Università
Ca'Foscari
Venezia

Master's Degree

in Economics and Finance

LM-56 – Economics

Final Thesis

A Cumulative Prospect Theory approach for portfolio optimisation: empirical investigations using PSO algorithms

Supervisor
Ch. Prof. Martina Nardon

Assistant supervisor
Ch. Prof. Marco Corazza

Graduand
Matteo Sammartin
868007

Academic Year
2018 / 2019

Acknowledge

I wish to express the most profound gratefulness to Professor Martina Nardon, for her endless knowledge, patience, and inspiring guidance, during the process of writing this dissertation. Without your continuous support, this thesis would not have been possible.

Besides my supervisor, I wish to acknowledge with gratitude Professors Marco Corazza and Diana Barro, who helped me in drafting the MATLAB model code.

I also express my most gratitude to Greta for her constant support and continuous motivation throughout my whole academic education in the last five years. Thanks to my friends of a lifetime, with whom I kept a great relationship despite the distance, and to all classmates and flatmates, who helped me feel less alone during my stay in Venice.

Finally, I wish to thank the care and love of my family, who allowed me to move to Venice to complete my education.

Abstract

One of the most important and influential economic theories is Cumulative Prospect Theory (CPT). Developed by Tversky and Kahneman, it is accepted worldwide as an alternative theory of Expected Utility Theory, which has faced several behaviourally criticisms in the last decades. After exploring how these two theories are implemented into the selection portfolio problem, the thesis fully describes both Index-Tracking and Cumulative Prospect Theory models in terms of the optimisation problem. Next, an application of the CPT model for index tracking is derived. Since the optimisation problem becomes quite complicated, the literature recommends taking advantage of metaheuristics, i.e., the Particle Swarm Optimisation. It allows finding the optimum solution to the portfolio selection problem. The fundamental assumption is that the behaviourally based portfolio should provide better results than traditional approaches employed in passive fund management. The project takes into consideration the Dow Jones Industrial Average, a stock market index that contains 30 publicly-owned companies listed on the NASDAQ and the NYSE. The employment of metaheuristic to the behaviourally based model allows managing a large number of securities in a short time. To improve the performances of Cumulative Prospect Theory model, an analysis of the data is performed as the sentiment of the prospect investor varies. Finally, the performance of the examined models is compared.

Table of Contents

Introduction

Decision-making problems under risks are widespread in everyday life. In the business world, some people have to make decisions that involve remarkably large sums of money (macro-decisions), e.g., those taken by politicians regarding the macroeconomics; on the other hand, some micro-decisions concern a pretty tight circle of persons, such as those taken by local firms that influence only employees. Besides, people make decisions when they handle with ordinary actions of life: buying a new car, selecting the appropriate home loan, or simply deciding in which supermarket to shop.

Each choice is supported both by own insight and available information at the time. Furthermore, people use different theoretical models that help them to make choices in specific situations. Since the 1950s, several models for making decisions have developed. The most well-known investment decision is undoubtedly Modern Portfolio Theory (MPT), designed by Henry Markowitz (1952), who proposes the expected mean-variance of returns rule. According to MPT, a portfolio of assets is composed to maximise the expected return, given a specific level of risk. Furthermore, the diversification of assets must be achieved to smooth out the specific risk. Indeed, an increasing portfolio size makes each investment smaller and smaller in percentages, muting the positive or negative performance of each asset. It is necessary to highlight that Mean-Variance (MV) rule implies the right kind of diversification, which entails a diversification across securities with low return covariances. Thus, for the first time, the concept of portfolio selection is distinctly formulated and solved. Before his work, the theory states that an investor has to choose assets to maximise discounted expected returns. However, the portfolio would be composed of only one well-performed stock, which contradicts the phenomenon of diversification.

Afterward, Tversky and Kahneman (1979) introduce Prospect Theory (PT), which assumes that human beings do not always make decisions consistent with the maximisation of expected utility, as much empirical evidence stress. According to this theory, people are risk-averse whenever they assess gains and risk-seeking concerning losses. Furthermore, they are loss averse because they are more sensitive to losses than to gains of equivalent amounts. Then, they do not estimate investment opportunities in terms of final wealth, but terms of potential gains and losses compared to a reference point. Finally, they use probabilistic decision weights that are different from objective probabilities. Consequently, they overestimate low probabilities and underestimate medium-high probabilities, and they are less sensitive to variations in the probability of the middle range.

To solve some PT's drawbacks, among which the main one is the violation of the stochastic dominance criterion, Tversky and Kahneman (1992) introduce Cumulative Prospect Theory. Recent researches (Bernard and Ghossoub, 2010; He and Zhou, 2011) employ it to derive the optimal portfolio choice.

This dissertation is composed of five chapters. The first two concerning the literature of Modern and Behavioural Finance and two portfolio selection theories; the third and fourth chapters respectively introduces the application of the selected models and the possible metaheuristic approaches that may be applied to find the optimal solution to asset allocation problem; the last chapter, finally, discusses the practical project, comparing the performances of CPT model and the benchmark.

The first chapter describes Modern Finance that relies on Expected Utility Theory, which was dominant in the last century. Then, after presenting the main exceptions arose in the second half of the 1900s, many schools of thought started to focus on different methods that take into consideration human behaviour, generating Behavioural Finance. Later, the chapter thoroughly discusses both works of Tversky and Kahneman, i.e., Prospect Theory and Cumulative Prospect Theory, giving a full explanation of the models, value, and weighting functions. Finally, it provides a list of the main heuristics and biases, describing how they influence decision-making process under risk.

The second chapter covers the portfolio selection theory, comparing Modern Portfolio Theory and Behavioural Portfolio Theory (BPT). The latter is introduced by Shefrin and Statman (2000), who develop a model based on PT and Security-Potential/Aspiration theory. BPT uses mental accounting, proposed by Thaler (1985) to coin a new model of consumer behaviour. Shefrin and Statman (2000) submit two versions: one for a single mental account and the other for multiple mental accounts.

The third chapter presents two models that will be implemented in Chapter 5: Index-Tracking and CPT models. A technical description of the structure and the implementation of each model in the optimisation problem are provided. At the end of the chapter, the CPT model for index tracking is defined. It realises that the optimisation problem becomes complicated to solve, and the only solution is to take into consideration a metaheuristic approach.

The fourth chapter examines three main metaheuristics approaches that help to find the optimum solution in optimisation problems: Particle Swarm Optimisation, Differential Evolution algorithm, and Genetic algorithms. These methods belong to evolutionary algorithms, widespread in the entire world for their capacity to find suitable solutions to computationally hard problems, even with incomplete information. There is a technical description that, along with figures and pseudo-code, allows to understand better how they work.

The fifth and last chapter treats the application of the CPT model for index tracking. The dissertation assumes that the model should have better performance than the passive management strategy. The project takes into consideration the Dow Jones Industrial Average, a stock market index that contains 30 publicly-owned companies listed on the NASDAQ and the NYSE. The use of metaheuristic to the behaviourally based model allows managing a large number of stocks in a short time. To improve the performances of Cumulative Prospect Theory model, an analysis of the data is performed as the sentiment of the prospect investor varies. Then, the performance of the examined models is compared with the benchmark through three relevant indicators: the mean absolute error (MAE), the root mean squared error (RMSE), and the information ratio (IR).

Finally, the conclusion exposes comments regarding the findings obtained, and it assesses if the validity of the starting assumption.

Chapter 1 - The dispute between Modern and Behavioural Finance

"Is it enough to assume people can be approximated by homo economicus, or do we need psychologically grounded assumptions?" (Cartwright, 2014, p. 14)

In the past century, Modern Finance was considered the theory par excellence, able to explain the events concerning financial economics. After exploring all available choices, individuals take rational decisions to maximise their optimal level of wealth. The rationality is a key point of this theory: human beings are always cautious and make logical choices that allow them to reach the highest personal utility. In these circumstances, emotions do not meddle in the decision-making process.

However, Modern Finance is based on unrealistic strong assumptions that cannot face empirical evidence and anomalies that emerge in the second half of the 20th century. Moreover, this theory is hardly applicable to the real world.

In this manner, a new school of thought rejecting the "homo economicus" assumption comes to light. A key assumption that is at the basis of Behavioural Finance. The main contributors of this research field are Daniel Kahneman e Amos Tversky, who, questioning how human take choices, they explained the decision-making process under risk.

In this chapter, the debate between Modern and Behavioural Finance is discussed. After inspecting Expected Utility Theory (EUT) and its developments in Sections 1.1, Section 1.2 investigates anomalies and empirical evidence in contrast with EUT. Section 1.3 introduces Behavioural Finance. Section 1.4 analyses in detail Prospect Theory and Cumulative Prospect Theory. Section 1.5 concludes the chapter with a review of the main heuristics in the decision-making process under risk: representativeness, availability and anchoring.

1.1 Expected Utility Theory

Only a few economists would disagree that EUT is dated back to Daniel Bernoulli in 1738, to the times when he was examining the St. Petersburg paradox related to the decision theory. This problem consists in a player that doubles the initial stake (set at the beginning of the game equal to 2 of some monetary unit) if the fair coin toss turns out to be "head". The goal is to compute the fair price that a player has to pay for entering the game. The possible payoff is 2^x, where x is the number of flips producing a head. It is easy to prove that the expected monetary payoff is infinite, albeit most players would just pay a small amount to enter it. Because of this gamble, Bernoulli recognizes that the value of the game is, in general, not equal to expected monetary value. He is the first economist who makes a clear distinction between "price" and "utility". Price is objective, equal for everybody and it depends on the thing itself; while the utility is the subjective value that varies among individual because it depends on personal estimates. Although Bernoulli had solved the St. Petersburg paradox, its theory did not find much favour among modern scholars due to the fact that it is based on cardinal utility: individuals assign a numerical value to the utility of their own choosing. Of course, it implies that the comparison between people choices is not possible (Bernoulli, 1954).

It has to pass at least two centuries before EUT is formally deepened and widened. John von Neumann and Oscar Morgenstern (1944) showed that the optimal decision is to maximise the expected value of the utility function, subject to an appropriate level of risk (probability). The theorem states that individual preferences between different prospects may be represented by a utility function in the following way:

$$V(q) = \sum_{i=1}^{n} p_i u(x_i), \tag{1.1}$$

where p_i is the probability of the outcome x_i, and n is the total number of prospects considered. Neumann and Morgenstern (1944) proved the expected utility hypothesis using compelling axioms on preference, which can be justified as sound principles of rational choice. The three core axioms are *continuity*, *ordering*, and *independence*.

Let a, b, c be different prospects, in other words, a list of events with an associated probability, and let p represent probabilities. The *continuity* axiom implies that for all prospects, if $a \succcurlyeq b$[1] and $b \succcurlyeq c$, then there exists a probability $p \in [0,1]$ such that: $p(a) + (1-p)(c) \sim b$.

The *ordering* axiom needs of both completeness and transitivity. The first allows only one of the following possible preferences: either $a \succcurlyeq b$ or $b \succcurlyeq a$ or both. Transitivity states that if $a \succcurlyeq b$ and $b \succcurlyeq c$, then $a \succcurlyeq c$.

Lastly, the *independence* axiom requires that: if $a \succcurlyeq b$ then $p(a) + (1-p)(c) \succcurlyeq p(b) + (1-p)(c)$. It supposes that, given a preference between two prospects, the adding of a third part on both sides does not change the initial preference.

If these axioms hold, preferences may be expressed by Equation (1.1), which helps to make the comparison between two different prospects, e.g. a and b, and choose option a if and only if the value $V(a)$ is equal or greater than $V(b)$.

1.1.1 Developments based on Expected Utility Theory

EUT led to a breakthrough in financial economics and dominated for at least four decades. The academic community was focused on analysing and exploring the mathematical probabilistic concept and developed new theories and optimisation models that represented key influences for the following years.

Referring to EUT, the well-known American economist Henry Markowitz (1952a) designed Modern Portfolio Theory (MPT), proposing the expected Mean-Variance (MV) of returns rule. MPT helps investors to take into consideration the best combination of assets, aiming either to maximise the portfolio expected returns fixed a specific level of risk, or to minimise the portfolio risk, given a level of expected returns. In this framework, risk is represented by variance. Furthermore, the diversification of assets must be achieved to smooth out the specific risk. Indeed, an increasing portfolio size makes each investment smaller and smaller in percentages, muting the positive or negative performance of each asset. It is important to highlight that MV-rule implies the right kind of diversification, which entails a diversification across securities with low return covariances. Thus, for the first time, the concept of portfolio selection was distinctly formulated and solved. Before his work, the theory stated that an investor has to choose assets to maximise discounted expected returns. However, the

---

[1] The notation $\succcurlyeq$ denotes the relation "is weakly preferred to".

portfolio would be composed of only one well-performed stock, which contradicts the phenomenon of diversification.

To better understand the portfolio selection problem under MPT, the MV-rule is discussed. People need a criterion that divides all investment choices into two subsets: an efficient subset and another inefficient. Markowitz (1952a) uses the MV-rule to identify as "efficient portfolios" those characterized either by the minimum variance at any given expected returns or, vice versa, the maximum expected returns at any given level of variance.

In mathematical terms, consider X and Y two assets with their respective mean $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ and variance $Var(X)$ and $Var(Y)$. $X \succcurlyeq_{MV} Y$[2] if and only if:

- $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $Var(X) \leq Var(Y)$.

Of course, one of the inequalities must be satisfied in the strict form; otherwise, investors would be indifferent in choosing X or Y ($X \sim_{MV} Y$). In other words, this criterion asserts that $X \succcurlyeq_{MV} Y$ in two situations: if the mean of X is higher than the mean of Y and the variance of X is lower or equal to the variance of Y; if the mean of X and Y are equivalent, but the variance of X is lower than the variance of Y.

The MV-rule is derived assuming that returns are normally distributed and considering the quadratic utility function (Feldstein, 1969; Hanoch and Levy, 1969).

The Markowitz model influenced the works of William Sharpe (1964) and John Lintner (1965), who proposed the capital asset price model (CAPM), which is the main pillar of Modern Finance and it is still even today used to evaluate an individual asset or portfolio.

In the 1970s, EUT reached its peak of dominance among academic researches. Assuming the rationality of human behaviour and efficient markets imply that stock prices reflect all available information and they change only for valid and sensible information since individuals are price-taker (Shiller, 2003). Eugene Fama (1970) published the article "Efficient Capital Markets: A Review of Empirical Work", maintaining that asset prices incorporate all available information and so stocks are traded at their fair value. In addition, Robert Merton (1972) expanded the CAPM developing the intertemporal capital asset pricing model (ICAPM), in which individuals are able to hedge against drops in consumption and against changes in investment opportunities. Finally, Robert Lucas (1978) examined a

2 This formula means that X dominates Y in the mean-variance sense.

theoretical method that is able to construct equilibrium prices, under the assumption that prices incorporate all information and rational behaviour.

1.2 Anomalies and empirical evidence against Expected Utility Theory

In the 1950s, several economists analysed in depth the determinants of individual choice behaviour, discovering more and more empirical evidence against EUT's axioms, especially independence axiom. A renowned example is provided by the Allais paradox (1953), which consists of a pair of choice problems that exhibits inconsistency of participants behaviour with EUT. Each problem is composed of two gambles (A and B, C and D).

Example 1.1. Consider the following lotteries:

Gamble A		**Gamble B**	
Outcome	Probability	Outcome	Probability
\$1 million	100%	\$5 million	10%
		\$1 million	89%
		nothing	1%
Gamble C		**Gamble D**	
Outcome	Probability	Outcome	Probability
\$1 million	11%	\$5 million	10%
nothing	89%	nothing	90%

Participants of the experiment were expected to choose one of the two choices before proceeding to the second problem. Looking carefully at the first lottery, it is possible to notice that there is a common outcome of \$1 million with probability of 89%. The second lottery is derived by subtracting that specific quantity from both prospects. This situation is so-called "common consequence" effect.

Nonetheless, empirical evidence shows that people opt to choose gamble A, lured by becoming a millionaire, and gamble D, because the possible winning is rather higher, and probabilities are quite similar. In other words, participants display a risk-aversion behaviour

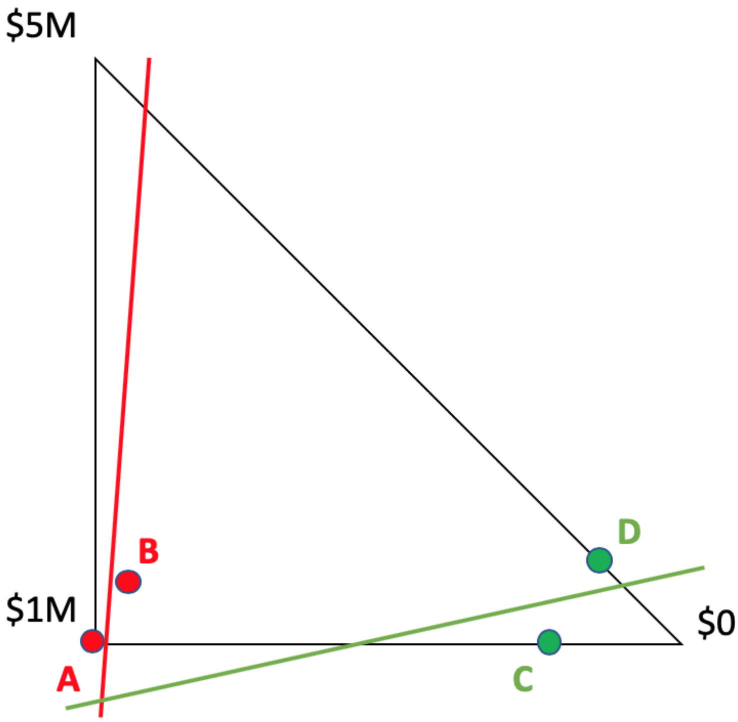


Figure 1.1. Allais paradox explained by MM-triangle
Source: personal processing.

for the first gamble and a risk-seeking behaviour for the second one. Of course, these changes in the behaviour are in contrast with EUT, specifically with independence axiom. The latter does not allow a behavioural change in people: they are either risk-averse or risk-seeking, no matter which situation they are facing. In the preceding example, independence axiom supports either A-C or B-D choices.

To better understand the reasons why independence axiom is violated, the Marschak-Machina probability triangle (Marschak, 1950; Machina, 1982) is introduced. It is a graphical technique that allows representing a set of gambles composed of three different outcomes $x_3 > x_2 > x_1$, with probabilities p_3, p_2 and p_1 respectively. In a right triangle, each angle symbolizes a possible outcome of the experiment, in particular: x_3 is placed on the highest corner, x_2 is placed on the right angle and x_1 is placed on the residual one. Moreover, the horizontal axis represents p_1, the vertical axis represents p_3, and the oblique axis is p_2. If a gamble is characterized by a combination of three possible outcomes (i.e. gamble B), its representation will be inside of MM-triangle; if it is a mix of two possible outcomes, it will be drawn on a side (i.e. gamble C and D). According to the independence axiom of EUT, indifference curves have to be upward sloping linear and parallel. This means that an individual must prefer either the "safer option" or the "riskier option" in both problems.

The Allais paradox is shown in Figure 1.1. The coloured segments are the indifference curves: steeper indifference curves mean risk-aversion, while flatter ones mean risk-seeking. As it is illustrated, the two curves are not parallel because people change their attitude toward risk. Thus, independence axiom is violated as Allais tested.

In the following years, Daniel Ellsberg (1961) theorized a new paradox that undermined EUT independence axiom more and more. He was analysing how individuals behave toward ambiguity. Suppose to have an urn with the following characteristics: 30 red balls and 60 black and yellow balls, whose proportion is not known. A ball will be drawn at random, and the following two choice problems are asked to the participants:

Example 1.2. Consider the following lotteries:

Gamble A	**Gamble B**
You win $100 if the ball drawn is red, otherwise nothing	You win $100 if the ball drawn is black, otherwise nothing
Gamble C	**Gamble D**
You win $100 if the ball drawn is red or yellow, otherwise nothing	You win $100 if the ball drawn is black or yellow, otherwise nothing

The results of this experiment show that people tend to prefer the gamble A to B, and D to C; less frequent happens that gamble B is preferred to A, and C to D. For both rejected options, the likelihood of winning is uncertain, ambiguous, and so people are reluctant to pick these gambles. They are said to be ambiguity-averse. According to EUT, if an individual prefers the gamble A to B, it means that he/she believes that the likelihood of drawing a red ball is greater than drawing a black one. It follows that he/she will opt for gamble C rather than D because the probability of drawing a red or yellow ball remains greater than drawing a black or yellow one.

A further EUT drawback is that it assumes human rationality, which implies consistency and coherence in the decision-making process. Therefore, it would possible to support that changes of the frame should not affect preference between options. However, human perception often fails to recognise frameworks with the same meaning but structured differently. Amos Tversky and Daniel Kahneman (1981) run an experiment in which two groups were supposed to answer to the following problem:

Example 1.3. Consider the following experiment (Tversky and Kahneman, 1981, p. 453):

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
>
> - if Program A is adopted, 200 people will be saved.
> - if Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.
>
> A second group of respondents was given the cover story of the previous problem with a different formulation of the alternative programs, as follows:
>
> - if Program C is adopted, 400 people will die.
> - if program D is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Results are quite surprising: 72% of first group subjects preferred Program A then B, while 22% of second group subjects opted for Program C rather than D. Tversky and Kahneman explained these findings claiming that people are risk-averse if choices implicates gains (first group), and they are risk-seeking when choices involve losses (second group).

In addition to these several empirical experiments, many studies from different fields brought to light further inconsistencies. In the 1980s, academic researchers expanded their econometric knowledge, which helps them to examine in details stock prices, dividends and earnings. Some excesses of volatility, unexplained by the efficient market, were discovered: the January effect and the weekend effect (Shiller, 2003). The American economist Donald Keim (1983) studied, month-by-month, the empirical link between inexplicable returns and common stocks belonging to NYSE and AMEX. He proved that the daily returns mean in January is greater than the other months, and the relationship between firm size and returns is negative and more evident in January: in other words, small stocks outperform larger stocks. Furthermore, Siegel and Coxe (2002) reported other calendar anomalies that efficient markets could not explain, i.e. stocks generally outperform on Friday than Monday, they do better in the first days of each month, and they do remarkably well in the days prior to a holiday, especially December 31.

1.3 The rise of Behavioural Finance

"There is still every reason to think that, while markets are not totally crazy, they contain quite substantial noise, so substantial that it dominates the movements in the aggregate market" (Shiller, 2003, p. 90).

In this sentence, Shiller affirms that people did not go insane and the market's behaviour may be explained by a new theory different from Modern Finance. The 1990s is the period of the so-called "blossoming of Behavioural Finance" (Shiller, 2003, p. 90) because economists began to focus on economic models that incorporate human psychology, analysing their effects on financial markets.

Nonetheless, a remarkable work regarding investors behaviour is traced at the beginning of the 20th century. George Selden (1912) asserts that some minor stock fluctuations depend on the actions of some traders who operate, "not on the basis of facts, nor on their own judgment as to the effect of facts on prices, but on what they believe will be the probable effect of facts or rumours on the minds of other traders" (Selden, 1912, p. 109). Even back then, there were some doubts that market does not strictly follow the law of supply and demand but can be subject to changes due to psychological human behaviour.

To prove the previous assertion, it is possible to recall the phenomenon of speculative "bubbles", in which a specific asset value enormously increases and then it dramatically drops. This process is initially driven by exorbitant expectations of future growth, which let price goes up, and first investors make successes; the word-of-mouth enthusiasm brings other investors realise this successful opportunity and start to invest in it. Of course, this phenomenon is not sustainable because it is pursued only by higher future expectations, and at any time the bubble may burst, causing an over-selling asset and falling price (Shiller, 2003). The most impressive examples are the tulip mania in Holland in the 1630s, the dot-com bubble in the 1990s and, more recently, the 2000s commodity bubble.

The evidence described above is also sustained by cognitive psychology, which reveals that people have subjective attitudes toward probabilities. Malcolm Preston and Philip Baratta (1948) run an experiment, in which they auctioned several gambles constituted by a low probability of winning and the remaining probability to gain nothing. They recognised that there is a linear relationship between people keen to spend money on gambles and the possible winnings, but a nonlinear relationship with the probability related to the outcome. They presumed that participants misperceived objective probabilities because people tend to understate large probabilities and overstate small probabilities. Similar works (Quiggin, 1993) proved their early conjectures regarding perceived probabilities. Moreover, Tversky and Kahneman (1974) discovered that if people have to guess which the occupation of someone is, of whom personality and hobbies are known, they tend to choose unusual works that fit with the description as much as possible. Rational individuals would opt for ordinary occupations since more people do them. Daryl Bem (1965) affirms that individuals have the inclination to misaddress bad luck to events that do not confirm their actions and ascribe events consistent with their actions to their own skills.

Thus, financial markets do not depend exclusively on mathematical or statistical laws, since individuals' actions are not predictable. Econometric models are not able to fully reproduce the variables that determine the price of a stock. These statements are also claimed by Gerald Loeb (1957), who reports that market values are based not only on income statements and balance sheets but also on emotions, such as hope, fear, greed, ambition and many others. Loeb is not the only one who takes into consideration such emotions. In the portfolio selection problem, Hersh Shefrin (2002) declares that investors experience a wide variety of emotions in different situations: during the ponderation of the alternatives, the choice of how much risk to bear, monitoring shares held, or while he evaluates if it is time to change strategy or not. Disagreeing the folklore according to which fear and greed influence financial markets, Shefrin maintains that fear and hope play crucial roles: fear encourages investors to pay attention to events, especially unfavourable ones, while hope helps to focus on favourable events.

In the last decades, many neoclassical economists change their inclination about rationality and begin to consider the emotional human side. The interaction between psychology and economics becomes stronger and stronger, leading to the creation of a new interdisciplinary research field that takes the name of Behavioural Finance. In the beginning, it arises with the need to explain those paradoxes and anomalies that were described previously; then, economists attempt to give consistency and a sound structure to their argumentations.

1.4 Prospect Theory

"I think one of the major results of the psychology of decision making is that people's attitudes and feelings about losses and gains are really not symmetric. So, we really feel more pain when we lose $10,000, than we feel pleasure when we get $10,000" (Kahneman, 2013).

Prospect Theory (PT), developed by Tversky and Kahneman (1979), takes into consideration the subjective preferences and it is able to provide a useful explanation of behavioural biases discovered at that time. Kahneman and Tversky managed to clarify the reflection effect and give a potential solution. This phenomenon leads people to evaluate gains and losses in a different way, determining a dramatic change of risk preference among similar gambles. They discovered that people tend to be more risk-averse when choosing gambles that involve gains and risk-seeking when choosing between losses. A typical example of this effect is proposed:

Example 1.4. Consider the following two lotteries:

Gamble A		**Gamble B**	
Outcome	Probability	Outcome	Probability
Win $240	100%	Win $1,000	25%
		Win nothing	75%
Gamble C		**Gamble D**	
Outcome	Probability	Outcome	Probability
Lose $750	100%	Lose $1,000	75%
		Lose nothing	25%

The majority of participants (84%) opts for the sure winning in the first problem (risk-averse), while 87% of participants choose the gamble, hoping to lose nothing (risk-loving). PT takes into consideration this attitude and states that people do not maximise their expected utility, but they are more sensitive to losses than gains. In each situation, individuals evaluate potential gains against potential losses, so they do not achieve the maximization of the final wealth as in EUT.

PT is originally thought for solving simple gambles with monetary outcomes and known probabilities. It entails two different phases: editing and valuation. In the editing stage, individuals analyse the available prospects, whose outcomes and probabilities are transformed to obtain a simpler representation of the whole problem. In this way, it is easier to make consequent evaluation and choice. The editing procedure implies the following main operations (Tversky and Kahneman, 1979):

i. *coding*: as discussed previously, the decision maker does not consider outcomes as final states of wealth, but as gains and losses. To this aim, the reference point is specified to assess if outcomes are considered gains or losses. Setting the reference point is not objective, it depends on the formulation of the problem and the individual expectations;
ii. *combination*: the decision maker combines those probabilities whose respective results are identical, making it faster to read the problem;
iii. *segregation*: some gambles include a risk-free element that is separated from the risky component;
iv. *cancellation*: the decision maker analyses all choices and detects the common components (either for outcomes or probabilities) between two or more prospects, and he/she eliminates them. This operation allows focusing on the main differences among prospects. To make *cancellation* clearer, suppose there are two prospects (500, 0.10; 200, 0.60; -100, 0.30) and (500, 0.10; 400, 0.60; -300, 0.30). They may be reduced by this operation to a choice between (200, 0.60; -100, 0.30) and (400, 0.60; -300, 0.30);
v. *simplification*: this procedure may include the outcomes and probabilities rounding, or the rejection of highly improbable outcomes;
vi. *detection of dominance*: the decision maker identifies if there are dominated prospects to discard them from the evaluation.

There are many drawbacks about the prospect editing phase: edited gambles may have a different interpretation if the activities above mentioned are applied in a different order; in addition, it may happen that simplification and cancellation eliminate small differences between choices (Tversky and Kahneman, 1979).

Subsequently, individuals need to evaluate the edited prospects determining their overall value V, which is expressed in terms of probability-weighted function $w(p)$ and the subjective value of the respective outcome $v(x)$, that corresponds to the departure from the reference point prefixed in the editing phase. Suppose two hypothetical gambles x and y,

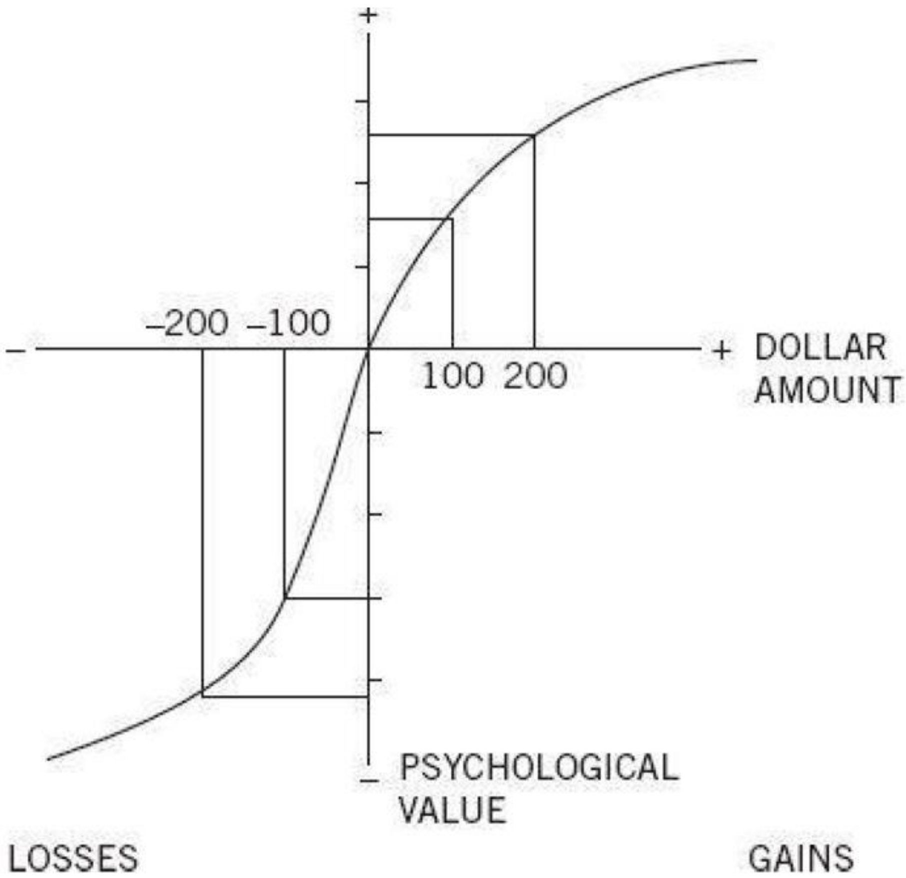


Figure 1.2 Psychological value of gains and losses
Source: Kahneman (2011).

their respective probabilities and $q = 1 - p$, the overall value of a prospect may be represented as:

$$V = \mathrm{w(p)}v(x) + \mathrm{w(q)}v(y), \tag{1.2}$$

where $w(0) = 0$, $w(0) = 1$ and $v(0) = 0$ (Tversky and Kahneman, 1979).
Under PT, Kahneman and Tversky aim to give their definition of value function assuming that "the carriers of value are changes in wealth or welfare, rather than final states" (Tversky and Kahneman, 1979, p. 277). An illustrative representation is exhibited in Figure 1.2. They suppose that the value function is convex below the reference point ($v''(x) > 0$, for $x < 0$) and concave above it ($v''(x) < 0$, for $x > 0$) [3]. Indeed, Kahneman and Tversky hypothesize

[3] Empirical evidence (exercise 1.4) shows that people is risk-averse when dealing with gambles involving gains, and risk-loving with gambles involving losses. Risk-aversion is represented by a concave curve because, as the payoff for the gambles increases, the slope of the utility function

that the marginal value of gains and losses commonly declines with their size. They believe that the difference between a gain (or loss) of 50 and one of 100 is perceived as greater than the difference between a gain (or loss) of 10,050 and one of 10,100. This assumption is the so-called *diminishing sensitivity* principle. Lastly, it is possible to notice that the curve is not symmetric and its slope changes at the reference point, which is zero. The reason is that losses appear worse than gains: the loss of an amount of money has a higher psychological value than the gain of the same sum. For this reason, people would not usually bet on symmetrical gambles (x, 0.50; -x, 0.50), considering them unappealing.

Even if PT was a revolutionary model at the time, it incurred two main problems concerning stochastic dominance and its application in the real world. According to dominance, if a prospect is better than another, the dominant prospect should be chosen. This concept is generally observed by PT in "transparent situations"[4], and it is often violated in "non-transparent ones"[5], when problems-choice are framed and stochastic dominance is not clear (Kahneman and Tversky, 1986). Mark Machina (1982, p. 292) stated that PT is "unacceptable as a descriptive model of behavior toward risk" since it violates first order stochastic dominance. His criticism was seen as trivial because empirical evidence supported PT (Kahneman and Tversky, 1986). Moreover, PT could not take into consideration prospects with a large number of outcomes, so it was hardly relevant in real-situations (Tversky and Kahneman, 1992).

In the "Advances in Prospect Theory: Cumulative representation of uncertainty", Tversky and Kahneman (1992) stepped forward and updated PT introducing Cumulative Prospect Theory (CPT), which faces the two drawbacks described above.

1.4.1 *Cumulative Prospect Theory*

Many theorists, including Quiggin (1982) and Schmeidler (1989), were able to solve successfully the PT's drawbacks, proposing the rank-dependent or cumulative functional. At

decreases. Similarly, risk-seeking is represented by a convex curve because, as the payoff for the gables increases, the slope of the utility function decreases.

4 To explain what "transparent situations" means, suppose there are the following two lotteries: (2, 0.90; 45, 0.06; -10, 0.03; -20, 0.01) and (2, 0.90; 45, 0.06; -5, 0.03; -20, 0.01). The latter lottery dominates the former one because all its outcomes are at least as appealing as the outcomes of the first lottery.

5 To better understand what "non-transparent situations" means, consider the example 1.4. The participants of the experiment have a strong preference for prospects A and D over B and C. Nonetheless, the rejected combination surprisingly dominates the favored one. The problem is structured in such a way that stochastic dominance is not evident.

a later stage, Tversky and Kahneman (1992) proposed Cumulative Prospect Theory (CPT), expanding their preceding work. First of all, CPT satisfies first-order stochastic dominance[6] and it is applicable to any finite prospect, both probabilistic and uncertain ones. It also allows to weight gains and losses in different ways, going beyond the original version according to which $w^{+}(p) = w^{-}(p)$[7]. Instead of transforming probabilities themselves, CPT imposes that cumulative and decumulative probabilities are transformed respectively for losses and gains. As a result, the weighting function displays the typical inverse-S shape.

The main idea of CPT is to implement Quiggin's rank-dependent representation distinctly to gain and losses and then sum the two consequent evaluations. Two different weighting functions are defined: w^{+} for probabilities related to gains and w^{-} for probabilities related to losses. As a result, there are distinctive attitudes toward probability for gains and losses.

Suppose an unknown prospect f is a function from a measurable set of states of nature S into a set of consequences (or outcomes) X, which designates each state of an outcome. To determine the cumulative function, Tversky and Kahneman (1992) sort all prospects' outcomes in increasing order. Then, a prospect f is described as a pairwise series (x_i, A_i) that produces x_i if A_i transpires, where $x_i > x_j$ if and only if $i > j$ and A_i is a partition of S. Positive subscripts indicate positive outcomes, negative subscripts indicate negative ones, and the zero subscript indicate the neutral outcome. The positive part of f, expressed by f^{+}, results to be $f^{+}(s) = f(s)$ if $f(s) > 0$, zero otherwise. On the other hand, f^{-} is the negative part of f and it is defined similarly.

According to CPT, there is a strictly increasing function $v: X \to \mathbb{R}$, such that $v(x_0) = v(0) = 0$, and capacities W^{+} and W^{-}[8], for $f = (x_i, A_i)$, $-m \leq i \leq n$ (Tversky and Kahneman, 1992):

[6] Let F_X and F_Y be two cumulative distribution functions respectively of X and Y, which correspond to returns of two portfolios. X dominates Y ($X \succ_{FSD} Y$) in the first dominance fashion if: $F_X(z) \leq F_Y(z)\ for\ all\ z$. An important relationship links the first order stochastic dominance criterion and the expected utility dominance. If $X \succ_{FSD} Y$ holds, then: $\mathbb{E}[U(X)] \geq \mathbb{E}[U(Y)]$ for all increasing U (Levy, 1992).

[7] w^{+} is a weighting function set for probabilities related with gains, while w^{-} is a weighting function set for probabilities related with losses. They both are strictly increasing functions mapping from $[0,1]$ to $[0,1]$, with $w^{+}(0) = w^{-}(0) = 0$ and $w^{+}(1) = w^{-}(1) = 1$.

[8] Tversky and Kahneman (1992) use the term *capacity* to indicate a nonadditive set function, which generalizes the traditional concept of probability. Denoted by W, it attributes to each $A \subset S$ a value $W(A)$ such that $W(\phi) = 0$, $W(S) = 1$ and $W(A) \geq W(B)$ when $A \supset B$.

$$V(f)=\sum_{i=-m}^{0}\pi_i^- v^-(x_i)+\sum_{i=0}^{n}\pi_i^+ v^+(x_i), \tag{1.3}$$

where $\pi^+(f^+)=(\pi_o^+,\ldots,\pi_n^+)$ and $\pi^-(f^-)=(\pi_{-m}^-,\ldots,\pi_0^-)$ are respectively the decision weight functions for gains and losses.

If the probability distribution $p(A_i)=p_i$ generates the prospect $f=(x_i,A_i)$, the latter can be seen as a probabilistic prospect (x_i,p_i). As a result, decision weights are represented by:

$$\begin{gathered}\pi_{-m}^- = w^-(p_{-m}),\\ \pi_i^- = w^-(p_{-m}+\cdots+p_i)-w^-(p_{-m}+\cdots+p_{i-1}), \qquad 1-m\le i\le 0,\\ \pi_i^+ = w^+(p_i+\cdots+p_n)-w^+(p_{i+1}+\cdots+p_n), \qquad 0\le i\le n-1,\\ \pi_n^+ = w^+(p_n),\end{gathered} \tag{1.4}$$

where w^+ and w^- are strictly increasing functions mapping from $[0,1]$ into $[0,1]$, with $w^+(0)=w^-(0)=0$ and $w^+(1)=w^-(1)=1$.

Testing preference-choices from laboratory experiments, Tversky and Kahneman (1992) were able to find parameters that allowed to explain the observed behaviour. The estimated model has the following structure:

$$v(x)=\begin{cases}v^+(x)=x^\alpha & x\ge 0\\ v^-(x)=-\lambda(-x)^\beta & x<0,\end{cases} \tag{1.5}$$

$$w^+(p)=\frac{p^\gamma}{(p^\gamma+(1-p)^\gamma)^{1/\gamma}}, \qquad w^-(p)=\frac{p^\delta}{(p^\delta+(1-p)^\delta)^{1/\delta}}, \tag{1.6}$$

where x is an outcome, the parameters α and β define the curvature of the value function v, while λ represents the degree of loss aversion and it is equal to 1 if the decision-maker is loss-neutral. Equation (1.5) is strictly increasing and continuous, it is convex below the reference point and concave above it, and $v'(x)<v'(-x)$ for $x\ge 0$, meaning that the value function is less steep for gains than losses. These assumptions imply that "losses loom larger than gains" and the principle of *diminishing sensitivity*. Then, Tversky and Kahneman applied the Equation (1.6) as the rank-dependent weighting function, which allows weighting

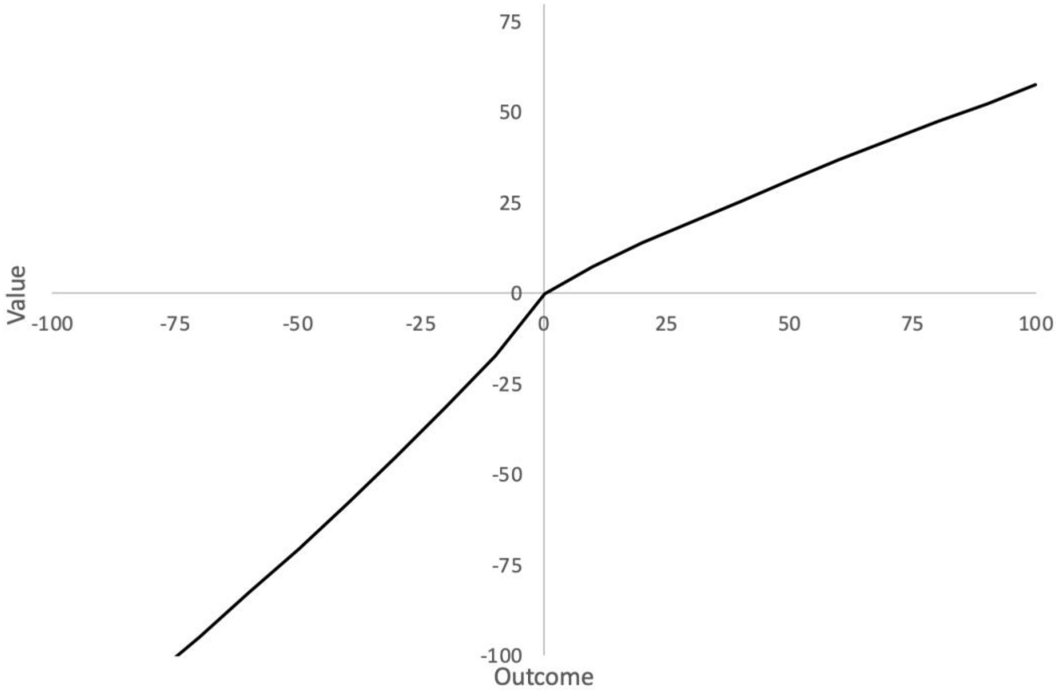


Figure 1.3 The value function (1.5)

positive and negative outcomes in different ways. According to their findings, "the shape of the weighting functions favors risk seeking for small probabilities of gains and risk aversion for small probabilities of loss, provided the outcomes are not extreme" (Tversky and Kahneman, 1992, p. 306).

The value and weighting functions are illustrated respectively in Figure 1.3 and Figure 1.4, taking into consideration that parameters estimated by the two economists are $\alpha = 0.88$, $\beta = 0.88$, $\lambda = 2.25$, $\gamma = 0.61$, and $\delta = 0.69$. The parameters α and β are consistent with the diminishing sensitivity principle while setting $\lambda = 2.25$ means that a prospect will be chosen if and only if the investor hopes to gain more than twice as large as the loss. The value function has an S-shape, and it highlights the different change in slopes in the reference point, equal to zero.

Figure 1.4 displays the inverse-S shaped estimated probability-weighting functions for gains (green curve) and losses (blue curve). Despite different coefficients, both functions share a similar shape, even if w^- is flatter than w^+ (i.e., $\delta > \gamma$), overweighting small probabilities and underweighting medium-high probabilities. As a result, people are less sensitive to differences in probability in the middle range. Besides, the diminishing sensitive principle implies that the weighting functions are concave close to 0 and convex close to 1. The 45-

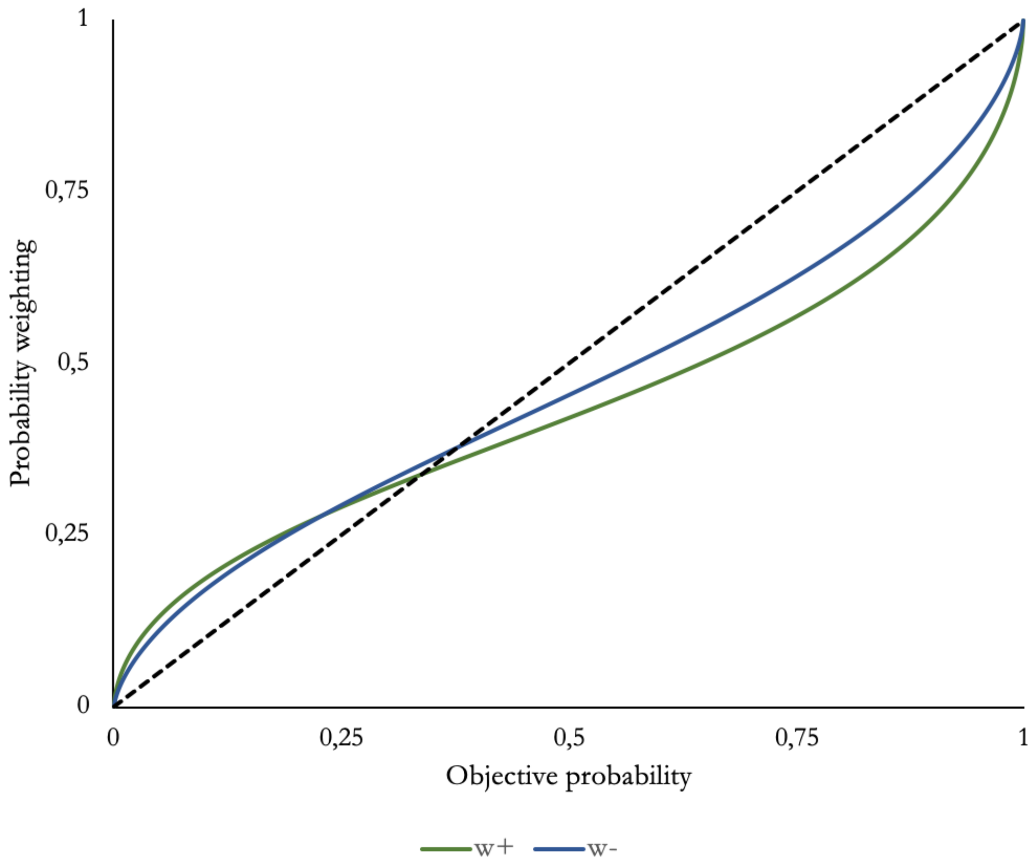


Figure 1.4 Probability weighting under CPT ($\alpha = 0.88$*,* $\beta = 0.88$*,* $\lambda = 2.25$*,* $\gamma = 0.61$*,* $\delta = 0.69$*)*

degrees line is crossed by the curves in two different points (0.33 for gains and 0.37 for losses), and this is the only notable difference.

1.5 Heuristics and biases in the decision-making process under risk

Tversky and Kahneman (1974) state that people trust in some heuristic principles which help them to reduce the complexity of decision-making processes under risk, appointing outcomes and probabilities to easier operations. Thaler and Sunstein (2008) define such processes as rules of thumbs: people cannot reflect and analyse everything in their lives, so they often recall some mental shortcuts that help them to reach an immediate solution, even if it is not perfect or optimal. Heuristics tend to monopolise the decision-making, and sometimes they lead to systematic biases.

In the following sections, the three main heuristics – representativeness, availability, and adjustment and anchoring – and their respective biases are discussed. In the last section, further heuristics will be discussed. It is important to be aware of them because they may have a dramatic impact on portfolio management. A limited or schematic point of view of problems leads to answers as fast as they are impartial and incomplete.

1.5.1 Representativeness

"Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail. How do people assess the probability that Steve is engaged in a particular occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely?" (Tversky and Kahneman, 1974, p. 1124).

This heuristic describes the behaviour for which people tend to judge the probability that an event X (Steve) belongs to process Y (specific occupation). The previous example was tested by Tversky and Kahneman, and participants used to order occupations by similarity, not by probability. This way of thinking leads to severe errors since representativeness is not affected by those factors that influence probability. To better understand this difference, consider the following concept: the "prior probability of the outcomes", one of the main determinants of probability. At the time, there were more farmers than librarians so, according to this principle, people should choose "farmer" as a solution to the problem. Nevertheless, individuals make their decisions based on the stereotypes of farmers and librarians, and, evaluating probability by representativeness, the majority of them used to choose "librarian". Moreover, further studies show that people use probabilities if and only if they have no further information, making no mistake (Tversky and Kahneman, 1974).
Concerning the finance world, representativeness guides investors to irrationally behave, impacting the stock prices and determining large price movements. Fuller (1998) illustrates how representativeness leads investors to wrongly believe that they have already handled information in the correct way before they make the choice. Generally speaking, Zhao and Fang (2014) recognise two different kind of representativeness: the horizontal representation bias leads investors to gather together similar stocks, judge them according to alike stocks trend and make predictions using the same rules; on the other hand, the vertical

representation bias entails that people evaluate and forecast an asset according to its historical data, overlooking the other related stocks.

1.5.2 Availability

Generally, people use availability heuristic when they are asked to evaluate the likelihood of an event. In these situations, if they easily remind these occurrences among either one's acquaintances or news read in previous days, they believe that such an event has a higher likelihood (Tversky and Kahneman, 1974). A notorious example is a comparison between homicides and suicides: people mistakenly believe that the first cause of death is more frequent than the second one (Thaler and Sunstein, 2008).

It is easily discerned that availability is influenced by several factors, which may lead to predictable biases: if an individual is *"familiar"* to a specific event (i.e., heart attack), he will believe that it has a higher frequency. The *"salience"* contributes to making more vivid in mind an event experienced at first hand: for example, the shock of seeing a burning house influences more the subjective perception rather than reading such event in the newspaper. Moreover, recent news proves to be more available than the older one (Tversky and Kahneman, 1974).

Availability heuristic also affects financial markets: Franklin Templetion Investment (2012) maintained that, after Global Financial Crisis (GFC), investors were reluctant to take in consideration investment opportunities because they had a negative perception of financial markets. They expected that the S&P500 index would have a down or flat trend in the years 2009-2011, suffering from the Crisis. Despite these judgments, the S&P500 performed well in that period. Moreover, Thakor (2015) showed a further link between GFC and availability biases. Before the speculative bubble burst, financial markets were characterized by high inflation and interest rates and reversed yield curve. Now, these conditions are seen as an alert by investors as a possible manifestation of a new crisis.

As shown in the previous cases, availability derives from streamlined reasoning, leading to wrong behaviours and choices.

1.5.3 Anchoring

Anchoring pushes people to evaluate a problem starting from a reference point (anchor), which could be either relevant information or recommended by the settings of the problem.

Then, adjustments are made to reach the appropriate solution. The bias is represented by the total reliability on the anchor since all subsequent judgements, arguments and estimates are made in relation to it. Moreover, this heuristic takes place when individuals base all their reasoning on partial information (Tversky and Kahneman, 1974).
This bias often occurs when people have to make numerical judgments under uncertainty, that is when they do not have all available information. Their estimate proves to be pretty similar to the reference point. Concerning the financial markets, suppose that a stock price drops and an investor aims to evaluate it using older trading price as the reference point. Of course, he will obtain unclear findings because important information would have not taken into consideration (Luppe and Fávero, 2012).
Further proofs confirm that anchoring takes the form of dispositional effect, a phenomenon that leads investors to sell stocks for which they earn and keep those for which they lose. This behaviour is inconsistent in statistical terms because a stock with a downward trend should increasingly fall, while a stock with an upward trend should increase its value. PT gives a possible solution to this abnormal behaviour: investors are unwilling to concede losses, they are risk-lovers when they face losses and they are risk-averse when they deal with profits. Furthermore, they misjudge likelihoods of price changes: they believe losing assets will increase and winning assets will decrease (Weber and Camerer, 1998).

1.5.4 Further cognitive bias

There are several additional heuristics accountable for incorrect processing of information: the *attribute bias*, the *self-attribution bias*, the *overconfidence*, the *confirmation bias*, the *home bias*, and the *status quo bias*.
The *attribute bias* refers to quantitative models used by investors when they select investments that have similar fundamental features. The securities chosen by a determined technique tend to have similar features because that technique allows identifying only those securities that have a specific set of characteristics. It is a common situation likely to happen unless the techniques used are designed not to include it. As a consequence, the portfolio composed in such a way will not well-balanced, and the investor will incur large losses if the specific market falls.
One possible solution to this bias is to use many different models in the selection asset process. Even if each model has this kind of bias, the investor will have a balanced portfolio because securities are chosen using different parameters.

It is important to highlight that *attribute bias* is different from *self-attribution bias*. The latter refers to systematic errors that people made when they attribute business successes to own skills and business failures to exogenous factors that are not under their control. In the finance world, it is expected that it strengthens the *overconfidence* of those successful investors who keep giving more and more trust in their trading skills, expanding their investments over time. Hoffmann and Post (2014) combine different survey data and matching trading experiences to demonstrate that *self-attribution bias* occurs among investors. If an individual portfolio obtains higher returns in the previous period, the investor will attribute that success to his personal trading skills.

A further heuristic mentioned above is the *overconfidence*, which is the tendency of people to believe that they are more proficient than what they really are. Of course, it may lead to blunders: the probabilities of a car accident or suffer major illnesses are undervalued by older people; although smokers know the likelihood of getting lung cancer and heart attack, they strongly believe that they are less disposed to suffer them than most non-smokers people; additionally, lotteries are partially thriving because of unreliable optimism (Thaler and Sunstein, 2008). Concerning the finance world, *overconfidence* may lead investors may accept an extreme risk if they believe to be able to bypass all market declines, and portfolio managers may not realize the value of holding bonds, which can be a cushion against stock market decays.

The *confirmation bias* refers to the typical behaviour of interpreting information so that supports one's pre-existing expectations or beliefs, overlooking and questioning that information that is in contrast with them. An additional reason why *confirmation bias* is so frequent is due to ambiguous information that requires interpretations. In these circumstances, people evaluate information respectively to their initial beliefs, strengthening their currently opinion.

The *home bias* describes the investor's preference for domestic securities rather than those from foreign markets. There are several reasons that explain this kind of behaviour: asymmetric information between domestic and foreign markets, closeness to the local one, and the presence of costs, different regulations and restrictions due to national boundaries. So, investors opt for allocating the majority of their portfolio in domestic securities, overlooking the advantages of a diversified portfolio. Actually, diversification may be achieved if investments are distributed among several different asset classes, businesses, and geographic areas. The main threat of *home bias* is the possibility of a country economy downturn.

Although the negative implications of *home bias* were discovered and explored in the previous century, this phenomenon is more widespread than once documented. Ke et al. (2010) analyse more than 3,000 mutual funds from different countries in the whole world (excluding the U.S.) and prove that they tend to invest more in U.S. firms only if they have subsidiaries in their home country than those that have not.

Finally, the *status quo bias* refers to the preference for the existing state rather than change it. In their experiments, Samuelson and Zeckhauser (1988) explain that people tend to choose the status quo as the number of alternatives rises. Moreover, this behaviour is visible in loss aversion because investors opt for keeping losing positions instead of selling and enter in a new state (Kahneman et al., 1991). Investors tend to hold the same trading strategy, even if that strategy does not provide optimal returns.

Chapter 2 – Toward a behavioural portfolio selection theory

This chapter discusses the problem that entails the decision of resources allocation among a set of securities to select the best portfolio, given the relative constraints. Since the 1950s, Modern Portfolio Theory (MPT) founded by Markowitz is the only theory able to face this problem. However, empirical evidence and new theories arise in the following decades, throwing up roadblocks and encouraging economists to discover more realistic models. Undoubtedly of remarkable reputation is Behavioural Portfolio Theory (BPT), founded by Shefrin and Statman in 2000. Section 2.1 introduces MPT to give a general idea of the topic. Then, the following sections focus on BPT. Section 2.2 introduces general concepts of BPT, followed by Section 2.3 that concerns Security-Potential/Aspiration Theory, which is one of the main pillars of BPT. Since BPT presents two versions, Section 2.4 and 2.5 separately describe them: they both examine the research of the efficient frontier, the comparison between the mean-variance model and structural issues. Lastly, Section 2.6 discusses the birth of mental accounts and their recent applications in the portfolio selection field.

2.1 Modern Portfolio Theory

Harry Markowitz (1952a) developed the Modern Portfolio Theory, a prescriptive theory that helps investors to take into consideration the best combination of assets, aiming either to maximise the portfolio expected returns fixed a specific level of risk, or to minimise the portfolio risk, given a level of expected returns. In this framework, variance represents the risk.

The MPT underlying assumptions rely both on individuals' behaviour and the efficient market hypothesis (EMH):

- *risk-aversion*: individuals invest in the portfolio that has the best risk-expected return profile. If two portfolios have the same expected returns, the portfolio with lower overall risk will be preferred;
- *rationality*: investors choose the portfolio with the highest expected returns to maximise their utility function;
- *investor's utility function*: it depends both on expected return $\mathbb{E}(r)$ and variability of return σ^2;
- *price-taker*: market prices are not affected by investors' actions;
- *frictionless market*: MPT does not consider transaction costs and taxes, and assets are infinitely divisible;
- *efficient markets*: markets rapidly assimilate correct information, which may be used by investors to compute expected returns and risks for any asset.

Given these assumptions, Markowitz (1952a) asserts that the overall portfolio risk is affected by three factors: the risk associated to each asset held in the portfolio, the proportion of wealth invested in each of them and the correlation between them.

In the next sub-sections, risks and returns, and portfolio selection are discussed in detailed to get a general idea about the strengths and the weakness of this theory.

2.1.1 Risk and return

MPT is also known as mean-variance analysis since portfolios are constructed considering both expected value and variance of returns (Markowitz, 1952a).

Consider the expected value $\mathbb{E}(X)$ and variance $Var(X)$ of individual asset. Setting $X = x_1, x_2, \dots, x_n$ a random variable representing possible returns of the asset X and $p = p_1, p_2, \dots, p_n$ relative probabilities, the following formulas are defined:

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_i, \tag{2.1}$$

$$Var(X) = \sum_{i=1}^{n} \left(x - \mathbb{E}(X)\right)^2 p_i. \tag{2.2}$$

From these formulas, it is easy to compute the mean and variance of the portfolio. Since a portfolio is defined as a weighted average of individual assets, the portfolio return is a weighted average of returns of each asset. Let R_p the portfolio rate of return, R_i the return on asset i and w_i the percentage invested in the asset i, then:

$$\mathbb{E}(R_p) = \sum_{i=1}^{n} w_i \mathbb{E}(R_i). \tag{2.3}$$

The second main feature is the variance. Consider σ_i^2 the variance of the individual asset i, the variance is defined as follows:

$$\begin{aligned} Var(R_p) &= \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} = \\ &= \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{i,j}\,, \end{aligned} \tag{2.4}$$

where $\sigma_{ij} = \sigma_i \sigma_j \rho_{i,j}$ is the covariance between asset i and asset j, and $\rho_{i,j}$ is the correlation coefficient between asset i and asset j. The covariance between two assets (X and Y) is a statistical measure that analyses their directional relation: if it is positive, it means that the asset returns move together in the same direction, while a negative covariance implies that returns move inversely. The covariance formula is the following:

$$\sigma_{ij} = \mathbb{E}\big[\big(X - \mathbb{E}(X)\big) - \big(Y - \mathbb{E}(Y)\big)\big]. \tag{2.5}$$

Of course, the covariance of an asset with itself is equal to its variance.

Since the covariance varies from units of measurement used, the deviations are written in terms of standard deviation units. From the covariance of deviations, the correlation coefficient between X and Y is obtained:

$$\rho_{X,Y} = \sigma_{XY} / \sigma_X \sigma_Y. \tag{2.6}$$

Regarding the portfolio selection problem, the correlation coefficient may be used to select assets correctly and achieve the best diversification.

The covariance among assets is a key point in the selection assets problem because it allows reaching the right-diversification. The combination of different types of assets, which respond in different ways to market trends, leads to reduce the overall portfolio risk. In this fashion, the portfolio should eliminate the unsystematic risk since the positive performance of some stocks neutralizes the negative performance of others. Generally, if a portfolio is composed of only firms within the same industry, a possible slowdown of the sector would have a high impact on portfolio performances (Markowitz, 1991). Before Markowitz diversification, the naïve diversification offered a less efficient method. Given N assets of interest, an individual should invest his wealth, splitting his current wealth equally among them, without considering any further information.

2.1.2 Portfolio selection

"We […] consider the rule that the investor does (or should) consider expected return a desirable thing and variance of return an undesirable thing" (Markowitz, 1952, p. 77)

The portfolio selection problem requires the selection of a satisfactory portfolio following the investor's preferences concerning expected return and risk.

From a geometrical point of view, Markowitz (1952a) determines the mean-variance efficient portfolio frontier, overlooking its analytical derivation. In the following years, Markowitz (1959) and Merton (1972) give an exhaustive mathematical derivation of the efficient frontier. Additionally, they consider various portfolio compositions, e.g., portfolios composed either of entirely risky assets or a combination between riskless and risky assets.

This dissertation aims to give a general idea of how MPT investors achieve efficient portfolios. Following the MV rule[9], they consider an investment as an acceptable one if they obtain either the minimum risk at any given level of expected return or the maximum expected returns at any given level of risk. Rigorous logical reasoning allows delineating the set of efficient portfolios. Figure 2.1 gives an illustrative representation. All portfolios along the curve between points E and F are efficient because they reach the maximum rate of return for each level of risk. Suppose that an MPT investor A aims to select an efficient

[9] In Chapter 1, Section 1.1.1 fully describes the MV rule.

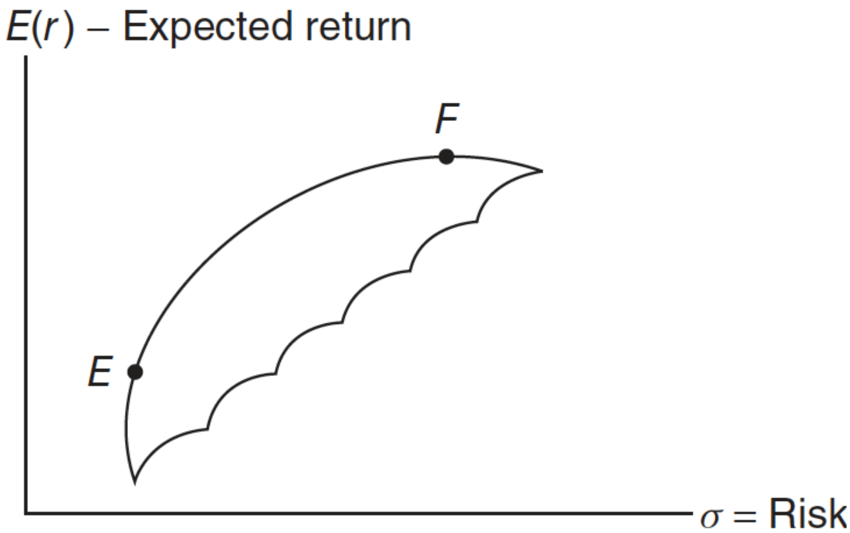


Figure 2.1 The efficient frontier. E(r) is the expected return, and the standard deviation, denoted by σ, represents the risk
Source: Francis and Kim (2013).

portfolio that provides a prefixed rate of return R_A. He is facing a minimisation problem, outlined as follows:

$$\min \sigma_P$$
$$\text{subject to} \begin{cases} \mathbb{E}(P) = R_A \\ \sum_i w_i = 1 \\ w_i \geq 0, \end{cases} \tag{2.7}$$

where σ_P represents the risk of the efficient portfolio and $w_i = w_1, w_2, \dots, w_n$ is the set of weight associated to each asset included in the portfolio, and this sum has to be equal to 1 to ensure that all money is invested. Finally, the last constraint does not allow short-selling. Figure 2.2 displays the efficient frontier, expressed by the concave curve, and two different kinds of indifference curves, which describe the preferences of investors A and B. Investor B accomplish his goals investing more in risky assets than investor A, who is more risk-averse.

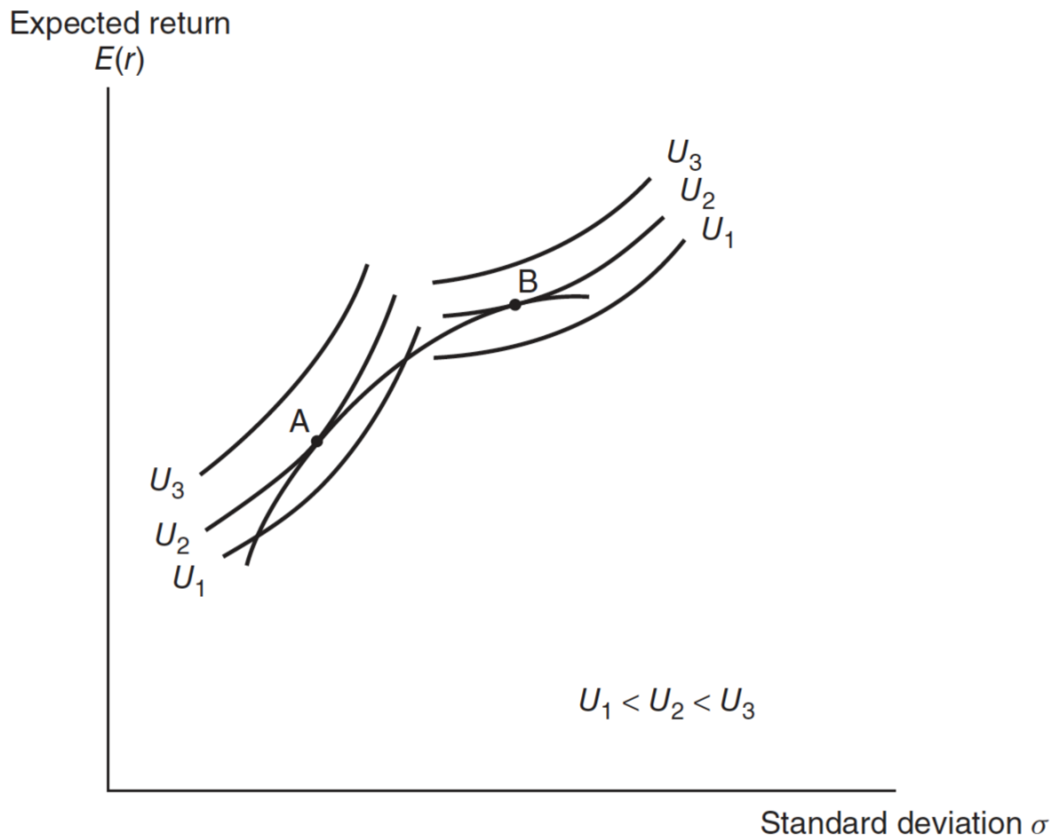


Figure 2.2 Different optimal portfolios for different MPT investors
Source: Francis and Kim (2013).

2.2 Behavioural Portfolio Theory

Behavioural Portfolio Theory was developed by Hersh Shefrin and Meir Statman (2000), who attempted to compensate for drawbacks of Expected Utility Theory (EUT). It is approved in the whole world as a descriptive theory able to define the investor's behaviour in decision-making under risk. BPT is capable of incorporate that empirical evidence in contrast to EUT, such as Ellsberg paradox.

BPT is based on two main pillars: Prospect Theory (Kahneman and Tversky, 1979) and Security-Potential Aspiration theory (Lopes, 1987), which is fully detailed in Section 2.3. In their work, Shefrin and Statman (2000) describe two different BPT frameworks: BPT-SA (Section 2.4), namely single mental account BPT, and BPT-MA (Section 2.5), which involves multiple mental accounts. BPT-SA investors, unlikely BPT-MA ones, take into account covariance, behaving similarly to mean-variance investors. Nonetheless, BPT-SA efficient portfolios are different from mean-variance efficient ones.

Shefrin and Statman (2000) make a clear distinction between mean-variance and BPT-MA investors: the former does not change their inclination toward risk, they are always risk-averse. Besides, they manage the portfolio as a whole, so they do not look at a single asset, and they strive to achieve the best trade-off between mean and variance. Thus, they take into consideration covariances when they select stocks to insert in the portfolio. On the other hand, BPT-MA investors are completely different: they consider portfolios as pyramids of securities, and they construct them layer by layer. Each layer has a specific purpose and a different inclination toward risk. This methodology is widely used by institutional pension funds, which collect money from many investors who have different risk inclination. Shefrin and Statman describe how the fund managers first define layers of the portfolio pyramid, then separately allocate capitals among securities and bond. In this way, BPT-MA investors do not take into account investment choices taken in the other layers of the portfolio, so they disregard covariances among assets.

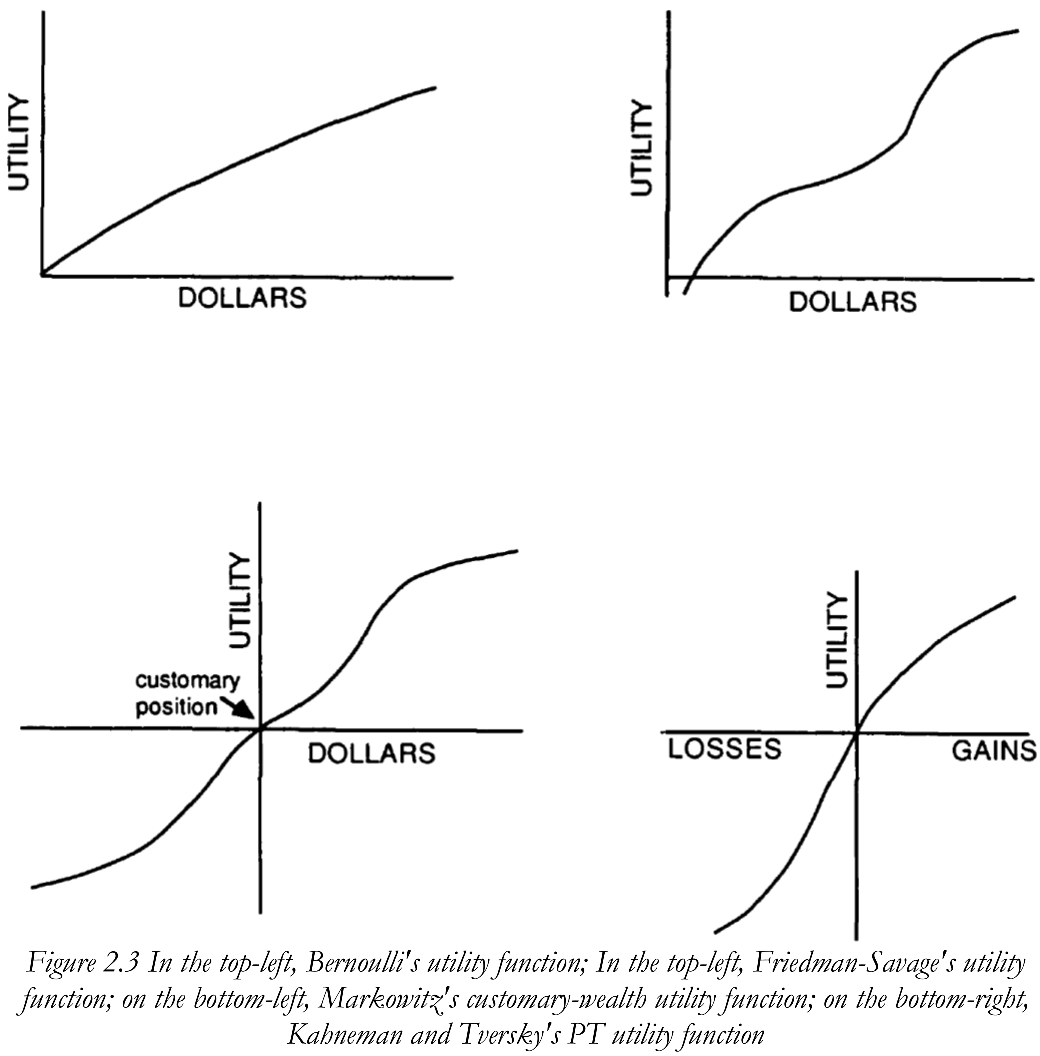


Figure 2.3 In the top-left, Bernoulli's utility function; In the top-left, Friedman-Savage's utility function; on the bottom-left, Markowitz's customary-wealth utility function; on the bottom-right, Kahneman and Tversky's PT utility function
Source: Lopes (1987).

Shefrin and Statman (2000) state that the Friedman-Savage (1948) puzzle had a significant impact on the search field and it stoked many researchers such as Lopes, Kahneman and Tversky, and many others. This issue observes that the same people bought both lottery tickets and insurance policies, in sharp contrast with EUT developed by Von Neumann and Morgenstern (1944) based on Bernoulli's work, which claims that people's attitude to risk never changes. Friedman and Savage suggest a utility function characterized by both concave and convex parts, as it is illustrated in the top-right of Figure 2.3. Markowitz (1952b) criticised their conclusions declaring that only a few people behave as Friedman-Savage investors, and highlighting that, according to their utility function, deprived individuals never buy lottery tickets and medium-income individuals do not cover themselves from reasonable losses. These implications seem wrong for Markowitz, who alleges that "even poor people, apparently as much as others, buy sweepstakes tickets, play the horses, and participate in other forms of gambling" (Markowitz, 1952b, p. 153). Keeping three inflection points, Markowitz (1952b) move the utility function to have the middle inflection point located at present wealth, as exhibited on the bottom-left of Figure 2.3.

2.3 Security-Potential/Aspiration theory

Security-Potential/Aspiration theory, denoted by SP/A theory, was developed by Lola Lopes (1987) to investigate the decision-making process under uncertainty. Shefrin and Statman (2000) consider this theory as an expansion of the model of Arzac and Bawa (1977), based on the safety-first portfolio model. The latter framework was defined by Roy (1952), who aims to minimise the likelihood of a dread event, namely $\mathbb{P}[W < s]$. Let returns be normally distributed, it means that an individual aspires to select a portfolio P such that minimises the following function:

$$\frac{s - \mu_P}{\sigma_P}, \tag{2.8}$$

where μ_P and σ_P are the expected returns and standard deviation of the portfolio P and s is the subsistence level.

SP/A theory (Lopes, 1987) implies that the decision maker's choice depends on two rational and psychological criteria: security-potential criterion ("dispositional factor") and aspiration criterion ("situational factor"). The dispositional factor explains the reasons why investors attempt to prevent poorness or bankruptcy (security) or taking advantage of getting rich (potential). Thus, this criterion shows how an individual looks at risks: people aiming for securities are risk-averse, while people driven by potential are risk-seeking. On the other hand, the situational factor defines how people behave to instant necessities and opportunities: even risk-averse people may expose themselves to take some risks if the situation is very favourable.

In Lopes' model, fear and hope interact on the intention to take risks. Investors need to take into account the danger that their wealth might fall below a subsistence level s. Thus, the probability of the safety can be defined as $\mathbb{P}[W \geq s]$, which is known to be a decumulative probability (Shefrin and Statman, 2000). The dispositional factor is patterned after a decumulative weighted value, which is as follows (Lopes and Oden, 1999):

$$SP = \sum_{i=1}^{n} w(p_i)(W_i - W_{i-1}), \tag{2.9}$$

where W_i represent outcomes ordered from the worst to the best, p_i is the decumulative probability of achieving an outcome at least equal to W_i, and w is a function that maps p_i within the range between 0 and 1. The latter is also useful for determining risk inclinations: if w is convex, the investor is risk-averse ("security-mindedness"), whereas if w is concave the investor is risk-seeking ("potential-mindedness"). Finally, if w assumes an inverse-S-shape, the investor's behaviour is comparable to the Markowitz (1959) pattern ("cautiously hopeful"), in which people purchase both lottery tickets and insurance policies.

The decumulative weighting function, w, is represented as:

$$w(p) = \delta p^{q_s+1} + (1-\delta)[1-(1-p)^{q_p+1}], \tag{2.10}$$

which works both for losses and gains. Equation (2.10) "is derived from the idea that subjects assess lotteries from the bottom up (a security-minded analysis), or the top down (a potential-minded analysis), or both (a cautiously hopeful analysis)" (Lopes and Oden, 1999, p. 290). The first addend of Equation (2.10) represents the effects of fear, which overestimates the likelihood related to the worst outcomes rather than the best outcomes. Driven by the need for security, fear influences the mentality toward risk, and so, for $q_s > 0$, the function p^{q_s+1} assigns greater weight to higher values of p. Hope takes a crucial role in the second addend of Equation (2.10), working in a similar way to fear, but opposite for meaning. Prompted by the need for potential, the function $[1-(1-p)^{q_p+1}]$ appoints greater weight to lower values of p. Thus, the non-negative parameter q_s defines the power of fear or, in other words, the need for security; q_p is also a non-negative parameter that defines the power of hope, in other terms, the need for potential. The parameter δ defines the weights of the relative S and P analysis; in other words, δ represents the strength of fear with hope. It is necessary to highlight that:

- if $\delta = 1$, the investor is security-minded, since he gives to security all the importance due to his risk-aversion. In this way, the decumulative weighting function has the following form: $w(p) = \delta p^{q_s+1}$;
- if $\delta = 0$, the investor is potential-minded, and the decumulative weighting function is equal to $(1-\delta)[1-(1-p)^{q_p+1}]$;
- if $0 < \delta < 1$, the investor is cautiously hopeful. The magnitude of fear and hope depends respectively to δ and $(1-\delta)$.

Moreover, SP/A theory works similarly to CPT because it lets to assume different values for q_s, q_p, and δ for gains and for losses (Lopes and Oden, 1999).

Finally, it is important to highlight that the expected value of wealth $\mathbb{E}_w(W)$ decreases if investors give more importance to fear; vice versa, hope acts incrementing $\mathbb{E}_w(W)$.

Regarding the situational factor, Lopes (1986) hypothesizes that people evaluate the attractiveness of gambles from their probability to obtain an outcome W equal to or greater than aspiration level α. In short:

$$A = \mathbb{P}(W \geq \alpha). \tag{2.11}$$

SP and A are combined to obtain the dual criterion model f, which is an increasing function in both SP and A criteria:

$$SP/A = f[SP, A]. \tag{2.12}$$

However, Lopes and Oden (1999) contend the aspiration level as something “fuzzy”, since not all outcomes are able to fulfil it.

An application of SP/A theory is proposed by Ginita Wall (1993): she shows how financial planners usually give advice to clients regarding the portfolio composition (Figure 2.4). For the purpose of a better understanding of Figure 2.4, assets become riskier and riskier moving from bottom to top; while moving from right to left implies higher yields. In the bottom of the pyramid, there are safe-securities (e.g. money fund and savings account) that avoid poverty, and so they accomplish security needs. Then, financial planners propose to use bonds (e.g. zero-coupon bonds) to provide a fund useful for the children's university education. Further up there are stocks, real estate, and other more speculative investments that are necessary to have a change of getting rich. Undoubtedly, the allocation of resources

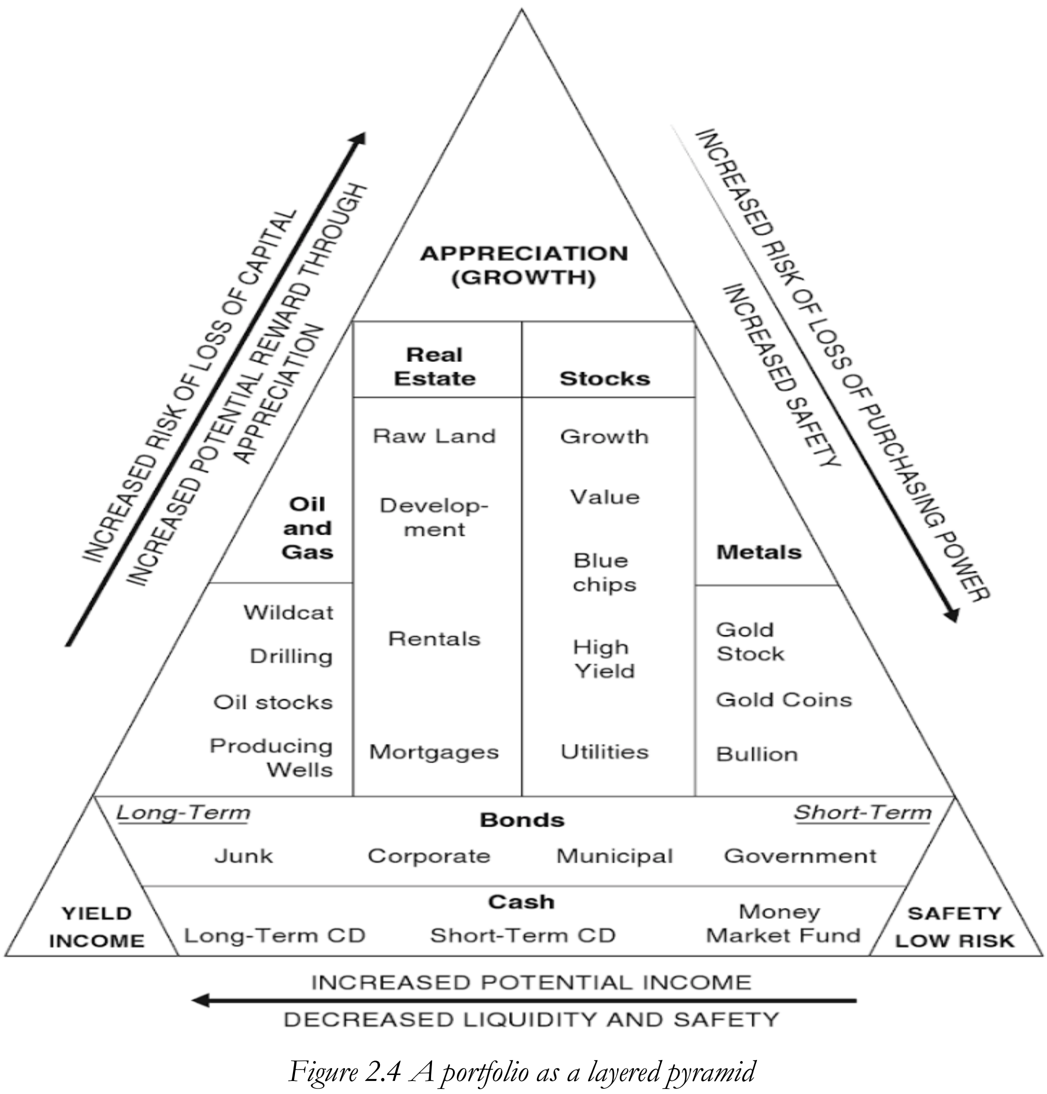


Figure 2.4 A portfolio as a layered pyramid
Source: Wall (1993).

between different layers of the pyramid is caused by an investor's risk tolerance, which is defined by fear and hope, combined with aspiration.

2.4 Single mental account Behavioural Portfolio Theory version

In this section, a single account version of BPT is discussed to get an idea of the general structure. In the next section, Lopes' SP/A theory and Kahneman-Tversky's mental accounting framework are joined, exploring the mental accounting version of BPT.

The concept of a single mental account means that investors view the portfolio as a whole: both BPT-SA and mean-variance investors behave in this way. The portfolio selection problem is very similar between the two frameworks because they consider covariances as Markowitz (1952a) described, and they opt for higher expected returns, keeping risk as lower as possible. Therefore, the efficient frontier for mean-variance investors is achieved maximizing the expected value μ, given fixed value of standard deviation σ, while the BPT-SA efficient frontier is achieved maximizing $\mathbb{E}_w(W)$[10], given $\mathbb{P}[W \leq A]$. Nevertheless, BPT-SA efficient portfolios are different from mean-variance efficient ones. To assess the latter statement, BPT-SA efficient portfolio is introduced, then the difference is presented.

Suppose financial markets are composed of contingent claims, which pay one at date one if state i happens, otherwise zero. Let sc_i the price of a state-i contingent claim, which is sorted in a way that the ratio between state prices and unit probability sc_i/p_i is always decreasing with increasing i (monotonically decreasing). An investor starts with W_0 money at date zero and aims to maximise his expected wealth $\mathbb{E}_w(W)$ at date one, keeping in mind a safety-first constraint that helps him not to go bankrupt. To do this, he buys a package of date one contingent claims $W_1, \dots, W_n$, whose total value $\sum sc_i W_i$ does not surpass his initial wealth (Shefrin and Statman, 2000). The investor exploits contingent claims to cover himself against the realization of an uncertain future event, e.g., bankruptcy.

[10] The expected value of wealth is not so trivial to compute because it takes into account the decumulative weighting function w, which is expressed by Equation 2.10.

Theorem 2.1. Any solution $W_1, \dots, W_n$ to

$$\max_{r_i} : \mathbb{E}_w(W) = \sum r_i W_i, \quad subject\ to: \mathbb{P}[W \le A] \le \alpha, \tag{2.13}$$

has the following form. There is a subset T of states, including the *n*-th state s_n such that:

$$\begin{aligned} &W_i = 0 \quad for\ i \notin T, \\ &W_i = A \quad for\ i \in T \setminus \{s_n\}, \\ &W_i = \left(W_0 - \sum_{i=1}^{n-1} sc_i W_i\right) \Big/ sc_n \quad which\ exceeds\ A\ when\ W_0 > sc_n A, \end{aligned} \tag{2.14}$$

where $\mathbb{P}\{T\} \ge \alpha$, but no proper subset T' of T features $\mathbb{P}\{T'\} \ge \alpha$. If all states are equiprobable, then there is a critical state i_c such that the optimal portfolio has the form:

$$\begin{aligned} &W_i = 0 \quad for\ i \le i_c, \\ &W_i = A \quad for\ i_c \le i < n, \\ &W_i = (W_0 - \textstyle\sum_{i=1}^{n-1} sc_i W_i)/sc_n \quad which\ exceeds\ A\ when\ W_0 > sc_n A, \end{aligned} \tag{2.15}$$

where i_c is the lowest integer for which $\sum_{i > i_c} p_i \ge \alpha$ (Shefrin and Statman, 2000, p. 133-134).

Theorem 2.1 defines an efficient BPT-SA solution, aiming at the maximisation of $\mathbb{E}_w(W)$, which represents the summation of probability wealth products $r_i W_i$. Primarily, suppose the unrestricted maximisation of $\mathbb{E}_w(W)$. To do so, it is necessary to have the cheapest price for buying contingent wealth, which is state n. Thus, sc_n/r_n is the smallest price of a contingent claim. This unrestricted maximisation occurs when $W_n = W_0/sc_n$ and $W_i = 0$ for all i. It entails that $\mathbb{E}_w(W)$ is equal to the whole initial wealth. The problem changes with the restriction submitted in Theorem 2.1. To obtain the most economical approach of meeting the constraint, let $\{T''\}$ be all sets, with s_n included, characterizes by $\mathbb{P}\{T''\} \ge \alpha$, albeit there is no subset T' of $\{T''\}$ that fulfils $\mathbb{P}\{T'\} \ge \alpha$. Then, Theorem 2.1 associates the sum $sc_A(T'') = A \sum_{i \in T''} sc_i$ to every set. Therefore, it selects the set characterized by the smallest value $sc_A(T'')$, which amount is the price for all contingent claims from those that pay merely s_n to those that pay precisely A units for states ranging in $T \setminus \{s_n\}$.

A trivial consideration regards the aspiration level A or α, which could not assume values too high since no optimal solution would be found.
To demonstrate that BPT-SA efficient portfolio is different from the mean-variance one, Shefrin and Statman (2000) suggest that if a BPT-SA efficient portfolio has at least three states that characterize positive consumption, with different values for sc_i/p_i, then it cannot be a mean-variance efficient one. The mean-variance portfolio is a strictly concave function of sc_i/p_i, while BPT-SA one is a strictly increasing with three different payoffs, one of which is equal to zero.

2.4.1 The BPT-SA efficient frontier

Mean-variance investors compose the efficient frontier by pinpointing portfolios with the highest expected value for any level of standard deviation. On the other hand, BPT replaces the standard deviation in exchange for the probability that final wealth may fall below α. In this way, behavioural investors fashion the efficient frontier selecting those portfolios with the highest expected wealth for any level of $\mathbb{P}[W \leq A]$.
Theorem 2.1 implies that BPT-SA investors tend to choose a portfolio whose return distribution is not normal, opting for portfolios composed of a combination of a risky bond and a lottery ticket. Nonetheless, Shefrin and Statman (2000) provide a comparison between low and high aspiration levels when behavioural investors deal with returns normally

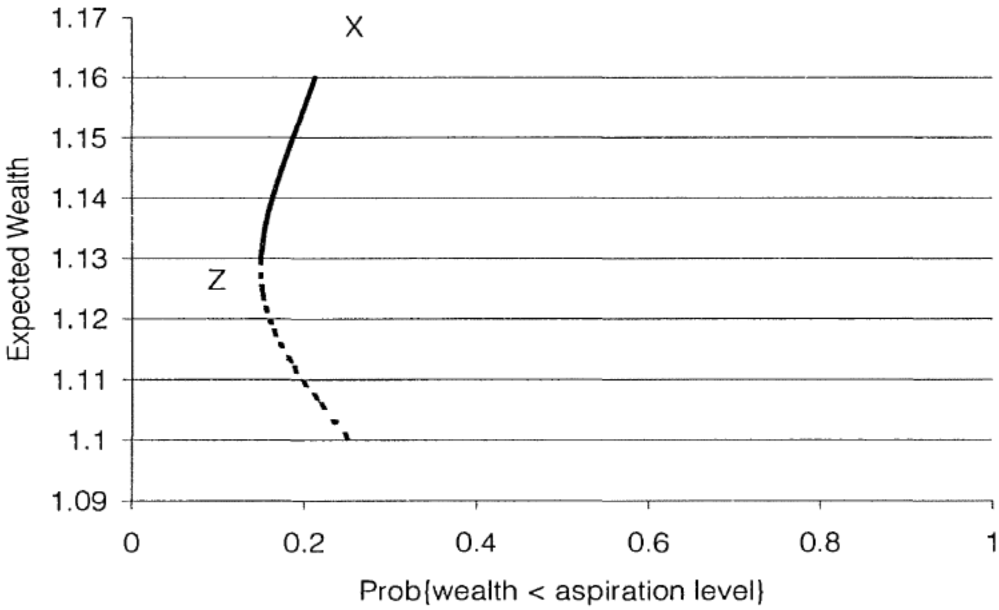


Figure 2.5 The BPT-SA efficient frontier for an investor with a low aspiration level
Source: Shefrin and Statman (2000).

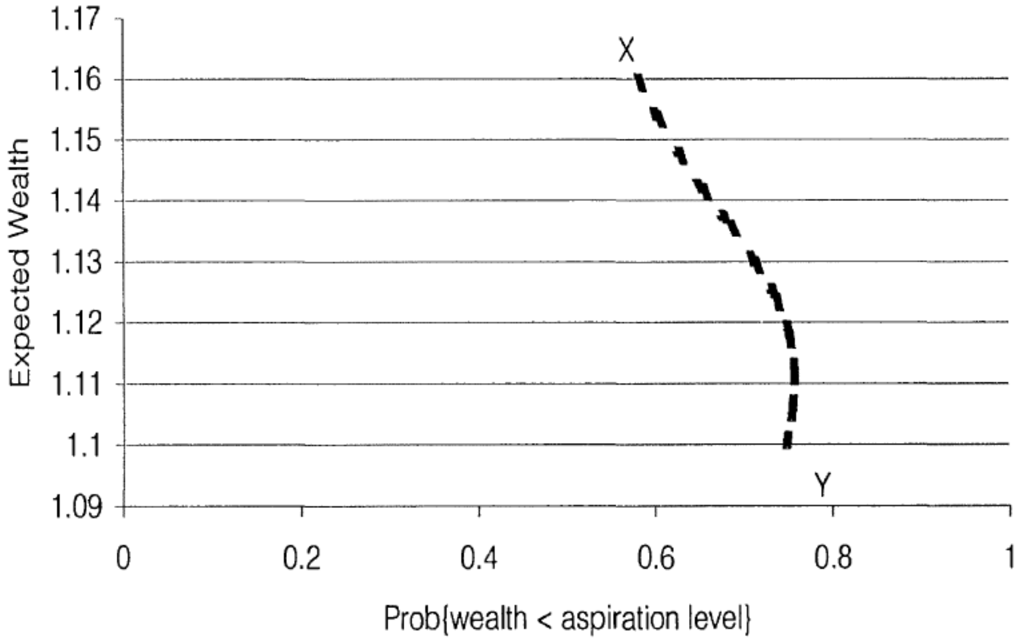


Figure 2.6 The BPT-SA efficient frontier for an investor with a high aspiration level
Source: Shefrin and Statman (2000).

distributed. Since behavioural investors view the portfolio as a whole, they choose diversified portfolios if they have low aspiration levels; on the other hand, they pick undiversified portfolios if they have high aspiration. The latter instance occurs when behavioural investor is similar to Durbins and Savage (1976) investor, who is hopeless in a casino with $1000, aiming to gain $10,000 by morning. The only solution is to make a single bet that allows him to reach his goal, even if odds are pretty unfavourable to him. Indeed, if he diversifies his bets, he will do worse because he has essentially no opportunity of winning the sum of money wished for.

Suppose there are two assets with the following features:

- $\mu_X = 16\%$ and $\sigma_X = 20\%$;
- $\mu_Y = 10\%$ and $\sigma_Y = 15\%$.

Two different investors with a low ($1) and a high ($1.20) aspiration level are compared: they both start with $1 wealth. Figure 2.5 illustrates the efficient frontier for the investor with low aspiration level: portfolio composed entirely of asset Y is not represented since it is dominated by Z, which is made up by a combination of both assets. The efficient frontier is represented by the solid curve. Concerning the high aspiration level case, the investor needs to make a single bet on security X because A is much higher than returns of both securities. Figure 2.6 shows the efficient frontier for the investor with high aspiration level.

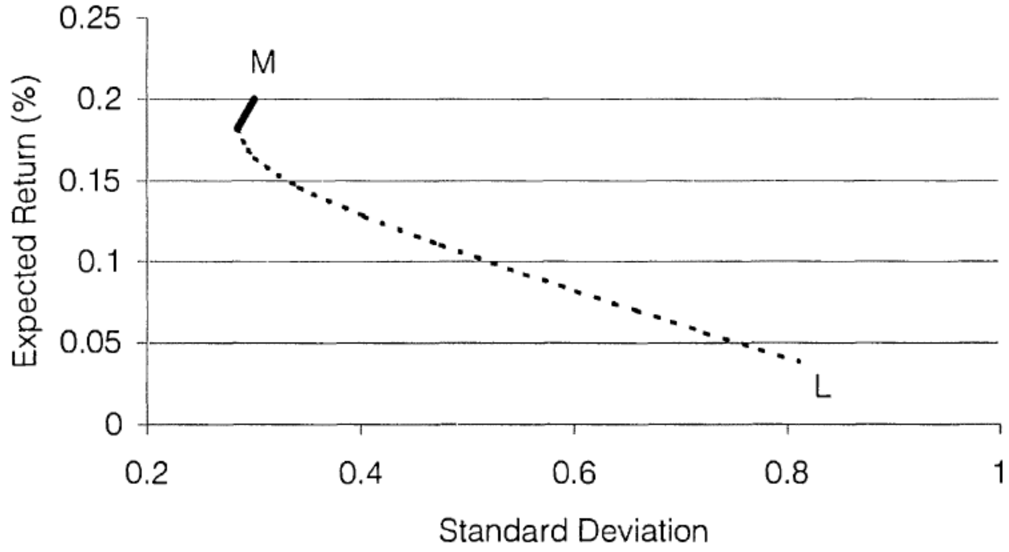


Figure 2.7 Mean-variance frontier
Source: Shefrin and Statman (2000).

In a further example, Shefrin and Statman (2000) show the remarkable difference between BPT-SA efficient frontier and mean-variance efficient one. Suppose there are two securities with the following characteristics:

- $\mu_L = 2\%$ and $\sigma_L = 90\%$;
- $\mu_M = 20\%$ and $\sigma_M = 30\%$.

For a mean-variance investor, the solution is trivial: applying the mean-variance rule, M is preferred to L since $\mu_M > \mu_L$ and $\sigma_M^2 < \sigma_L^2$ (Markowitz, 1952a), in other words, M dominates L in the mean-variance dominance sense (Figure 2.7). However, a portfolio composed totally of security L, which recalls casino-type security, is on the efficient frontier for a behavioural investor whose aspiration level is very high, e.g., \$1.30: this is the only way to reach his goal (Figure 2.8). It is necessary to highlight that this behaviour cannot be associated with a risk-loving one. They pick the risky L asset only because it grants the highest likelihood of obtaining what they want; they are not driven by the willingness of making risky choices.

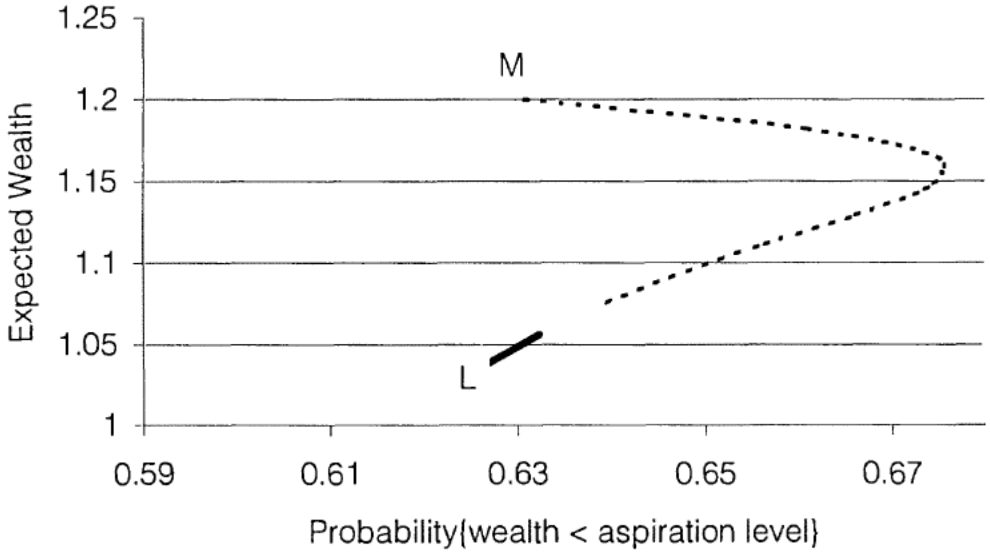


Figure 2.8 The BPT-SA frontier for an investor with a $1.30 aspiration level
Source: Shefrin and Statman (2000).

2.5 Multiple mental account Behavioural Portfolio Theory version

"People keep their money in separate pockets" (Shefrin and Statman, 2000, p. 142).

BPT-SA is a tool used by Shefrin and Statman (2000) to explain different levels of aspiration that affect investors in the decision-making process. Now, a further step is done by submitting multiple mental accounts. The latter is an economic concept analysed by Thaler (1985, 1999), who argues that people do not follow a logical and rational way to categorize their funds and so they make fallacious decisions in their investment and spending habits. Besides, Das et al. (2010) illustrate that investors tend to construct their portfolios with distinct mental accounts according to the wealth they aim and their personal goals that must be reached within the date prearranged.

The inclusion of mental account version allows splitting the portfolio into different subsets, with different roles and different aspiration levels. Unlike BPT-SA, BPT-MA investors overlook covariances. This phenomenon is proved by empirical evidence, which confirms that people do not use such information to construct their portfolios (Kroll et al., 1988). Additional proofs come from institutional investors, who often separate the management of assets and currencies, relying on the handling of one security to another colleague (Jorion, 1994).

In their work, Shefrin and Statman (2000) describe the framework for two mental accounts. Suppose that an investor holds three "psyches" (Thaler and Shefrin, 1981). Two "doers" characterized by distinct aspiration levels (a low and a high) and each of them is attached to one mental account and a "planner" who coordinates the two doers to maximise the utility function of the investor. The latter has the task to split the initial wealth between the two doers. The Cobb-Douglas function is used to represent the utility function of the low aspiration doer U_{la} and the high aspiration doer U_{ha}:

$$U_{la} = P_{la}^{1-\gamma} E_h(W_{la})^{\gamma}, \tag{2.16}$$

$$U_{ha} = P_{ha}^{1-\beta} E_h(W_{ha})^{\beta}, \tag{2.17}$$

where P_{la} and P_{ha} are, respectively, the probabilities of falling below of the low aspiration level A_{la} (\$0.20) and A_{ha} (\$1.20), W_{la} and W_{ha} are the final wealth of the respective low and high aspiration doers, and γ and β are two non-negative weighting parameters equal to 0.1. The planner's utility function U_p is a combination of the previous two equations, in short:

$$U_p = \left[1 + K_{ha}\left(P_{ha}^{1-\beta} E_h(W_{ha})^{\beta}\right)\right] K_{la}\left[P_{la}^{1-\gamma} E_h(W_{la})^{\gamma}\right], \tag{2.18}$$

where K_{ha} (10,000) and K_{la} (1) are, respectively, the weights associated with the high and low aspiration doers. The planner has to distribute current wealth to both doers, but how is the first dollar spent? Looking carefully at Equation (2.11) it can be noticed that the planner's utility function is equal to zero if and only if $U_{la} = 0$, but it is not zero when $U_{ha} = 0$. This means that the investor needs first to accomplish the low aspiration goal, and, in the second time, he may focus on high aspiration one.

2.5.1 Structural issues

Recalling Theorem 2.1, investors choose portfolios composed by possibly risky bonds and a lottery ticket, regardless of their aspiration level. Nonetheless, mental accounts push each type of investor to his extreme: the high aspiration investor focuses more on a lottery ticket, in contrast to low aspiration investor, who concentrates more on riskless bonds. If an additional sum of money is given to the low aspiration investor, it will be used to reduce the risk of the portfolio, so decreasing the lottery component percentage. In the opposite way, a high aspiration investor would invest extra allocations in the risky component, aiming to maximise the expected payoff. In this way, BPT-MA investors opt for risk-free bonds for the low aspiration layer and for an incredibly risky asset (e.g. a lottery ticket) for the high aspiration layer. These inclinations are influenced by risk tolerance, which is caused by the same parameters used for BPT-SA, which are: degrees of fear and hope, and the intensity of fear in relation to hope. If BPT investors have a high degree of fear, they will be prone for investing in riskless bonds; on the contrary, if they are characterized by a high degree of hope, they will tend to invest in more risky assets to maximise their payoff (Shefrin and Statman, 2000).

Moreover, Shefrin and Statman (2000) demonstrate that BPT-MA portfolios are neither mean-variance efficient nor BPT-SA efficient. It is proved by the following theorem:

Theorem 2.2. If there are at least five states that feature positive consumption in a BPT-MA portfolio P_M, and distinct values for v_i/p_i, then P_M is neither mean-variance efficient nor BPT-SA efficient (Shefrin and Statman, 2000, p. 146).

It is important to highlight that BPT-SA and BPT-MA portfolios are different not because the former has one aspiration level and the second has two, but for the dearth of assimilation between the two accounts by BPT-MA investors. As it is analysed in Section 2.3, the possible payoff values for BPT-SA portfolios are zero, the aspiration level, and the unlikely payoff from the winning of a lottery ticket. This is different from what BPT-MA investors behave in the portfolio selection problem.

2.6 Further progress on mental accounting

Kahneman and Tversky (1984) investigate which are the key factors that drive people to make decisions in risky and riskless circumstances. In contrast with rational choice, people behave in different ways based on how they frame decision problems. Kahneman and Tversky use mental accounting to describe some features of consumer behaviour. Soon after, Thaler (1985) uses mental accounting to coin a new model of consumer behaviour, which strives to explain the process how people code the combination of gains and losses by Prospect Theory value function and then evaluate the purchases using the idea of *transaction utility*[11].

In the following years, several economists run experiments to understand better the mental accounting process. Krouse (1986) conducts experiments on children, realising that older children employ mechanisms alike to those reported for adults; Singer et al. (1986) test the *theatre ticket* problem proposed by Kahneman and Tversky to find the real trigger of the problem framing. Finally, Henderson and Peterson (1992) formally describe how mental accounts are composed and work, and what is incorporated in them.

Concerning the portfolio selection problem, Shefin and Statman (2000) use mental accounts to describe how investors do not consider their portfolios as a whole but composed of several different sub-portfolios, each characterized by a specific goal (see Section 2.2). Recently, Das et al. (2010) reach a significant milestone regarding portfolio optimisation. They introduce a new mental accounting (MA) framework capable of combining characteristics both of MPT and BPT. In details, the MA framework involves: a layer-by-layer structure, the risk is defined as the likelihood of missing to exceed the threshold level in every layer, and different approaches toward risk that change by mental accounts. They prove that the MVT efficient frontier always holds MA optimal portfolios. Alexander and Baptista (2011) extend the aforementioned framework, introducing the possibility for the investor to delegate his riches among portfolio managers. Finally, since Das et al. (2010) take into account merely the portfolio risk, Baptista (2012) also decides to consider the so-called background risks, i.e., those deriving from real estate and labour income. Investors cannot fully ensure themselves from this kind of risk in financial markets. Unlike Das et al., Baptista assumes that, for each layer, the investor maximises his objective function, taking into account the background risk.

11 Thaler (1985) introduces the concept of transaction utility to describe the satisfaction that a consumer feels from the perceived value of an agreement. In other words, it is the difference between the actual price and a reference point.

Chapter 3 - Rational and behavioural models in comparison

Following Modern Portfolio Theory, the mean-variance model controls the portfolio risk considering the standard deviation of stocks and the covariance between them. Harry Markowitz detects two kinds of portfolio risk, systematic and unsystematic, which last may be eliminated by a proper diversification of the portfolio. Instead, systematic risk is considered as market risk, i.e., it is undiversifiable and influences the whole market, not only a singular stock or business. Therefore, if an investor strives to find the optimal portfolio with the smallest risk concurrently with any feasible return, he should opt for the market portfolio, which might be the market index. Section 3.1 thoroughly explains the Index-Tracking problem, a passive strategy that aims to replicate the market portfolio using a fewer amount of stocks held by the benchmark. Later, Section 3.2 illustrates the He and Zhou's model, which implements the Kahneman and Tversky's Cumulative Prospect Theory (CPT). The chapter provides an accurate description of both processes and their main features, with a generalized formulation of the respective objective function. The purpose is to understand which the main differences between these two models are so that it will be clearer to follow the comparison and findings discussed in chapter 5.

3.1 Index-Tracking model

Index-Tracking is a passive fund management strategy, which aims at the replication of the performance of a stock market index, without acquiring all of the stocks that compose the index.

The purpose of fund managers is to produce both capital growth and income over the medium or long-term. The primary approaches chosen by fund managers can be generally divided as active management and passive management. The former occurs when fund managers strive to beat a benchmark index or target return, picking those "winning stocks" that are going to outperform other stocks. To do so, they have a considerable degree of flexibility, and they believe they can select performing stocks due to their expertise and judgment. Nonetheless, this kind of management has higher fixed costs due both to payments of the management team and high transaction costs resulting from the frequent trading (Beasley et al., 2003).

On the other hand, passive management implies that fund managers[12] replicate the investment weighting and returns of a benchmark index, following a well-defined set of guidelines. They have more limited flexibility. Usual guidelines presuppose that the fund should perform almost the corresponding return of a market index, i.e., S&P 500 in New York, by investments in a properly set of stocks, included in the same index. The advantage of this strategy is that it implies lower fixed and transaction costs, but it has the drawback that if the market collapses, then unavoidably the return will drop (Beasley et al., 2003).

It is now clear that, in active management, investors face both market and company risk[13], while, in passive management, investors face only market risk.

Last but not least, there is a relevant feature concerning passive management. Suppose that a trucker fund aims to invest in a single index. It could buy all stocks included in the index, in their proper amounts, to correctly reproduce the index. This approach is so-called as full replication. Notwithstanding his simplicity, it has a few crucial drawbacks. Some stocks in the index may be purchased (proportionally) in short quantities. This behaviour is inconvenient in terms of costs, especially if there is a restricted market that limits their purchasing. Besides, if the composition of the index is reviewed and changed, the weightings of the trucker fund portfolio must be modified to match their new weightings. As a result, full-replication implies enormous transaction costs. For these motives, several passively fund managers purchase fewer stocks than those involved in the tracked index (Beasley et al., 2003).

---

12 A passively managed fund whose purpose is to replicate the return of an index is so-called as an index fund or tracker fund.

13 An investor faces company risk when he holds stocks in a distinct company. Only diversification may contribute to reducing this risk, acquiring stocks in further firms and uncorrelated securities.

A possible solution to fulfil the Index-Tracking is to hold an exchange-traded fund (ETF)[14]. It is regularly traded at any time, and also for relatively small amounts. In this way, investors hedge from company risk because they can invest in the index when it is low and sell when it is high (Beasley et al., 2003).

In the previous decade, Beasley et al. (2003) discussed the Index-Tracking model, focusing on the research of a way to replicate index performances, limiting transaction costs and the number of stocks traded. They also took into account the revision of the portfolio, not just its creation.

In the following sections, the Index-Tracking problem is formulated, outlining notation, the two parameters of interest, i.e., tracking error and excess return, and the generalised objective function with constraints.

3.1.1 Problem formulation

To better understand the index-tracking model, it is necessary to define all the variables used. (Beasley et al., 2003). Let:

N – be the number of assets which form the portfolio index,

K – be the acceptable number of assets in the tracking portfolio,

ε_i – be the minimum proportion of portfolio invested in stock i,

δ_i – be the maximum proportion of portfolio invested in stock i. It allows restricting the exposure to stock i. As a result, any stock i is held so that $0 \leq \varepsilon_i \leq \delta_i \leq 1$ is preserved;

X_i – be the number of units of stock i in the present tracking portfolio,

T – be such as stocks and the benchmark have been managed along the period $0, 1, 2, \ldots, T$. So, T is the decision date when the present tracking portfolio $[X_i]$ may turn to a new one;

V_{it} – be the value of a unit of stock i $(i = 1, \ldots, N)$ at time t $(t = 0, \ldots, T)$,

I_t – be the value of the index at time t,

R_t – be the single-period continuous time return produced by the index: $R_t = ln\left(\frac{I_t}{I_{t-1}}\right)$ (with $t = 1, \ldots, T$),

14 Traded on stock exchanges, an ETF is an investment fund that aims to replicate an index to which refers (benchmark) through totally passive management.

C_{cash} – be the cash adjustments in the tracking portfolio: if it is positive, it means that there is new cash to invest; otherwise, cash already invested must be disinvested,

C – be the total value of the present tracking portfolio $[X_i]$ at time T plus cash adjustments: $C = \sum_{i=1}^{N} V_{it}X_i + C_{cash}$. The former addend calculates the summation of the value of all assets included in the existing tracking portfolio, multiplying the value of an asset V_{it} with the relative number of units X_i owned;

$F_i(\zeta, \theta, t)$ – be the transaction cost function that describes how transaction costs are caused by stock i in moving from holding ζ to θ units of the stock at time t. If $\theta \geq \zeta$, new units of stock i are purchased; if $\theta \leq \zeta$, units of stock i are sold. No stocks are bought or sold if $\theta = \zeta$, i.e. $F_i(\zeta, \theta, t) = 0$;

γ – be the quantity of C that may be used by transaction cost $(0 \leq \gamma \leq 1)$.

Then, there are two decision variables used in the model (Beasley et al., 2003):

x_i – denotes the number of units of stock i held in the new portfolio,

z_i – takes the value of 1 if the stock i is held in the new portfolio, 0 otherwise.

Then, the total transaction cost C_{trans} associated with the changing from the present portfolio $[X_i]$ to the new portfolio $[x_i]$ at time T is denoted as:

$$C_{trans} = \sum_{i=1}^{N} F_i(X_i, x_i, T), \tag{3.1}$$

where C_{trans} represents the total quantity of transaction costs of all stocks at time T.

Moreover, the single-period continuous time return[15] r_t of the new portfolio $[x_i]$ at time t $(t = 1, \dots, T)$ is linked to the preceding variables by (Beasley et al., 2003):

$$r_t = \ln\left(\frac{\sum_{i=1}^{N} V_{it}x_i}{\sum_{i=1}^{N} V_{it-1}x_i}\right). \tag{3.2}$$

3.1.2 Tracking error and excess return

The passive manager pursues to minimise the tracking error TE, which is represented by a function of the difference between returns r_t in the controlled portfolio and returns R_t in

15 The definitions of r_t and R_t assume that successive observations of stock/index values are made at equal time intervals.

the index. If r_t and R_t correspond for each value, i.e., $TE = 0$, then the passive manager's portfolio replicates well the benchmark. Beasley et al. (2003) differ among numerous economists, who consider that TE is interpreted as the variance of $\{(r_t - R_t)|t = 1, \dots, T\}$. To claim their definition, Beasley et al. take advantage of the so-called reduction ab absurdum: supposing TE is defined in terms of variance, it presumes that a portfolio, which regularly underperforms the benchmark, i.e., $r_t = R_t - X\ \forall t$ with $X > 0$ is a constant, has a TE equal to zero. The result would be meaningless because it overlooks the role of bias $(r_t - R_t)$.

Beasley et al. (2003) represent tracking error as:

$$TE = \frac{(\sum_{t\in S}|r_t - R_t|^{\alpha})^{1/\alpha}}{T}, \tag{3.3}$$

where[16] S is a proper interval of time in which r_t is in comparison with R_t and $\alpha > 0$ represents the power by which differences between portfolio returns and benchmark returns are penalised.

Beasley et al. (2003) aim to obtain a well-performing portfolio both in-sample and, especially, out-of-sample. Determine the most optimal value of α in-sample does not mean that it leads to suitable performance out-of-sample. For this reason, a weighting factor Δ_t is introduced to give more importance to recent time periods rather than the oldest ones. Equation (3.3) becomes:

$$TE = \frac{(\sum_{t\in S}\Delta_t|r_t - R_t|^{\alpha})^{1/\alpha}}{T}. \tag{3.4}$$

Then, Beasley et al. (2003) define excess return as a return exceeding the return on the benchmark. In short:

$$r^* = \sum_{t=1}^{T}(r_t - R_t)/T, \tag{3.5}$$

16 If $S = [t|r_t < R_t\ t = 1, 2, \dots, T]$, the interval of time reflects the downside risk because it includes all those times for which the index return is higher than the portfolio return.

where R_t is always a known constant over all the time. Overlooking R_t, the excess return r^* matches to the average return by period obtained by the tracking portfolio. Substituting both r_t from Equation (3.2) and the definition of R_t, Equation (3.5) may be re-written as:

$$r^* = \left(\ln\left[\frac{\sum_{i=1}^{N} V_{iT}x_i}{\sum_{i=1}^{N} V_{i0}x_i}\right] - \ln\left[\frac{I_T}{I_0}\right]\right)/T, \tag{3.6}$$

where r^* is tantamount to a comparison between the total return obtained by the portfolio and the total return obtained by the benchmark, covering the period of time $[0, T]$.

To sum up, tracking error and excess of return are two parameters of interest considered in the implementation of the objective function. If the Index-Tracking problem takes into account only the former parameter, different portfolios will compete only based on costs incurred by passive funds. The addition of excess of return allows losing a degree of TE to have a portfolio past returns of which are above the benchmark (Beasley et al., 2003).

3.1.3 Generalised formulation

Before proceeding, it is necessary to highlight that Equations (3.5) and (3.6) are specified so that TE and r^* share the same unit.

So, the Index-Tracking problem can be formulated in terms of TE and r^* as (Beasley et al., 2003):

$$\min_{\lambda} \lambda TE - (1-\lambda) r^* \tag{3.7}$$

subject to the following constraints:

$$\sum_{i=1}^{N} z_i \leq K, \tag{3.8}$$

$$\varepsilon_i z_i \leq V_{iT}x_i/C \leq \delta_i z_i \quad i = 1, \dots, N, \tag{3.9}$$

$$C_{trans} \leq \gamma C, \tag{3.10}$$

$$\sum_{i=1}^{N} V_{iT} x_i = C - C_{trans}, \tag{3.11}$$

$$x_i \geq 0 \quad i = 1, \dots, N, \tag{3.12}$$

$$z_i \in [0, 1] \quad i = 1, \dots, N. \tag{3.13}$$

Equation (3.7) represents the generalised objective function, where λ, ranging in the interval $[0, 1]$, indicates an implied compromise between TE and r^*. Two cases for λ may be identified. Fixing $\lambda = 0$ leads to the maximisation of excess return[17], while setting $\lambda = 1$ leads to the minimisation of tracking error.

Concerning the restrictions, Equation (3.8) is a cardinality constraint because it sets the number of stocks owned in the new portfolio $[x_i]$. In this way, $\sum_{i=1}^{N} z_i$ may be different from the number of stocks held in the present portfolio $[X_i]$ and be lower or equal to[18] K.

Equation (3.9) ensures two conditions: if the stock i is in the new portfolio, i.e. $z_i = 1$, then the number of units of the stock i is limited appropriately between the two boundaries ε_i and δ_i. Besides, if the stock i is not held in the new portfolio, i.e. $z_i = 0$, then the number of units of the stock i is equal to zero.

Equation (3.10) confines the total transaction costs contracted (both γ and C are known values). These costs allow obtaining that portfolio, which is characterized by the best historically combination of tracking error and excess return.

Equation (3.11) defines the cash balance constraint, which sets the value of the new portfolio equal to the difference between the total value of the present portfolio and the total transaction costs. This Equation, in addition to Equation (3.2), implies that the return of the new portfolio takes into account both cash change and transaction costs contracted.

Finally, Equation (3.12) fixes the number of units of the stock i greater or equal to zero, indicating that no short-selling is allowed. Furthermore, Equation (3.13) sets that the decision variable z_i cannot take values different from 0 or 1.

17 If λ is equal to zero, Equation (3.7) converts to $\min -r^* = \max r^*$.

18 Beasley et al. (2003) set the sum of z_i equal to K. This dissertation aspires to release this strict constraint, allowing the number of assets is equal to or lower than cardinality limit. In this way, a better and more suitable result is achievable.

To sum up, some final considerations are discussed. The above-proposed formulation refers to as a *rebalancing*, or *revision* problem, that is reviewing the existing portfolio and seeking for a new one with the best characteristics. However, a *creation* problem is possible to take into account, setting $X_i = 0 \ \forall i$ and using C_{cash} to buy stocks and construct the portfolio. Moreover, the analysed pattern is a strategic model that shows how the tracking portfolio should be composed. It does not provide further information on how and when to buy and sell stocks in the market, so it is not considered a tactical trading model (Beasley et al., 2003).

3.2 Cumulative Prospect Theory model

He and Zhou (2011) study an analytical framework of single-period portfolio selection model supporting the Kahneman and Tversky's cumulative prospect theory (CPT). To better understand the problem, it is necessary to recall some main preconditions. Behavioural investors identify themselves as agents who assess stocks in comparison with specific benchmarks than on final wealth conditions. Moreover, they evaluate gains and losses in different ways, are not always risk-averse, and are much more sensitive to losses rather than to gains. Finally, they often overestimate small probabilities and underestimate high ones.

These behaviours outline the formulation of a portfolio selection model. Investors set a specific reference point, which may be either a neutral outcome or a benchmark, that divides gains from losses. Besides, they are characterized by a value function that is concave for gains and convex for losses, and less steep for gains rather than for losses; they have a probability weighting function, which is not a linear transformation of probability measure and overweights small probabilities and underweights medium-high probabilities. So, people are less sensitive to differences in probabilities in the middle range.

He and Zhou (2011) hypothesise a market composed of one risky asset and one riskless. The former could be read as the market portfolio or an index fund. Let $[0, T]$ be an investment period, $r(T)$ be the riskless total return, and $R(T)$ be the stock total return. As a result, the risky asset total excess return, defined as $R(T) - r(T)$, is a random variable that follows a cumulative distribution function (CDF), indicated by F_T.

Imagine there is an investor identified by CPT preferences. So, let v^+ and v^- be the value function for gains and losses respectively, and w^+ and w^- from $[0,1]$ to $[0,1]$ be two probability weighting functions (occasionally also called *distortions*) respectively for probabilistic of gains and losses. The value function and weighting functions follow the Tversky-Kahneman's functional form introduced in the Chapter 1 (see Equations (1.5) and (1.6), respectively). Besides, the investor has a reference point, denoted by RP, which helps to divide gains from losses (He and Zhou, 2011).

In the beginning ($t = 0$), the investor has an amount of capital, denoted by M_0, to invest such that he maximises his objective function at time T. A part of the money, θ, is used for risky asset and the remainder for the riskless asset, let $x_0 = r(T)M_0 - RP$ be the departure of the riskless payoff from the reference point. If x_0 is positive, it means that the return of the total capital is better than the reference point (or benchmark) taken into consideration. Therefore, the final capital is the sum of two components: the product between the risky asset total excess return and the portion of the money invested in the risky asset; the return of the total capital of the riskless asset, that is $r(T)M_0$. In short:

$$\mathrm{X}(\mathrm{x}_0, \theta, T) = [R(T) - r(T)]\theta + x_0 + RP. \tag{3.14}$$

The value function of X may be rewritten using the Lebesgue-Stieltjes integral:

$$\begin{aligned} V(X) = \int_{RP}^{+\infty} v^+(x - RP)d\left[-w^+\big(1 - F(x)\big)\right] \\ - \int_{-\infty}^{RP} v^-(RP - x)d\left[w^-\big(F(x)\big)\right], \end{aligned} \tag{3.15}$$

where $F(x)$ is the CDF, namely $F_X(x)$, of X. The first integral represents the value function for gains, where the distortion function has decumulative probabilities. The second integral defines the value function for losses, where the distortion function exhibits the cumulative form (Tversky and Kahneman, 1992). Remind that Tversky and Kahneman (1992) define CPT value function (Equation (1.3)) merely for discrete values of X. Equation (3.15) is an expansion of Equation (1.3) that implements both discrete and continuous values of X, and it is equivalent to Tversky-Kahneman's formula if X is discrete (He and Zhou, 2011).

He and Zhou (2011) aim to estimate the CPT value of Equation (3.14) by applying Equation (3.15). It leads to the CPT value function of θ, which is denoted by $V(\theta)$. If an investor invests no money in risky asset, i.e., $\theta = 0$, then:

$$V(\theta) = \begin{cases} v_{+}(x_0) & if\ x_0 \geq 0 \\ -v_{-}(-x_0) & if\ x_0 < 0. \end{cases} \tag{3.16}$$

In this situation, the CPT value function of θ coincides with the Tversky-Kahneman's formula (see Equation (1.5)). However, if the investor spends some money on the risky asset, i.e., $\theta > 0$, then the investor's CPT value function is achieved by substituting some variables in Equation (3.15):

$$\begin{aligned} V(\theta) = &\int_{-\frac{x_0}{\theta}}^{+\infty} v^{+}(\theta t + x_0) d\left[-w^{+}\left(1 - F_T(t)\right)\right] \\ &- \int_{-\infty}^{-\frac{x_0}{\theta}} v^{-}(-\theta t - x_0) d\left[w^{-}\left(F_T(t)\right)\right], \end{aligned} \tag{3.17}$$

where $-\frac{x_0}{\theta}$ is the reference point that bounds both integrals, and it represents the ratio between the deviation of the riskless payoff from RP and the sum of money invested in the risky asset. $\theta t + x_0$ and $-\theta t - x_0$ are the arguments of value functions of gains and losses, and they take into consideration the deviation of the riskless payoff from the reference point and the amount of money invested in the risky asset (He and Zhou, 2011).

The last step is the formulation of the CPT portfolio selection model, which is summed up as:

$$\max_{\theta \geq 0} V(\theta), \tag{3.18}$$

where the constraint $\theta \geq 0$ ensures that there is no non-negative amount of money invested in the risky asset.

To conclude, remind that if $V(\theta)$ has no probability weighting, it degrades to the usual expected value, even though the whole value function is still S-shaped rather than globally concave (He and Zhou, 2011).

3.2.1 Further representative functions

Risk attitude to probabilities influences the shape of the weighting function. Empirical evidence recommends that, given a specific value $0 < p^* < 1$, the curvature of probability weighting functions tends to be concave for probabilities in the interval $(0, p^*)$, and convex in the interval $(p^*, 1)$. Besides, small probabilities are overestimated, and large ones are overestimated. These features indicate that the probability weighting function has the inverse-S shape, as illustrated in Figure 1.4 in the Chapter 1 (Tversky and Kahneman, 1992). It is necessary to highlight that Tversky and Kahneman are not the only economists who seek to provide typical probability weighting functions. Many others strive to find new functions that better explain probabilistic risk behaviour with the referred characteristics. Most of them measure parameters through various empirical experiments. Karmarkar (1978), Tversky and Fox (1995) and Wu and Gonzalez (1996, 1999) propose a weighting function form that is similar to Tversky and Kahneman's one.

Karmarkar (1978) proposes the following function:

$$\mathrm{w(p)} = \frac{\mathrm{p}^{\gamma}}{p^{\gamma} + (1-p)^{\gamma}}, \tag{3.19}$$

where $0 < \gamma < \infty$ and w ranges between zero and one. If $\gamma = 1$, weights match with the true probabilities. Setting $\gamma \neq 1$, the mapping has three common points: 0, 0.5, and 1. Figure 3.1 illustrates how the weighting function changes by varying the parameter γ. If γ tends to zero, all weights are considered to be equally likely (see plot on the top-right). For $\gamma > 1$, the Equation (3.19) transforms probabilities such that events with probability greater than 0.5 are more and more certain, and events with probability lower than 0.5 are quite doubtful. This effect is more emphasised if γ increases.

Later, Tversky and Fox (1995) suggest a further function that allows weighting positive and negative outcomes in different ways:

$$\begin{aligned} \mathrm{w}^{+}(\mathrm{p}) &= \frac{\delta^{+}\mathrm{p}^{\gamma^{+}}}{\delta^{+}\mathrm{p}^{\gamma^{+}} + (1-p)^{\gamma^{+}}}, \\ \mathrm{w}^{-}(\mathrm{p}) &= \frac{\delta^{-}\mathrm{p}^{\gamma^{-}}}{\delta^{-}\mathrm{p}^{\gamma^{-}} + (1-p)^{\gamma^{-}}}, \end{aligned} \tag{3.20}$$

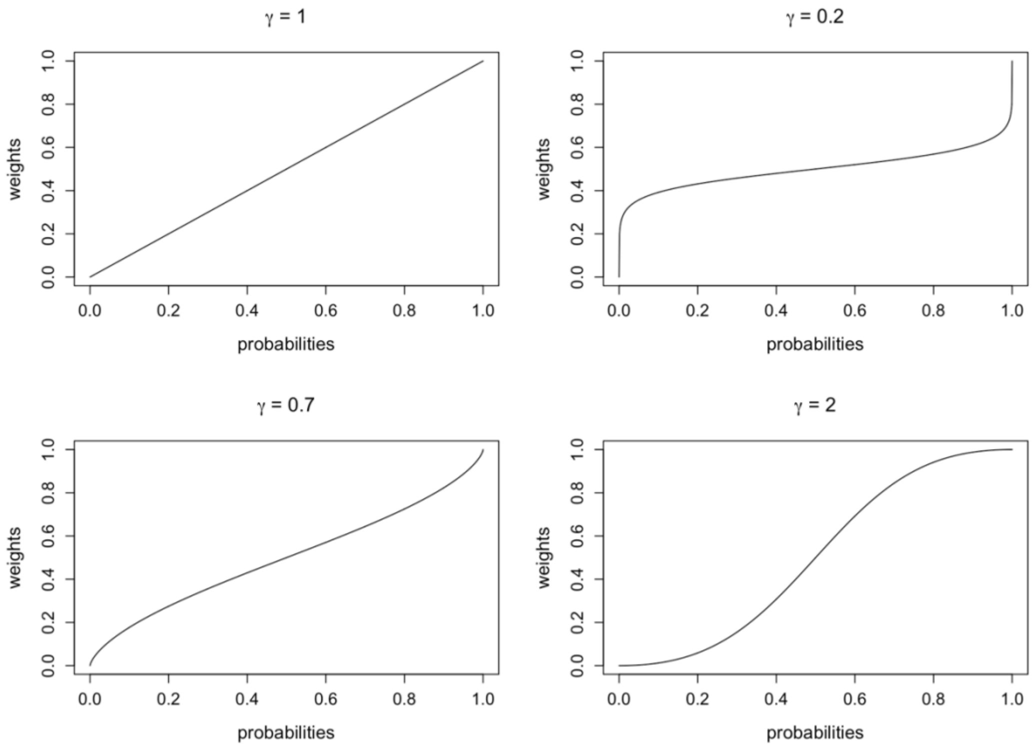


Figure 3.1 Probability weighting under Karmarkar (1978) function
Source: personal processing.

where $\delta^+, \delta^- > 0$, $\gamma^+ > 0$, and $\gamma^- < 1$.

Wu and Gonzalez (1996) resume the work of Karmarkar, and offer the following two weighting functions:

$$w(p) = \frac{p^\gamma}{[p^\gamma + (1-p)^\gamma]^\delta}, \tag{3.21}$$

which is a general case of Equation (3.19), where $\delta = 1$.

Afterward, an alternative variant is still introduced by Gonzalez and Wu (1999), who attempt to model w in such a way that a parameter describes its degree of curvature and another parameter describes its elevation. The former property, called *discriminability*, regards the behaviour of discrimination towards to probabilities in the 0-1 interval. The latter, named *attractiveness*, relates to the size of overweighting and underweighting. In short:

$$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}, \tag{3.22}$$

where the parameter δ measures elevation, while γ measures curvature. Equation (3.22) is a widely-spread version, used before them by Lattimore et al. (1992), Goldstein and Einhorn (1987), Tversky and Fox (1995), and many others.

Furthermore, many other economists who strive to provide innovative probability weighting functions. For example, Lattimore et al. (1992) considers the following function:

$$w(p) = \frac{\delta p_i^{\gamma}}{\left[\delta p_i^{\gamma} + \sum_{k=1}^{n} p_k^{\gamma}\right]}, \tag{3.23}$$

where $\delta, \gamma > 0$ are arbitrary parameters, n is the number of outcomes, and $i, k = 1, 2, \dots, n$, with $k \neq i$. Setting $\delta = \gamma = 1$, it is possible to obtain the restricted version, i.e., $w(p) = p_i / \sum_{i=1}^{n} p_i$, and it is consistent with the hypothesis of Expected Utility Theory. Furthermore, regardless of the values of δ and γ, $w(0) = 0$ and $w(1) = 1$, which is consistent with CPT model proposed by Tversky and Kahneman (1992). The parameter γ describes the inflection of weighting function and its direction, for example, concave-convex. In his experiments, Lattimore et al. (1992) observes that people overweight small probabilities and underweight large ones, determining $\delta \leq 1$ and $\gamma < 1$.

A further alternative form is the one studied by Prelec (1998), called *compound invariant* function[19]:

$$w(p) = e^{-(-\ln p)^{\gamma}}, \tag{3.24}$$

where γ, mapping from $(0, 1)$, is the only parameter that characterized the function and $p \in (0, 1)$. The main features of Equation (3.24) are the following: it is regressive, S-shaped, and it does not depend by γ only when $p = 1/e$, which is the inflection point. Figure 3.2 gives an illustrative representation of the latter property. It is remarkable to asses that, for $\gamma \to 1$, w tends to approx the linear expected utility case. If $\gamma \to 0$, the weighting function approx a flat curve, except for both extremes of the probability interval.

---

19 Prelec (1998) also introduces two additional probability weighting functions: the exponential power function, i.e., $w(p) = e^{-\frac{\delta}{\gamma}(1-p^{\gamma})}$, with $\gamma \neq 0, \delta > 0$; the projection invariance power hyperbolic logarithm function, i.e., $w(p) = (1 - \gamma \ln p)^{-\frac{\delta}{\gamma}}$, with $\gamma, \delta > 0$.

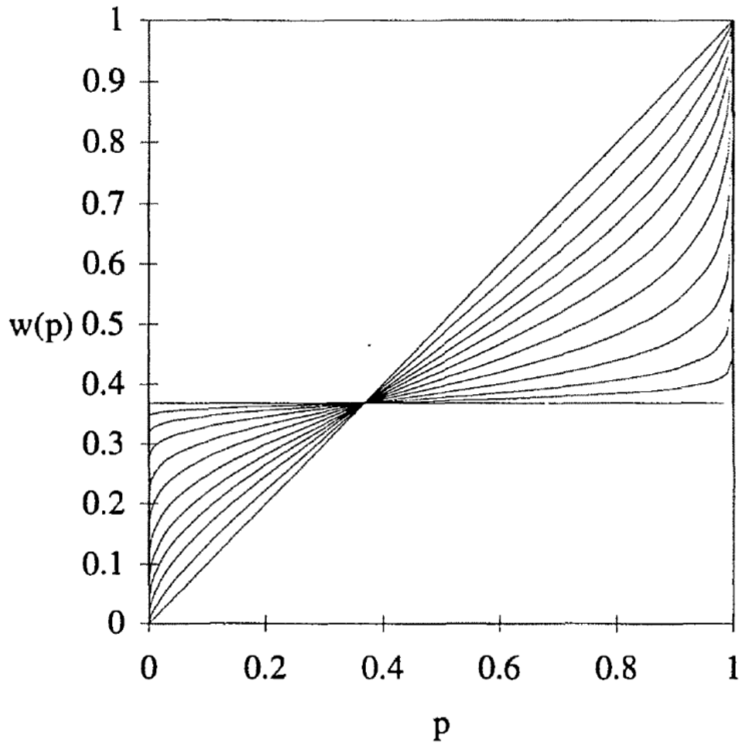


Figure 3.2 Probability weighting under Prelec (1998) function, plotted for different γ values
Source: Prelec, 1998.

Prelec (1998) expands Equation (3.24), providing an additional form:

$$w(p) = e^{-\delta(-\ln p)^{\gamma}}, \tag{3.25}$$

where the parameter δ measures the elevation of w, while γ measures the curvature.

3.3 Cumulative Prospect Theory model for index tracking

This section explains a model that combines CPT and portfolio selection based on the tracking error. It follows and develops the work of Grishina et al. (2017), who implement only the value function of the Prospect Theory model for index tracking, overlooking the weighting function of the model.

Let be the following lottery:

$$(r_1, p_1), (r_2, p_2), \ldots, (r_{RP}, p_{RP}), \ldots, (r_{M-1}, p_{M-1}), (r_M, p_M) \tag{3.26}$$

where $m = 1, 2, \dots, RP, \dots, M-1, M$ is the number of prospects, RP is the reference point and based on the preferences of the decision-maker, and (r_m, p_m) indicates that the decision-maker gains r_m with probability p_m. Besides, it sets the restriction regarding the sum of all probabilities equal to 1. In such a lottery, the decision-maker faces three different scenarios:

1. if $m < RP$, then $r_m < r_{RP}$ and the decision-maker loses;
2. if $m = RP$, then $r_m = r_{RP}$ and the decision-maker gains nothing;
3. if $m > RP$, then $r_m > r_{RP}$ and the decision-maker wins.

From the perspective of a behavioural investor, it is necessary to transform p and r into the CPT probability weighting function w and the value function v (see the respective Equations (1.5) and (1.6)). The prospect value, defined in terms of π and v, is as follows:

$$V(f) = \sum_{m=1}^{M} \pi_m v(r_m), \tag{3.27}$$

where v is concave if $r > r_{RP}$, and convex if $r < r_{RP}$. Section 1.4 describes all features and comments regarding π, v, and V.

The CPT model strives to determine the optimal portfolio that maximises Equation (3.27), where decision variables are the weights of stocks, denoted as x, held in the portfolio. The portfolio selection problem is formulated as follows (Grishina et al., 2017):

$$\max_{\text{x}} V(f) = \sum_{i=1}^{M} w_i v\left(\sum_{j=1}^{N} r_{ij} x_{ij}\right). \tag{3.28}$$

Note that Grishina et al. (2017) do not transform the probabilities. Equation (3.28) is replaced in the application by:

$$\max_{\text{x}} V(f) = \sum_{i=1}^{M} \pi_i v\left(\sum_{j=1}^{N} r_{ij} x_{ij}\right). \tag{3.29}$$

The last objective function is subject to the following restrictions:

$$\sum_{j=1}^{N} x_j = 1, \tag{3.30}$$

$$x_j \geq 0, \qquad j = 1, \dots, N. \tag{3.31}$$

The product between stocks returns, indicated by r_{ij}, and their respective portfolio weights x_{ij} defines the value function v, which represents the return of the portfolio. N indicates the number of stocks held. Equation (3.30) commits the investment of all money into the portfolio, while Equation (3.31) does not allow short-selling.

It is necessary to highlight that the CPT model utilizes a reference point that distinguishes gains from losses in each period, which is similar to the IT model that employs the benchmark as a reference point. Therefore, it is possible to realise a framework where CPT and the index tracking problem coexist by adjusting the value of the reference point. The problem formulation then becomes as follows:

$$\max_{\mathrm{x}} V(f) = \sum_{i=1}^{M} \pi_i v \left(\sum_{j=1}^{N} r_{ij} x_{ij} - rb_i \right), \tag{3.32}$$

where rb is the return of the benchmark. If before the problem formulation focuses on the maximisation of the portfolio returns (Equation 3.29), now the new problem formulation focuses on the maximisation of the deviation of the CPT portfolio returns from the benchmark returns in each period (Equation 3.32). Thus, the purpose of the portfolio is not to have a positive return, but to be able to beat the benchmark returns over time. Restrictions remain the same considered before.

It is evident to recognise that the intricacy of the problem is substantial due to the non-linearity and the non-convexity of the objective function. For this reason, there are no exact methods that allow solving the problem in a short time and with limited information. Even if it is hard to obtain a satisfactory solution to this problem, several economists and traders adopt metaheuristics, inexact methods that find an answer to this kind of portfolio optimisation problems (Grishina et al., 2017).

The following chapter illustrates three metaheuristic approaches widespread for portfolio optimisation problems: Particle Swarm Optimisation (Eberhart and Kennedy, 1995), Bee

Algorithm (Pham et al., 2005) and Ant Colony Optimisation (Dorigo and Di Caro, 1999). This dissertation uses Particle Swarm Optimisation to solve the CPT model for index tracking.

Chapter 4 - Evolutionary algorithms for portfolio optimisation problems

Evolutionary algorithms (EA) include a range of optimisation methods that follow the laws of natural evolution. Economists employ them to give appropriate solutions to problems, which are unlikely solved by other techniques. This category includes several optimisation problems, such as portfolio optimisation. In certain situations, it is computationally hard to obtain the correct solution, but EA allows to detect a close-optimal solution if it exists. EAs are part of the metaheuristic approaches, which comprises Particle Swarm Optimisation (Eberhart and Kennedy, 1995), Bee Algorithm (Pham et al., 2005) and Ant Colony Optimisation (Dorigo and Di Caro, 1999).

It is necessary to define metaheuristics. They consist of a series of steps on specific objective functions, and they apply to arbitrary problems. Of course, both objective functions and the operations need to be adjusted to the problem. This chapter discusses three kinds of metaheuristics: Particle Swarm Optimisation, Differential Evolution Algorithm, and Genetic Algorithm. Section 4.1 introduces Particle Swarm Optimisation, which is a bio-inspired stochastic algorithm that solves optimisation problems. It gets inspiration from the unpredictable choreography of flocks of birds, and it aims to determine the rules that explain their way of flying in sync. This thesis employs Particle Swarm Optimisation to solve the portfolio selection problem illustrated in Section 3.2.2. Sections 4.2 and 4.3 discuss respectively Differential Evolution algorithm (Storn and Price, 1997) and Genetic Algorithm (Holland, 1975). Both methods are widespread for the capacity to handle with highly complex optimisation problems.

4.1 Particle Swarm Optimisation

To solve the portfolio optimisation problem, in which the tracking error and CPT consider, the metaheuristic Particle Swarm Optimisation (PSO) applies. A metaheuristic algorithm is a high-level procedure that aims to produce, detect, or choose a satisfactory solution to an optimisation problem. Commonly, it requires few assumptions on the optimisation problem, and it achieves promising results even with rough or incomplete information.

The term bio-inspired stands for the idea of replication of behaviour of animals (e.g., birds, fish), which hunt for food and avoid predators in flocks or schools. This hunt leads these animals to explore surroundings near the group up to a certain distance, and then return to the group with new information about discovered food sources. In this way, the entire group may go to the food source (Kruse et al., 2013). Social cooperation is at the basis of the PSO. Birds within the flock learn from each other and, because of knowledge acquired, move to locate in the "better" beforehand occupied position. In other words, the best particle (best problem-solution) discovered so far lures all the other particles, who then emulate it (Khanesar et al., 2007).

Suppose to take into consideration a flock of birds (or particles). Two vectors characterize each member: the former is the *position vector* and indicates the location in the multidimensional search-space; the other is the *velocity vector* and defines the succeeding movement. Additionally, the PSO algorithm updates the velocity of a member depending on both current speed and the most suitable location it has examined up to now and additionally based on the global most suitable position examined by the flock. Then, the PSO process repeats a set amount of times or till it satisfies another predefined termination criterion (Khanesar et al., 2007).

In terms of quantity, PSO establishes that every member of the group, denoted by sw, represents a potential solution to the optimisation problem. Then, let N be the number of variables of the optimisation problem, Ω be the search-space in real values, i.e., $\Omega \subseteq \mathbb{R}^n$ and f be the fitness function such that $f\colon \mathbb{R}^n \to \mathbb{R}$ (Kruse et al., 2013).

In every iteration, each *i*-th particle (where $i = 1, \dots, sw$) characterizes by a position $\boldsymbol{x}_i$ in the search-space and a velocity $\boldsymbol{v}_i$. The starting location of each particle is randomly chosen in Ω; the starting speed is randomly selected and helps to determine the first movement (Kruse et al., 2013).

At each iteration, PSO adjusts speed and location of the *i*-th particle to seek for an adequately better solution, according to (Kruse et al., 2013):

$$\begin{aligned}&\boldsymbol{v}_i(j+1) = \alpha\boldsymbol{v}_i(j) + \beta_1\left(\boldsymbol{x}_i^{(local)}(j) - \boldsymbol{x}_i(j)\right) + \beta_2\left(\boldsymbol{x}^{(global)}(j) - \boldsymbol{x}_i(j)\right),\\ &\boldsymbol{x}_i(j+1) = \boldsymbol{x}_i(j) + \boldsymbol{v}_i(j+1).\end{aligned} \tag{4.1}$$

Concerning the first equation, it composes of three parts: the first part represents the past speed of the *i*-th particle; the second and third parts, respectively called *cognition* and *social*, are variables that adjust the future velocity of the *i*-th particle. In details, $\boldsymbol{x}_i^{(local)}$ is the *local memory* of the particle, in other words, the best location explored by *i*-th particle up to iteration j. In mathematical terms, it is defined as[20] (Kruse et al., 2013):

$$\boldsymbol{x}_i^{(local)} = \boldsymbol{x}_i\left(\arg\max_{u=1}^{j} f\left(\boldsymbol{x}_i(u)\right)\right). \tag{4.2}$$

By analogy, $\boldsymbol{x}^{(global)}$ is the *global memory* of the entire flock or the best location explored by each particle up to iteration j. It is defined as (Kruse et al., 2013):

$$\boldsymbol{x}^{(global)}(j) = \boldsymbol{x}_k^{(local)}(j) \quad with\ k = \arg\max_{i=1}^{sw} f\left(\boldsymbol{x}_i^{(local)}\right). \tag{4.3}$$

From Equation (4.1), the parameters β_1 and β_2 manage the potential pull of $\boldsymbol{x}_i^{(local)}$ and $\boldsymbol{x}^{(global)}$. At the drawn of PSO, Eberhart and Kennedy (1995) make several simulations to find proper values for β_1 and β_2. If both parameters set high, the flock immediately reaches the local minima; on the contrary, the flock wanders around the goal, slowly moving closer, and lastly arrives on the target. Moreover, they estimate that if β_1 set with a much higher value than β_2, many isolated particles excessively wander in the search-space; otherwise, with moderately high values β_2, the flock hastens precipitately to local minima. Eberhart and Kennedy soon realise that there is no good reason whether β_1 and β_2 should be higher, so they substitute both terms by 2. In this way, the algorithm does not become unstable due to

20 The mathematical function $\arg max$ stands for arguments of the maxima, and it selects the elements of the domain such that the function outputs maximise.

increments with no particle speed control. Nonetheless, Kruse et al. (2013) propose to randomly select β_1 and β_2 at each iteration from the interval $[0, a)$, where a is an arbitrary integer number.

Shi e Eberhart (1998) introduce the inertia weight α, which is necessary to offset the trade-off derived by the local and global searches[21]. Following their simulations, they recommend choosing α in the range $[0.9, 1.2]$ because PSO has more possibility to locate the global optimum in a moderate number of iterations.

Nonetheless, recent studies prefer using linearly decreasing functions over time to update α. As time progresses, the impact of $\boldsymbol{v}_i$ of the *i*-th particle contracts, therefore the relative pull of $\boldsymbol{x}_i^{(local)}$ and $\boldsymbol{x}^{(global)}$ increases. Thus, at the beginning the particles have a high speed due to the attraction of the global memory; later, the velocity decreases more and more until particles ultimately almost stop at the best position found by the flock (Kruse et al., 2013).

Khanesar et al. (2007) provide a point summary of a PSO:

1. initialise the population of particles on the search-space, randomly selecting velocity and position;
2. estimate the fitness function f of each individual, employing its current location;
3. compare the performance of every particle to its most suitable performance so far: if $f(\boldsymbol{x}_i) < f\left(\boldsymbol{x}_i^{(local)}\right)$, then $f\left(\boldsymbol{x}_i^{(local)}\right) = f(\boldsymbol{x}_i)$, $\boldsymbol{x}_i^{(local)} = \boldsymbol{x}_i$;
4. compare the performance of every particle to the global most suitable particle so far: if $f(\boldsymbol{x}_i) < f\left(\boldsymbol{x}^{(global)}\right)$, then $f\left(\boldsymbol{x}^{(global)}\right) = f(\boldsymbol{x}_i)$, $\boldsymbol{x}^{(global)} = \boldsymbol{x}_i$;
5. adjust the speed and location based on Equation (4.1);
6. repeat the procedure from step 2 until any predefined termination criterion meets.

The following box represents a plausible representation of the pseudo-code (Kruse et al., 2013, p. 272):

[21] The Eberhart-Kennedy's velocity equation overlooks any parameter that modifies the past speed.

function pso (*sw*: int, *a*, *b*, *c*: real) : array of real;
begin (∗ *sw*: number of particles ∗)
$j \leftarrow 0$; $q \leftarrow -\infty$; (∗ *a*, *b*, *c*: update parameters ∗)
for $i \in \{1, \ldots, sw\}$ **do begin** (∗ initialize the particles ∗)
$\boldsymbol{v}_i \leftarrow 0$; (∗ initialize velocity and position ∗)
$\boldsymbol{x}_i \leftarrow$ choose a random point of $\Omega = \mathbb{R}^n$;
$\boldsymbol{x}_i^{(local)} \leftarrow \mathbf{x}_i$; (∗ initialize the local memory ∗)
if $f(\boldsymbol{x}_i) \geq q$ **then begin** $\boldsymbol{x}^{(global)} \leftarrow \boldsymbol{x}_i$; $q \leftarrow f(\boldsymbol{x}_i)$; **end**
end (∗ compute initial global memory ∗)
repeat (∗ update the swarm ∗)
$j \leftarrow j + 1$; (∗ count the update step ∗)
for $i \in \{1, \ldots, sw\}$ **do begin** (∗ update the local and global memory ∗)
if $f(\boldsymbol{x}_i) \geq f(\boldsymbol{x}_i^{(local)})$ **then** $\boldsymbol{x}_i^{(local)} \leftarrow \boldsymbol{x}_i$; **end**
if $f(\boldsymbol{x}_i) \geq f(\boldsymbol{x}^{(global)})$ **then** $\boldsymbol{x}^{(global)} \leftarrow \boldsymbol{x}_i$; **end**
end
for $i \in \{1, \ldots, sw\}$ **do begin** (∗ update the particles ∗)
$\beta_1 \leftarrow$ sample random number uniformly from [0, a);
$\beta_2 \leftarrow$ sample random number uniformly from [0, a);
$\alpha \leftarrow b/j^c$; (∗ example of a time dependent α ∗)
$\boldsymbol{v}_i(j+1) \leftarrow \alpha \cdot \boldsymbol{v}_i(j) + \beta_1 \left(\boldsymbol{x}_i^{(local)}(j) - \boldsymbol{x}_i(j)\right) +$
$\beta_2 \left(\boldsymbol{x}^{(global)}(j) - \boldsymbol{x}_i(j)\right)$;
$\boldsymbol{x}_i(j+1) \leftarrow \boldsymbol{x}_i(j) + \boldsymbol{v}_i(j)$; (∗ update velocity and position ∗)
end
until termination criterion is fulfilled;
return $\boldsymbol{x}^{(global)}$; (∗ return the best point found ∗)
end

4.2 Differential evolution algorithm

Storn and Price (1997) developed a reliable and versatile function optimizer, which is so-called the differential evolution (DE) algorithm. It is a revolutionary and useful approach that well-spreads all over the world due to its innovative features. First of all, DE is meant to be a stochastic direct search method, which may handle with nonlinear, non-differentiable, and multimodal cost functions[22]. Then, it is relevant to consider the computational time demanded by the solver to complete the task: some evaluations may take minutes, hours, or even days. The possible solution to this problem is to resort to either a network of computers

22 Cost functions refer to those objective functions represented by minimization problems.

or a parallel computer[23]. DE can guarantee remarkable results in a short period because the stochastic perturbation of the population vectors considered in the problem may compute independently. Moreover, DE is ease of use because the user needs to input only a few robust control variables to reach the minimization problem and it is necessary less than 30 lines of C-code to write the core search engine. Last but not least, DE has suitable converge properties that help to find consistent solutions under various kind of conditions.

Price et al. (2006) divide the entire process into four phases: *initialization*, *mutation*, *crossover*, and *selection*. After describing all steps in detail, *pseudocode* will be shown at the end to give an idea of the process behind it.

The first phase consists in defining the population structure taken into consideration. DE employs two vector populations, which both include Np D-sized vectors of real-valued parameters. The present population $P_{\boldsymbol{x}}$ consists of the so-called *target vector*, denoted by $\boldsymbol{x}_{i,g}$, which had been considered to be satisfactory as initial values. In short:

$$\begin{gathered} P_{\boldsymbol{x},g} = \left(\boldsymbol{x}_{i,g}\right), \quad i = 0,1,\dots,Np-1, \quad g = 0,1,\dots,g_{\max}, \\ \boldsymbol{x}_{i,g} = \left(x_{j,i,g}\right), \quad j = 0,1,\dots D-1, \end{gathered} \tag{4.4}$$

where $g = 0,1,\dots,g_{\max}$ is the generation to which a vector refers, $i = 0,1,\dots,Np-1$ is the population index ascribed to each vector, and $j = 0,1,\dots D-1$ is the parameter index within the two vectors. Price et al. (2006) propose a simple way to obtain the initial value $x_{j,i,g}$. First, they define the lower and upper bounds for all parameters D entered by the user. Then, they gather these values into two vectors D-sized, namely $\boldsymbol{b_L}$ and $\boldsymbol{b_U}$, which correspond respectively to the lower and upper bounds. Lastly, a value within the prefixed interval is ascribed to all parameters by a random number generator, which takes the following form:

$$x_{j,i,0} = rand_j(0,1)\left(b_{j,U} - b_{j,L}\right) + b_{j,L}, \tag{4.5}$$

[23] Parallel computing is a kind of process in which many computations are conducted simultaneously. Massive problems are usually subdivided into smaller ones, which may be resolved at the same time.

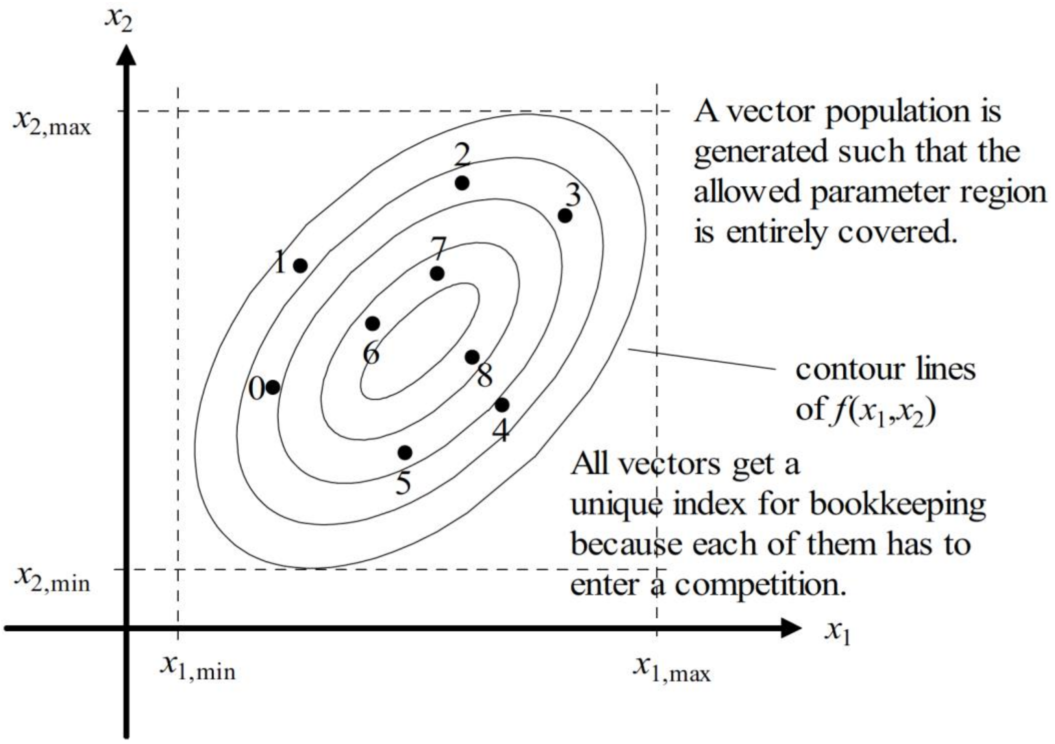


Figure 4.1 Initializing the DE population
Source: Price et al. (2006).

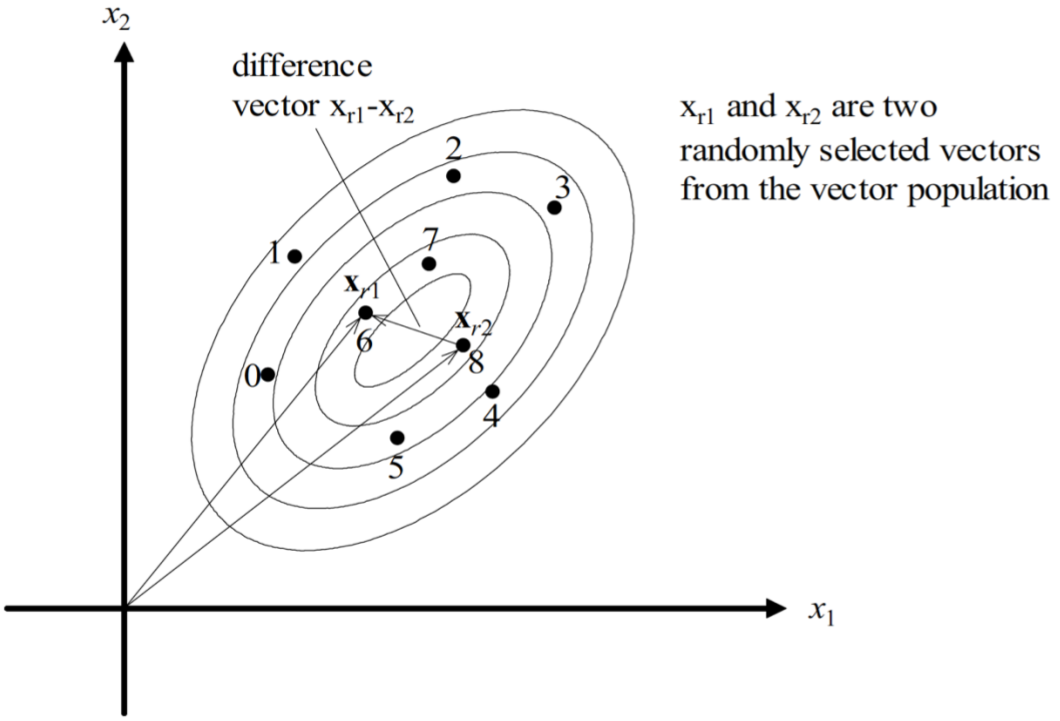


Figure 4.2 Generating the perturbation: $\boldsymbol{x}_{r1} - \boldsymbol{x}_{r2}$. *The index generation is omitted to simplify the notation*
Source: Price et al. (2006).

where $x_{j,i,0}$ is the value attributed to the j^{th} parameter of i^{th} vector of generation 0. The function $rand_j(0,1)$ gives back a value included in the interval $[0,1)$, which is multiplied by the difference between the upper and lower bounds of each j^{th} parameter; the product is then summed by the respective lower bound. As a result, each parameter obtains a new random value within the prefixed interval. Figure 4.1 represents the *initialization* phase.

As soon as the initialization finishes, DE mutates and rearranges the population to generate a population of Np trial vectors. DE randomly selects two population vectors $\boldsymbol{x}_{r1,g}$ and $\boldsymbol{x}_{r2,g}$, which are then subtracted each other, obtaining the so-called *difference vector* (see Figure 4.2). Equation (4.22) displays the combination of three different, randomly selected vectors, which generates a mutant vector $\boldsymbol{v}_{i,g}$:

$$\boldsymbol{v}_{i,g} = \boldsymbol{x}_{r3,g} + F\left(\boldsymbol{x}_{r1,g} - \boldsymbol{x}_{r2,g}\right), \tag{4.6}$$

where $F \in (0, 1+)$ is the scale factor represented by a constant, positive and real number, which regulates the speed at which the population grows[24], $r_1, r_2, r_3 \in \{1,2, \dots, Np\}$ are random, integer and respectively different indexes, and $\boldsymbol{x}_{r3,g}$ is the *base vector* randomly selected, distinct from the target vector index i. Consider that the indices $r1$, $r2$, and $r3$ must be different from each other and the target vector index. As a consequence, $Np \geq 4$ to allow the previous condition.

To complete the differential mutation strategy, DE uses the *uniform crossover*. The latter needs to improve the heterogeneity of the perturbed parameter vectors. To do so, DE compares each vector with the mutant vector:

$$\boldsymbol{u}_{i,g} = u_{j,i,g} = \begin{cases} v_{j,i,g} \;\; if\; \left(rand_j(0,1) \leq Cr \; or \; j = j_{rand}\right) \\ x_{j,i,g} \;\; if \; \left(rand_j(0,1) > Cr \; and \; j \neq j_{rand}\right) \end{cases}, \tag{4.7}$$

where $Cr \in [0, 1]$ is the crossover probability prefixed by the user, and it compares to the result of a random number generator, $rand_j(0, 1)$. If the crossover probability is greater than or equal to the random number, the trial vector copies the parameter from the mutant $\boldsymbol{v}_{i,g}$; otherwise, it inherits the parameter from the vector $\boldsymbol{x}_{i,g}$. Furthermore, j_{rand} represents the trial parameter whose index is randomly picked, and it guarantees that the trial vector

[24] Even if there is no upper bound, F rarely assumes value greater than 1.

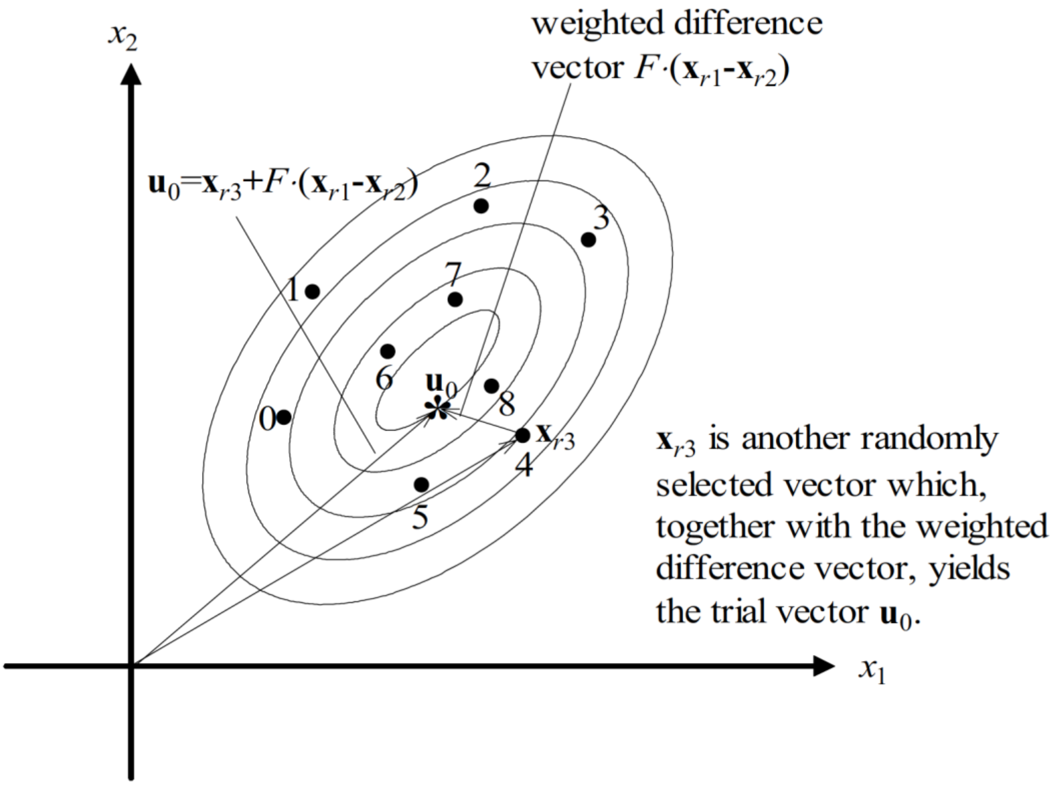


Figure 4.3 Mutation and crossover phases
Source: Price et al. (2006).

obtains at least one parameter from the mutant vector. In this way, $\boldsymbol{u}_{i,g}$ does not duplicate $\boldsymbol{x}_{i,g}$. Figure 4.3 illustrates the passages previously described.

The last step is the *selection* phase. To determine if the trial vector should turn into a member of the next generation $g+1$, the trial vector obtained compares with the target vector, using the greedy criterion[25]. If $\boldsymbol{u}_{i,g}$ yields a lower or equal objective function value than that of $\boldsymbol{x}_{i,g}$, then it takes the place of the target vector; otherwise, the current target vector keeps its place for one further generation. Equation (4.8) summarises this process:

$$\boldsymbol{x}_{i,g+1} = \begin{cases} \boldsymbol{u}_{i,g} & if\ f(\boldsymbol{u}_{i,g}) \leq f(\boldsymbol{x}_{i,g}) \\ \boldsymbol{x}_{i,g} & if\ f(\boldsymbol{u}_{i,g}) > f(\boldsymbol{x}_{i,g}) \end{cases}. \tag{4.8}$$

25 Used by several direct search methods, it helps to accept a new parameter vector if it decreases the cost value function (Storn and Price, 1997).

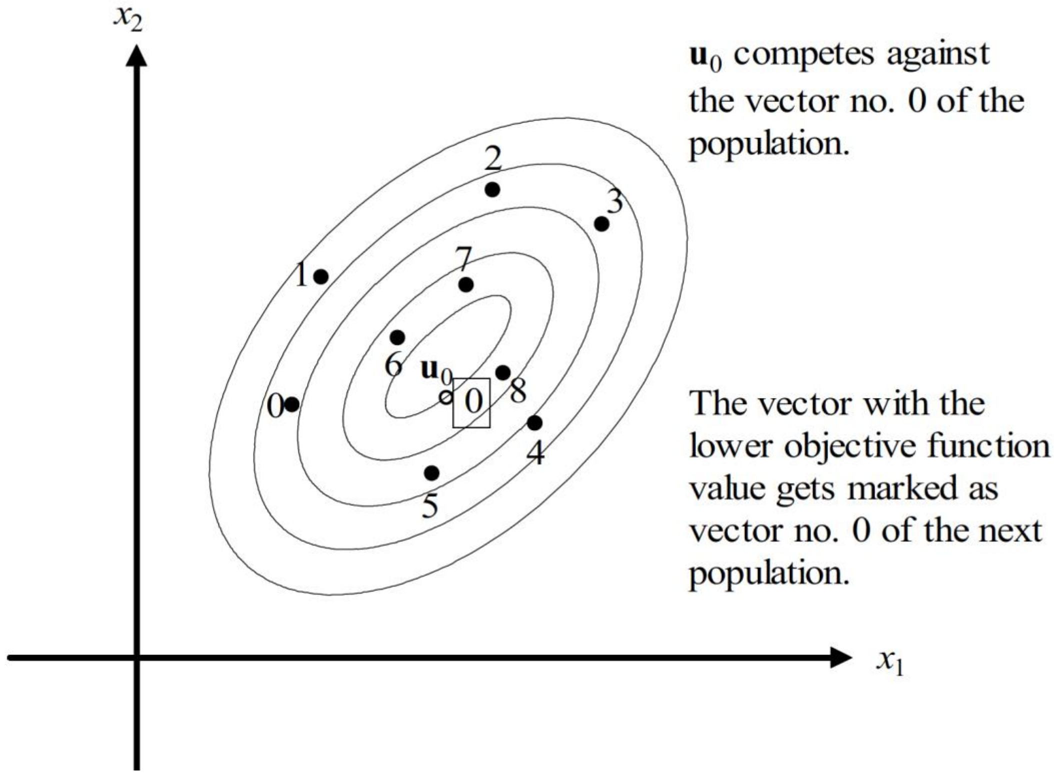


Figure 4.4 Selection phase: $\boldsymbol{u}_0$ *takes the place of the target vector with index 0 in the next generation because it has a lower cost function value*
Source: Price et al. (2006).

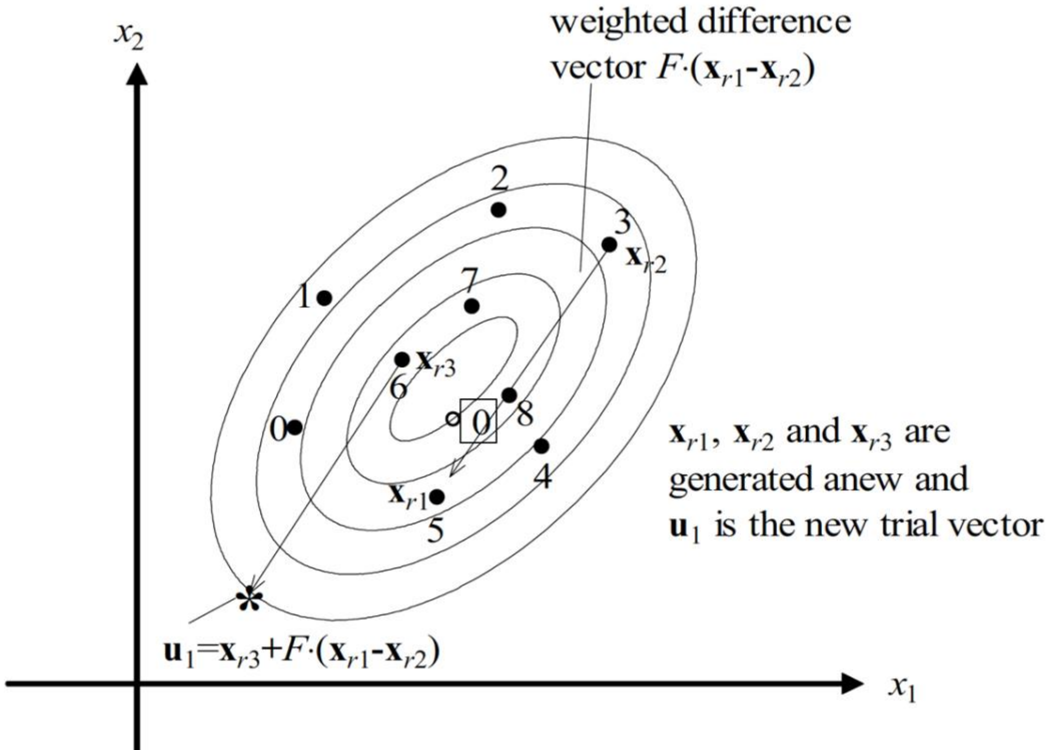


Figure 4.5 DE randomly mutates again, and then it produces another trial vector $\boldsymbol{u}_1$
Source: Price et al. (2006).

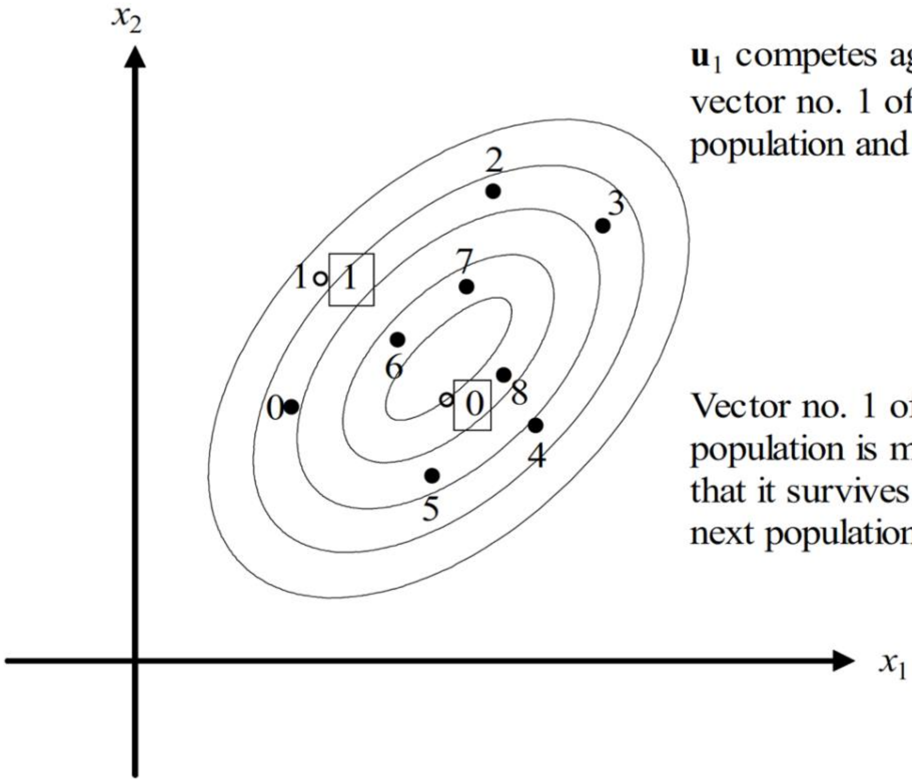


Figure 4.6 Selection phase: this time, $\boldsymbol{u}_1$ does not have a lower cost function value, so the target vector keeps its place
Source: Price et al. (2006).

Figure 4.4 illustrates the selection of a trial vector that successfully replaces the target vector, while Figures 4.5 and 4.6 show a new random trial vector, which, however, does not generate a lower cost function value than the target vector.

After this step, the process starts again with mutation, followed by crossover and selection, and it replicates until the optimum solution is determined, or a predefined termination criterion achieves, e.g., the number of generations attains $g_{\max}$.

There are several ways to represent the DE algorithm. This dissertation exhibits C-style pseudo-code. Let $r1$, $r2$, and $r3$ be different indices from each other and the target index i and suppose the process keeps going until the trial population is complete. The pseudo-code takes the form of (Price et al., 2006, p. 42):

```
// initialize...
do // generate a trial population
{
      for (i=0; i<Np; i++) // r1!=r2!=r3!=i
      {
```

```
            do r1=floor(rand(0,1)*Np); while (r1==i);
            do r2=floor(rand(0,1)*Np); while (r2==r1 or r2==i);
            do r3=floor(rand(0,1)*Np); while (r3==r2 or r3==r1 or
            r3==i);
            jrand=floor(D*rand(0,1));
            for (j=0; j<D; j++) // generate a trial vector
            {
                  if (rand(0,1)<=Cr or j==jrand)
                  {
                        u_{j,i}=x_{j,r1}+F*(x_{j,r2}-x_{j,r3}); //check for out-of-
                        bounds ?

                  }
                  else
                  {
                        u_{j,i}=x_{j,i};
                  }
            }
      }
      // select the next generation
      for (i=0; i<Np; i++)
      {
      if ( f(u_i)<=f(x_i) ) x_i=u_i;
      }
} while (termination criterion not met);
```

4.3 Genetic algorithms

Holland was an American scientist who first formulated the concept of the genetic algorithm (GA) in 1975. Frequently implemented to large and complicated search spaces, it is a robust and efficient searching and optimisation process that depends on evolutionary principles of natural genetic systems. In this process, the genetic information of individuals, denominated as *gene*, is the potential solution to the problem, and structures named *chromosomes* incorporate them. From a collection of points denominated as population, GA explores each chromosome looking for those with the best-related fitness value. The latter represents how much the chromosome is suitable for the solution of the problem (Bandyopadhyay and Pal, 2007).

Concerning the whole process, GA summarises the basic operations in three steps: first, each chromosome is evaluated to obtain the fitness value, then chromosomes are selected, and

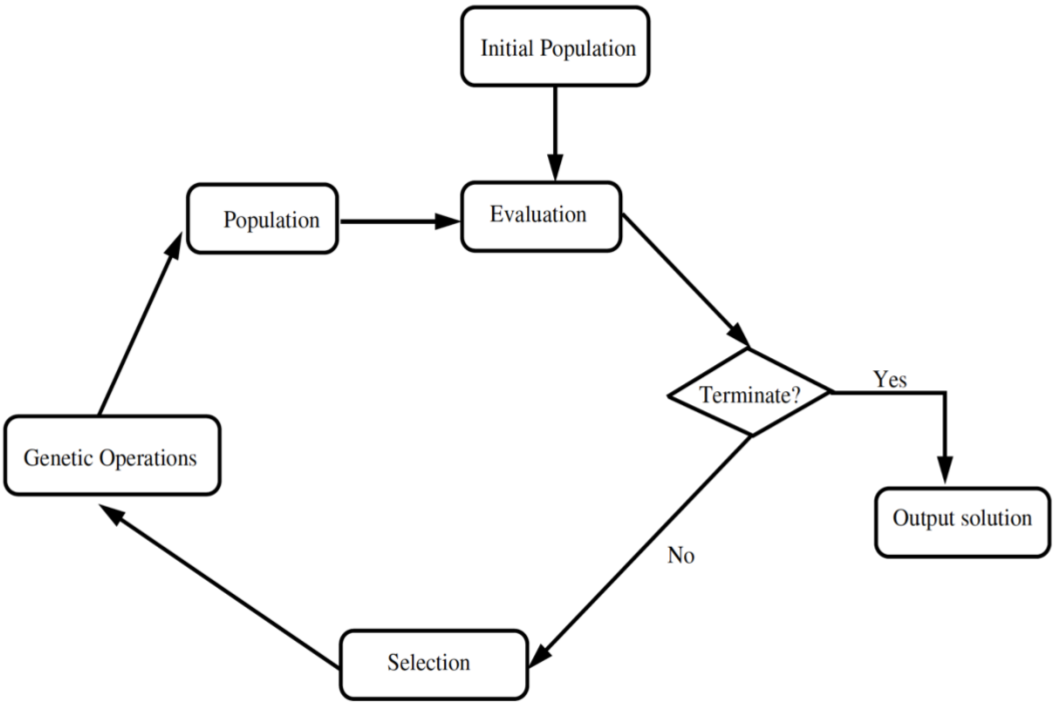


Figure 4.7 Basic steps of GA
Source: Bandyopadhyay and Pal (2007).

finally, a new population of chromosomes generates through genetic manipulations (see Figure 4.7). This cycle is repeated several generations as far as a termination criterion is reached, e.g., the number of repetitions reaches a prefixed number, or the optimum solution is determined, or the average fitness value of a population is roughly constant over several subsequent generations (Bandyopadhyay and Pal, 2007).

As for DE, GA begins with the initialisation of the population. A *population* contains a set of different chromosomes. The latter are commonly represented as a series (or string) of a combination of 0 and 1 (e.g., 10100011). Supposing l is the series length, the number of distinct strings (or chromosomes) is 2^l. Each series may outline a possible solution to the problem.

The evaluation phase merely consists of the choosing of the fitness or objective function, which depends on the problem that must be solved.

In the selection phase, GA replicates individual strings, termed *parent chromosomes*, into a trial population, called *mating pool*, for genetic operations. Ensuring the imitation of the natural selection process, GA suggests that the number of copies that a chromosome takes for the subsequent generation is proportional to its fitness value. This system is generally named the *proportional selection scheme*. *Roulette wheel parent selection* is one of the most commonly applied

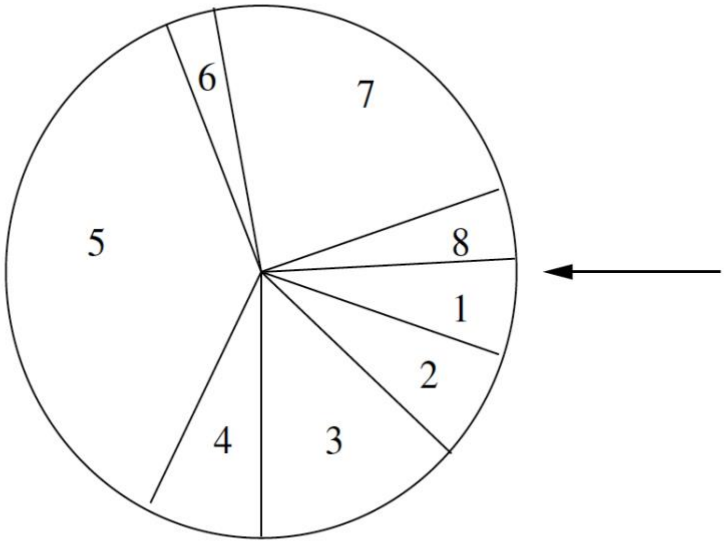
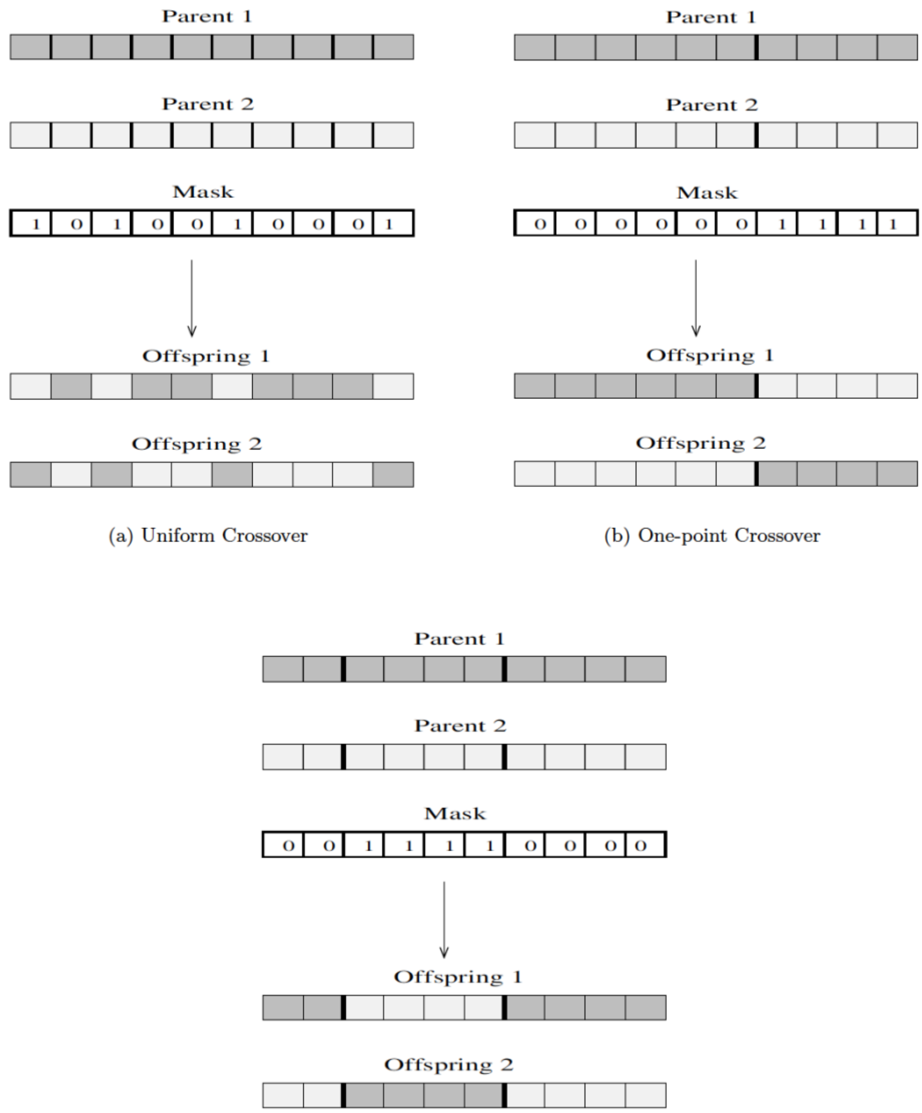


Figure 4.8 Roulette wheel parent selection
Source: Bandyopadhyay and Pal (2007).

selection methods. Figure 4.8 provides an illustrative representation. The population size (8) determines how many slots are contained by the wheel, and the size of each slot is due to the fitness of the associated chromosome. Then, a chromosome is chosen by rotating the roulette and remarking the position of the label at the roulette stops. In this way, the amount of times that a chromosome is chosen is proportional to its fitness in the population (Bandyopadhyay and Pal, 2007).

At this point, genetic operations enter the game. Crossover is the process of generating new chromosomes by the conjunction of genetic material randomly picked from two or more parent chromosomes. In other words, offspring for the future generation is created by the combination of parts from two or more parent chromosomes. In the algorithm, the *mask* defines which pieces of the parent chromosomes should be switched to generate the offspring. There exist several different kinds of crossover, as Figure 4.9 illustrates. The most commonly accepted method is the *one-point crossover*, which randomly chooses a crossover point k in the interval $(1, l-1)$[26]. So, the bit-strings after k are switched between the two parent chromosomes. Figure 4.9 (c) shows the *two-point crossover*, very similar to the previous one but with two crossover points. Finally, Figure 4.9 (a) exhibits the uniform crossover,

26 Remember that l is the string length.



(a) Uniform Crossover (b) One-point Crossover

(c) Two-point Crossover

Figure 4.9 Crossover operators
Source: Engelbrecht (2007).

whose mask randomly generates. Let p_{bs} be the bit-switching probability, if $p_{bs} = 0.5$, then every bit has the same likelihood to be switched (Engelbrecht, 2007).

The subsequent genetic operation is the mutation, which randomly alters the genetic composition of a chromosome. Thus, it adds genetic variety into the population, and it may lead to the optimal solution that dwells in a part of the search space not included in the current population (Bandyopadhyay and Pal, 2007). Each gene of the offspring $\widetilde{x_i}(t)$ has the likelihood p_{mr} of incurring in a mutation. Consequently, the mutation generates the mutated offspring $x_i'(t)$. The mutation probability, also called mutation rate, is a small number ranged

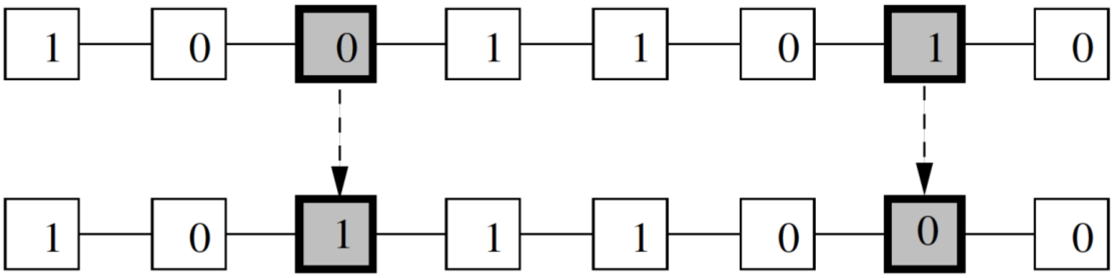


Figure 4.10 Binary bit-by-bit mutation
Source: Bandyopadhyay and Pal (2007).

between 0 and 1, and it assures mutation does not deteriorate excessively satisfactory solutions. Engelbrecht (2007) provides a formula for computing the chance that a chromosome will be mutated:

$$Prob(\widetilde{x}_i(t)\ is\ mutated) = 1 - (1 - p_{mr})^l, \tag{4.9}$$

where l is the number of genes contained by a chromosome.

There are many different kinds of mutation. Figure 4.10 illustrates the binary bit-by-bit mutation, where mutation takes place with the probability p_{mr}. The 3rd and 7th genes of the chromosome exhibited are subjected to the mutation.

To sum up, there are essential factors that mostly affect the GA performances: population size, mutation rate, and crossover rate, denoted by p_{cr}. Recent studies opt for selecting very high values for p_{cr}, and moderately low values for p_{mr}, both held constant. These two parameters are potentially helpful to reach the optimal solution, so it is necessary to set them carefully. However, it is not trivial to find such optimal parameters through empirical evidence. To solve the problem, Engelbrecht (2007) suggests the Fogarty's studies about mutation rates that dynamically adjust. Observing successful findings, Fogarty affirms that p_{mr} is a large number in the earliest moments of the search, and it exponentially diminishes over time as follows:

$$p_{mr}(j,t) = \frac{0.4026}{2^{t+j-1}}, \tag{4.10}$$

where $j = 1, \dots, lb$ is the bit of a chromosome, and lb indicates the least significant bit.

Chapter 5 – Computational investigations on portfolio optimisation problem

The dissertation aims to analyse two different approaches to the optimisation problem: Index-Tacking and Cumulative Prospect Theory models. In particular, it assumes that the behavioural based-model works better and gives better performance than the passive fund strategy. This chapter treats the application of the CPT model for index tracking, and it strives to enhance the performance of the CPT model varying the sentiment of the prospect investor. Moreover, this work assumes that IT model is correctly performed, obtaining comparable performances of the index. So, the comparison will be between CPT model and the benchmark.

The chapter is composed of three main sections. Section 5.1 illustrates all features of the data-set taken into account. Following, Section 5.2 reports the application of the Particle Swarm Optimisation. Since literature does not specify how to set all parameters, this section performs a tuning analysis on some of them to find the best value for the study. Finally, Section 5.3 analyses the performance of the CPT model for index tracking. The purpose is to expose the comparison between these models, both from a graphic and a statistical point of views and highlight the most significant findings. To improve the performances of this model, an analysis of the data is performed as the sentiment of the prospect investor varies. The study estimates three indicators to make a reliable comparison: the mean absolute error (MAE), the root mean squared error (RMSE), and the information ratio (IR). They are useful to assess both the magnitude of errors and performance of CPT model and verify if it beats the benchmark.

5.1 Basic descriptive statistics

The selection of a proper set of data assumes a crucial role. The comparison requires to occur for identical periods, with homogeneous statistical surveys, to avoid distortion in the evaluation.

The project uses data relating to Dow Jones Industrial Average (DJIA) index, available from Bloomberg software. As described on Borsa Italiana's website, DJIA is a stock market index that contains 30 publicly-owned companies listed on the NASDAQ and the NYSE. It is not associated with any particular sector, and so it includes securities belonging to different productive industries, both traditional and new techs. The Wall Street Journal is the decision-maker that chooses which share include or exclude from the benchmark. It regularly selects securities of US companies that assume the role of leadership in their industry.

The sampling period taken in consideration refers to 368 observations of weekly closing price, ranging from 15-06-2012 to 24-06-2019. Since daily observations may be subjected to high peaks of volatility and monthly observation do not entirely catch stock price trends, the project opts for weekly observations, following the example of a similar work conducted by Grishina et al. (2017).

The optimisation problem applies the cardinality constraint, i.e., $K = 10$, which limits the number of securities in the portfolio. It allows having a more concentrated portfolio, reducing risk and, especially, transaction costs. Since the stocks' selection is usually random, this dissertation aims to pick those securities which have had the highest expected returns over the period considered. Table 5.1 lists all stocks from DJIA with their respective expected log-returns (third column), highlighting in bold those included in the portfolio optimisation problem. It is universally recommended to work with the logarithm of the price ratio to determine the rate of returns rather than applying the absolute stock price method (Longerstaey and Spencer, 1996). Therefore, let p the stock price for each period t, the log-returns r have the following form:

$$r_i = \ln\left(\frac{p_{i,t}}{p_{i,t-1}}\right), \quad i = 1, \dots, N, t = 1, \dots, T, \tag{5.1}$$

Name	Symbol	Expected log-returns
3M Co	MMM	0.19%
American Express Co	AXP	0.21%
Apple Inc	**AAPL**	**0.24%**
Boeing Co	**BA**	**0.44%**
Caterpillar Inc	CAT	0.12%
Chevron Corp	CVX	0.05%
Cisco Systems Inc	**CSCO**	**0.32%**
Coca-Cola Co	KO	0.08%
Dow Inc	DOW	0.13%
Exxon Mobil Corp	XOM	-0.02%
Goldman Sachs Group Inc	GS	0.21%
Home Depot Inc	**HD**	**0.38%**
Intel Corp	INTC	0.15%
International Business Machines Corp	IBM	-0.10%
Johnson & Johnson	JNJ	0.20%
JPMorgan Chase & Co	**JPM**	**0.32%**
McDonald's Corp	MCD	0.23%
Merck & Co Inc	MRK	0.21%
Microsoft Corp	**MSFT**	**0.41%**
Nike Inc	**NKE**	**0.33%**
Pfizer Inc	PFE	0.18%
Procter & Gamble Co	PG	0.15%
Travelers Companies Inc	TRV	0.23%
United Technologies Corp	UTX	0.15%
UnitedHealth Group Inc	**UNH**	**0.39%**
Verizon Communications Inc	VZ	0.07%
Visa Inc	**V**	**0.48%**
Walgreens Boots Alliance Inc	WBA	0.15%
Walmart Inc	WMT	0.13%
Walt Disney Co	**DIS**	**0.30%**

Table 5.1 List of stocks from DJIA, with their respective expected log-returns

where N is the number of stocks, and T is the total number of periods. Therefore, the expected value uses the following formula:

$$\mathbb{E}(r_i) = \frac{1}{T}\sum_{t=1}^{T} r_{i,t}. \tag{5.2}$$

Table 5.2 reports the general statistics about benchmark, i.e., first row, and the ten stocks selected. It is necessary to highlight two key points. The variance of DJIA is the lowest, and the cause is quite trivial. DJIA is a basket of 30 securities and, as suggested by the literature, a well-diversified portfolio should have the lowest variance, especially if it compares with a

Name	Mean	Variance	Semivariance	Downside risk in %
DJIA	0.2000%	0.0301%	0.0173%	57.60%
AAPL	0.2400%	0.1313%	0.0703%	53.53%
BA	0.4416%	0.1017%	0.0555%	54.59%
CSCO	0.3170%	0.0933%	0.0467%	50.09%
HD	0.3784%	0.0664%	0.0344%	51.84%
JPM	0.3162%	0.0827%	0.0431%	52.09%
MSFT	0.4075%	0.0892%	0.0489%	54.83%
NKE	0.3253%	0.0883%	0.0423%	47.93%
UNH	0.3873%	0.0782%	0.0421%	53.88%
V	0.4819%	0.0591%	0.0291%	49.21%
DIS	0.2962%	0.0703%	0.0336%	47.90%

Table 5.2 General statistics about benchmark and the stocks selected

single asset. The second relevant consideration regards semivariance and downside risk. The semivariance measures the prospective downside risk of an investment portfolio. It is possible to compute it by determining the dispersion of all observations that drop below the mean. As illustrated in Table 5.2, all stocks have a downside risk (in percentage) lower than the benchmark, suggesting that departures from the mean are usually more "positive" and so more profitable for stocks than for the benchmark. Generally, risk-averse investors can decrease the portfolio's risk by investing in stable securities, which returns do not usually contract significant variations over time.

Before moving to the application of the model to the data-set, the Particle Swarm Optimisation requires some considerations regarding both the structure and the parameters settings. It is necessary to understand better how the code works. Concerning the latter, MATLAB R2019a runs the optimisation problem code, and the experiment is performed on a MacBook Pro with an Intel Core i5 processor, with 4GB ram.

5.2 Particle Swarm Optimisation parameters setting and general structure

This section defines which are the main parameters of PSO, and it investigates their appropriate values running some tests. Later, it discusses the PSO general framework, providing a more concrete idea of how the metaheuristic approach works and implements the CPT model for index tracking, early discussed in Section 3.2.2.

First, it is necessary to prefix parameters such as inertia weights and acceleration coefficients β_1 and β_2. Li and Engelbrecht (2007) recommend using $\alpha = 0.7298$ and $\beta_1 = \beta_2 = 1.49618$ because they demonstrate that PSO provides satisfactory results.

After, the thesis conducts a tuning analysis about the penalty parameter, the number of particles, and the number of iterations. To allow the PSO to learn better from mistakes, it is necessary to add a penalty parameter ϵ to the objective function. It represents the degree of the penalty that applies if the optimisation problem violates restrictions. There is no available literature that estimates the most satisfactory value of ϵ for this kind of application. For this reason, the only solution is to conduct some empirical tests. For each value of ϵ, it tests ten instances of time, for each of which PSO runs ten times[27] and selects the simulation that reaches the best fitness function. These tests consider 250 iterations and 20 particles in the swarm[28]. To determine the best value of ϵ, it is computed the average value of the normalized fitness functions of each period and the respective standard deviation. Table 5.3 summarizes the results. The best value is $\epsilon = 0.1$ because the standard deviation is the lowest. In such a situation, PSO obtains a satisfactory fitness function with the smallest variation from an instance of time to another. In other words, PSO can achieve continually adequate results over time.

Next, it is necessary to set the number of particles sw to employ. This test works similarly to the previous test, but the changing variable is not ϵ anymore, but sw. In particular, it

ϵ	Normalized Fitness	Standard Deviation
0.1	0.872271162	0.095572078
0.01	1.006778894	0.219593967
0.001	0.970714652	0.226959423
0.0001	0.98532873	0.110587958
0.00001	1.044493471	0.215604517

Table 5.3 Tuning analysis of the penalty parameter

27 As described in Chapter 4, the PSO algorithm is a metaheuristic approach that attempts to obtain the best solution whenever it runs. For this reason, it does not provide the same solution every time. It may happen that, during the search of the optimum, it discovers a local optimum solution, which does not correspond to the global optimum solution of the problem. For this reason, from now on the code could be run more times to be able to find a more satisfactory solution.

28 The portfolio optimisation problem recommends setting the number of particles equal to or greater than 20 because there are 20 variables: 10 variables represent the weights of held stocks, and the other 10 are binary variables that assess if the portfolio includes or not stocks. The next step is to test the best number of particles.

Particles	Normalized Fitness	Standard Deviation	CPU time (s)
20	0.872271162	0.095572078	745
40	1.060823616	0.171515113	1055
60	1.216620692	0.344348198	1484

Table 5.4 Tuning analysis of the number of particles

examines three different values of sw, which are multiples of 20. These tests still consider 250 iterations and $\epsilon = 0.1$. Table 5.4 reviews the results. Although the increase of particles should help to find a better solution, it does not occur for this kind of problem. The best solution between all alternatives lasts the model with 20 particles. The last column shows the computational time, which increases as the number of particles increases.

The last tuning analysis regards the number of iterations. It is the amount of time that PSO repeats all operations to reach the optimum solution. So, it is the termination criterion of PSO. It is necessary to make some considerations. If the number of iterations is too short, the particles will not be able to reach the optimum because they have not enough time. On the other hand, a higher number of iterations should achieve better results; nevertheless, PSO requires more time, being at the expense of its capability to obtain satisfactory solutions in a short time. The analysis assesses four scenarios with 25, 250, 500, and 1000 iterations. Figure 5.1 shows the research of the optimum solution for a different number of iterations in a random instance of time. It clearly illustrates that the fitness function improves as the number of iterations increases. The best convergence is given by the higher number of iterations, i.e., 1000 iterations (violet line), while the worst is the one with the lowest number of iterations, i.e., 25 iterations (blue line). Nevertheless, the CPU time increases too: PSO with 250 iterations spends about 10 minutes, while PSO with 1000 iterations spends more than 40 minutes. Considering that 250 iterations provide a good convergence too, this work will examine this choice.

After setting all initial parameters, consider the PSO structure. To run the PSO, it is necessary to create a varied number of scenarios for each stock and benchmark returns considered. In other words, it simulates several trends for each time series. These scenarios are useful to PSO to learn how both stocks price and the benchmark may evolve in time. Based on this knowledge, PSO acts to find an optimal solution to the optimisation problem. To this purpose, the application of the model starts with the simulation of 1000 scenarios S for each stock and benchmark returns. Therefore, it creates a matrix $(S * nsh)\ x\ (N+1)$, where the

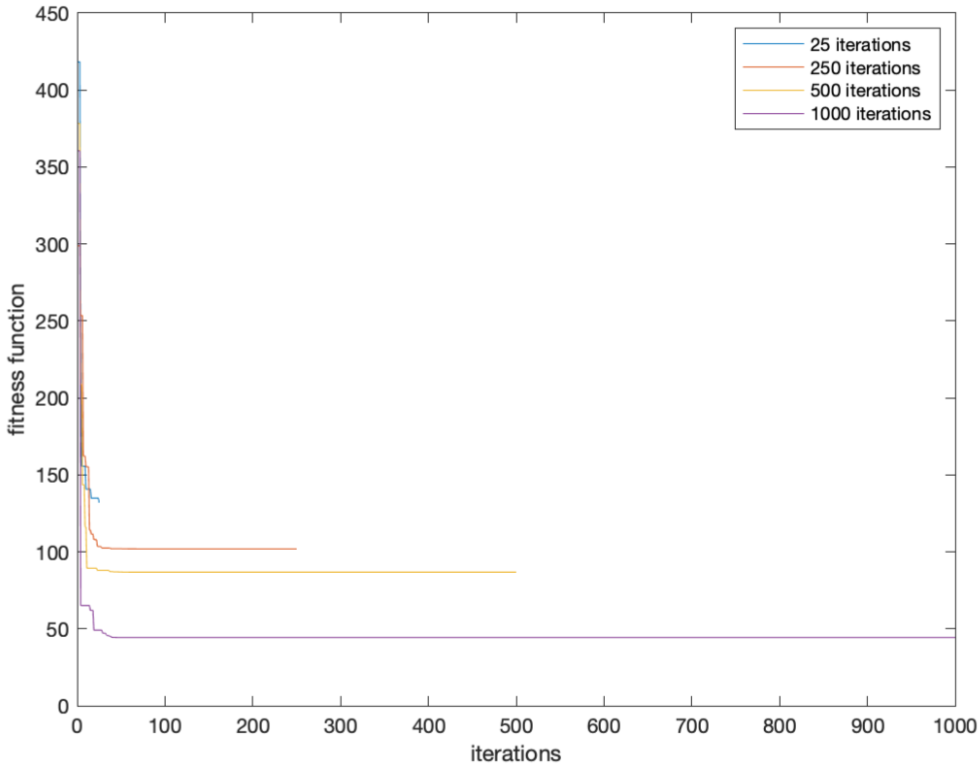


Figure 5.1 PSO with different number of iterations

first column represents benchmark returns, and the following columns are the returns of the risky stock. The notation nsh represents the number of rolling steps. In other words, they are the number of time-periods in which the CPT portfolio compares with the index. In each instance of time, the CPT portfolio could overcome the performance of the benchmark. It depends on the capability of PSO of finding satisfactory weights for each stock held in the portfolio. At this point, the number of runs for PSO, denoted by $runs$, comes into play. It is the number of times for which PSO runs, and $runs$ are equal to 10. So, after running ten times, the PSO sorts the results in descending order by the fitness function. The weights that have obtained the best result are adopted in the portfolio optimisation, to be ready for comparison with the benchmark. The work set itself to compare ten time-periods, so $nsh = 10$.

The following box outlines the crucial keys of the PSO general framework for better compensation:

```
[…] % setting initial parameters and uploading data

for i = 1:nsh % number of rolling steps

      […]

      for ii = 1:runs % number of runs for PSO

            […] % initialise the population of particles

            for iii = 1:iterations % number of iteration

            1) computation of the range for the maximum velocity
            2) estimate the fitness function
            3) compare the performance of every particle to x_i^(local)
            4) compare the performance of every particle to x^(global)
            5) adjustments of the velocity and location
            6) repeat the procedure from step 2

            end

      […]

      end

      […] % computations for comparison with benchmark

end
```

Before proceeding to the application of the model, it is necessary to highlight some important features regarding the implementation of PSO into the CPT model. The metaheuristic approach works if and only if it deals with a minimisation problem. Nevertheless, the problem formulated in Equation (3.32) is a maximisation problem. To solve the problem, Equation (3.32) takes the form of:

$$\min_{\mathrm{x}} V(f) = \left\{ -\sum_{i=1}^{M} w_i v \left(\sum_{j=1}^{N} r_{ij} x_{ij} - rb_i \right) + \frac{1}{\epsilon} \left(\left| \sum_{j=1}^{N} x_j - 1 \right| + \sum_{j=1}^{N} \max(0, -x_j) \right) \right\}, \tag{5.3}$$

where the ratio $1/\epsilon$ penalises the violation of the constraints: the absolute value has to be equal to zero to ensure that all capital is invested in the portfolio, while the second sum checks that all securities have not negative weights, i.e., the sum is equal to zero. Therefore, if both constraints are equal to zero, it means that no violations incur, otherwise the ratio $1/\epsilon$ multiplied by the value in the round brackets will penalise the minimisation problem, increasing the total fitness function.

Moreover, since the comparison between the CPT model and the benchmark returns occurs in ten different instances of time, the model requires to take into consideration a changeable reference point over time. Remind that the benchmark is the reference point. So, the model strives to find a satisfactory combination of weights of assets in each considered instance of time such that it can beat the benchmark returns, which, of course, vary for each period.

5.3 Application of CPT model for index tracking and significant results

This section focusses on the application of the model described in Equation (5.1), subject to the restrictions illustrated by Equations (3.30) and (3.31).

In the beginning, the model runs with the parameters of the CPT model equivalent to those discovered by Tversky and Kahneman (1992), i.e., $\alpha = 0.88$, $\beta = 0.88$, $\lambda = 2.25$, $\gamma = 0.61$, and $\delta = 0.69$. Statistical indicators evaluate the performance of this model compared to the benchmark. Results are not completely satisfactory due to the CPT parameters. Therefore, PSO runs several times to improve portfolio performance until it can achieve satisfactory results. Next, some investigations regarding the sentiment of the prospect investor are performed to discover dissimilarities between different kinds of investors. Since CPT is a behaviourally-based model, results may be considerably different if investor's preferences alter. It is necessary to highlight that the sentiment of the prospect investor does not vary over time. A possible suggestion to expand this thesis would be considered a sentiment of the prospect investor that changes based on the past portfolio performance.

5.3.1 Models comparison through statistical indicators

Before comparing the performance of CPT portfolio with the benchmark, it is necessary to describe the indicators that this thesis employs, which are: the mean absolute error (MAE), the root mean squared error (RMSE), and the information ratio (IR).

MAE and RMSE are quite similar. The former measures the average of the absolute errors e_i defined by the difference between prediction y_i and actual observation x_i. In short (Willmott and Matsuura, 2005):

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{T} = \frac{\sum_{i=1}^{n} |e_i|}{T}, \tag{5.4}$$

where T is the number of observations.

The RMSE is the square root of the average of squared errors e_i. The formula is (Willmott and Matsuura, 2005):

$$RMSE = \left(\frac{\sum_{i=1}^{n} |e_i|^2}{T}\right)^{1/2}. \tag{5.5}$$

Both measures share several similarities. They vary from zero to ∞ and do not take into account the direction of errors due to the presence of absolute value. Then, despite the square root and the division by T only change the scale of total square error, they are directly proportional to the variance of errors. As a result, they are *negatively-oriented scores*, and so investors prefer to obtain lower values of both measures. The lower the values of MAE and RMSE, the higher the convergence of the portfolio to the performance of the index. If the portfolio exhibits high values of MAE and RMSE, it implies two situations: the portfolio significantly outperforms the index or vice versa. IR helps to check which is the circumstance in question, such that it removes this ambiguity.

Nevertheless, there are some differences between MAE and RMSE. RMSE implies that errors are first squared, and then the average is taken. It determines that the RMSE gives moderately great importance to high errors. In other words, high errors have a more

Runs	MAE	RMSE	IR	Deviation (+)	Deviation (-)	Total deviations
1	0.94%	1.34%	17.04%	5.91%	-3.54%	2.37%
2	1.00%	1.21%	-1.62%	4.90%	-5.11%	-0.21%
3	1.53%	1.66%	-3.14%	7.38%	-7.93%	-0.55%
4	0.98%	1.19%	28.37%	6.60%	-3.18%	3.42%
5	0.74%	0.91%	3.31%	3.88%	-3.56%	0.32%

Table 5.5 The CPT model for index tracking with $\alpha = 0.88$, $\beta = 0.88$, $\lambda = 2.25$, $\gamma = 0.61$, *and* $\delta = 0.69$

significant impact on the total square error than do smaller ones. It means the RMSE should be advantageous if high errors are particularly undesirable.

The IR is an indicator achievable by the ratio between the excess return of the portfolio over the benchmark (or index) and its standard deviation. The formula is:

$$IR = \frac{r_p - R_p}{TE}, \tag{5.6}$$

where r_p is the portfolio return, and R_p is the benchmark return, and TE is the standard deviation of excess return or, simply, the tracking error. Therefore, the IR is crucial to estimate the portfolio manager's skills to produce excess returns than the index. Consequently, as managers use their abilities to exceed the benchmark, so the PSO use its operations to beat the benchmark. Remember that a passive management observes a null ratio because the difference in the numerator is equal to zero (Cogneau and Hübner, 2009).

Finally, tracking error requires some considerations. It examines to what extent the portfolio tracks the performance of the index. If the tracking error is low, the portfolio is continually exceeding the index over time. If the tracking error is high, the portfolio returns present more variability over time, and it is not regularly beating the benchmark.

At this point, the CPT model for index tracking runs for five times. The comparison occurs for ten rolling steps, i.e., $nsh = 10$, and ten runs for PSO, i.e., $runs = 10$. Table 5.5 summarises the above-described indicators. The fifth column represents the positive deviation of the CPT portfolio from the index, while the penultimate column represents the negative deviation of the CPT portfolio from the index. The last column shows the total deviation of the CPT portfolio from the benchmark, i.e., how much CPT portfolio is better performing than the benchmark in terms of deviations of returns. Results are quite weak: CPT portfolio not always overcomes the benchmark, and negative values for IR means that

Runs	MAE	RMSE	IR	Deviation (+)	Deviation (-)	Total deviations
1	0.87%	1.39%	9.32%	5.05%	-3.69%	1.36%
2	1.29%	1.42%	47.24%	9.64%	-3.29%	6.35%
3	1.24%	1.45%	32.00%	8.49%	-3.86%	4.63%
4	1.16%	1.64%	32.56%	8.47%	-3.14%	5.33%
5	0.95%	1.05%	12.63%	5.46%	-4.08%	1.39%

Table 5.5 The CPT model for index tracking with $\alpha = \beta = 0.88, \lambda = 1.125,$ *and* $\gamma = \delta = 1$

the model has poor abilities to generate excess returns than the index. This problem regards not the PSO parameters, but the parameters of the CPT model. Varying the parameters α, β, λ, γ, and δ changes the investor preference and leads to better results.

After several attempts, the following setup, i.e., $\alpha = \beta = 0.88$, $\lambda = 1.125$, and $\gamma = \delta = 1$, provides better results, as Table 5.6 shows. Total deviations are always positive and more stable than before, meaning that the new parameters of the CPT model properly work. IR confirms the last conjecture because it is always positive and so portfolios can produce excess

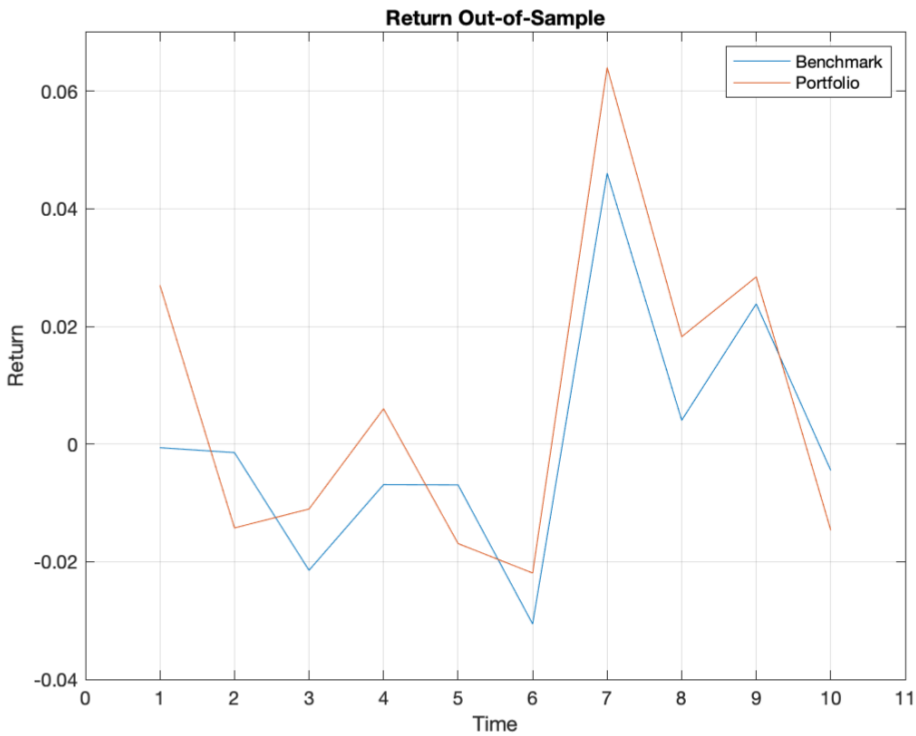


Figure 5.2 Comparison between CPT portfolio and benchmark returns

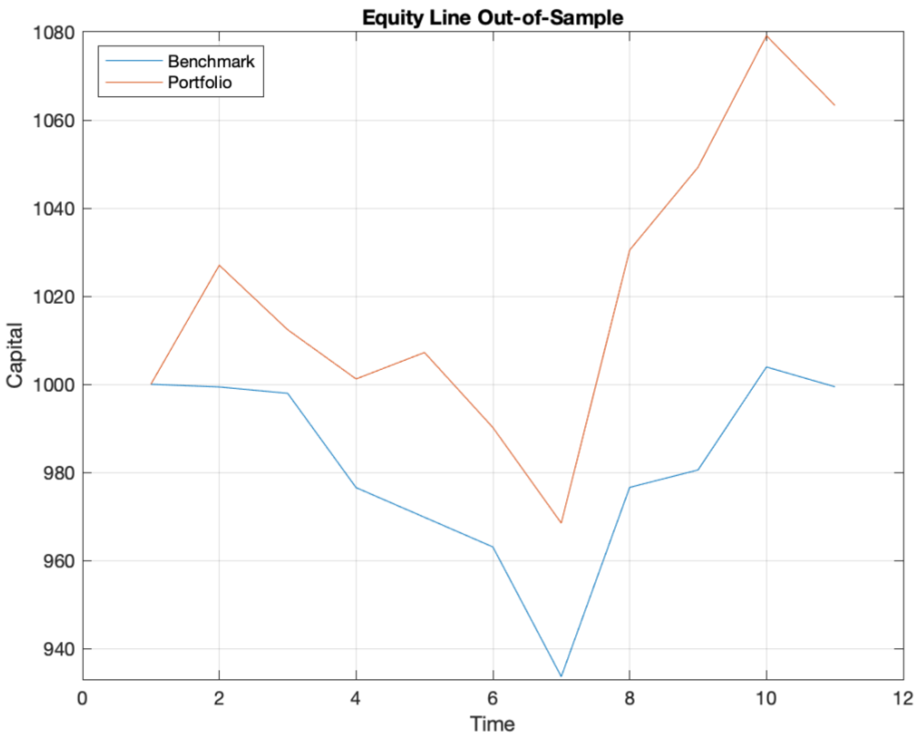


Figure 5.3 Comparison between CPT portfolio and benchmark performance, supposing to have 1000 of capital to invest

returns than the index. Concerning the MAE and RMSE, they work as previously supposed. The first and last runs have lower MAE values because the portfolios tracks more closely the index. Besides, there is always a negative deviation of about 3.5%, which hints that the portfolio does not always beat the benchmark in every rolling step.

From now on, the analysis focuses only on the best performing portfolio, i.e., the second in the list, because it conspicuously delineates the advantages of the behavioural portfolio. Figure 5.2 illustrates the comparison between the CPT portfolio and benchmark returns over 10 rolling steps. In most cases, the portfolio exceeds the benchmark, except for $t = 2, 5, 10$. Figure 5.3 clearly shows the supremacy of CPT portfolio, displaying the consequences of the investment of the initial capital of 1000 in both the portfolio and the benchmark. The CPT portfolio regularly follows the trend of the benchmark, but it can keep a higher total wealth over the entire period.

Finally, Table 5.6 summarises all findings regarding the comparison seen in the previous two figures, while Table 5.7 exhibits the respective weights of the stocks included in the portfolio over the whole period. Concerning the last table, it is possible to notice that the optimisation problem does not violate the restrictions: the sum of the weights of stocks is equal to one in each rolling step, and there are no negative weights (short-selling is not available)[29].

[29] It is necessary to make a consideration regarding the PSO code. The implementation of the CPT model with the restriction of no short-selling into the PSO algorithms is very complicated. For the sake of simplicity, PSO does not meet this constraint, generating some stocks with negative weights. To manage this problem, the code includes a special check that transforms all negative weights to zero weights, and it assigns those negative weights to all the other stocks contained in the portfolio. In this way, there are no negative weights, and the sum of all weights is equal to one.

	Rolling steps	0	1	2	3	4	5	6	7	8	9	10
Returns out of sample	Benchmark		-0.06%	-0.14%	-2.15%	-0.69%	-0.69%	-3.06%	4.60%	0.41%	2.38%	-0.45%
	Portfolio		2.70%	-1.43%	-1.10%	0.60%	-1.69%	-2.19%	6.40%	1.83%	2.84%	-1.46%
	Δ		2.76%	-1.28%	1.04%	1.29%	-1.00%	0.87%	1.80%	1.42%	0.46%	-1.02%
Equity out of sample	Benchmark	1000	999.39	997.94	976.53	969.80	963.06	933.61	976.58	980.54	1003.92	999.43
	Portfolio	1000	1027.02	1012.38	1001.21	1007.21	990.18	968.47	1030.47	1049.29	1079.12	1063.33
	Δ	0	27.63	14.44	24.67	37.42	27.12	34.86	53.89	68.74	75.20	63.90

Table 5.6 The top of the table exhibits the returns out of the sample of both benchmark and CPT portfolio. The bottom of the table shows the fluctuations of 1000 of capital invested in both the benchmark and the portfolio. In each section, the third line represents the respective deviation of the portfolio from the benchmark.

Rolling steps	AAPL	BA	CSCO	HD	JPM	MSFT	NKE	UNH	V	DIS
1	0.00%	0.00%	39.21%	0.00%	15.53%	0.00%	1.08%	44.18%	0.00%	0.00%
2	0.00%	31.12%	0.00%	24.04%	13.72%	1.37%	29.74%	0.00%	0.00%	0.00%
3	0.00%	0.00%	0.00%	0.00%	0.00%	32.55%	25.21%	1.96%	19.41%	20.87%
4	0.00%	1.27%	9.59%	30.95%	1.44%	17.47%	33.54%	0.00%	0.00%	5.74%
5	0.00%	33.75%	2.42%	0.00%	9.32%	0.00%	50.90%	0.00%	3.61%	0.00%
6	0.00%	0.00%	0.00%	0.00%	27.64%	0.00%	7.24%	17.59%	35.66%	11.87%
7	12.06%	19.37%	10.06%	0.00%	0.00%	0.00%	36.24%	0.00%	22.26%	0.00%
8	0.12%	0.00%	0.00%	36.50%	7.79%	20.23%	32.62%	0.00%	0.00%	2.74%
9	0.00%	0.00%	26.59%	8.25%	0.00%	11.88%	31.22%	0.00%	18.24%	3.83%
10	0.00%	0.00%	0.00%	31.88%	0.00%	21.23%	27.71%	3.63%	0.00%	15.55%

Table 5.7 Composition of the CPT portfolio for index tracking over all rolling steps

5.3.2 Investigations about the sentiment of the prospect investor

This subsection examines how the optimization problem changes as the investor preferences vary. In other words, it realizes a tuning analysis of the parameters of the CPT model. Starting from the setup previously found, i.e., $\alpha = \beta = 0.88$, $\lambda = 1.125$, and $\gamma = \delta = 1$, it seeks to analyse the consequences of altering one parameter and keeping the others constant.

Before starting, it is necessary to clarify. The following tuning analysis sometimes looks ambiguous and does not produce significant results. As previously mentioned, PSO is not a static approach that provides a single solution: every time it runs, it may give a different solution based on the research of the optimum.

After setting the record straight, the investigation starts with those parameters, i.e., λ, α, and β, that influence the value function of Cumulative Prospect Theory. Table 5.8 illustrates the effects of the variation of λ, α, and β parameters. Concerning λ, remind that it represents the degree of loss aversion and, if the decision-maker is loss-neutral, it is equal to one. The increase of loss aversion should lead to more cautious investment decisions, wiling to avoid losses. It is possible to verify that variations of λ do not entail significant differences between the scenarios, except for a decreasing of the negative deviation for $\lambda = 1.5$ and $\lambda = 2$. The increased loss aversion leads to the tenacity of the model to be careful not to suffer losses, worsening the general situation. The increasing of the loss aversion entails an increase of the negative deviation, so the benchmark returns more often exceed the CPT portfolio returns. A low value of λ was picked in the previous subsection for this reason.

Regarding α and β, they affect how the decision-maker respectively evaluate gains and losses. They follow the *diminishing sensitivity* principle, for which people usually perceive the difference between a gain (or loss) of 50 and one of 100 as greater than the difference between a gain (or loss) of 10,050 and one of 10,100. Thus, the further α and β move away from 1, the higher is the strength of this principle. The middle part of Table 5.8 regards the variation of α, which results are quite ambiguous from an economic point of view. It suggests that CPT portfolio obtains higher performance if α is closer to 1 or 0.5. In other words, these results are not so much consistent and sound.

On the other hand, the results of β are prone to prefer a value close to 0.5. In this way, positive deviations of the CPT portfolio from the benchmark are rather high. In other words, these findings indicate that if the investor underestimates the losses, he will obtain positive performance.

λ	1	1.5	2	2.5	3	3.5
α	0.88	0.88	0.88	0.88	0.88	0.88
β	0.88	0.88	0.88	0.88	0.88	0.88
MAE	1.13%	0.94%	0.89%	1.09%	1.00%	1.21%
RMSE	1.33%	1.37%	1.15%	1.43%	1.27%	1.51%
IR	10.76%	16.25%	18.56%	19.16%	13.56%	20.47%
Deviation (+)	6.38%	5.88%	5.56%	6.85%	5.92%	7.66%
Deviation (-)	-4.87%	-3.56%	-3.36%	-4.03%	-4.12%	-4.48%
Total deviations	1.50%	2.31%	2.20%	2.82%	1.80%	3.19%

α	0.5	0.6	0.7	0.8	0.9	1
λ	1.125	1.125	1.125	1.125	1.125	1.125
β	0.88	0.88	0.88	0.88	0.88	0.88
MAE	1.07%	1.26%	0.82%	0.66%	1.00%	0.98%
RMSE	1.46%	1.42%	1.06%	0.83%	1.27%	1.27%
IR	19.35%	17.81%	3.60%	12.40%	13.56%	45.02%
Deviation (+)	6.83%	7.59%	4.29%	3.83%	5.92%	7.59%
Deviation (-)	-3.92%	-4.97%	-3.89%	-2.75%	-4.12%	-2.16%
Total deviations	2.92%	2.62%	0.40%	1.07%	1.80%	5.42%

β	0.5	0.6	0.7	0.8	0.9	1
λ	1.125	1.125	1.125	1.125	1.125	1.125
α	0.88	0.88	0.88	0.88	0.88	0.88
MAE	1.47%	1.21%	1.26%	1.23%	0.79%	0.76%
RMSE	1.88%	1.57%	1.45%	1.60%	1.03%	0.89%
IR	37.96%	37.07%	24.77%	31.04%	28.17%	5.69%
Deviation (+)	10.82%	8.91%	8.12%	8.65%	5.40%	4.06%
Deviation (-)	-3.85%	-3.19%	-4.45%	-3.68%	-2.46%	-3.53%
Total deviations	6.98%	5.72%	3.67%	4.97%	2.94%	0.53%

Table 5.8 Differences in the application of the CPT portfolio for tracking error. The top, middle and bottom parts of the table show which are the differences caused by a variation of the respective parameters λ, a, and β. Those parameters individually vary. Parameters γ and δ are always kept constant.

The subsequent tuning analysis regards the parameters γ and δ, which affect the weighting function of Cumulative Prospect Theory. Table 5.9 exhibits the consequences of the variation of these parameters. Concerning γ, it influences the positive probability weighting function. Between the values of 0.75 and 0.85, the parameter provides the best performing portfolios. As a result, an investor should overestimate small positive probabilities and underestimate medium-high positive probabilities. So, investors are less sensitive to differences in probability in the middle range. The adoption of this kind of behaviour allows obtaining portfolios with better performance.

γ	0.7	0.75	0.8	0.85	0.9	0.95
δ	1	1	1	1	1	1
MAE	1.06%	1.23%	1.60%	0.90%	1.03%	1.01%
RMSE	1.15%	1.64%	2.10%	1.17%	1.33%	1.43%
IR	10.87%	32.61%	20.96%	42.62%	20.19%	16.81%
Deviation (+)	5.94%	8.82%	10.27%	6.92%	6.54%	6.27%
Deviation (-)	-4.63%	-3.50%	-5.74%	-2.12%	-3.78%	-3.79%
Total deviations	1.31%	5.32%	4.53%	4.80%	2.76%	2.49%

δ	0.7	0.75	0.8	0.85	0.9	0.95
γ	1	1	1	1	1	1
MAE	0.96%	1.10%	1.30%	0.95%	1.40%	1.27%
RMSE	1.26%	1.35%	1.80%	1.20%	1.67%	1.47%
IR	51.92%	58.41%	48.45%	11.69%	31.78%	17.87%
Deviation (+)	7.82%	9.04%	10.59%	5.46%	9.64%	7.70%
Deviation (-)	-1.75%	-1.97%	-2.41%	-3.99%	-4.32%	-4.97%
Total deviations	6.07%	7.07%	8.18%	1.47%	5.32%	2.73%

Table 5.9 Differences in the application of the CPT portfolio for tracking error. The upper and lower parts of the table show which are the differences caused by a variation of the respective parameters λ, a, and β. Those parameters individually vary. Parameters λ, α, and β are always kept constant.

Finally, the tuning analysis on δ generates the portfolios with the best performances obtained so far. Set δ between 0.7 and 0.8, IR proves that those portfolios produce high excess returns than the benchmark. Therefore, the winning strategy is achievable if the investor overestimates the smaller negative probabilities more and underestimates the medium-high negative probabilities more than the respective positive probabilities.

To conclude the whole study, a last further comparison occurs between the CPT portfolio collected in Table 5.7, indicated as Portfolio 5.7, and the CPT portfolio obtained in Table 5.9 with $\delta = 0.8$, denoted as Portfolio 5.9. To better estimate the differences, Figure 5.4 illustrates the investment of 1000 of capital in both portfolios and index. For the whole period, Portfolio 5.9 proves its supremacy over Portfolio 5.7 as supposed and, of course, over the benchmark.

The investigation on the sentiment of the prospect investor is so concluded. The final findings are extremely satisfactory, and dissertation succeeds to prove its purpose.

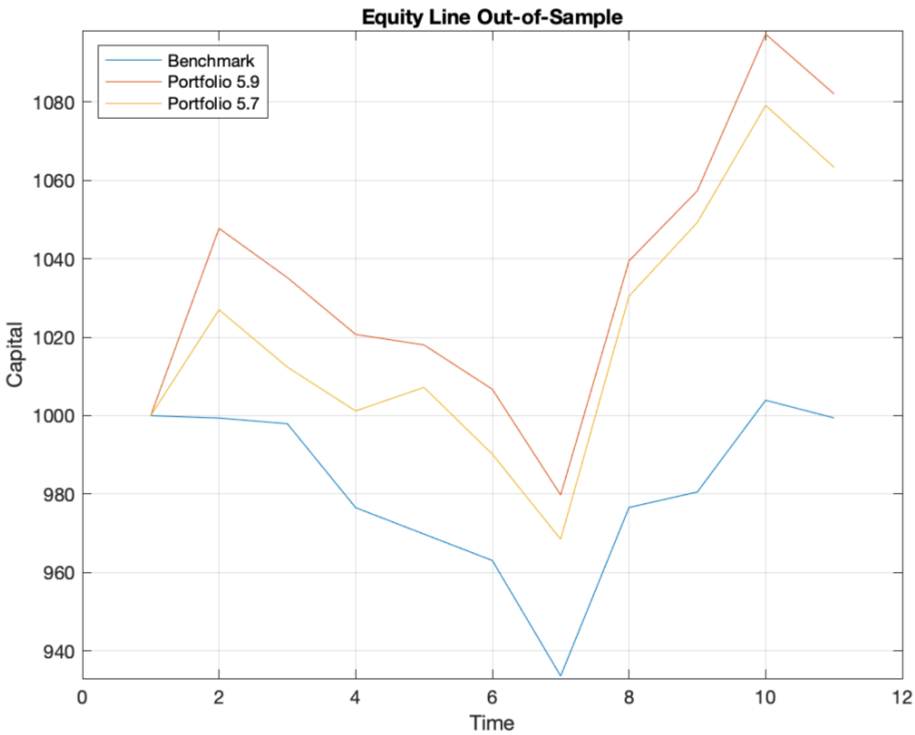


Figure 5.4 Comparison among benchmark, Portfolio 5.9 and Portfolio 5.7, supposing to have 1000 of capital to invest

Conclusion

The thesis aims to develop a portfolio selection model based on a behavioural approach suggested by the literature. The main purpose is to demonstrate that the behaviourally based model has better performances compared to a specific benchmark. The thesis applies the Cumulative Prospect Theory model for index tracking, where the objective function is represented by the deviation of the portfolio returns from the index. So, it maximises the objective function, taking into account two significant constraints: no short selling is allowed, and all capital needs to be invested in the portfolio. These restrictions make the analysis more reliable.

The complexity of the optimisation problem implies the use of metaheuristics approaches, which help to detect satisfactory solutions in a short time and with limited information. They are inexact methods, but they can solve the optimisation problem considered by the thesis. The Particle Swap Optimisation is the metaheuristic approach used for this analysis. It is widely known for its ability to achieve good results for portfolio selection problems.

The thesis focusses on the Dow Jones Industrial Average (DJIA) index, which includes 30 large companies listed on the NASDAQ and the NYSE. The Cumulative Prospect Theory portfolio holds only 10 securities of the DJIA index to reduce transaction costs and make the analysis more actual and reliable.

The final findings collected are remarkably good. The behavioural portfolio proves to be better performing than the passive management that could be achieved by either with a replication of the benchmark or through the index-tracking model.

The tuning analysis of the PSO algorithm demonstrates that an extension of computational time allows achieving CPT portfolios with better performances. In particular, it proves that the increase in the number of iterations from 250 to 1000 generates better portfolio performances in the investment period considered.

The results of this thesis should be considered for future inquiries and researches regarding CPT portfolio optimisation problems. As my concern, I recommend spending more time on analysing the Cumulative Prospect Theory parameters to discover connections with portfolio performance. It would be interesting to analyse different markets and indices, evaluating if the CPT portfolio keeps its supremacy over the benchmark considered.

Bibliography

Alexander, G. J., & Baptista, A. M. (2011). Portfolio selection with mental accounts and delegation. *Journal of Banking & Finance, 35(10)*, 2637-2656.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 503-546.

Arzac, E. R., & Bawa, V. S. (1977). Portfolio choice and equilibrium in capital markets with safety-first investors. *ournal of Financial Economics, 4(3)*, 277-288.

Bandyopadhyay, S., & Pal, S. K. (2007). *Classification and learning using genetic algorithms: applications in bioinformatics and web intelligence.* Springer Science & Business Media.

Baptista, A. M. (2012). Portfolio selection with mental accounts and background risk. *Journal of banking & Finance, 36(4)*, 968-980.

Beasley, J. E., Meade, N., & Chang, T. J. (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research, 148(3)*, 621-643.

Bem, D. J. (1965). An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology, 1*, 199-218.

Bernard, C., & Ghossoub, M. (2010). Static portfolio choice under cumulative prospect theory. *Mathematics and financial economics, 2(4)*, 277-306.

Bernoulli, D. (1954). Exposition of a new theory on the measurement. *Econometrica, 22(1)*, 23-36.

Borsa-Italiana. (n.d.). *Glossario Finanziario - Indice Dow Jones Industrial Average.* Retrieved from Borsa Italiana Web site: https://www.borsaitaliana.it/borsa/glossario/indice-dow-jones-industrial-average.html

Cartwright, E. (2014). *Behavioral economics.* Routledge.

Cogneau, P., & Hübner, G. (2009, January 11). *The 101 ways to measure portfolio performance.* Retrieved from http://ssrn.com/abstract=1326076

Das, S., Markowitz, H., Scheid, J., & Statman, M. (2010). Portfolio optimization with mental accounts. *ournal of Financial and Quantitative Analysis, 45(2)*, 311-334.

Dorigo, M., & Di Caro, G. (1999). Ant colony optimization: a new meta-heuristic. *1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406). 2*, pp. 1470-1477. IEEE.

Dubins, L. E., & Savage, L. J. (1976). *How to gamble if you must: Inequalities for stochastic processes.* New York: 1965 edition McGraw-Hill.

Eberhart, R., & Kennedy, J. . (1995). Particle swarm optimization. *IEEE international conference on neural networks. 4*, pp. 1942-1948. IEEE.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, 643-669.

Engelbrecht, A. P. (2007). *Computational intelligence: an introduction.* John Wiley & Sons.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance, 25(2)*, 383-417.

Feldstein, M. S. (1969). Mean-variance analysis in the theory of liquidity preference and portfolio selection. *The Review of Economic Studies, 36(1)*, 5-12.

Francis, J. C., & Kim, D. (2013). *Modern portfolio theory: Foundations, analysis, and new developments* (Vol. 795). John Wiley & Sons.

Franklin Templetion Investment. (2012, October 6). *Investors Should Beware The Role Of 'Availability Bias'.* Retrieved from https://www.businessinsider.com.au/the-availability-bias-is-driving-investor-decisions-2012-10

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of political Economy, 56(4)*, 279-304.

Fuller, R. J. (1998). Behavioral finance and the sources of alpha. *Journal of Pension Plan Investing, 2(3)*, 291-293.

Goldstein, W. M., & Einhorn, H. J. (n.d.). Expression theory and the preference reversal phenomena. *Psychological review, 94(2)*, 236-254.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive psychology, 38(1)*, 129-166.

Grishina, N., Lucas, C. A., & Date, P. (2017). Prospect theory–based portfolio optimization: an empirical study and analysis using intelligent algorithms. *Quantitative Finance, 17(3)*, 353-367.

Hanoch, G., & Levy, H. (1969). The efficiency analysis of choices involving risk. *Review of Economic studies, 36*, 335-346.

He, X. D., & Zhou, X. Y. (2011). Portfolio choice under cumulative prospect theory: An analytical treatment. *Management Science, 57(2)*, 315-331.

Henderson, P. W., & Peterson, R. A. (1992). Mental accounting and categorization. *Organizational Behavior and Human Decision Processes, 51(1)*, 92-117.

Hoffmann, A. O., & Post, T. (2014). Self-attribution bias in consumer financial decision-making: How investment returns affect individuals' belief in skill. *Journal of Behavioral and Experimental Economics, 52*, 23-28.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: University of Michigan Press.

Jorion, P. (1994). Mean/variance analysis of currency overlays. *Financial Analysts Journal, 50(3)*, 48-56.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: New York: Farrar, Straus and Giroux.

Kahneman, D. (2013, January 24). Nobel Prize Winner Daniel Kahneman: Lessons from Hitler's SS and the danger in trusting your gut. (Forbes, Interviewer)

Kahneman, D., & Tversky, A. (1984). Choices, frames and values. *The American Psychologist, 39(4)*, 341-350.

Kahneman, D., & Tversky, A. (1986). Rational choice and the framing of decisions. *Journal of business, 59(4)*, 251-278.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives, 5(1)*, 193-206.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational behavior and human performance, 21(1)*, 61-72.

Ke, D., Ng, L., & Wang, Q. (2010). Home bias in foreign investment decisions. *Journal of International Business Studies, 41(6)*, 61-72.

Keim, D. B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of financial economics, 12(1)*, 13-32.

Khanesar, M. A., Teshnehlab, M., & Shoorehdeli, M. (2007). A novel binary particle swarm optimization. *2007 Mediterranean Conference on Control & Automation* (pp. 1-6). IEEE.

Kroll, Y., Levy, H., & Rapoport, A. (1988). Experimental tests of the separation theorem and the capital asset pricing model. *The American Economic Review*, 500-519.

Krouse, H. J. (1986). Use of decision frames by elementary school children. *Perceptual and Motor Skills, 63(3)*, 1107-1112.

Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., S., Steinbrecher, M., & Held, P. (2013). *Computational intelligence: a methodological introduction.* Springer.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of economic behavior & organization, 17(3)*, 377-400.

Levy, H. (1992). Stochastic dominance and expected utility: survey and analysis. *Management science, 38(4)*, 555-593.

Li, X., & Engelbrecht, Andries P. . (2007). Particle swarm optimization: An introduction and its recent developments. *Proceedings of the 9th annual conference companion on Genetic and evolutionary computation* (pp. 3391–3414). London: ACM.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance, 20(4)*, 587-615.

Loeb, G. M. (1957). *The Battle for Investment Survival.* USA: Simon & Schuster.

Longerstaey, J., & Spencer, M. (1996). *Riskmetricstm—technical document.* Morgan Guaranty Trust Company of New York: New York.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. *Advances in experimental social psychology, 20*, 255-295.

Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of mathematical psychology, 43(2)*, 286-313.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, 1429-1445.

Luppe, M. R., & Fávero, L. P. L. (2012). Anchoring heuristic and the estimation of accounting and financial indicators. *International Journal of Finance and Accounting, 1(5)*, 120-130.

Machina, M. J. (1982). "Expected Utility" Analysis without the Independence Axiom. *Econometrica: Journal of the Econometric Society*, 277-323.

Markowitz, H. (1952a). Portfolio selection. *The journal of finance, 7(1)*, 77-91.

Markowitz, H. (1952b). The utility of wealth. *Journal of political Economy, 60(2)*, 151-158.

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments.* New York: New York: Wiley.

Markowitz, H. (1991). *Portfolio selection: efficient diversification of investments.* Cambridge: Mass: B. Blackwell, Cambridge.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica: Journal of the Econometric Society*, 111-141.

Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis, 7(4)*, 1851-1872.

Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour.* New York: Princeton University Press: Princeton.

Pham, D. T., Ghanbarzadeh, A., Koc, E., Otri, S., Rahim, S., & Zaidi, M. (2005). *Bee Algorithm A Novel Approach to Function Optimisation.* Technical Note: MEC, 501.

Prelec, D. (1998). The probability weighting function. *Econometrica-Evanston III, 66*, 497-528.

Preston, M. G., & Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *The American journal of psychology, 61(2)*, 183-193.

Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: a practical approach to global optimization.* Springer Science & Business Media.

Quiggin, J. (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior & Organization, 3(4)*, 323-343.

Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank-Dependent Model.* Netherlands: Springer.

Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica: Journal of the econometric society*, 431-449.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty, 1(1)*, 7-59.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica: Journal of the Econometric Society*, 571-587.

Selden, G. C. (1912). *Psychology of the Stock Market: Human Impulses Lead to Speculative Disasters.* New York: New York: Ticker Publishing.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance, 19(3)*, 425-442.

Shefrin, H. (2002). *Beyond greed and fear: Understanding behavioral finance and the psychology of investing.* Oxford University Press on Demand.

Shefrin, H., & Statman, M. (2000). Behavioral portfolio theory. *Journal of financial and quantitative analysis, 35(2)*, 127-151.

Shi, Y., & Eberhart, R. (1998). A modified particle swarm optimizer. *1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360)* (pp. 69-73). IEEE.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of economic perspectives, 17(1)*, 83-104.

Siegel, J. J., & Coxe, D. G. (2002). *Stocks for the long run (Vol. 3).* New York: McGraw-Hill.

Singer, A. E., Singer, M. S., & Ritchie, G. (1986). Role of transactions in mental accounting. *Psychological reports, 59(2)*, 835-838.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization, 11(4)*, 341-359.

Thakor, A. (2015). Lending booms, smart bankers, and financial crises. *American Economic Review, 105(5)*, 305-309.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing science, 4(3)*, 199-214.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making, 12(3)*, 183-206.

Thaler, R. H., & Shefrin, H. M. (1981). An economic theory of self-control. *Journal of political Economy, 89(2)*, 392-406.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven: Yale University Press.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological review, 102(2)*, 269-283.

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

Tversky, A., & Kahneman, D. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica, 47*, 263-292.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211(4481)*, 453-458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty, 5(4)*, 297-323.

Wall, G. (1993). *The Way to Save.* New York: Henry Holt.

Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization, 33(2)*, 167-184.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research, 30(1)*, 79-82.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management science, 42(12)*, 1676-1690.

Wu, G., & Gonzalez, R. (1999). Nonlinear decision weights in choice under uncertainty. *Management science, 45(1)*, 74-85.

Zhao, D., & Fang, Y. (2014). Representation Bias, Return Forecast, and Portfolio Selection in the Stock Market of China. *Mathematical Problems in Engineering*.