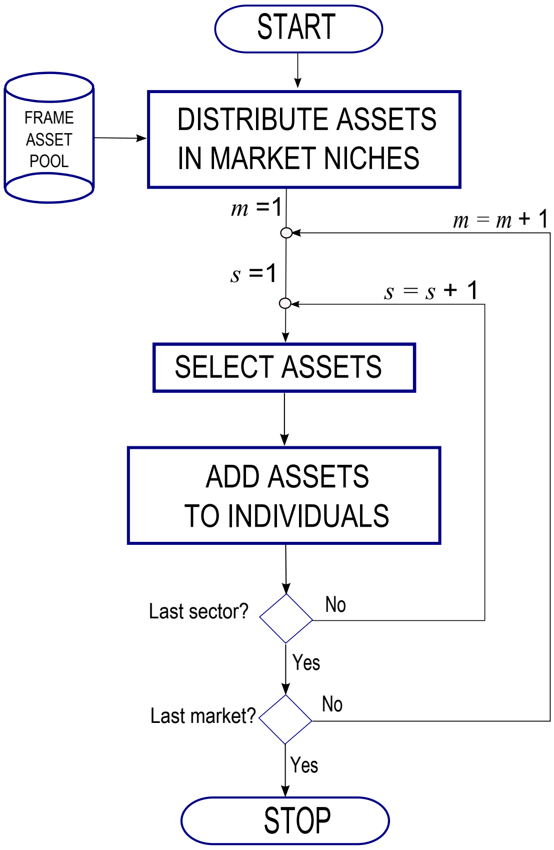


Figure 6.2: Initialization of vs-GRA individuals

contribution to diversification benefits, so that nodes are added to or deleted from the individual probabilistically rather than randomly.

Selection

The selection operator preserves a group of the better *individuals* from the *individual pool*, where the *individuals* of the population are updated continuously generation by generation(53). Better individuals are selected by tournament selection. Thus, the selection operator is a dynamic memory tracking the better *individuals* in potential areas of the search space.

Crossover

In a similar manner to sexual reproduction in biology, the crossover operator generates new offspring considering parent *individuals*. The crossover of *vs-GRA* is based on principles of speciation adaptation genetic algorithm(SAGA)(65) and synapsing variable-length crossover(SLCV)(66), which are also variable size evolutionary algorithms that perform variable size crossover to enhance the exploration ability in the search space. SAGA is based on *individual* mutual matching, where randomly chosen nodes in each *individual* are mutually tested to decide what nodes should be exchanged. However, in SAGA the selection of the crossover point is chosen at random, and its genomes are still inflexible rigid arrays of data, having a small effect on the improvement of the exploration ability. On the other hand, SLCV is a more *restrictive* operator in the sense that it chooses the initial crossover point based on similarity among parent genomes, in which only different nodes among parent *individuals* are exchanged and similar nodes are preserved. Consequently, genomes in SLVC are flexible arrays of data. However, the exchanging procedure is still executed at random, which might have an effect only on increasing the genome size, but not on improving the *individual* fitness.

Concretely speaking, Algorithm 3 describes the procedure to generate two *individuals*. Instead of taking a randomized approach to decide which nodes are to be exchanged, we take a probabilistic approach based on fundamental concepts of *restrictiveness* of SLCV and *mutual matching* of SAGA. An example of the crossover procedure is shown in Fig. 6.3.

- *Restrictiveness* means that only different nodes in both parent individuals are considered as the candidates of exchange nodes. In Fig. 6.3, the nodes $\{a_1, a_2, ..., a_6\}$ and $\{b_1, b_2, b_3\}$ are the assets which P_1 and P_2 do not share in common and thus are considered as potential candidates to exchange. Meanwhile, the nodes A and B are the assets which P_1 and P_2 share in common, thus not considered by the restrictive crossover.

- *Mutual matching* means that every candidate node in each parent individual is tested in the other parent individual, thus, every node i has a probability P_i^E to be selected as an exchanging node. Which nodes are to be exchanged is decided by comparing each P_i^E and a threshold P_*^E in Algorithm 1. For example, in Fig. 6.3, $\{a_1, a_2, a_3\}$ and $\{b_1\}$ are selected as exchanging nodes in P_1 and P_2, respectively.

The probability P_i^E of exchanging node i is based on the contribution to individual diversity, and is calculated using the following equation:

$$P_i^E = w_1 {\alpha_i}^{u_1} + w_2 {\phi_i}^{u_2}, \tag{6.1}$$

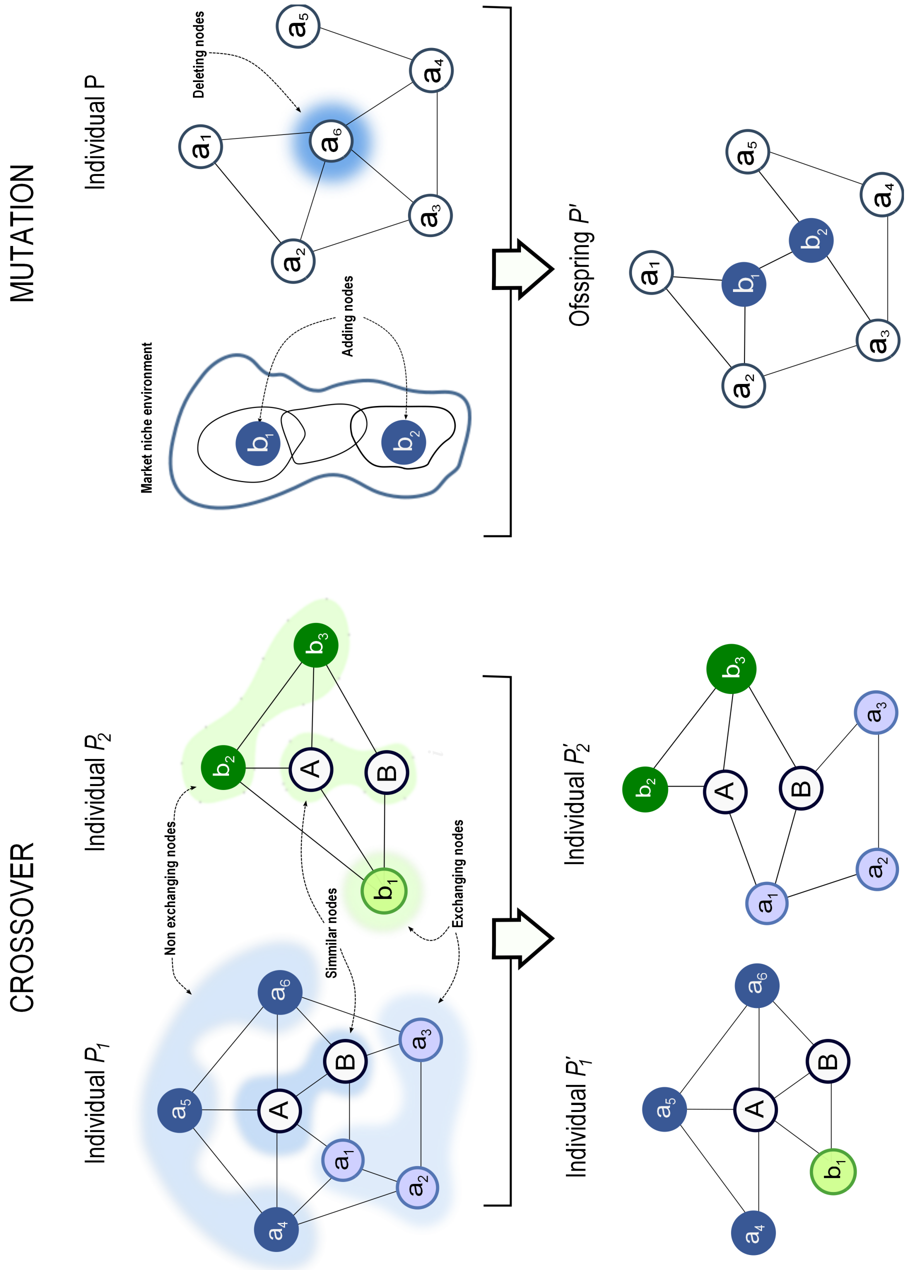


Figure 6.3: Crossover and Mutation of vs-GRA

Algorithm 3: Crossover operator for vs-GRA

Select two individuals P_1 and P_2 executing tournament selection twice
for *each different node i in each individual* **do**
  if $P_i^E \geq P_*^E$ **then**
    Move node i to another individual
  else
    Take no action
New individuals become available for the population in the next generation

where,

α_i : normalized diversity change after excluding node i (asset i) from individual P_1,

ϕ_i : normalized diversity change after including node i (asset i) in individual P_2,

u_1, u_2 : constant power for crossover probability,

w_1, w_2 : equally distributed importance weight, like $w_1 = w_2 = 0.5$.

The value of α_i shows the normalized variation $\triangle E_i$ of entropy values in P_1 when asset i is excluded from its composition. Similarly, ϕ_i shows the normalized variation $\triangle E_i$ of entropy values in P_2 when asset i is included in its composition. The variation of entropy values is calculated using the following equation:

$$\triangle E_i = \frac{E_P^* - E_P}{E_P}, \tag{6.2}$$

where,

E_P^* : diversity measure of individual P after excluding/including asset i in P,

E_P : diversity measure of individual P before excluding/including asset i in P.

$$E_P = \sum_{c \in C} x_c \ln \frac{1}{x_c} + \sum_{s \in S} x_s \ln \frac{1}{x_s}, \tag{6.3}$$

where,

C : set of suffixes of countries,

S : set of suffixes of sectors,

x_c : proportion of the money that individual P allocates to country c,

x_s : proportion of the money that individual P allocates to sector s.

The above is a compounded measure for portfolio diversification, in which the diversity to the country and sector allocation is considered. Thus, the proportions x_c and x_s are the total money allocation that portfolio P makes to country c and sector s, respectively. As we can see, the diversity measure is based on entropy concepts borrowed from Information Theory. Although there exists other metrics such as the *Herfindahl* or *Gini* Index that quantify how diversified a given portfolio P is, previous studies have shown that the entropy based metric provides a reasonable and effective approximation to the benefits of diversifying investments in international finance context(67, 68). Although recent studies favor country diversification rather than the industry based diversification approach(69, 70), we consider the diversification over countries to be as important as the diversification over industries(sectors), for which the optimal structure would be determined by the evolution of *vs-GRA*.

Mutation

Simulating asexual reproduction of individuals in biology, the mutation procedure generates an offspring from a single parent *individual*. Concretely speaking, Algorithm. 4 shows the procedure to generate an *individual* using a parent *individual*.

The mutation procedure not only changes the information parameters of the *individual*, but also alters its size by probabilistically shrinking (expanding) the genosize of the *individual* by deleting (adding) a single node from (to) the *individual* as Fig. 6.3 shows. The probability P_i^{add} and P_i^{del} of adding and deleting node i, respectively, depends on the extent of the diversity contribution to the *individual*. The following equations are used to calculate these values:

Algorithm 4: Mutation operator for vs-GRA

```
Select an individual P using tournament selection
Change connections: Select connections with the
  probability of P_m and reconnect to a different node
Change function: Select node functions F_i and change
  to other function with the probability of P_m
for asset i ∈ AP do                              /* Add node */
    Select i from the frame asset pool AP with the
      probability of P_i^add
    Add i to the individual P
    Set its connections to other nodes at random

for asset i ∈ P do                            /* Delete node */
    Select i with the probability of P_i^del
    Delete i from the individual P

New individual become available for the population in the
  next generation
```

$$P_i^{add} = {\gamma_i}^{v_1}, \tag{6.4}$$

$$P_i^{del} = {\eta_i}^{v_2}, \tag{6.5}$$

where,

γ_i : normalized diversity change when including node i from the *frame asset pool* AP into individual P,

η_i : normalized diversity change when excluding node i from individual P,

v_1 : constant power for adding probability,

v_2 : constant power for deleting probability.

Mutation and crossover have complementary roles throughout generations. By testing each crossover node in each parent *individual*, crossover not only combines beneficial mutations that occur in different parent *individuals*, but also avoids detrimental mutations that occur in one parent *individual* and not in the others.

Similarly, the mutation operator aims at avoiding detrimental crossover by adding and deleting nodes. The *add* procedure picks the candidate assets from the *frame asset pool AP* to incorporate into the *individual*, so that portfolio diversity is improved. Although we use the AP as a set of candidate assets, it can be changed easily to any other asset set that the investor has interest in. The *delete* procedure removes needless nodes in the *individual* that have poor contribution to the portfolio diversity.

6.4 Simulation Results

6.4.1 Problem

A *Selection Mechanism* is performed to first identify a set of prospective assets into a *Frame Asset Pool* by using the asset selection algorithm proposed in Chapter 2. The resulting set, along with relevant bonds and currencies in the *Investment Universe*, are used to build flexible portfolio diversification structures to allocate a capital K into the given subset of asset classes. The resulting portfolios are hypothetically invested using a buy and hold strategy over 1 month.

6.4.2 Investment Universe

Three assets classes are used in the proposed approach, i.e. stocks, bonds and currencies.

- The *stock market index M* consists of 2372 assets listed in the *Russell Developed Index.*
- As for bonds, rates and yields of Treasury Bills from U.S.A., Japan, Germany and France with 3-month and 6-month coupons are used.
- As for currencies, initial spot and end prices from foreign exchange rates for dollar, euro and yen are used.

6.4.3 Time Span

To validate the effectiveness of the proposed approach, we perform simulations from 2005 to 2009. As we focus on the *buy side* of the investment cycle in this paper, the testing phase is executed for one month period, in which we *buy and hold* the obtained portfolio in the evolution phase. Thus, the training and testing periods are shown in Table 6.1, implying 48 periods of the training and testing.

Table 6.1: Dates for simulation

Period	Starting date	Ending date
Training period	01/03/2005	11/27/2009
Testing period	01/02/2006	12/31/2009

6.4.4 Parameters

A simulation run consists of executing the training and testing under the sliding time windows approach. The training phase consists of the evolution of vs-GRA using data of one year period, in which the the parameters for evolution is shown in Table 6.2. The initial capital is \$10,000. The base currency for the fitness evaluation is the dollar. Other parameters include a risk averse investor, policy for reinvesting profits and no tax expenses.

Table 6.2: Parameters of vs-GRA and GRA

Item	Description	Value
GE	the number of generations for GRA evolution	300
I	the number of individuals in GRA	200
I_C	the number of individuals by crossover	75
I_M	the number of individuals by mutation	120
I_E	the number of elite individuals	5
P_m	probability of mutation	0.25
P_*^E	threshold for crossover	0.25
TS	tournament size	7
u_1, u_2, v_1, v_2	constant power for crossover and mutation	0.5

6.4.5 Performance

As for the training period, the convergence rate of the average best fitness values over 20 independent runs is shown in Fig. 6.4. Bars represent the standard deviation with 2σ. We can observe that vs-GRA obtains better values during the training. We believe this occurs due to the fact that vs-GRA is able to allow the variable size during the evolution, implying that the search space is enlarged rather than fixed in the conventional GRA.

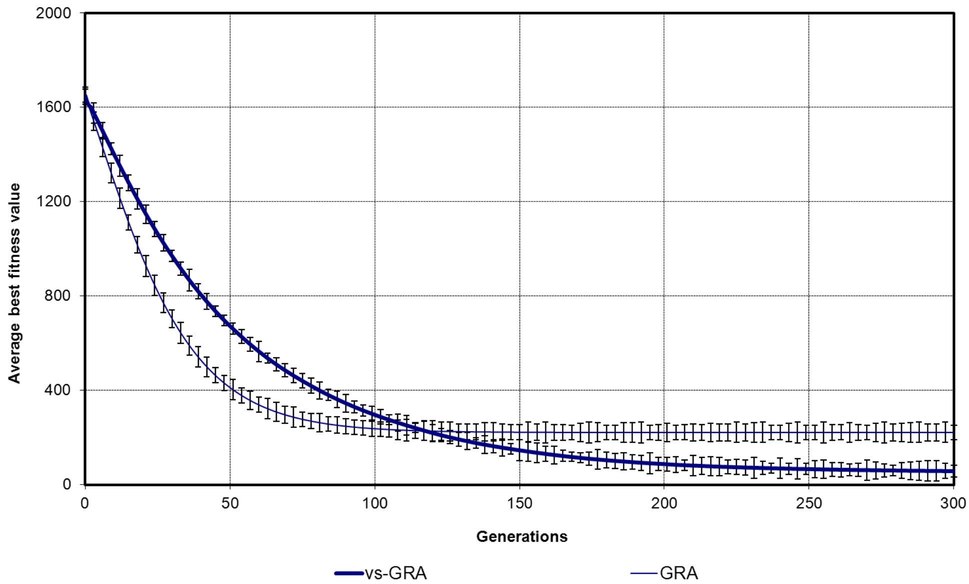


Figure 6.4: Convergence of the average best fitness values

Fig. 6.5 shows the averaged monthly accumulated return performance, i.e., the economic performance of the proposed method in the testing period over 48 months. Dec-05 represents the initial date when investors hold the capital of 100%; subsequent gains/losses are added/deleted from the initial capital until the last testing month in Dec-09. We can see from Fig. 6.5 that vs-GRA accumulates return considerably well during 2006 and a large part of 2007, which clearly represents upward trends in financial markets. On the other hand, vs-GRA is also affected by the systemic crisis in 2008 and 2009, however, even in such an event, it is able to keep positive but low accumulated return rates.

To show the performance of both systems in a period of financial crisis, Feb-09 is chosen for the analysis because, during this period, the majority of financial markets had the worst return performance as shown in Fig. 6.5. Fig. 6.6 shows the performance comparison, asset class, sector composition, countries and equity holdings of the elite individuals in Feb-09. We can see from Fig. 6.6 that the performance of GRA with different sizes is dominated by the performance of vs-GRA with 238 assets, which is automatically determined for Feb-09. This gives an idea that merely changing the indi-

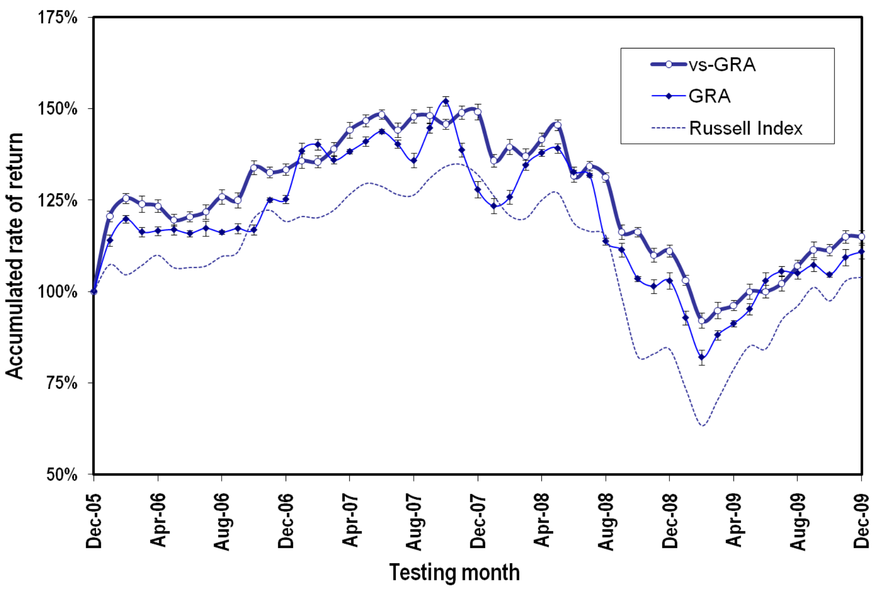
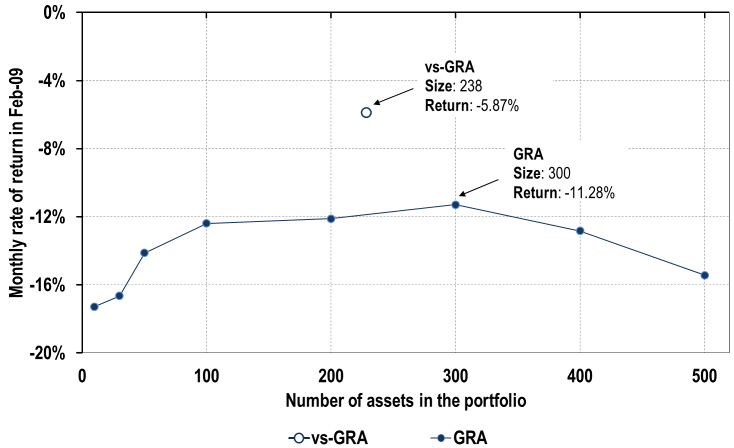


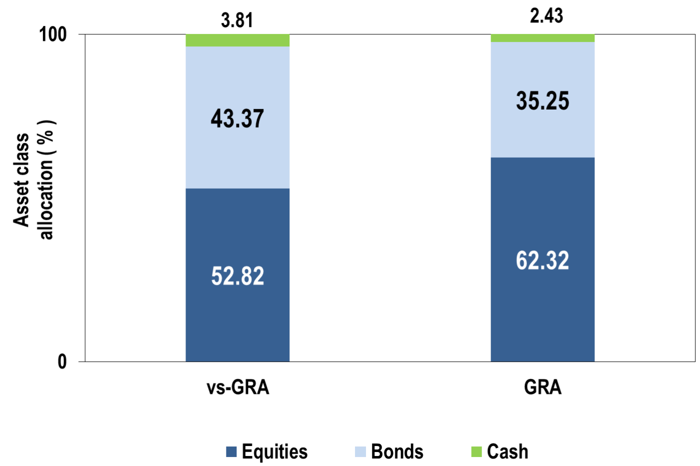
Figure 6.5: Average accumulated return rate in the testing period

vidual size in the standard GRA is not enough to build robust portfolios. Additionally, we can see that the standard GRA tends to have overconcentrated allocations on top equities, top performing sectors or leading countries. This is natural since GRA lacks of an explicit diversification mechanism and the size is fixed by the user choice. Thus, the risk of overconcentration of investments on asset classes, countries or sectors leads to the suboptimal portfolios in GRA.

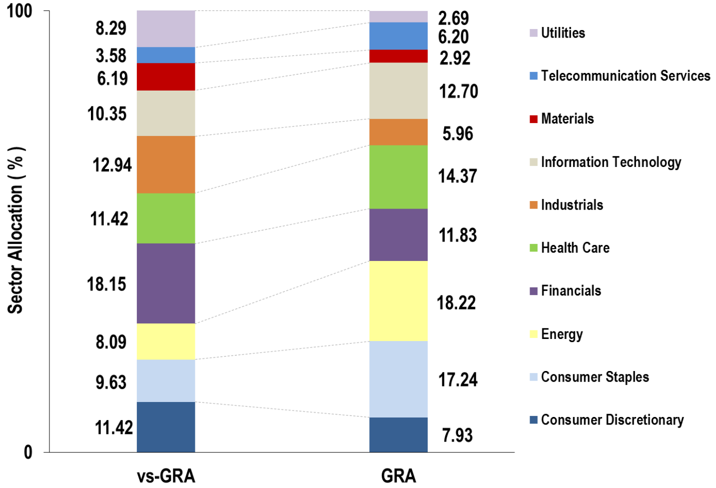
To show the performance of both systems in the whole testing period, Fig. 6.7 compares vs-GRA against the standard GRA considering different portfolio sizes in terms of average return, volatility and Sharpe Ratio. The Sharpe ratio is a measure of the return per unit of risk, which is obtained by averaging the division of the monthly rate of return by the monthly standard deviation(71). We can see from Fig. 6.7 that GRA reduces its volatility when the size is larger, which is consistent with similar findings(2, 3, 59). However, vs-GRA obtains better returns than the standard GRA even if GRA increases its size to 500 assets. This result is because vs-GRA keeps the individual size flexible to optimize the diversification through variable length crossover and mutation, i.e., the robust spread of the investment over asset classes, sectors and countries is decided by evolution. Since GRA lacks of explicit diversification mechanism, it tends to have highly concentrated allocations as shown in Fig. 6.6.



(a) Monthly Return Performance



(b) Asset Class Allocation



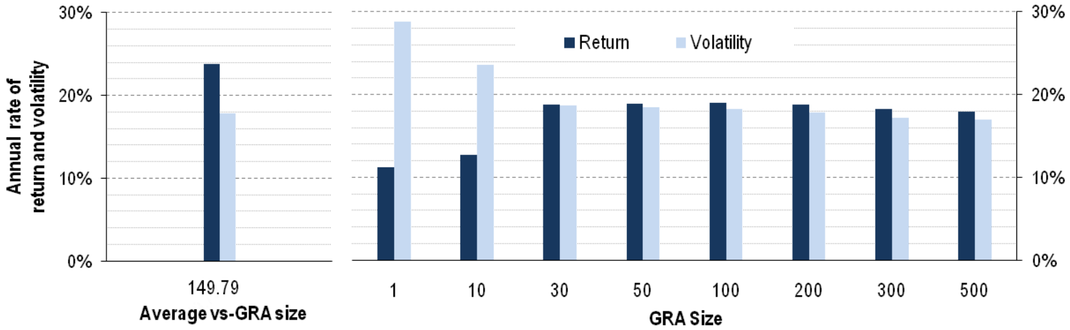
(c) Sector Allocation

Top 20 holdings			
vs-GRA	%	GRA	%
EXXON MOBIL CORP	0.95	EXXON MOBIL CORP	4.63
ROYAL DUTCH SHELL PLC-B	0.95	JOHNSON&JOHNSON	3.60
HOME DEPOT INC	0.94	PROCTER&GAMBLE CO	3.19
SANOFI-AVENTIS	0.92	NESTLE SA-REG	2.80
KRAFT FOODS INC-A	0.92	INTL BUSINESS MACHINES	2.74
ENI SPA	0.92	MICROSOFT CORP	2.73
MEDTRONIC INC	0.92	CHEVRON CORP	2.72
ALTRIA GROUP INC	0.88	VODAFONE GROUP PLC	2.70
MITSUBISHI UFJ FINANCIAL G.	0.88	BP PLC	2.64
BAXTER INTERNATIONAL INC.	0.88	TOTAL SA	2.50
3M CO	0.87	WAL-MART STORES INC	2.44
E.ON AG	0.87	NOVARTIS AG-REG	2.14
EXELON CORP	0.83	GENERAL ELECTRIC CO	1.97
THE WALT DISNEY CO.	0.83	TELEFONICA SA	1.95
BANCO SANTANDER SA	0.82	JPMORGAN CHASE&CO	1.88
COLGATE-PALMOLIVE CO	0.82	HSBC HOLDINGS PLC	1.87
BG GROUP PLC	0.81	PFIZER INC	1.83
BRITISH AMERICAN TOBACCO	0.80	TOYOTA MOTOR CORP	1.82
KIMCO REALTY CORP	0.78	COCA-COLA CO/THE	1.80
REGENCY CENTERS CORP	0.78	GLAXOSMITHKLINE PLC	1.75
ELI LILLY&CO	0.72	PEPSICO INC	1.65

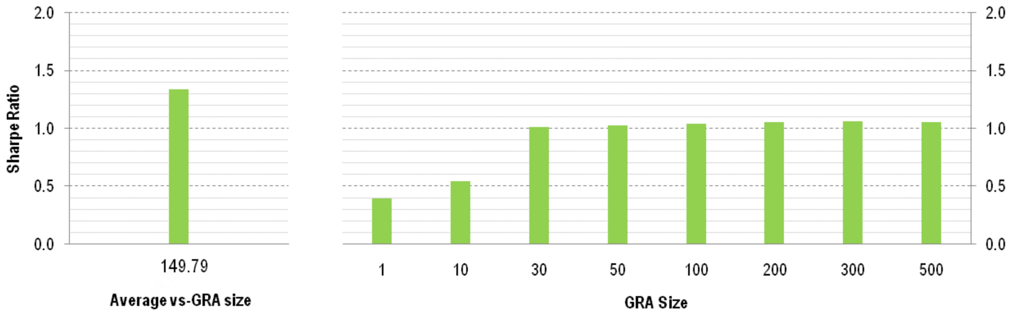
Top 10 country allocation (%)		
Country	vs-GRA	GRA
United States	25.82	46.65
United Kingdom	12.38	15.63
France	6.53	5.92
Germany	8.46	7.52
Switzerland	6.54	5.72
Japan	4.93	3.29
Spain	3.58	2.49
Italy	5.82	1.97
Netherlands	2.31	1.72
Canada	2.61	1.67

(d) Top Holdings and Top Conutries

Figure 6.6: Performance of elite genotypes in vs-GRA and GRA during the testing period of Feb-09



(a) Average rate of return and volatility



(b) Average Sharpe Ratio

Figure 6.7: Performance of vs-GRA and GRA with different sizes

In terms of the performance comparison, the average monthly Sharpe ratios is shown in Fig. 6.8 and the average annualized/monthly return and volatility rates are shown in Table 6.3. We can note easily that vs-GRA is able to keep higher Sharpe ratios than the conventional GRA in the asset allocation context. As vs-GRA manages a flexible number of assets in the portfolio structure, it is able to minimize the volatility risk expressed by return standard deviations.

Table 6.3: Economic comparison in the testing period

Metric	Russell Index	GRA	vs-GRA
Average monthly return(%)	1.42	1.59	1.64
Average monthly volatility(%)	5.87	5.39	4.96
Average annualized return(%)	15.78	18.81	21.55
Average annualized volatility(%)	20.36	18.67	17.19

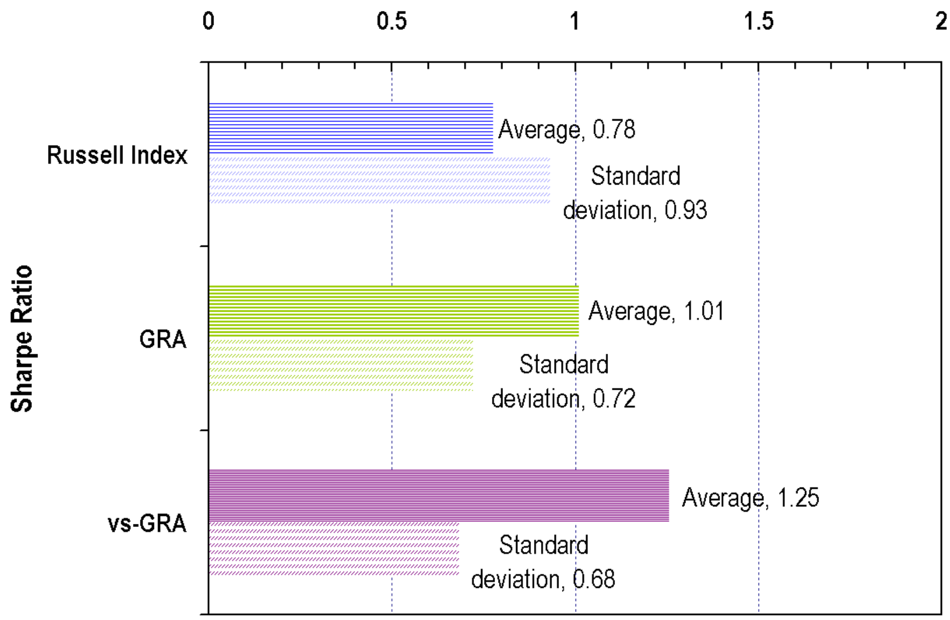


Figure 6.8: Average Sharpe ratios in the testing period

6.5 Summary

This chapter has introduced a novel method for diversifying investments in globally located assets. vs-GRA has the role of building portfolio structures considering the variable size during the evolution, which are guided probabilistically using portfolio diversity metrics in the portfolio. In this sense, the optimal structure for portfolio diversification is decided by the evolution.

It has been observed from simulations that considering the optimal structures for diversification reduces the impact of losses even when financial markets turn distressed.

However, as our method relies on entropy based measures for mapping the risk exposure of portfolios, other sources of the risk need to be considered, and the future research should aim at designing holistic risk management frameworks.

CHAPTER

7

Conclusions

This thesis has provided algorithms for making investment models, which are able to diversify the risks while allocating the scarce economic resources in multiple asset classes spread in developed financial markets, by using the principles of evolution of *Genetic Network Programming(GNP)* and pricing of *Value Investing*

Chapter 2 has proposed an algorithm to build optimal *asset selection models* in the form of *risk pricing mechanisms* by using the *evolutionary computing* principles of *Genetic Network Programming* and *value investing* principles; where the *judgment* and *processing* nodes in the network structure of GNP use the *intrinsic* and the *extrinsic* risk factors to decide on the *asset selection* decision. Moreover, the number, type and combination of required/relevant risk factors are decided by the evolution structure of GNP. The resulting output of the GNP-based algorithm is a subset of prospective assets that can be invested with equal proportion over a defined period of time using a simple buy and hold and strategy. The proposed algorithm has suggested through simulations that making *risk pricing mechanisms* considering the *intrinsic* and *extrinsic* factors embedded in evolutionary networks of GNP brings advantages in return performance over the standard *value*, *growth* and *capitalization* indexing strategies. It has implications on improving the resiliency of the conventional indexing strategies in finance by relaxing the assumption that ranking and fixed number of risk factors are relevant to find *misspricing opportunities*.

Chapter 3 has proposed an algorithm to stress-test the *asset selection models* by using the juncture of *evolutionary* and *robustness* principles in *Robust Genetic Network Programming(r-GNP)*; where the fitness evaluates not only the the main performance function, but also the deviation of performances over multiple and divergent environments(scenarios generated by block bootstrapping technique with noises). The resulting output of the r-GNP algorithm is a *robust asset selection model* with improved *generalization* ability, which means the avoidance of the extrapolation of historical risk factors when applied to future horizons. From simulations, it has been observed that the *generalization* ability of r-GNP has advantages not only in terms of *return*, but also in terms of *risk* and *liquidity*, implying better prospects to avoid the overfitting issues in the conventional GNP. The results bring relevant implications in finance, that is, stress-testing *risk pricing models* by using the robust evolutionary framework of r-GNP, brings benefits to avoid extrapolating historical performance into future horizons(behavioral bias).

Chapter 4 has proposed an algorithm to enhance the adaptability of the *asset selection models* based on *Genetic Network Programming with Changing Structures (GNP-cs)*, which implements *control* and *operational* functions. The *control* function monitors the occurrence of environmental changes, in terms of *economic fluctuations*, and the *operational* function devises strategies, in terms of *asset selection models*, to deal with the detected changes. The evolution of GNP-cs is executed using jointly evolved functionally distributed systems, where the fitness function considers the accuracy of the *control* function and the economic performance of the *operational* function. It has been observed from simulations that GNP-cs has benefits in return enhancement. It has implications not only on building adaptive economic systems that can better consider the fluctuations in the markets, but also on finding the *misspricing opportunities* in macroeconomic factors, while building an asset selection model.

Chapter 5 has proposed an algorithm to build the optimal asset allocation models by using the evolutionary undirected network structures of *Genetic Relation Algorithm(GRA)*, where each *node* in the network of GRA models financial *assets*, such as *stock*, *bond* or *currency*, and each *relationship* measures the *systematic risk* between a pair of assets. The evolution of GRA considers accumulative strategies through recent generations/time frames to aid the search for the optimal portfolios when the recent past resembles the characteristics of the future horizons. It has been observed from simulations that the GRA-based asset allocation models outperformed other portfolio optimization algorithms and conventional strategies in Finance, such as Makowitz and CAPM. It means that considering a partial set of *systematic risk* relationships when building portfolios by using GRA brings benefits on minimizing the portfolio's market risk, expressed in the portfolio *beta*, and has implications on lessening the portfolio's correlation with the market indexes.

Chapter 6 has proposed an algorithm to build *portfolio diversification* models using *Genetic Relation Algorithm with Variable Size(vs-GRA)*, which *shrink/expand* the structure of individuals to enhance the survivorship during the *evolution* process. The shrinkage/expansion is implemented through variable size crossover and mutation, which are guided probabilistically to guide toward the contribution of *diversification benefits* during the evolution process. It has been observed from simulations that GRA-vs has benefits on building flexible asset allocation structures, which in turn diversify the risk over multiple asset classes, such as stocks, bonds and currencies, economic sectors and geography more effectively than the standard GRA approach. In finance it has implications on leaving the conventional view that large portfolio sizes diversify risk effectively.

Further studies should be addressed. Although the proposed scheme relies on market factors and fundamentals as sources of risk, other sources of risk should be systematically evaluated. As financial integration opens new uncertain markets, such as the case of emerging markets, the future research should aim at designing holistic risk management frameworks.

REFERENCES

[1] B. Graham and D. L. Dodd, "Security Analysis", The Mapple Press Company, 1951. 5, 6, 8, 82

[2] H. Markwoitz, "Portfolio selection: Efficient diversification of investments", New York, John Whiley and Sons, 1959. 5, 6, 8, 61, 82, 84, 97

[3] W. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk", Journal of Finance, Vol. 19, No. 3, pp. 425-442, 1964. 5, 8, 61, 76, 83, 97

[4] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities", Journal of Political Economy, Vol 81, No. 3, pp. 637-654, 1973. 5, 61

[5] E. Fama, "Multifactor portfolio efficiency and multifactor asset pricing", The Journal of Finance, Vol. 31, No. 4, pp. 441-465, 1996. 5

[6] A. Lo and P. Patel, "130/30: The New Long-Only", The Journal of Portfolio Management, Vol. 34, No. 2, pp. 12-38, 2008. 5, 8

[7] E. Fama and K. R. French, "Multifactor explanation of asset pricing anomalies", The Journal of Finance, Vol. 51, No. 1, pp. 55-84, 1996. 6

[8] R. Sullivan, "Taming global village risk II: Understanding and mitigating bubbles", Journal of Portfolio Management, Vol. 35, No. 4, pp. 131-141, 2009. 6, 61, 62, 83

[9] J. M. Hill, "A Perspective on Liquidity Risk and Horizon Uncertainty", The Journal of Portfolio Management, Vol. 35, No. 4, pp. 60-68, 2009. 6, 15, 61, 62, 65, 76

[10] United Nations. "The global economic crisis: systemic failures and systemic remedies", The report of the United Nations Conference on Trade and Development, New York, 2009. 6, 62, 83

[11] S. Mabu, K. Hirasawa and J. Hu, "A graph-based evolutionary algorithm: Genetic Network Programming and its extension using reinforcement learning", Evolutionary Computation, MIT Press, Vol. 15, No. 3, pp. 369-398, 2007. 6, 47, 48, 52, 66, 84

[12] W. Chan, R. Frankel and S. Kothari, "Testing behavioral finance theories using trends and consistency in financial performance", Journal of Accounting and Economics, Vol. 38, pp. 3-50, 2004. 8

[13] I. Casanova, "Tradinnova-LCS: Dynamic stock portfolio decision-making assistance model with genetic based machine learning", In Proc. of the 2010 IEEE World Congress on Evolutionary Computation. 8, 46

[14] W. Yan and C. D. Clack, "Behavioural GP diversity for adaptive stock selection", In Proc. of the 11th annual conference on Genetic and evolutionary computation, pp. 1641 - 1648, 2009. 8, 24, 46

[15] R. McGee, M. O'Neill and A. Brabazon, "The syntax of stock selection: Grammatical Evolution of a stock picking model", In Proc. of the 2010 IEEE World Congress on Evolutionary Computation. 8, 46

[16] E. Fama and K. French, "Value versus Growth: The International Evidence", The Journal of Finance, Vol. 53, No. 6, pp. 1975-1999, 1998. 9, 18

[17] J. Brush, "Value and Growth, Theory and Practice", The Journal of Portfolio Management, Vol. 33, No. 3, pp. 22-30, 2007. 9, 13, 18, 20

[18] D. Blitz and P. Vliet, "Global Tactical Cross-Asset Allocation: Applying Value and Momentum Across Asset Classes", The Journal of Portfolio Management, Vol. 35, No. 1, pp. 23-38, 2008. 9, 13, 18

[19] M. L. Leibowitz and A. Bova, "Allocation Betas", Financial Analysts Journal, Vol. 61, No. 4, pp. 70-82, 2005. 14, 68, 69

[20] Y. Amihud and H. Mendelson, "Stock and Bond Liquidity and its Effect on Prices and Financial Policies", Financial Markets and Portfolio Management, Vol. 20, No. 1, pp. 19-32, 2006. 15, 76

[21] L. Chang and J. Lakonishok, "Value and Growth Investing: A Review and Update", Financial Analysts Journal, Vol. 60, No. 1, pp. 71-86, 2004. 18, 20

[22] V. Nissen and J. Propach, "On the Robustness of Population-Based Versus Point-Based Optimization in the Presence of Noise", IEEE. Transactions on Evolutionary Computation, Vol. 2, No. 3, pp. 107-119, 1998. 23

[23] J. Visser, J. Hermisson and GP. Wagner, "Perspective: Evolution and detection of genetic robustness", Evolution, Vol. 57, No. 9, pp. 1959-1972, 2003. 23

[24] H. Beyer and B. Sendohff, "Robust Optimization A comprehensive survey", Computer Methods in Applied Mechanics and Engineering, Vol. 196, No. 33-34, pp. 3190-3218, 2007. 23, 24

[25] Y. Jin and J. Branke, "Evolutionary optimization in uncertain environments - A survey", IEEE. Transactions on Evolutionary Computation, Vol. 9, No. 3, pp. 303-317, 2005. 23

[26] F. Fabozzi, D. Huang and G. Zhou, "Robust portfolios: contributions from operations research and finance", Analysis of Operations Research, Vol. 176, No. 1, pp. 191-220, 2010. 23

[27] V. Parque, S. Mabu and K. Hirasawa, "Asset selection in global financial markets using Genetic Network Programming", In Proc. of the 2010 IEEE International Conference on Systems, Man, and Cybernetics, pp. 677-683, 2010. 23

[28] K. Hirasawa, J. Murata, J. Hu and C. Jin, "Universal Learning Network and its Application to Robust Control", IEEE Trans. on Systems, Man and Cybertnetics, Part C, Vol 30, No. 3, pp. 419-430, 2000. 23

[29] I. Kushchu, "Genetic Programming and Evolutionary Generalization", IEEE. Transactions on Evolutionary Computation, Vol. 6, No. 5, pp. 431-442, 2002. 24

[30] L. Panait and S. Luke, "Methods for evolving robust programs", In Proc. of the 2003 Genetic and Evolutionary Computation Conference, Vol. 2724 of LNCS, pp. 1715-1728, 2003. 24

[31] BIS. "Basel II: Revised international capital framework", Banking for International Settlements,
http://www.bis.org/publ/bcbs128.htm, 2010. 24, 28, 61, 62

[32] P. Bühlmann, "Bootstrap for Time Series", Statistical Science, Vol. 17, pp. 52-72, 2002. 28

[33] P. Hall, J. Horowitz and B. Jing, "On blocking rules for the bootstrap with dependent data", Biometrika, Vol. 82, pp. 561-574, 1995. 28

[34] G. Wilson and W. Banzhaf, "Soft Memory for Stock Market Analysis using Linear and Developmental Genetic Programming", In Proc. of the 9th annual conference on Genetic and Evolutionary Computation, pp. 2234 - 2241, 2007. 46

[35] G. Hassan and C. Clack, "Robustness of multiple objective GP stock-picking in unstable financial markets", In Proc. of the 9th annual conference on Genetic and Evolutionary Computation, pp. 1513 - 1520, 2009. 46

[36] W. Yan and C. D. Clack, "Evolving Robust GP Solutions for Hedge Fund Stock Selection in Emerging Markets", In Proc. of the 9th annual conference on Genetic and Evolutionary Computation, pp. 1633 - 1640, 2009. 46

[37] L. Charbonneau and N. Kharma, "Evolutionary Inference of Rule-based Trading Agents from Real-World Stock Price Histories and Their Use in Forecasting", In Proc. of the 9th annual conference on Genetic and Evolutionary Computation, pp. 1459 - 1466, 2009. 46

[38] S. K. Lee and B. R. Moon, "A New Modular Genetic Programming for Finding Attractive Technical Patterns in Stock Markets", In Proc. of the 9th annual conference on Genetic and Evolutionary Computation, pp. 1459 - 1466, 2010. 46

[39] A. Ghandar, Z. Michalewicz and R. Zurbruegg, "Interpretable multi-criteria fuzzy rule based decision models for hedge fund management", In Proc. of the 2010 IEEE World Congress on Evolutionary Computation. 46

[40] V. Parque, S. Mabu and K. Hirasawa, "Evolving the asset selection with Genetic Network Programming", IEEJ Transactions on Electrical and Electronic Engineering, to appear, 2011. 47

[41] OECD, "Reference chronology of turning points", Organization for Economic Co-Operation and Development, http://www.oecd.org, 2011. 54

[42] F. Modigliani and M. Miller, "The Cost of Capital, Corporation Finance and the Theory of Investment", American Economic Review, Vol. 48, No. 3, pp. 261-297, 1958. 61

[43] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities", Journal of Political Economy, Vol 81, No. 3, pp. 637-654, 1973. 5, 61

[44] S. Ross, "The arbitrage theory of capital asset pricing", Journal of Economic Theory, Vol 13, pp. 341-360, 1976. 61

[45] J. Cox, S. A. Ross and M. Rubinstein, "Option Pricing: A Simplified Approach", Journal of Financial Economics 7, pp. 229-263, 1979. 61

[46] K. A. Froot and J. Stein, "Risk Management, Capital Budgeting and Capital Structure Policy for Financial Institutions: An Integrated Approach", Journal of Financial Economics 47, No. 1, pp. 55-82, 1998. 61

[47] A. Fernández and S. Gómez, "Portfolio selection using neural networks", Computers & Operations Research, Vol. 34, Issue 4, pp. 1177-1191, 2007. 61, 79

[48] P. Lipinski, "Evolutionary Strategies for Building Risk-Optimal Portfolios", Studies in Computanional Intelligence, Vol. 100, pp. 53-65, Springer Berlin, 2008. 61

[49] A. Ghandar, Z. Michalewicz, M. Schmidt, T. D. To and R. Zurbrugg, "Computational intelligence for evolving trading rules", IEEE Transactions on Evolutionary Computation, Vol. 13, No. 1, pp. 71-86, 2009. 61

[50] P. Jckel, Monte Carlo Methods in Finance, Wiley, 2002. 61

[51] E. Gonzales, K. Taboada, K. Shimada, S. Mabu and K. Hirasawa, Evaluating Class Association Rules using Genetic Relation Programming, In Proc. of the IEEE congress on evolutionary computation 2008, pp. 731-736, 2008. 66

[52] V. Parque, S. Mabu and K. Hirasawa, "Evolving asset portfolios by Genetic Relation Algorithm", Journal of Advanced Computational Intelligence and Inteligent Informatics, Vol. 14, No. 5, pp. 464-474, 2010. 68, 85, 86

[53] H. Zhou, W. Wei, K.Shimada, S. Mabu and K. Hirasawa, "Time Related Association Rules Mining with Attributes Accumulation Mechanism and its Application to Traffic Prediction", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 12, No.5, pp. 467-478, 2008. 70, 88

[54] N. Topaloglou, H. Vladimirou and S. A. Zenios, "A dynamic stochastic programming model for international portfolio management", European Journal of Operational Research, No. 185, pp. 1501-1524, 2008. 79

[55] T. Chang, S. Yang and K. Chang, "Portfolio optimization problems in different risk measures using genetic algorithm", Expert Systems with Applications, Vol. 36, Issue 7, pp. 10529-10537, 2009. 79

[56] A. Fernández and S. Gómez, "Portfolio selection using neural networks", *Computers & Operations Research*, Vol. 34, Issue 4, pp. 1177-1191, 2007. 61, 79

[57] F. Black and R. Litterman, "Global Portfolio Optimization", Financial Analysts Journal, Vol. 48, No. 5, pp. 28-43, 1992. 79

[58] H. Grubel, "Internationally diversified portfolios: welfare gains and capital flows", American Economic Review, Vol. 58, pp. 1299-1314, 1968. 83

[59] A. Hastings and H. Nordby, "Benefits of Global Diversification on a Real Estate Portfolio," Journal of Portfolio Management, Vol. 33, No. 5, pp. 53-62, New York, 2006. 83, 84, 97

[60] W. Bennet and C. Conan, "Risk management lessons worth remembering from the credit crisis of 2007-2009", Journal of Portfolio Management, Vol. 36, No. 3, pp. 21-44, 2010. 84

[61] J. H. Holland, Adaptation in Natural and Artificial Systems Ann Arbor, University of Michigan Press, 1975. 84

[62] J. R. Koza, Genetic Programming, on the programming of computers by means of natural selection, Cambridge, Mass.: MIT Press, 1992. 84

[63] A. Bennet, "Benefits of a Population: Five Mechanisms That Advantage Population-Based Algorithms", IEEE. Transactions on Evolutionary Computation, Vol. 14, No. 4, pp. 500-517, 2010. 84

[64] Y. Jin and J. Branke, "Evolutionary optimization in uncertain environments - A survey", IEEE. Transactions on Evolutionary Computation, Vol. 9, No. 3, pp. 303-317, 2005. 86

[65] I. Harvey, R. Manner and B. Manderick, "The SAGA cross: The mechanics of crossover for variable-length genetic algorithms", Parallel Problem Solving from Nature, Vol. 2, pp. 269-278, 1992. 88

[66] B. Hutt and K. Warwick, "Synapsing variable-length crossover: meaningful crossover for variable-length genomes", IEEE. Transactions on Evolutionary Computation, Vol. 11, No. 1, pp. 118-131, 2007. 88

[67] P. Krokhmal, M. Zabarankin and S. Uryasev, "Modeling and optimization of risk", Surveys in Operations Research and Management Science, Vol. 16, pp. 49-66, 2011. 92

[68] K. Gembat and F. Kodama, "Diversification dynamics of the Japanese industry", Research Policy, Vol. 30, pp. 1165-1184, 2001. 92

[69] R. Elsas, A. Hackethal and M. Holzhaüser, "The anatomy of bank diversification", Journal of Banking and Finance, Vol. 19, pp. 373-386, 2010. 92

[70] L. Leaven and R. Levine, "Is there a diversification discount in financial conglomerates?", Journal of Financial Economics, Vol. 85, pp. 331-367, 2007. 92

[71] W. Sharpe, "The Sharpe Ratio", The Journal of Portfolio Management, Vol. 21, No. 1, pp. 49-58, 1994. 97

Acknowledgements

I would like to acknowledge the people who have taken part in the elaboration of this thesis. Chief of these has been my advisor Professor Kotaro Hirasawa, whose wise counsel and warm aid have been invaluable.

I would like to acknowledge Professor Yoshie, Professor Fujimura and Professor Lepage for their time and their insightful comments to improved the quality of this work. Also, to Dr. Mabu for their thoughtful comments and valuable suggestions.

To my friends at Hirasawa Lab, for the precious time and valuable experiences we faced together.

List of Publications

<Journals>

(1) V. Parque, S. Mabu and K. Hirasawa, "Diversifying risk in portfolios using variable size Genetic Relation Algorithm", IEEJ Transactions on Electrical and Electronic Engineering, 2011, accepted.

(2) V. Parque, S. Mabu and K. Hirasawa, "Evolving Asset Selection by Genetic Network Programming", IEEJ Transactions on Electrical and Electronic Engineering, 2011, accepted.

(3) V. Parque, S. Mabu and K. Hirasawa, "Evolving Asset Portfolios by Genetic Relation Algorithm", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 14, No. 5, pp. 464-474, 2010.

<International Conference Papers>

(1) V. Parque, S. Mabu, and K. Hirasawa, "Guided Genetic Relation Algorithm on the Adaptive Asset Allocation", SICE International Annual Conference, Tokyo, September, 2011, accepted.

(2) V. Parque, S. Mabu, and K. Hirasawa, "Genetic Network Programming for a novel Asset Selection Model", Genetic and Evolutionary Congress, Ireland, July, 2011, accepted.

(3) V. Parque, S. Mabu, and K. Hirasawa, "Asset selection in global financial markets using Genetic Network Programming", In Proc.

of the IEEE International Conference on Systems Man and Cybernetics, pp. 677-683, Turkey, October, 2010.

(4) V. Parque, S. Mabu, and K. Hirasawa, "Robust Genetic Network Programming on Asset Selection", In Proc. of the IEEE Tencon 2010, pp. 1021-1026, Fukuoka, November, 2010.

(5) V. Parque, S. Mabu, and K. Hirasawa, "Enhancing Global Portfolio Optimization using Genetic Network Programming", In Proc. of the SICE International Annual Conference, pp. 3078-3083, Taiwan, August, 2010.

(6) V. Parque, S.Mabu, and K. Hirasawa, "Global Portfolio Optimization by Genetic Relation Algorithm", In Proc. of SICE International the Annual Conference, pp. 2567-2572, Fukuoka, May, 2009.