Pasquale De Luca

Analytical Corporate Valuation

Fundamental Analysis, Asset Pricing, and Company Valuation

Springer

Analytical Corporate Valuation

Pasquale De Luca

Analytical Corporate Valuation

Fundamental Analysis, Asset Pricing, and Company Valuation

Springer

Pasquale De Luca
Department of Law and Economics
of Productive Activities
Sapienza University of Rome
Rome, Italy

ISBN 978-3-319-93550-8 ISBN 978-3-319-93551-5 (eBook)
https://doi.org/10.1007/978-3-319-93551-5

Library of Congress Control Number: 2018954045

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

A Dadi,
che mi ha fatto scoprire la più bella e
misteriosa delle equazioni,
quella dell'amore.

Preface

Without a doubt, finance has undertaken a central role in the strategic governance of the company. In the recent decades, finance has been characterized by a strong development in the passage from a marginal and subordinate role compared with the other primary divisions to a central role in company governance. Today, finance is not only strictly related to these classical divisions by influencing their choices and their operating processes, but also it is increasingly able to influence the company's strategies and it is not unusual that it can in itself represent the starting point of new strategies and business models.

This central role of finance in the life of a company is undoubtedly the result of the globalization process and the correlated strong, fast and deep development of capital markets that have changed the development of economic models and then the company's business model.

The role of finance and its contribution is increasingly important for the survival and development of the company over time. Finance has already changed and will probably change more and more together with its paradigms, strategies and operating processes. Today, and even more in the future, it is difficult to imagine doing business without doing finance as well.

This book is based on a shareholder-oriented capitalism. Consequently, the company can thrive only if it is able to create value for the shareholders over time. In the Great Recession period as derived from the financial crisis of 2007–2008, this approach could be unpopular. Indeed, in common sense and in a part of academics the recent financial crisis is usually charged to the shareholder-oriented capitalism. This is especially true in Europe for economic and business cultures. I believe that the financial crisis cannot be attributed to a shareholder-oriented capitalism but to a particular distortion. The problem is the transfer from the long-term perspective to a short-term perspective. It is in contrast with one of the most relevant fundamental principles in the shareholders' value creation perspective that states the company's capability to create shareholders' value in the long-period is not the same as the maximization of its short-term profits. Often the choice of maximizing the shareholders' value in the long-term perspective is irreconcilable with the choice of maximizing the shareholders' value in the short-term period. Consequently, if the value creation in the long-term is confused with the profit in the short-term, it generates a great problem capable of damaging shareholders' interests as well as the stakeholders' interests. Therefore, the main problem of the financial crisis is not the shareholders' value perspective but the short-term perspective of some managers.

The company's ability to create value for the shareholders over time is strictly related to the deep understanding of the business model of the company as well as the investors' behaviour in the capital markets. The company valuation can be considered one of the most relevant fields in which the classical paradigms of the company meet the paradigms of the capital markets. Indeed, the right company's valuation requires high competence in the fields of strategy, financial management, corporate finance and capital markets.

The basic equation of the value is based on a principle that dates back to Alfred Marshall: a company creates value if and only if the return on capital invested exceeds its cost of capital.

The amount of value is equal to the difference between cash-in flows derived from the investment and the cost of capital invested able to reflect the time value of money and the risk premium. Consequently, to create value over time, the company must invest the capital raised at a rate of return higher than its cost of capital. Therefore, there are two main variables of value creation:

– the *return on capital invested*;
– the *cost of capital*.

In this book, the company's ability to invest the capital raised by obtaining a high return is investigated through an analysis of the company's fundamentals with regards to its business model and its economic and financial performance over time.

Specifically, the return on capital invested in the business is function of the company's business model and the quantitative effects of the strategic choices on its economic and financial dynamics. Specifically, the company's ability to create profit over time requires an analysis based on two main parts:

– the qualitative analysis of the business model;
– the quantitative analysis of the company's performance which regards the effects of the business model choices on the economic and financial dynamics over time.

Otherwise, the company's cost of capital invested in its business is derived from the investors' behaviour and their analysis of the risk-return profile of the company in the capital markets. The cost of capital for the company is one of the most relevant topics for managers and financial economists, and it plays a central role in the valuation models of the company. For decades, several studies have focused on the relationship between capital structure, cost of capital and company value. Despite a broad experience approach in both academic and practices, it should not be surprising that the method for estimation of the cost of capital is still under intensive discussion.

An estimation of the cost of capital for the company is based on the investors' behaviour and expectations in the capital market. It requires the knowledge of their models about the risk valuation and the expected returns estimation. The greater the managers' skill to understand the investors' behaviour and their choices, the greater the company's probability to satisfy the investors' expectations by acquiring the capital required for its development at favourable conditions.

Specifically, the cost of capital is function of the asset pricing in the capital markets. It is the function of the investors' models about risk diversification and returns maximization, and thus, it can be derived by general equilibrium model in the capital market.

Based on these two variables, return on capital invested and the cost of capital, the company's ability to create value over time for its shareholders is the function of the effectiveness of the Company Strategic Formula to create expected cash-flows as well as investors' models to diversify the risk and maximize expected returns in order to estimate the cost of capital.

The basic equation of value states that the company creates value if and only if the return on capital invested exceeds its cost of capital. The explicit application of the basic equation can be realized through several methodologies. Among them, the Discounted Cash-Flows model (DCF) is the best. It is commonly used in the financial community. It is relevant since all members of the international financial community use a common criteria and language.

By using the DCF approach, the company value is equal to the current value of expected future cash-flows and the cost of capital is used as a discount rate. Therefore, there are three main variables:

- *Time*: the referenced time is the future. The value of the company is strictly related to future performance rather than past performance;
- *Cash-flows*: company's performance is measured in cash-flow terms. Specifically, the expected future cash-flows from operations and to equity;
- *Cost of capital*: it is the cost of debt and the cost of equity, and it defines the discount rate for expected future cash-flows.

The *General Equation of Value* can be defined based on these three main variables as follows:

$$W_F = \sum_{t=1}^{\infty} \frac{CF_t}{(1+k)^t} \tag{1}$$

where W_F is the company's value; t is the period-time of valuation; CF_t is the expected future cash-flows for each year (t); k is the cost of capital used as a discounted rate.

Equation (1) has a great theoretical relevance. It estimates the value of the company based on expected cash-flows, arising from the fundamental analysis of the company and the cost of capital. Also, the equation defines the relationship

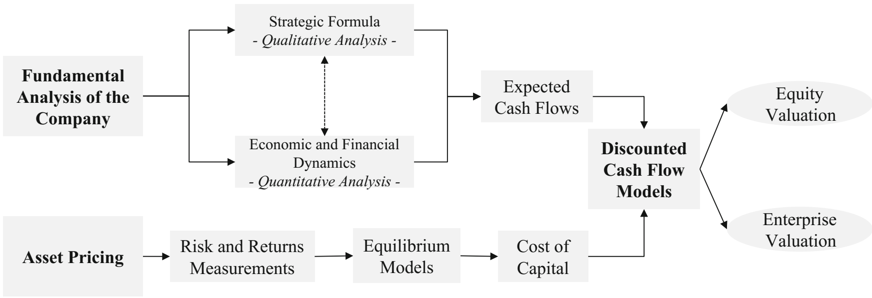


Fig. 1 Methodological approach

between company value, the expected cash-flows and the cost of capital in the time of valuation.

The integration between models on company's fundamentals from which the expected cash-flows are derived and the investors' models about risk and return in the capital market by which the cost of capital for the company is derived can be summarized as follows (Fig. 1).

The integration between the company's fundamental analysis and the investors' models of risk and returns in the capital markets is essential for the company's success over time. It is not possible to fully understand the company's ability to create value over time and to measure this value without the simultaneous deep knowledge of these models and their integration.

Consequently, the managers must define their strategies and operational processes by considering the business and industrial logics with regard to customers, suppliers, competitors, as well as the financial criterion with regard to investors in equity and debt.

Therefore, clear thinking about drivers of the company's value creation as well as a right approach to its measurement requires two main skills: (i) the analysis and evaluation of the company's fundamentals with regards to its business model and its performances over time; (ii) the knowledge of the investors' models about risk diversification and returns maximization from which the cost of capital for the firm is derived. To integrate the company's fundamental analysis and the investors' models about risk and return in the capital markets with reference to company valuation, the book is characterized by a large recourse to a rigorous quantitative analysis. Specifically, the methodological approach used in this book is based on:

– mathematics, to assure the consistency of models in its construction;
– graphics, to provide intuition;
– words, to explain the results and the economic significance.

The large use of a rigorous quantitative analysis to integrate the company's fundamental analysis and the investors' models about risk and return in the capital markets in order to the company valuation is not to complicate the analysis but, on the contrary, to simplify the discussion. There are three main reasons.

- first, models are easier to understand if they are studied in their formal construction. The mathematical form allows us to further understand the models in their construction, assumptions and, then, in their clear capabilities and limits;
- second, the mathematical form does not allow inappropriate manipulation of the equations and, consequently, an incorrect use of the models. Every equation is the result of a rigorous formal process and their modification can be realized only by following the same rigorous formal process;
- third, the mathematical form does not allow for attribution of the equation meanings that are not supported by their strict formal derivation. Every equation acquires form and meanings strictly related to the mathematical process of derivation. The clear derivation step-by-step of each equation does not allow errors in the equations interpretation and, consequently, in the use of models.

Rome, Italy
June 2018

Pasquale De Luca

Contents

Part I
Fundamental Analysis of the Company

Chapter 1
Company Business Model Analysis

Abstract This first step of the company analysis focuses on the business model. The company's "health condition" in a defined period of time and its ability to create profit over time requires an analysis based on two main elements:

- a qualitative analysis of the business model;
- a quantitative analysis of the effects of the business model choices on the economic and financial dynamics over time.

A qualitative analysis of the company's business model focuses on the *Company Strategic Formula* (*CSF*).

The CSF defines the strategic profile of the company on the basis of two different strategic fronts:

- *internal strategic front* referring to the internal structure of the company;
- *external strategic front* referring to the structural relationships between the company and the players of its environment classified into two main groups: business players and financial players.

The CSF allows for simultaneous optimisation of the companies operating in the Strategic Business Area and Capital Market. The internal and the external strategic fronts are strictly connected on the basis of systemic and dynamic bidirectional relationships. In this sense the CSF must be "continuant": it is achieved only if the relationships between all of its elements are *Systemic-Structural-Dynamic*.

A quantitative analysis of the company focuses on the economic and financial dynamics over time. Several analytical schemes can be used. In this context the analysis is developed on the basis of Operating and Net Income, Capital Invested and Capital Structure, and Free Cash-flow from Operations and Free-Cash Flow to Equity.

The qualitative and quantitative analyses are strictly related. The competitive advantage of the company, on the basis of its business model, must be reflected in the economic and financial values over time. Consequently, it is not possible to investigate into the company by only taking into consideration the analysis of its business model without considering the effects of the strategic choices on the

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_1

economic and financial dynamics. At the same time, it is impossible to investigate into the company's ability to perform by considering the economic and financial dynamics without clearly understanding the source of the strategic choices.

1.1 Strategic Formula of the Company

The company can be defined as a dynamic and open system towards the environment with which it maintains several types of relationships in order to pursue the economic and financial equilibrium to be valid over time.

The company can be considered as a production system, with regards to the combination and coordination of its production factors, as well as a system of relationships with the environment with which it exchanges materials and products, flows of information and financial values on the other side (Bertini 1990; Bianchi Martini 2009).

Although the company is part of the environment, from which it can never be separated, it qualifies itself as unique, different and independent from the environment. The company defines its profile and the business, and it identifies the players with which it establishes bi-directional relationships (Giannessi 1979; Bertini 1990; Coda 1988).

Relationships between the company and the environment are constantly changing due to the dynamic company-environment paradigm. The dynamic nature of the competitive context drives the company towards constant renewal. It requires a "*strategic attitude*" from the company rather than a defined strategy (Bertini 1995; Garzella 2005, 2006; Markides 2008; Galeotti and Garzella 2013). Therefore, the definition and development of the strategy should not be considered as a separate and unique moment in the life of the company, but an on-going process.

Based on these considerations, the company's government requires a model characterized by internal efficiency and effectiveness on the one side, and a coherent and balanced system of relationships with all external players on the other side.

The company's "health condition" in a defined period of time and its ability to create profit over time requires an analysis based on two main elements:

- the first is a qualitative analysis of the business model;
- the second is a quantitative analysis of the effects of the business model choices on the economic and financial dynamics over time.

These two parts are strictly related. In reality, it is not possible to investigate into the company through the sole analysis of its business model without considering the effects of the strategic choices on the economic and financial dynamics. In the same way, it is impossible to investigate into the real company's ability to perform by taking into consideration the economic and financial dynamics without fully understanding the source of the strategic choices.

Therefore, the competitive advantage of the company on the basis of its business model must be reflected in the economic and financial values over time.

This paragraph focuses on a qualitative analysis of the company's business model, while the other paragraphs focus on the analysis of the economic and financial dynamics over time.

The qualitative analysis of the company's business model proposed in this context is defined as a *"Company Strategic Formula"* (*CSF*) (De Luca 2013a, 2015; De Luca et al. 2016, 2017).

The CSF is the business model of the company as defined in its strategic profile. Specifically, the CSF defines the way in which the company is organised internally and how it manages the relationships with external players for self-development over time.

The CSF can be considered as the ideal conceptual place in which, on the basis of a systemic and dynamic paradigm:

(a) the *ideas* are developed;
(b) the *decisions* are made;
(c) the *operations* are defined and planned.

On the basis of a systemic and dynamic perspective, the CSF allows for the transformation of the "*system of ideas*" into the "*systems of operations*" by means of the "*systems of decisions*" in order to achieve and maintain economic-financial equilibrium over time. Therefore, CSF takes form and substance in a unique, systemic and dynamic way to the entrepreneurship and managerial skills of the company (Bertini 1995). It allows the company to acquire and develop over time a certain level of superiority over competitors, both qualitative (with regards to the acquisition, development and conservation of a defensible competitive advantage) and quantitative (with regards to the achievement of economic and financial performance levels higher than those of their competitors) (Porter 1985; Grant 1991; Invernizzi 2008; Galeotti and Garzella 2013).

The CSF defines the strategic profile of the company, by considering two different "*strategic fronts*":

– *internal strategic front*: it refers to the internal structure of the company;
– *external strategic front*: it refers to the structural relationships between the company and the players of its environment classified into two main groups of business players and financial players.

Internal Front of the CSF

The *Internal Strategic Front* (*ISF*) refers to the internal structure of the company. It is defined from all elements, tangible and intangible, needed for the production of goods and services. The internal structure defines the company's specific characteristics by generating its uniqueness. It gives form and substance to the CSF by establishing the uniqueness of the thinking and operation of the company. Therefore, it is the main reason for which one company is different from another.

The internal structure of the company must be contemporarily both "*stable in the moment*" and "*dynamic over time*". It must be constantly seeking a form of balance between the internal characteristics of the company and the needs of the markets in order to develop and defend the competitive advantage.

The internal structure of the company is defined on the basis of three main elements:

(a) *Corporate governance*
(b) *Organizational architecture*
(c) *Strategic resources*

(a) ***Corporate Governance***

The corporate governance of the company refers to the rules and the procedures by which the decision-making processes in the governmental area and how the managerial and operating activities of the company are defined (Bertini 2009).

Corporate governance plays a key role in the CSF by activating processes that can be virtuous or vicious as the case may be (Bianchi Martini 2009).

It is important to highlight how the corporate governance quality is a function of its operating efficiency rather than compliance with the rules as defined by law. Corporate governance involves "*substance*" rather than "*form*"; it is a "*system of government*" rather than a "*system of rules*" (Bertini 2009).

In this sense the corporate governance processes and rules that involve all managers must necessarily be characterised by a high level of professionalism and competence (Bianchi Martini 2009). Therefore, the corporate governance model is defined based on the specific characteristics of the company (Fiori et al. 2004) mainly with regards to its well-defined elements of "*entrepreneurship*" and "*managerial*" skills. While the first refers to the company's ability to project itself towards the future looking for new opportunities, the second refers to the knowledge and competence needed for fulfilment of the opportunities by connecting the entrepreneurial intuition and its execution. This combination allows for facing of the customers' expectations by creating value for the company over time (Bertini 1995, 2009).

The mechanism by which the entrepreneurial idea is converted into the company's strategies on the basis of the resources available is function of the corporate governance (Bertini 1995; Bianchi Martini 2009). Corporate governance defines the mechanism of dynamic and bidirectional relationships between the company and its environment on the basis of the entrepreneurial idea and the specific decision-making process of the company.

The decision-making processes must be considered able to affect the entrepreneurial idea, by modifying it in several forms and degrees, based on the knowledge arising from the systemic and dynamic bidirectional relationships between internal and external strategic fronts (Mintzberg 1994).

It is worth noting that the corporate governance of the company is also influenced by the structure of the economic and financial system and the culture of the country references of the company (Bianchi Martini 2009). In this sense, it can be distinguished between two dimensions of the corporate governance: the first is based on the unique nature and specific characteristics of the company while the second is function of the specific characteristics of the institutional, economic, financial and culture system of the country references of the company. With regards to this second dimension, there are two main models: the model of "market-oriented" and "bank-oriented" capitalism.

The model of "market-oriented" capitalism is based on the separation between management and shareholders of the company.

In this model the main problems of governance refer to the agency conflicts (Jensen and Meckling 1976; Jensen 1986). The agency conflicts are carried out between management and shareholders, as well as, shareholders and debtholders. Managers, shareholders and debtholders are all focused on the company value creation but they are characterized by different utility functions and different information owned due to the different role in the company's government (Dallocchio et al. 2011).

With regards to the conflicts between managers and shareholders, the different utility function and the information asymmetries lead the first to maximize the enterprise value while leading the second to the equity value. The agency costs on equity are due to the introduction of control mechanisms on management activities, as well as the introduction of a managers compensation system to align the aims of managers with those of the shareholders.

With regards to the conflicts between shareholders and debtholders, the different utility functions and the information asymmetries lead the first to maximize the dividends and investment policies even at the expenses of debtholders. In this sense high risk investments are preferred. In this case, by considering the limited liability of the shareholders and the fixed remuneration of debtholders, if the investment is successful, the shareholders obtain the benefits while if the investment fails the debtholders claim a part of the cost if the company cannot pay interest on debt and debt reimbursement. Therefore, the debtholders do not participate in the investment benefits but they may have to bear the costs in the event of failure (Harris and Raviv 1991; Diamond 1989).

Furthermore, the shareholders lead managers to maximize the dividend policy. A policy of high dividends reduces the company's self-funding by reducing sources to be invested in the business. Over time it can reduce the company's ability to face customer expectations. It increases the operating and financial risk of the company by reducing the company's ability to face debt obligations.

The cost of an agency on debt is mainly due to the debtholders policies to reduce the risk of shareholders' opportunistic behaviour. The adoption of these polices by debtholders increases the cost of debt ex-ante due to the real application of opportunistic behaviour due to the low level of confidence of debtholders in the company (Myers 1977; Dallocchio and Salvi 2004).

The "bank-oriented" capitalism model is based on the strong relationship between bank systems and industrial systems. In this context the banks play a key role in the capital structure of the companies. Therefore, the bank can play an important role in the company's corporate governance and thus on the company's choices regarding strategies and operations. The bank's control on management and shareholders reduces agency costs on equity and debt. On the contrary, there are two main problems: the first refers to the conflict between majority and minority shareholders based on the asymmetric information, while the second refers to annulment of the capital market discipline effects. The capital market is unable to measure the company's performance and the price is no longer the reference point of the company's value and its operativeness. In this context only a few large companies are financed on the basis of policy choice rather than on the capital market's selection based on the effectiveness and efficiencies of the companies with all imaginable negative effects.

(b) ***Organizational Architecture***

The organizational architecture of the company refers to how the company's resources are combined and coordinated between them for company operations.

The company's ability to compete in the business is strictly related to its organization structure and operations.

There are two main levels involved:

- the *organizational structure* of the company with regards to both hard and soft elements that give form, substance and operation to all parts of the organization;
- the *operations* with regards to the processes and procedures that cross the company vertically and horizontally.

The *organisational structure* defines the organizational context. It defines the decision-making environment in which company strategies are defined. It is based on hard and soft components. While the hard components define the mechanical operating nature of the organisational structure, the soft components define the operating mode based on formal and informal flows of information.

The hard component refers to the variables that define the organizational model. It defines the work model of the company, the levels of hierarchy, the relationship mechanisms among all parts of the organizational structure, both formal and informal.

There are three main elements that define the hard components (Invernizzi 1999):

- *organizational model*: this choice is a function of the company's characteristics and the decision-making processes. There are two main logics: a vertical logic, based on hierarchy relationships among the different levels, or a horizontal logic based on the processes and activities that cross the company;
- *organizational microstructure*: this choice involves the definition of the central, staff and line functions with regards to their internal structure, roles, positions and skills;
- *mechanisms of relationship*: they involve definition of the characteristics and structure of the relationships among all parts of the organizational structure, both formal and informal.

The soft component refers to the intangible variables such as culture, values and managerial approach (Invernizzi 2011). They give substance to the relationship mechanisms in the organizational structure on the basis of the behaviour acquired by the company as function of its value and culture (Hofstede 1993).

There are two main elements to define the soft component of the organizational structure (Invernizzi 1999):

– *the degree of entrepreneurship*: there are two main archetypes of organizational context. The first is the entrepreneurial model based on a high level of informal relationships in the company in order to develop a new idea thanks to the contribution of all parts of the company. The second is the bureaucratic model based on a high level of formal relationships in the company in a rigid structure based on strictly hierarchy levels;
– *the degree of discipline*: there are two main archetypes of organizational context. The first is the discipline model characterized by a high level of cooperation among all parts of the organization leading to maximisation of efficiency. The second is the undisciplined model characterized by a high level of confusion and opportunistic behaviours in all parts of the company.

Therefore, while the hard elements define the operating units of the company and their relationships based on rational mechanics, the soft elements give life and substance to the company in a unique way based on the value, culture and knowledge that define all company members.

The *operations* refer to the processes and procedures that cross the company vertically and horizontally.

The company can obtain a competitive advantage if it can fulfil the processes and activities in a much more efficient manner than its competitors or if it is able to create new ones (Porter 1985).

In this context the key role is filled by the make or buy decisions (Williamson 1985; Porter 1985; Grant 1991; Thompson et al. 2006). It requires clear identification of the strategic processes and activities that cannot be outsourced.

It is worth pointing out that if alignment between the organisational structure and company strategies is necessary. It implies that the organizational structure must be allowed the organizational effectiveness in the direction defined by company strategies.

This alignment must be dynamic since both the internal structure of the company as well as its environmental structure are characterized by dynamic development. Therefore, the relationship between organizational structure and the strategies must be circular rather than linear (Grant 1991). In this sense an optimal organisation context should be defined rather than an optimal organizational structure as it allows for response to a change in the environment.

Definition of the Organization Structure and Operations must be related to the company strategies in its competitive environment. The alignment between strategies and organization and operating model is the result of the bidirectional relationships based on an iterative logic approach (Grant 1991).

(c) ***Strategic Resources***

The strategic resources of the company refer to the company's tangible and intangible assets and the human skills necessary for their coordination.

The company's strategic resources represent the most important way of competing in the business (Grant 1991; Hamel and Prahalad 1990; Quinn 1992).

Generally, the strategic resources provide the company with uniqueness and are able to protect its competitive advantage from imitation processes by generating "*isolation mechanisms*" (Rumelt 1987).

Consequently, management of strategic resources over time has the highest strategic relevance for the company. The company-environment paradigm is characterized by high dynamicity and complexity. The company has to acquire a strategic approach able to modify the internal structure faster than competitors by following market changes (Bertini 1995; Coda 1988). The strategic resources allow the company to modify its internal structure by aligning it with business changes. In this sense, the company's policy for the increase and development of its strategic resources must be developed in two main directions:

- focusing attention on the processes of training, accumulation and control of the implemented resources;
- constant re-thinking on the composition of resources by integrating them for their re-vitalization due to the dynamic relationship between company and its environment. The re-thinking of the strategic sources must be made by imagining what resources will be distinctive in the future in order to satisfy market changes.

It is worth noting that the core competences based on strategic sources do not change in core rigidities over time. It happens when strategic resources cannot be renewed in order to follow market changes. In this case the resources able to generate a competitive advantage in a given time become the first cause of company failure.

In this sense, by considering that the strategic resources of the company are the cause and the effect of its competitive advantage, it is necessary to develop and maintain over time a circular relationship between strategic sources and competitive advantage: the competitive advantage of the company is function of its strategic resources that, in turn, are developed over time based on the effects of the strategic choices.

Finally, as far as the company is concerned, it is important to imagine the source that will become distinctive in the future for market changes.

Not all sources can be defined as strategic resources. Not all resources are able to build and defend the company's competitive advantage over time. To be strategic, a resource must be relevant for value creation and must be characterized mainly by:

- *scarceness*: the higher the source of scarceness, the higher the relevance;
- *uniqueness*: the higher the ability of the source to generate company's uniqueness, the higher the relevance;
- *inimitability*: the higher the source inimitability, the higher the relevance;

– *capabilities over time*: the higher the duration over time of source capabilities, the higher the relevance;
– *depreciation*: the lower the depreciation rate, the higher the relevance;
– *specificity*: the higher the number of times of use, the higher the source relevance;
– *combinability*: the higher the source capability to combine with others, the higher the relevance.

Based on these considerations, tangible assets can rarely be considered as a strategic resource because they can be easily replicated by competitors.

Intangible assets are more difficult to replicate by the competitors than tangible assets due to their specific nature. Generally, they arise from the internal processes of the company and therefore they are unique. Also, unlike tangible assets, their value does not decrease over time and sometimes it increases thanks to use (Grant 1991; Collis et al. 2012).

Human capital (or intellectual capital) refers to the people whose skills, knowledge, culture, ideas and values allow for combination and coordination between tangible and intangible assets in order to achieve the strategic target of the company (Quagli 2001).

In this context, human capital is the main strategic resource for the company (Grant 1991; Itami 1988; Saloner et al. 2001; Thompson et al. 2006) and it represents the real source of the company's competitive advantage. People at all levels of the organization represent the main and deeper reason of strategic thinking and operation of the company as well as its ability to create value over time (Bertini 1995). Therefore, human capital is the most relevant element of diversification among the companies. Consequently, it gives uniqueness and non-imitability to the company's competitive advantage and is defendable over time (Barney 1991; Rumelt 1984; Wernefelt 1984; Hamel and Prahalad 1990; Itami 1988).

External Front of the CSF

The *External Strategic Front* (*ESF*) refers to the structural relationships between the company and external players (Coda 1988; Galeotti 2008; Galeotti and Garzella 2013). These relationships can be exchange related, if they refer to products, services and money, or conditioning if they refer to constraints, limitations and opportunities. Also, they may have both an economic content (revenues, costs, price, etc.) as well as a non-economic content (Galeotti and Garzella 2013). Company competitiveness is due to its ability to create value for all of its players simultaneously.

The external players of the company can be classified into two main groups: business players and financial players. Based on differences in their nature, interests and behaviour, the external strategic front can be divided in two main parts (Coda 1988; Galeotti 2008; Galeotti and Garzella 2013):

(a) *Strategic Business Area*;
(b) *Capital Market.*

(a) ***Strategic Business Area***

Strategic Business Area (*SBA*) the company refers to the real market in which it carries out the business. The company can operate in more than one business. In any case, any SBA can be defined on the basis of two main elements (Porter 1985):

- *competitive players*: it refers to the players with which the company defines relationships. Specifically, they are customers, suppliers, competitors (both existing competitors and potential competitors entering the business, as well as the producers of potential substitute products). The relationship between the company and customers is characterized by continuity (due to the sale process) and stability (arising from the possibility to identify a hard core of customers). More and more the relationship between the customers and the company is based on emotional elements that go beyond the technical characteristics of the product. Therefore, the politics of the company with regards to customers are focused on these emotional elements. The relationship between the company and supplier is characterized by continuity (due to the buy process) and stability (deriving from the possibility to identify a hard core of suppliers). The need for the company to achieve and maintain a high-quality level, drives the company to establish strong relationships with suppliers especially if they are considered strategic.
  The relationship between the company and competitors, is characterized by continuous and systematic competition in order to acquire and to maintain a competitive advantage. The relationship is based on strategic interaction arising from the dynamic mechanisms of actions-reactions by generating the competitive dynamics in the market and its average profitability.
  The interactions among these players define the competitive system in the SBA (Porter 1985);
- *product system*: it refers to the product offered by the company with regards to its material and immaterial elements, service components and economic and non-economic terms. The product must also incorporate the image and reputation of the company. Therefore, it should be defined as a "*product system*" rather than a product because it incorporates the technical elements, as well as the image, the value, and the history of the company (Bertini 1990; Coda 1988; Bianchi Martini 2009).

In each SBA the company competes by means of a defined business strategy (De Luca 2013b) in order satisfy customer requirements and expectations better than competitors. It allows the company to acquire a competitive advantage in the business and greater profitability than competitors (Porter 1985).

The company's business strategy is defined according to its internal characteristics and the structural characteristics of its business area. The company defines its business strategy in order to satisfy customer requirements better than its competitors. In this case the company acquires a competitive advantage in the business, allowing the company to undertake a dominant role in the market and higher profitability than the average level of the market.

If there is more than one SBA, the company has to develop a business strategy for each SBA on the one side and a Multi-Business Strategy on a corporate level in order to coordinate the different business strategies and to define the incoming SBA's, those in which to continue to operate and those from which to exit (Invernizzi 2011; Garzella 2006).

It is worth noting that the competitive advantage of the company is bound to get lost in time due to competition. Therefore, the CSF must be continually renewed over time. In this sense, the company's ability to compete in the business is function to the quality of its strategic thinking rather than a well-defined competitive advantage. Therefore, a dual logic is necessary: on the one hand, the company has to develop its current strategy in order to fulfil a given target; on the other hand, the company has to constantly re-think the strategy for its renewal and development of new future competitive advantages.

The development of a new business strategy must be defined by considering the CSF in its entire structure. It implies that business strategies must be aligned in a dynamic way with internal characteristics of the company and its strategies in the capital markets.

The business strategy must be defined by considering jointly the internal characteristics of the company and the customers' needs and expectations.

Specifically, the business strategy requires the definition of the competitive advantage of the company.

On the basis of analysis of the business structure and characteristics, and by considering the internal characteristics the company must define its competitive advantage to be fulfilled in the business.

Generally, the company acquires a competitive advantage if the product meets the customers' expectations at a higher level than the competitors. The product superiority can be defined on the basis of its material characteristics or emotional characteristics due to its immaterial elements.

It is worth noting that the competitive advantage is specific for the business area and it refers to the product to satisfy customer expectations. It should also be effective and defendable over time by the company.

In this context, the company should correctly evaluate its internal characteristics and the quality of sources available and sources available in a short-time. There are two main competitive advantages that the company can pursue: the *cost advantage* and the *premium-price advantage* (Porter 1985).

The *cost advantage* is fulfilled through a cost strategy that allows for cost structural efficiencies. It is fulfilled whenever the company can create a product at a structural cost (and thus sustainable over time) lower than competitors. The product cost reduction must be achieved without the product quality reduction compared to competitors.

Cost leadership in the business is unique. However, the second leader and the followers can fulfil a cost strategy by obtaining a return on capital investment lower than the cost leader but higher than the market average. The most relevant advantage of the cost leader is that it is also the price leader. The lower price on the market can be defined and therefore it can influence market competition.

The cost strategy can be achieved on the basis of two main directions:

- structural reduction of the cost of processes and activities of the company: in this case the company must manage the cost drivers such as economies of scale, learning processes, access to unique resources, interrelations, integrations, connections, synergies and institutional factors. They usually are combined between them where each one reinforces the other through coordination and maximization;
- reengineering of processes and activities: structural cost reduction requires the re-definition of the company's processes and activities based on a new production method. In this case, it is not pursuing an incremental improvement of the processes and activities with cost reductions, but new processes and activities with a new level of cost.

These two directions can be achieved jointly as well as separately.

Thanks to structural cost reduction, the company can translate this advantage in its price or margin directly.

The cost advantage can be used to:

- reduce the product price: in this case the unit profit margin per product does not change. The price reduction is equal to the cost reduction;
- increase the product unit profit margin: in this case the product price does not change. The reduction in the cost of product increases the profit margin;
- reduce the product price lower than cost reduction: in this case there are the joint effects of the price reduction and the unit profit margin increases.

The *price premium advantage* is fulfilled through a differentiation strategy. It is only fulfilled if the company is able to differentiate its product and the customers are willing to pay a price (premium-price) higher than that of similar products.

The advantage differentiation is achieved if the company can achieve a price (premium-price) higher than competitors based on higher quality of product. The differentiation strategy is successful only if the premium-price is higher than differentiation costs borne by the company.

The advantage differentiation is based on differentiation drivers. They refer to conditions that give the products its unique nature. Also in this case they usually are related among them and they operate jointly.

The differentiation strategy is more useful, the greater the heterogeneity of customers' needs by requesting non-standard products (Thompson et al. 2006).

The main sources of price-premium are the following:

- innovative products: they are protected by patents or, however, they are difficult to copy, or both. Without either two of these protections, the innovative product can be copied and then it is not able to generate high returns over time;
- quality: the product must be characterized by a real or perceived difference with the others and the customers are willing to pay a higher price;
- brand: it refers to product perception of the customers based on the company's brand. The customer choice is based on the brand first of all and then on the

product. However, the price-premium based on brand is difficult to distinguish from the price-premium based on quality. The two are usually highly correlated.

The cost advantage and price premium advantage cannot be achieved jointly. The two strategies are different in their structure and require different product characteristics that can be achieved with different internal structure characteristics of the company.

The company's ability to defend the advantage over time, both cost and premium-price, is function of the quality of the drivers used and their effectiveness in the building of barriers to enter the business.

The choice of a competitive advantage must be made by the company by considering its internal characteristics and the structure and characteristics of the business.

The business structure and characteristics can be defined on the basis of three main pillars (De Luca 2013c):

– market structure;
– market cycle-life;
– market competitive dynamics.

It is worth noting that each one of them implies a specific business element that the company must evaluate for definition of the business strategy. Indeed, perfect alignment between the business structure and characteristics and the company's business strategy is necessary.

The *market structure* refers to the level of concentration. The business can be distinguished between: highly concentrated market, fragmented market, niche market.

The highly concentrated market is characterized by few companies with a high market share. There are two main forms: the oligopolistic structure in which there are few big companies but no one is dominant; the structure with a dominant company, is when there is one big company and few smaller companies that share the residual market.

The fragmented market is characterized by a lot of companies with a small market share. In this market competition among the companies is high and the average profit could be low.

The niche market is characterized by customers with homogeneous preferences among them and highly inhomogeneous preferences with all of the others. The market niche is a small part of the market. The competition level in a market niche is usually different from competition in the market.

Once the market structure has been defined, it is necessary to fully understand the stage of the life-cycle in which the market is located. The *market cycle-life* can be divided into four main phases that are well-known in literature: introduction, development, maturity and decline. Obviously, it is not necessary for the market to cross all of these stages. They represent an ideal scheme of the cycle-life of market.

In this context, it is important to fully understand the stage in which the market is located as each stage is characterised by specific elements that must be considered in order to define the business strategy of the company in a current and future perspective. There must be coherence between the company's business strategy and the stage of the cycle-life in which the market is located.

In the introduction stage, the business is characterized by a structure that has not yet been fulfilled completely. There is normally radical innovation in the technology of products and/or processes to generate a new business. In this phase, the business is characterized by uncertainty of the product due to low consumer knowledge and then uncertainty in demand and expectations, uncertainty in defensibility of the business in its future development. However, the uncertainty of this phase can also generate positive elements. The absence of clear rules allows for experimentation of the company strategy. The company can acquire a market share due to an increase in demand function of customers' increase, as well as a low competition level (Thompson et al. 2006).

In the development stage, the business is characterized by an increase in demand with regards both to volume and revenues. Transition from the introduction phase to the development phase is usually fast and it generates an increase in competition level in the business. The main problem of the company is to acquire a good position in business as well as to defend it over time. Development of the business attracts new competitors by increasing the competition pressure.

In the maturity stage, the business is characterized by stable demand with regards both to volume and revenues. The market share of companies is normally stable. The competition level among companies is high because the increase in the market share of each company can only be fulfilled through a reduction in that of the others.

In the decline stage, the business is characterized by a structural decline of demand with regards both to volume and to revenues. The product is no longer attractive to customers and it requires regeneration. A new development of the business can only be achieved through its regeneration based on a new concept of product.

Finally, once the market structure and the stage of its lifecycle in which it is located have been defined, it is important to clearly understand the *competitive dynamics* of the market (De Luca 2013d). Generally, the intensity of the strategic interactions in the business are defined according to the action-reactions scheme among companies. The company must, subsequently, face on the one hand customer expectation and, on the other hand, the strategies of competitors with regards to the same demand. In this sense the company's business strategy must be defined by also considering the effects on the business strategies of its competitors and their reactions. The strategic interactions change among the businesses and for the same business in different times.

Therefore, the company must evaluate its business according to its structural characteristics as well as the intensity of the strategic interaction among the companies.

(b) ***Capital Market***

Capital Market refers to the ideal place in which a company looks for the capital, in equity and debt, needed for its survival and development over time.

The capital market can be defined according to two main elements:

- *financial players*: it refers to the investors in equity and debt. The relationship between company and debtholders is of stable nature. The company normally defines a long-term debt level and parts of it are constantly being replaced in the short-term period. The relationship with debtholders becomes complex if the debt level is too high by generating structural financial imbalance. In this case, the debtholders can acquire, directly or indirectly, a role in the company's government with relevant effects on their strategies. The relationship between company and shareholders is of a structural nature. They invest in equity, then they undertake full risk of the company, and they are entrusted the company's government. The relationship undertakes specific characteristics depending on whether or not they are majority or minority shareholders;
- *financial company profile*: it refers to the risk return profile of the company. It is function of the company's expected cash flows on one side and investor expectations about risk and returns on the other.

In the capital market the company competes through its financial strategy in order to acquire the capital needed, in equity and debt, at profitable conditions (Galeotti 2008).

Consonance of the Corporate Strategic Formula

The CSF, as defined in its internal and external strategic front, allows for simultaneous optimisation of the company operating in the Strategic Business Area and Capital Market. The internal and the external strategic fronts are strictly connected on the basis of systemic and dynamic bidirectional relationships.

The CSF based on its structural elements as defined, can be schematically represented as in Fig. 1.1.

The internal and external strategic fronts are two parts of a whole. The success of the company is function of their joint quality. They are subsequently strictly connected by systemic and dynamic relationships.

Therefore, the CSF must be characterized by a "*consonance*" between all structural elements of the internal and external strategic fronts. This consonance must be (De Luca 2013a, De Luca et al. 2017):

- *Systemic*: all elements of the internal and external strategic fronts must be aligned between them;
- *Structural*: there must always be correspondence between the characteristics of each element of the strategic fronts, both internal and external, based on well-defined and structural bidirectional relationships;
- *Dynamic*: the systemic-structural relationships between elements of the internal and external strategic fronts must be dynamic over time and never static.

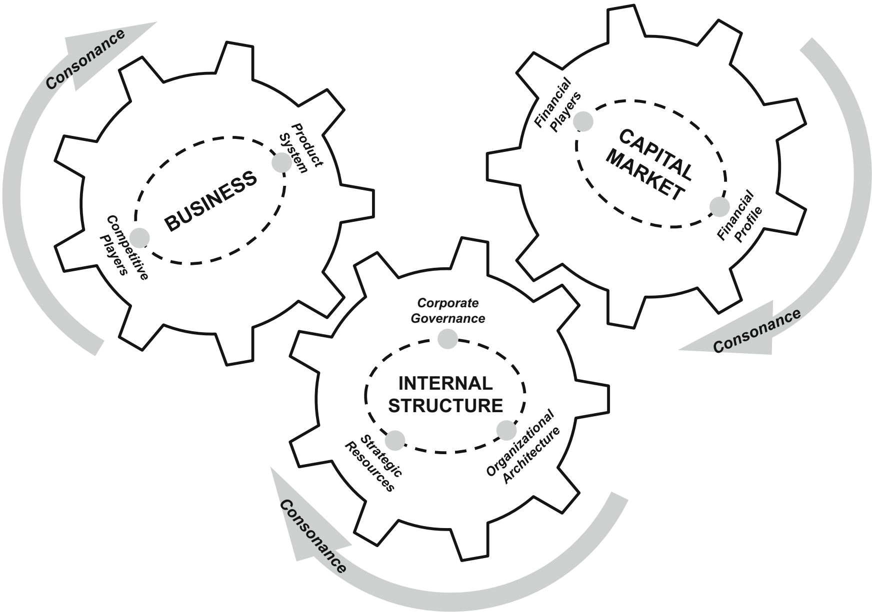


Fig. 1.1 Corporate strategic formula (CSF)

Therefore, the CSF can be defined as "*consonant*" only if the relationships between all of its elements can be defined as *Systemic-Structural-Dynamic*.

The search for consonance must be considered as a strategic approach based on change rather than a specific target to be achieved in a given period of time. Indeed, each element of both internal and external strategic fronts, as well as the relationships between them, change over time because they change the company's environment (business area and capital market) as well as its internal characteristics. The success of the company can never be defined as a given condition in a given time, but as a process of value creation over time (Bertini 1995).

Generally, if there is a difference in speed between internal and external, the faster the external changes than the internal ones, the greater the probability of failure of the company.

Based on this the internal structure of the company allows for the achievement and consolidation of a competitive advantage in the business area and capital market. Specifically, these competitive advantages require continuous adaptations of the internal structure of the company due to the constant changes in the business and capital markets according to the logics of the strategic change.

The success of the company should not be interpreted as a temporary situation but as a way of being. In this sense, rather than representing a target to be achieved at given periods of time, the *Systemic-Structural-Dynamic Consonance* of the CSF should be considered as a strategic approach based on change.

The internal structure of the company plays a key role in the CSF. It is the ring that connects the business and financial strategies allowing for the creation of a virtuous circle between them. Specifically, the distinctive resources allow for the ideation, the design and the development of a product system in line with the needs and expectations of customers in a better way than competitors. The ability of the product system to satisfy the critical factors of strategic business area, allows the company to generate, to develop and to defend a competitive advantage over competitors in the SBA over time by generating company value.

The company's ability to generate value in the SBA together with an internal structure characterized by efficiency and effectiveness of the organizational and operating model and by good governance, allows the company to present itself to capital market with a risk-return profile in line with investor expectations. The positive assessment of investors allows the company to generate, to develop and to defend a competitive advantage over competitors in the capital market increasing the company's ability to attract the capital required for its development at profitable conditions.

The company's ability to raise capital increases its investments in SBA. It reinforces its competitive advantage over competitors and increases its value in the business. In turn, the increase in cash-flows increases the company's ability to raise capital in capital markets reinforcing its competitive position. Therefore, it creates a virtuous circle between business areas and capital markets where one increases the other based on an increase of the expected cash-flows.

Finally, the company's ability to raise capital at profitable conditions in the capital markets together with the ability to invest in the business in a profitable way, allows the company to attract and invest in strategic human capital and technologies. It reinforces the internal structure of the company that, in turn, increases the company's ability to generate product systems in line with the changing needs of customers by generating value in the business that, in turn, increases the company's ability to raise capital at profitable conditions in the capital markets to invest in the business.

Therefore, the *Systemic-Structural-Dynamic Consonance* of the CSF generates a virtuous circle between the internal structure, business and capital markets where the one feed the others. It allows the company to generate, to develop and to defend a competitive advantage in the business by generating value for investors.

It is worth noting that a key role in the company's internal structure is played by the quality of its resources and specifically human resources. They have to change continuously to keep up with environmental changes. It could be a problem on the short-term, because the internal structure elements tend to be stable. Consequently, a company culture based on changing rather than structure change in a given period of time is more effective.

1.2 Analytical Schemes for the Analysis of Company Performance

The company's ability to create profit over time requires an analysis based on two main parts:

- the qualitative analysis of the business model;
- the quantitative analysis of the effects of the business model choices on the economic and financial dynamics over time.

These two parts are strictly related. It is not possible to investigate into the company by only considering the analysis of its business model without considering the effects of the strategic choices on the economic and financial dynamics and vice versa (Graham and Dodd 1940).

In the analysis of these two parts jointly, three are the main caveats to keep in mind:

- *first, there must always be full consistency between the business model of the firm and its economic and financial dynamics over time*. The economic and financial analysis measures the quantitative effects of the business model on economic and financial dynamics with regards to the three dimensions of Operating and Net Income, Capital Invested and Capital Structure, and Cash-flow from Operations and Cash-flow to Equity. Therefore, while the analysis of the business model is a qualitative analysis, the analysis of economic and financial dynamics is a quantitative analysis. The two types of analysis cannot be separated and they are normally used together in the definition and assessment of the company's business planning. Consequently, an estimate of the expected economic and financial performance must be a coherent and consistent translation of the business model adopted by the company.
  In the business planning process, the definition of the business model is the first step, while an estimation of the expected quantitative effects on economic and financial dynamics is the second. Between them there is a two way relationship. If the business model is coherent in all parts, but the expected economic and financial dynamics are not satisfactory, it is necessary to redefine the business model. Only if the business model of the company is coherent in all parts and the expected economic and financial performances are satisfying and the baseline assumptions are individually reliable and coherent with each other, the planning process and the business plan generated can be defined as reliable;
- *second, the future is the reference time*. The value of the company is function of its ability to generate value in the future. In an analysis of the past, with regards both to the business model adopted and the financial dynamics that it has fulfilled, it is important to understand if the future expectations of the company, as defined in the business model to be implemented and in the estimation of the expected cash-flows, are really reasonable or unreasonable;

- *third, the assumptions are the key variable of the forecast*. The business model implemented and the estimation of the expected economic and financial dynamics are based on assumptions. Then, the quality of the forecast is function of the quality of the baseline assumptions.
  Generally, an analysis of the assumptions requires strict coherence or a reasonable relationship based on personal elements of the company or straightforward to acquire. In this sense, the reference assumptions should be clearly defined for each variable, also in their relationship with other assumptions. Each assumption must be individually reliable and coherent with each other.

The financial analysis of company performance proposed in this context follows a marked financial approach.

In order to simplify the comparison between the past and the future for the same company and between different companies over time, the same analytical schemes should be used. Several analytical schemes should be used. They are defined on the basis of the specific purpose of management according to the decision-making process.

In this contest, the analytical schemes used are defined based on the financial approach to company assessment and they are defined in order to investigate the three main pillars:

- *Operating Income and Net Income;*
- *Capital Invested and Capital Structure;*
- *Free Cash-flow from Operations and Free Cash-flow to Equity;*

While the first defines the economic dynamic, the second and the third define the financial dynamic of the company.

Using the analytical schemes proposed in this context, the following should be borne in mind:

- they are defined with a view to the financial community rather than the accounting one. Therefore, they must not be confused with the schemes used for balance sheet analysis and for definition of the classic accounting ratios. Moreover, terminology is not strictly based on the accounting rules;
- they are strictly connected between them. Therefore, the definition of each one is strictly related to the composition of each other;
- they are defined based on non-financial companies. Furthermore, they can also be used for financial companies after some changes in their structures;
- they are used to analyse the expected future economic and financial dynamics for an estimate of company value. Therefore, their application to past data is necessary to link the past and future in a coherent manner.

Furthermore, for greater understanding of the economic and financial dynamics of the company over time, past values should be aligned with expected future values. The alignment procedure between past and future values regarding Operating and Net Income, Capital Invested and Capital Structure, and Free

Cash-flows from Operations and Free Cash-flows to Equity, as represented in the analytical schemes used, can be achieved by a procedure based on three main steps:

- *the first step, is the collection and recognition of past values*: the aim is to build Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity of the company in the past. For this objective, the analyses should be based on the balance sheet, income statement and cash flows statement on the one side, and on the internal management accounts of the company on the other side. The combination of these two data sources allows for an analysis of the real condition of the company. An analysis of the management accounts is necessary for three main reasons: (i) they are built to support management in the decision-making phase; (ii) they are characterized by both monetary and non-monetary quantitative data; (iii) they are well known in their composition and dynamics thanks to the technique of the variance analysis implemented constantly;
- *the second step, is the "adjustment" of past values:* the aim of this step is to obtain the "*normalized*" value of Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity of the company in the past. The aim of the process is to define these values in stand-alone conditions of the company. Therefore, their effects on extraordinary events in the broadest sense are not considered;
- *the third step, is the estimate process of value in the future*: the aim of this step is to build estimates on Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity of the company in the future. A company business plan should be defined in order to achieve this objective. It is created by defining the Company Strategic Formula and by estimating its effects on future economic and financial dynamics.

Based on these three steps, the origin of the company, where it is and where it plans to go, should be clear. Thanks to normalization of the past economic and financial dynamics their values can dialogue with those expected for the future. Consequently, it is easier to highlight the jumps between the past and the future and to evaluate whether or not they can be fulfilled in the future based on the strategies that will be implemented.

The analytical schemes allow for analysis of composition of the Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity. An analysis of the economic and financial dynamics requires the definition and the knowledge of the connections between them.

1.3 Operating and Net Income

The analysis of the economic dynamic requires an analysis of the Operating Income and Net Income over time. It is the starting point for an analysis of company performance in the business over time.

The aim of the analysis is to investigate into the drivers of the economic engine of the company and the existence of a long-lasting competitive advantage. Specifically, an analysis of costs and revenues over time must be focused on in order to understand if the company has a real competitive advantage and if it lasts over time. The competitive advantage must be identified on the basis of its effects on revenues and costs. Only if the company confirms a dynamic of revenues and costs that is better than that of its competitors for a long period, the company has a lasting competitive advantage over time. Consequently, an analysis of the revenues and costs must consider a long period of time ranging from 7 to 10 years.

The *analytical scheme* adopted to analyse the income of the company can be divided into four main sections:

– *Section* (1) *Operating area* (or *Operations section*): it refers to the operating revenue and costs due to activities of the company's core business. This is the most important section of income because it defines the economic results of the company's core business. In this section, it is important to distinguish between cash operating costs and non-cash operating costs (such as amortizations, depreciations, accruals);
– *Section* (2) *Non-Operating area* (or *Non-Operations section*): in this section two different types of revenues and costs are considered. The first are the revenues and costs due to the execution of activities different from the company's core business; the second are the revenues and costs related to the core-business but not repeated in time and thus considered as a one-off. It is worth noting that in this context definition of the perimeter of the company core-business is of a strategic assessment only. Therefore, in both cases, the difference between operating and non-operating revenues and costs is based on the strategic analysis only and it is independent of the accounting rules;
– *Section* (3) *Financial area* (or *Financial section*): it refers to the financial revenues and costs. The first refers mainly to the dividend received and interest on financial credit while the second refers mainly to the interest on debt and other financial costs linked to it. Generally, for a non-financial company, the financial revenues are low and the entire financial area refers to the cost of debt;
– *Section* (4) *Tax area* (or *Tax section*): it refers to corporate taxes due to the company activities. Usually, taxes are considered entirely in this part. But, more appropriately, they should be divided among the three sections (Operating, Non-Operating, Financial) as taxes arising from them. Often, in practice tax splitting is difficult to perform and therefore corporate taxes are considered entirely in this specific section.

The aim of the analytical scheme used, as reported in Table 1.1, is to distinguish between strictly operating activities and others it can only be done with a deep analysis of each item in a financial analysis perspective. In this sense, the EBITDA and EBIT refer to Revenues and Costs strictly operating.

The *Gross Profit* (or *Gross Margin*) is the difference between the revenues from the sale of products, goods and services and the related production costs.

Table 1.1 Analytical scheme of the Operating and Net Income

Net sales revenues
Net operating revenues
(Costs of raw materials, parts of products and products)
(Costs of production of goods and services and distribution services)
(Costs of direct labour on production)
Direct operating costs of goods and services sold
Gross profit
(Costs of research and development)
(Costs of marketing and sales)
(Costs of administration)
(Costs of advisory)
(Costs of employees)
(Costs of leasing and rent)
(Other general operating costs)
Indirect operating costs (cash)
EBITDA
(Amortization of intangible operating assets)
(Depreciation of tangible operating assets)
(Accruals for employees)
(Accruals for risk, charges and taxes)
(Changes in value of operating assets)
Operating costs (non-cash)
EBIT
Operating financial revenues
(Interest on debts)
(Other financial costs)
Financial profit/(loss)
EBT—operating
Non-operating and non-current operating revenues
Non-operating financial revenues
(Non-operating and non-current operating costs)
(Amortization and depreciation of intangible and tangible surplus assets)
(Changes in value of surplus assets)
Non-operating profit/(loss)
EBT
(Current operating taxes)
(Current corporate taxes)
Net income

Net sales revenue (or more simply net sales) are the starting point. Revenues are the result of the sale of goods and services to customers. The word “net” refers to the difference between gross sales (the total invoice of goods and services) and the sales returns and allowances that refer to the sales value of goods that were returned by customers and reimbursements to customers due to faulty goods or for some other reason related to the product and service. This amount can be subtracted from sales directly.

Sales discounts (referring to the amounts of discounts used by customers for payment) and trade discounts (referring to the amounts of discounts from the actual selling price as published in the official price lists) can be included in net revenues or in commercial costs. If they are included in net revenues, they are not traceable. Otherwise, if they are considered in the commercial costs they are traceable and it is possible to measure the effects of the commercial policy on the revenues. In this contest, for a more in-depth analysis, they are separately considered in the commercial costs.

It is worth noting that at times entrepreneurs look to revenues only to measure their growth. But an increase in revenues does not necessary imply an increase in profit. If the increase in revenues is accompanied by such a rise in costs, profit does not increase. Similarly, by structurally reducing the costs, profit increases for the same revenues. Therefore, a good company growth requires an increase in profit that can be achieved by increasing the revenues to a greater level than costs, or by structurally reducing the costs for equal or decreasing revenues.

The *Direct Operating Cost of Goods or Services Sold* (or more simply *Cost of Sales* or *Cost of Goods Sold*) refer to the costs due to direct fulfilment of the goods and services sold. Therefore, they can be defined as production costs.

These production costs must be defined on the basis of the type of company with regards to its activities. Three types of companies can be presumed (Anthony, Hawkins, Merchant 2011): (i) merchandising company, (ii) manufacturing company, and (iii) service organization. Before starting the analysis, it can be useful to solve a terminological problem. It is worth noting, that because both merchandising and manufacturing companies sell tangible goods, the term “cost of goods sold” is usually used rather than “cost of sales”. However, the two terms can be used interchangeably.

Merchandising companies sell tangible goods. Specifically, they sell goods in substantially the same physical form in which they acquires them. Therefore, its cost of sales is the acquisition costs of goods that are sold. For these companies, the cost of goods sold must consider the merchandise inventory and thus the costs of goods that have been acquired but not yet sold at a defined date. In this case the change in inventory in the cost of goods sold must be considered.

Specifically, the cost of goods sold can be divided into two main parts: (1) purchase cost of the goods (invoice costs) less the goods returned, allowance and discounts, plus any other related costs made to make the goods ready for sale (such as shipping costs, freight-in, unpacking costs, etc.); (2) the inventories with regards to the change between the beginning and the end of inventory. By following the regular inventory method, the cost of goods sold is equal to the purchase of goods plus the beginning of inventory less the end of inventory as follows (Table 1.2).

Table 1.2 Cost of goods sold in merchandising companies

Purchase cost of goods
Other related costs necessary for the sale
(Goods returns, allowance and discounts)
Net purchase of goods
Beginning inventory
Goods available for sale
(Ending inventory)
Cost of goods sold

Manufacturing companies sell tangible goods. Specifically, they sell goods after an industrial process by converting raw materials and purchased parts into finished goods. Therefore, its sales costs include conversion costs as well as raw material and parts of goods that it sells. Also in this case it is necessary to consider the inventory in order to define the cost of goods sold. The inventory account can be divided into three main parts: (1) materials inventory, referring to raw materials that are to become a part of the ultimately sellable goods resulting from the manufacturing process. In this case the cost of goods sold is defined as in the case of a merchandise company; (2) work in progress inventory, referring to goods that have started through the manufacturing process but have not yet been finished. The cost is defined as the materials plus the conversion costs incurred on these items up to the end of the accounting period; (3) finished goods inventory, referring to goods that have been manufactured but have not yet been shipped to customers. In this case, the cost of goods sold is based on the total costs incurred in their production. Therefore, the cost of goods sold is defined as in the merchandising company. The only difference is that, in this case the items are recorded at their production cost rather than at their acquisition cost.

Table 1.3 Cost of goods sold in manufacturing company

Purchase cost of materials
Other related costs of materials
(Materials returns, allowance and discounts)
Net purchase of materials
Beginning inventory of materials
Amount of materials available for use
(Ending inventory of materials)
Costs of materials used
Beginning inventory of work in progress
Amount available goods available for use
(Ending inventory amount of work in progress)
Cost of goods manufactured
Beginning finished goods inventory
Total amount of goods available for sale
(Ending finished goods inventory)
Cost of goods sold

By using the periodic inventory method and by considering all three types of inventory, the cost of goods sold can be defined as follows (Table 1.3).

Service organizations sell intangible services. In this case the cost of sales includes all services needed for the execution of services sold to the clients. There is also an inventory in this case. There are two main types: materials inventories with regards to all of the materials required for production of the service; intangible inventories with regards to the costs of services that have been incurred on behalf of clients but that have not yet been invoiced to clients (as in the case of professional service companies such as legal, consulting, etc.) and thus called jobs in progress or unbilled costs. The service organization, therefore, does not have finished goods inventories.

It is necessary to consider that a company may have other inventory accounts for suppliers apart from the inventory of goods directly involved in the merchandising or manufacturing process. They refer to tangible items that must be consumed during normal operations such as repair parts for equipment, lubricants, etc. Suppliers are generally distinguished from a merchandise company because they are not sold, and they are distinguished from materials because they are not accounted for as an element of the cost of goods manufactured. Also in this case, the logical scheme is equal to merchandising company.

The Costs of production of goods and services and distribution services refers to the all of the costs due to the service production and distribution. It includes both the service for the production of goods and all elements for execution of the service to be sold.

The Cost of direct labour only refers to the costs of workers for the production of goods to be sold.

Therefore, Gross Profit measures the part of revenues that remain after the coverage of production costs that must be used to cover all of the rest. Therefore, it is the heart of the company by measuring the company's ability to perform in the business on the basis of the revenues and production cost of the goods to be sold.

It is important to know that by considering the behaviour of costs on the basis of the output of products, all operating costs can be also defined variable costs. Indeed, all costs of goods sold vary directly and proportionately with volume. Therefore, the Gross Profit can be the Contribution Margin of the company at the same time.

Usually, to acquire more information about Gross Profit, it is defined as a percentage of the Revenues as follows:

$$Gross\,Profit\,Margin = \frac{Gross\,Profit}{Net\,Operating\,Revenues} \tag{1.1}$$

The Gross Profit Margin is a good preliminary indicator of the existence of a competitive advantage of the company. Normally, companies with a competitive advantage are characterized by a Gross Profit Margin higher than competitors over time. Without a competitive advantage, the competition reduces the Gross Profit Margin of the company.

A general rule (practical rather than scientific) assumes that the company is characterised by Gross Profit Margin (Buffet and Clark 2008):

– equal or higher than 40% continuously over time: it has some kind of competitive advantage;
– lower than 40%: the business of the company is characterized by a high level of competition, reducing the company's profitability;
– equal or lower than 20%: the business of the company is characterized by a very high level of competition where no company is characterized by a competitive advantage on competitors sustainable over time. If the company in this type of business does not have a competitive advantage, it normally has a low level of profitability.

In order to understand further the Gross Profit Margin, a historical track record can be used considering a range between 7 and 10 years. The competitive advantage must exist and it must be sustainable over time. In this sense, the key word about the competitive advantage is "continuity" over time.

The EBITDA (*Earnings Before Interest Tax Depreciation and Amortization*) is equal to the difference between operating revenues and direct and indirect operating costs (cash) due to the execution of the core business activities.

The operating costs positioned between the Gross Profit and the EBITDA can be defined as the hard costs of the company and they refer to the operating costs required for company operations; without them there is no company and there is no product. Specifically, they are the first components of the hard costs while the second refers to the non-cash operating costs.

These costs can be aggregated in several ways. In this context, they are grouped according to their relevance. Their relevance can be defined on the basis of two main parameters:

– first, with regards to an increase in the company's ability to compete in the business: in this sense, the costs of goods with regards to the production function are defined (Vernimmen et al. 2014) as well as the costs of research and development with regards to the function of research and development, marketing and sales costs with regards to the commercial function and finally, administration costs with regards to all of the other functions of the company including advisors;
– second, with regards to their relevance in terms of amount: in this sense, employee costs and leasing and rental costs are defined as well as other general operating costs that are not relevant in terms both of function and amount.

The amount of operating costs can be very different among the companies due to the specific characteristics of the business.

The relevance of these costs can be measured on the basis of the EBITDA percentage of net operating revenues or on Gross Profit.

Generally, if the market is characterized by a high level of competition and if the company does not have a defined competitive advantage, these costs tend to be high. Generally, these costs absorb between 30 and 70% of the Gross Profit. If it is equal or lower than 30% over time, the company can be considered as high-performing. Otherwise, the closer it get to 100% the worse the condition of the company.

The EBIT (*Earnings Before Interest and Taxes*) is the net operating income because it incorporates non-cash operating costs such as amortization, depreciation and accruals. These non-cash operating costs define the second part of the hard costs and define the distance between EBITDA and EBIT.

There are two main elements with regards to the difference between EBITDA and EBIT that should be kept in mind:

- amortization and depreciation: the distance between EBITDA and EBIT can be a measurement of the investment policy in intangible and tangible assets of the company. The higher the investments in assets, the greater the amortizations and depreciations, and then the greater the distance between EBITDA and EBIT. In a dynamic perspective, the reduction of this distance over time due to lower amortization and depreciation, indicates the reduction of the company's investments in the business. Over time, a reduction of investments in assets generates a reduction of the company's ability to compete in its business and therefore a reduction in both operating and net expected income and consequently a reduction of the company's value creation.
  If the company invests in operating and non-operating assets, and the investments in non-operating assets are so important as to disturb reading of the EBIT, the amortization and depreciation relate these non-operating assets can be located out of the EBT-Operating. In this case, it is necessary to distinguish between amortization and depreciation operating and non-operating.
- accruals for risk and charges: the distance measures the risks undertaken by the company in time. In a dynamic perspective, relevance is mainly due to the probability that they are not enough to cover future costs if they will be executed. Generally, the greater the provision for risk and charges, the higher the probability that the events referenced may be fulfilled and therefore, the higher the risk that they will not be enough. Therefore, the real problem is to understand if the provisions are enough to cover the costs derived from execution of the future events considered to be feasible.

It is worth noting that the difference between EBITDA and EBIT is relevant in the economic perspective only. Indeed, in the financial perspective they are equal because non-cash operating costs do not affect cash flows.

Finally, it is worth noting that by using tax splitting and therefore by distinguishing between operating and corporate taxes (the first are due to the operating activities, while the second are due to the non-operating and financial activities), EBIT is equal to NOPAT (*Net Operating Profit After Taxes*). Otherwise, if tax splitting is not used and operating taxes as well as corporate taxes are considered jointly in the tax area, EBIT is the operating income before taxes while the NOPAT is the operating income after taxes.

The EBT (*Earnings Before Taxes*) defines the operating income before corporate taxes. If there are non-operating revenues and costs, it could be interesting to highlight the components of EBT that refer to the operating revenues and costs only by defining the EBT-Operating.

If there are no non-operating revenues and costs, the difference between EBIT and EBT is due to the financial revenues and costs. Generally, the difference is mainly due to debt costs. Then the distance between EBIT and EBT indirectly measures the relevance and the risk level of debt in the capital structure: the greater the distance, the higher the costs on debt and therefore the higher the amount and the risk level of debt in the capital structure.

The Net Income is due to the difference between EBT and the corporate taxes. The Net Income is frequently expressed per share of equity, that is Earnings per Share.

It is worth noting that usually the final performance of the company is measured on the basis of the EBT rather than the Net Income. It allows for a comparison of different companies in different countries and therefore subject to different taxes.

It is worth noting that the "*congruity*" of the company's Net Income for the investor in equity is measured in the financial markets according to the return request by investor for the same risk-class. Part 2 of the book focuses on this analysis.

1.4 Capital Invested and Capital Structure

The aim of the analysis of the Capital Invested and Capital Structure is to investigate into the sources of capital and their use.

The Capital Invested (CI) defines the amount of capital invested in company activities, while the Capital Structure (CS) defines the sources of capital used to finance these activities.

The aim of the analytical scheme proposed is to highlight the main figures whose variations can be interpreted in terms of cash-flows immediately and therefore whose previsions are relevant for company value. The analytical scheme proposed is based on separation between financial and non-financial assets-liabilities and, with regards to this second category between operating and surplus assets-liabilities with regards to those that are not linked with operating activities. There are two main implications of these separations:

- first, they allow for definition of the investments directly in company's assets, both operating and surplus as well as relative capital sources;
- second, they allow for an assessment of variations in the assets-liabilities directly in terms of changes in cash flows.

Table 1.4 illustrates the analytical scheme used for analysis of the Capital Invested and Capital Structure.

Table 1.4 Analytical scheme of the capital invested and capital structure

Net intangible operating assets
Net tangible operating assets
Financial operating assets
Inventory stable over time
Net operating capital expenditures (CAPEX)
Trade receivables net
(Trade payables)
Trade working capital (TWC)
Inventory
Others operating receivables net
(Others operating payables)
Net working capital (NWC)
Net operating capital invested (NOCI)
Net intangible surplus assets
Net tangible surplus assets
Financial surplus assets
Non-operating and non-current operating receivables
(Non-operating payables and non-current operating payables)
Surplus assets (SA)
(Provision for employee)
(Provisions for risk, charges and taxes)
(Provisions)
Capital invested (CI)
Share capital
Realized retained earnings
(Treasury shares)
Net profit (loss)
Equity (E)
Long-term financial debts
(Long-term financial credits)
Long-term net financial position (L-NFP)
Short-term financial debts
(Short-term financial credits)
(Marketable securities)
(Cash and cash-equivalent)
Short-term net financial position (S-NFP)
Net financial position (NFP)
Capital structure (CS)

The Capital Invested (CI) consists of investments in Capex, Net Working Capital, Surplus Assets, less Provisions. The Capital Structure (CS) is defined by Equity and Net Financial Debt.

Specifically, the sum of Capex and Net Working Capital (of which the Trade Working Capital is the difference between trade receivables and trade payables only) defines the Net Operating Capital Invested (NOCI). It is the capital invested in the operating assets fixed and working. The total of NOCI and Surplus Assets (that is the amount of capital invested in non-operating assets fixed and working) defines capital invested in all assets of the company both operating and non-operating. The difference between the capital invested in all assets and Provisions defines the Capital Invested (CI). It is important to note that the CI is net of Provisions. In this sense it can be also defined as Net Capital Invested.

Also if the Provisions can be considered as debt, in this context they are considered in the Capital Invested with a negative sign in order to define the Capital Structure based only on the two main capital sources: Equity and Financial Debt.

This analytical scheme allows for the definition of capital invested to be financed costly in terms of debts and equity.

The CAPEX (*Capital Expenditures*) refers only to the operating investments. Therefore, it can be defined as Net Operating Capital Expenditures and it represents the investments needed for the execution of Operating Income.

They refer mainly to the operating investments in fixed assets (tangible and intangible) of the company necessary to execute the products sold in the business net of their amortization and depreciation funds.

Investments in Financial Assets are included in the Capex only if they have a strategic industrial relevance. Otherwise, the investments for fulfilment of financial income are included in the Surplus Assets that includes non-operating investments.

Sometimes, the Capex also includes the inventory. If the company needs a constant and stable stock of inventories in order to guarantee production activity (stocks of raw materials, semi-finished and finished products) and to satisfy customer requirements in time (stocks of goods to be sold), it represents a stable investment. Therefore, part of the inventories that in monetary value must be stable in the company for the needs of the business can be approximated to an investment in tangible assets and therefore it is included in the Capex.

The Net Working Capital (*NWC*) refers to the investments in working capital of the company arising from repetitive operations (cycle of buying, processing, sales). It is equal to the difference between current assets and current liabilities arising from the company's operating activities. Therefore, it does not include financial and surplus assets and liabilities (with regards to non-operating company activities). Specifically, the NWC consists of:

- Trade Working Capital (TWC): it is the difference between trade receivables and trade payables arising from the trade activities of the business with customers and suppliers. The trade receivables are net of the allowance for doubtful accounts and bad debt;
- Inventories: it is the value of the inventory in a time *t*. They can be assessed based on FIFO (first in, first out), LIFO (last in, first out) and the Weighted Average.

– Other operating receivables less other operating payables: they refer to the receivables and payables arising from operating activities of the company different from the strictly trade (grouped in Trade Working Capital) and the financial receivables and payables (grouped in Net Financial Position).

The NWC entity is function of the operating income with regards to the operating revenues and costs on the one side, and the time of cash-in and cash-out of the operating revenues and costs on the other side. Therefore, the higher the NWC, the higher the receivables and inventories than payables, and therefore the lower the cash-in. On the contrary, the lower the NWC, the lower the receivables and inventories that payables, and therefore the higher the cash-in.

The NWC measures the resources used in operating current activities. Therefore, presuming equal revenues, the higher the NWC, the greater are the financial needs. In this case, the receivables increase and the lack of cash-in with their displacement in the future time, increases the financial needs to cover the cash-out.

It is worth noting that in ordinary conditions NWC is always positive. If it is negative, debts are higher than credits, and therefore the company funds its activities by using debts. It can be considered a degenerated condition of the firm.

The Surplus Assets (*SA*) refers to the investments in a non-operating area. It consists of:

– tangible and intangible assets for the creation of non-operating activities leading to the non-operating income;
– financial assets, such as shares and financial credits, leading to the financial income;
– receivables and payables due to the time of cash-in and cash-out related to the non-operating revenues and costs.

The Provisions refer to the funds accrued for risks and charges, for employees and for taxes. They refer to the amount of costs accrued but not paid, that will be paid in the future for execution of the reference event.

In this context, they are registered in the Capital Invested (CI) with a negative sign. They can be interpreted as obligations deriving directly from the operating area. It includes:

– provision for employees: it represents costs achieved but not liquidated yet due to legal constraints. Therefore, they are part of the operating costs with deferred payments;
– provision for taxes: it refers to taxes in company activities matured but not yet liquidated. If there is contemporaneity between tax maturity and tax payment, there are no provisions;
– provision for risks and charges: it refers to the costs accrued for company activities executed but whose negative effects will be in the future. Therefore, they represent costs deriving from company activities.

In order to provide a clear picture without damaging the NWC, the provisions can be presented separately and away from the Surplus Assets.

An introduction of the provisions in the Capital Invested (with a negative sign), allows for definition of the Capital Structure based only on the two main capital sources: Equity and Net Financial Debt.

Nevertheless, they can be considered as debt for the company and thus as financial sources. In this case, they are considered in the Capital Structure (CS). Specifically, it is easy to presume the provisions for the employees as company financing from employees. It is a fund based on the payment matured for the employees but that will be paid in the future. Similarly, the provision for taxes, refers to the payments matured that will be paid in the near future. Therefore, they represent a debt for the company. Finally, provisions for risk and charges, refer to the payments that probably will be made in the future based on current events. Therefore, they can be considered as a debt in order to the future payments matured now.

The Equity refers to the personal sources of the company. There are two main types:

- capital invested by the stockholders;
- self-financing of the company due to the cash-flows generated and not distributed in dividends form, by increasing reserves.

The Net Financial Position (*NFP*) refers to the net financial debt of the company. It is equal to the difference between financial debts only (both long and short term) and liquidity, marketable assets and financial credits. The NFP can be divided in two parts:

- Long-term NFP: it is equal to the difference between long-term financial debt and long-term financial credit;
- Short-term NFP: it is equal to the difference between short-term financial debt and the sum of short-term financial credits, marketable assets and liquidity.

The non-financial debts are included in the NWC. Therefore, the NFP's construction requires the NWC's construction jointly. It is not possible to define the NFP without defining jointly the NWC.

The general equation on capital requires that the Capital Invested (CI) must be equal to the Capital Structure (CS), as follows:

$$CI = CS \tag{1.2}$$

and then:

$$CAPEX \pm NWC \pm Surplus\,Assets - Provisions = Equity + NFP \tag{1.3}$$

It is important to note that Net Financial Position (NFP) plays a central role in the equation. There are two main caveats to be kept in mind.

First, the NFP defines the net financial debt (financial debt less the sum of financial credit, marketable securities and liquidity). Therefore, the company's Leverage (L), can be measured on the basis of the Net Financial Position (NFP) (and thus by considering its net financial debt) or the financial debt (FD) as follows:

$$L(\%) = \frac{NFP}{E + NFP} = \frac{NFP}{CS} \quad \text{or} \quad L(\%) = \frac{FD}{E + FD} \tag{1.4}$$

Therefore, only in the first case the denominator is the Capital Structure (CS). In the second case it is not the CS because it creates a misalignment between Net Working Capital and Net Financial Position and, consequently between Capital Invested and Capital Structure.

Second, the Net Financial Position (NFP) can be defined as a "mobile" item. It can be considered as a source of capital and then classified in the Capital Structure or as an investment and then classified in the Capital Invested.

Specifically, the Net Financial Position (NFP) can be:

– *negative*: the financial debts are greater than the sum between financial credits, marketable securities and liquidity. In this case, the NFP is a source of capital and then it is classified in the Capital Structure and:

$$CAPEX \pm NWC \pm Surplus\,Assets - Provisions = Equity + NFP \tag{1.5}$$

– *positive*: the financial debts are lower than the sum between financial credits, marketable securities and liquidity. In this case, the NFP is an investment and then it is classified in the Capital Invested and:

$$CAPEX \pm NWC \pm Surplus\,Assets - Provisions + NFP = Equity \tag{1.6}$$

1.5 Free Cash Flow from Operations and Free Cash Flow to Equity

The aim of the analysis of the Free Cash-flow from Operations (FCFO) and Free Cash-flow to Equity (FCFE) is to investigate into how the operating cash flows and dividends over time are defined.

It is based on an analysis of cash-in and cash-out arising from the company's activities leading to the Free Cash-flow from Operations and Free Cash-flow to Equity.

The definition of cash-flows is based on the items related to the Operating and Net Income on the one side, and the items related to the Capital Invested and Capital Structure on the other side, must be considered in different way.

The items of Operating and Net Income must be considered with regards to their value in the same year of the cash-flows determination and with the same sign. Therefore, for the definition of cash-flow in time t_1 revenues and costs must be considered at the same time (t_1) as follows:

- revenues: all different types of revenues (operating, non-operating and financial) generate cash-in;
- costs: all types of costs (operating, non-operating, financial and taxes) generate cash-out. Only the non-cash operating costs (amortization, depreciation and accruals) must not considered because they do not generate cash-flow movements.

The items of Capital Invested and Capital Structure must be considered with regards to their changes between two different years. Therefore, for the definition of the cash-flow in time t_1 the change of the item between t_0 and t_1 must be considered.

For cash-flow determination, each item of Capital Invested and Capital Structure must be considered according to the movements in terms of cash-flow regardless of their nature, as follows:

- *credit*: it is the same for trade receivable net, other operating receivable net, other non-operating receivables, financial credits (short and long term). The cash-flow movement is the following:

$$\downarrow(\uparrow)\Delta \mathit{Credits} \Rightarrow \downarrow(\uparrow)\mathit{Capital\,Invested} \Rightarrow \uparrow(\downarrow)\mathit{Cash\,In} \Rightarrow \uparrow(\downarrow)\mathit{Cash\,Flows}$$

$$\downarrow(\uparrow)\Delta \mathit{Financial\,Credits} \Rightarrow \uparrow(\downarrow)\mathit{Net\,Financial\,Position} \Rightarrow \uparrow(\downarrow)\mathit{Cash\,In} \Rightarrow \uparrow(\downarrow)\mathit{Cash\,Flows}$$

- *debt*: it is the same for the trade payable, other operating payable, other non-operating payable, financial debts (short and long term). The cash-flow movement is the following:

$$\downarrow(\uparrow)\Delta \mathit{Debts} \Rightarrow \uparrow(\downarrow)\mathit{Capital\,Invested} \Rightarrow \downarrow(\uparrow)\mathit{Cash\,In} \Rightarrow \downarrow(\uparrow)\mathit{Cash\,Flows}$$

$$\downarrow(\uparrow)\Delta \mathit{Financial\,Debts} \Rightarrow \downarrow(\uparrow)\mathit{Net\,Financial\,Position} \Rightarrow \downarrow(\uparrow)\mathit{Cash\,In} \Rightarrow \downarrow(\uparrow)\mathit{Cash\,Flows}$$

- *inventory*: it is the same for the inventory and inventory stable over time. The cash-flow movement is the following:

$$\downarrow(\uparrow)\Delta Inventories \Rightarrow \downarrow(\uparrow) Capital\,Invested \Rightarrow \uparrow(\downarrow) Cash\,In \Rightarrow \uparrow(\downarrow) Cash\,Flows$$

- *Net Assets*: it is the same for the tangible and intangible assets. The cash-flow movement is the following:

$$\Delta[NetAsset_{t+1} - (NetAsset_t \pm Change\ Value_{t+1} - Amortization_{t+1})] \Rightarrow \begin{cases} = 0 \Rightarrow \downarrow Capital\,Invested \Rightarrow Cash\,Flows = 0 \\ > 0 \Rightarrow \uparrow Capital\,Invested \Rightarrow Cash\,Out \Rightarrow \downarrow Cash\,Flows \\ < 0 \Rightarrow \downarrow Capital\,Invested \Rightarrow Cash\,In \Rightarrow \uparrow Cash\,Flows \end{cases}$$

- *Financial Assets*: it is the same for the financial assets operating and non-operating. The cash-flow movement is the following:

$$\uparrow(\downarrow)\Delta Financial\,Asset \Rightarrow \uparrow(\downarrow) Capital\,Invested \Rightarrow \downarrow(\uparrow) Cash\,In \Rightarrow \downarrow(\uparrow) Cash\,Flows$$

It is worth noting that if the increase of financial asset is due to a reassessment of the asset, there is no cash-out but the increase of a reserve in equity. In this case, this reserve is not considered in the determination of dividends. Similarly, if the decrease of financial assets is due to the reduction in value, there is no cash-out but operating no-cash costs. In these cases, the movements of the financial assets can be summarized as follows:

$$\Delta[Financial\,Asset_{t+1} - (Financial\,Asset_t \pm Change\ Value_{t+1})] \Rightarrow \begin{cases} = 0 \Rightarrow \uparrow Capital\,Invested \Rightarrow Cash\,Flows = 0 \\ > 0 \Rightarrow \uparrow Capital\,Invested \Rightarrow Cash\,Out \Rightarrow \downarrow Cash\,Flows \\ < 0 \Rightarrow \downarrow Capital\,Invested \Rightarrow Cash\,In \Rightarrow \uparrow Cash\,Flows \end{cases}$$

- *Provisions*: it is the same for provisions for risks and charges, taxes and for employees. The cash-flow movement is the following:

$$\Delta[Provision_{t+1} - (Provision_t + Accruals_{t+1})] \Rightarrow \begin{cases} = 0 \Rightarrow \downarrow Capital\,Invested \Rightarrow Cash\,Flows = 0 \\ < 0 \Rightarrow \uparrow Capital\,Invested \Rightarrow Cash\,Out \Rightarrow \downarrow Cash\,Flows \end{cases}$$

- *Equity*: only the changes achieved in monetary terms must be considered. The cash-flow movement is the following:

$$\uparrow(\downarrow)\Delta Equity\,in\,money \Rightarrow \uparrow(\downarrow)Capital\,Structure \Rightarrow \uparrow(\downarrow)Cash\,In \Rightarrow \uparrow(\downarrow)Cash\,Flows$$

- *Liquidity*: it is the same for the marketable securities, cash and cash equivalents. The cash-flow movement is the following:

$$\uparrow(\downarrow)\Delta Liquidity \Rightarrow \downarrow(\uparrow)Net\,Financial\,Position \Rightarrow \downarrow(\uparrow)Cash\,In \Rightarrow \downarrow(\uparrow)Cash\,Flows$$

  It is important to know that the movements of liquidity, and therefore its effects on Net Financial Position and Free Cash-flow to Equity, refer to the amount of capital that the company wants to invest in Liquidity as measured at the end of the period analysed. Therefore, all movements on liquidity used to balance all other movements on items that generate cash-in and cash-out, they are transitory only.

The cash-flows movements due to Operating and Net Income and Capital Invested and Capital Structure, can be summarized in an analytical scheme, as reported in Table 1.5 as follows:

The main items that must be investigated are:

– *Free Cash Flow from Operations* (*FCFO*): they are the *Free Cash-flows from Operating Area* as derived by the company's operating activities and they are designed to pay all investors both in equity and debt. They represent the monetary component of the Operating Income of the company. Therefore, the FCFF is function of the Operating Income (there is no difference between EBITDA and EBIT because the non-monetary costs are not considered) and dynamics in the NWC and CAPEX. They are defined "*free*" because they represent the cash that the company is free to distribute to debtholders and shareholders and to pay taxes by having already covered the needs for Investments and NWC;
– *Free Cash Flow to Equity* (*FCFE*): they are the remaining free cash flows after having covered all company requirements including payments on debt, and therefore they are designed to pay the shareholders in terms of dividends. They represent the monetary component of the net income of the company. They show how the FCFF's are divided between bondholders, stockholders and taxes. If the FCFE are negative, they represent the company's capital need to continue its business. Therefore, it is the amount of the capital required for recapitalization.

Table 1.5 Analytical scheme for Free Cash-Flow From Operations and Free Cash Flow to Equity

EBITDA
(Increase)/decrease—trade receivables net
Increase/(decrease)—trade payables
(Increase)/decrease—trade working capital (TWC)
(Increase)/decrease—inventory
(Increase)/decrease—others operating receivables net
Increase/(decrease) – Others Operating Payables
(Increase)/decrease—net working capital (NWC)
(Increase)/decrease—net tangible and intangible operating assets
(Increase)/decrease—financial operating assets
(Increase)/decrease—inventory stable over time
(Increase)/decrease—Capex
(Decrease)—provisions for employees
(Decrease)—provision for risk and charges
(Decrease)—provision for taxes
(Decrease)—provisions
(Current operating taxes)
(Operating taxes)
Free cash flow from operations (FCFO)
Increase/(decrease)—share capital in money
Increase/(decrease)—Equity
Increase/(decrease)—long-term financial debts
(Increase)/decrease—long-term financial credits
Increase/(decrease)—long-term net financial position
Increase/(decrease)—short-term financial debts
(Increase)/decrease—short-term financial credits
(Increase)/decrease—marketable securities
(Increase)/decrease—cash and cash-equivalents
Increase/(decrease)—short-term net financial position
Increase/(decrease)—net financial position (NFP)
Operating and non-operating financial revenues
(Interest on debts)
(Other financial costs)
Financial profit/(loss)
Non-operating and non-current operating revenues
(Non-operating and non-current operating costs)
Non-operating profit/(loss)
(Increase)/decrease—net tangible and intangible surplus assets
(Increase)/decrease—financial surplus assets
(Increase)/decrease—non-operating and non-current operating receivables

(continued)

Table 1.5 (continued)

Increase/(decrease)—non-operating and non-current operating payables
(Increase)/decrease—surplus assets
(Current corporate taxes)
(Corporate taxes)
Free cash flow to equity (FCFE)

References

Anthony RN, Hawkins DF, Merchant KA (2011) Accounting: text and cases, 13th edn. McGraw-Hill, New York
Barney J (1991) Firm resources and sustained competitive advantage. J Manag 17:99–120
Buffet M, Clark D (2008) Warren buffet and the interpretation of financial statements. Simon & Schuster Inc., New York
Bertini U (1990) Il sistema d'azienda. Schemi di analisi. Giappichelli, Torino
Bertini U (1995) Scritti di politica aziendale. Giappichelli, Torino
Bertini U (2009) Modelli di governance, aspettative degli stakeholder e creazione del valore. In: Corporate governance: governo, controllo e struttura finanziaria. AIDEA, il Mulino, Bologna
Bianchi Martini S (2009) Introduzione all'analisi strategica dell'azienda. Giappichelli, Torino
Coda V (1988) L'orientamento strategico dell'impresa. Utet, Torino
Collis DJ, Montgomery CA, Invernizzi G, Molteni M (2012) Corporate level strategy. McGraw-Hill, Milano
Dallocchio M, Salvi A (2004) Finanza d'azienda. Egea, Milano
Dallocchio M, Tzivelis D, Vinzia A (2011) Finanza per la crescita sostenibile. Egea, Milano
De Luca P (2013a) La formula strategica dell'azienda. In: Gaeleotti M, Garzella S (eds), Governo Strategico dell'Azienda. Giappichelli Editore, Torino
De Luca P (2013b) La strategia di business. In: Gaeleotti M, Garzella S (eds), Governo strategico dell'azienda. Giappichelli Editore, Torino
De Luca P (2013c) Strategia e dinamica del settore. In Gaeleotti M, Garzella S (eds), Governo strategico dell'azienda. Giappichelli Editore, Torino
De Luca P (2013d), La dinamica competitiva. In: Galeotti M, Garzella S (eds), Governo strategico dell'azienda. Giappichelli Editore, Torino
De Luca P (2015) Il risanamento della formula strategica dell'azienda. Un modello di analisi per la risoluzione della crisi aziendale, Collana di Studi e Ricerche di Economia Aziendale, Giappichelli Editore, Torino
De Luca P, Ferri S, Galeotti M (2016) The entrepreneurial strategic formula of the firm: a theoretical business model. Int J Bus Perform Manag 17(4):447–465
De Luca P, Ferri S, Galeotti M (2017) The company's business model and its valuation: a theoretical approach. Int J Acad Res Acc Financ Manag Sci 7(3):139–150
Diamond DW (1989) Reputation acquisition in debt markets. J Polit Econ 97:828–862
Fiori G, Tiscini R, Donato F (2004) Corporate governance, evoluzione normativa ed informazione esterna d'impresa. In: Salvioni DM (ed), Corporate governance e sistemi di controllo della gestione aziendale. Franco Angeli, Milano
Galeotti M (2008) Le strategie competitive dell'azienda nei mercati finanziari. Aracne, Roma
Galeotti M, Garzella S (2013) Governo strategico dell'azienda. Giappichelli, Torino
Garzella S (2005) Il sistema d'azienda e la valorizzazione delle "potenzialità inespresse. Giappichelli, Torino
Garzella S (2006) Il governo delle sinergie. Sistematicità e valore nella gestione dell'azienda. Giappichelli, Torino

Giannessi E (1979) Appunti di economia aziendale. Libreria Scientifica Giordano Pellegrini, Pisa
Graham B, Dodd D (1940) Security analysis. McGraww-Hill Book Company Inc., New York
Grant RM (1991) Contemporary strategy analysis. Concepts, techniques, applications. Blackwell, Oxford
Hamel G, Prahalad C (1990) The core competence of the corporation. Harvard Bus Rev, 1–15 (May-Giune)
Harris M, Raviv A (1991) The theory of capital structure. J Financ 1:297–355
Hofstede G (1993) Cultural constraints in management theories. University of Limburg, Maastricht, The Netherlands
Invernizzi G (1999) Il sistema delle strategie a livello aziendale. McGraw-Hill, Milano
Invernizzi G (2008) Strategia aziendale e vantaggio competitivo. McGraw-Hill, Milano
Invernizzi G (2011) La strategia delle imprese multibusiness. McGraw-Hill, Milano
Itami H (1988) Le risorse invisibili. Gea-Isedi, Milano
Jensen MC (1986) Agency cost of free cash flow, corporate finance and takeovers. Am Econ Rev 76:323–329
Jensen MC, Meckling W (1976) Theory of the firm: Managerial behavior, agency costs, and ownership structure. J Financ Econ 3:305–360
Markides P (2008) Game changing strategies. Jossey Bass, San Francisco
Mintzberg H (1994) The rise and fall of strategic planning. Prentice Hall, Hertfordshire
Myers SC (1977) Determinants of corporate borrowing. J Financ Econ 5:147–175
Porter ME (1985) Competitive advantage. Creating and sustaining superior performance. Free Press, New York
Quagli A (2001) Knowledge management – La gestione della conoscenza aziendale. Egea, Milano
Quinn JB (1992) Intelligent enterprise: a knowledge and service based paradigm for industry. The Free Press, New York
Rumelt RP (1984) Toward a strategic theory of the firm. In: Lamb R (ed), Competitive strategic management. Prentice-Hall, Englewood Cliffs, New York
Rumelt RP (1987) Theory, strategy, and entrepreneurship. In: Teece DJ (ed), The competitive challenge. Ballinger, Cambridge, MA
Saloner G, Shepard A, Podolny J (2001) Strategic management. Wiley, New Jersey
Thompson AA, Strickland AJ, Gamble JE (2006) Strategic management. McGraw-Hill, Higher Education
Vernimmen P, Quiry P, Dallocchio M, Le Fur Y, Salvi A (2014) Corporate finance—theory and practice, 4th edn. Wiley, New Jersey
Wernerfeldt B (1984) A resource based-view of the firm. Strateg Manag J 5:171–180
Williamson OE (1985) The economic institutions of capitalism. The Free Press, New York

Chapter 2
Company Profitability Analysis

Abstract In this second step of the fundamental company analysis, attention is focused on the company's profitability measurements. Several tools can be used. The choice is usually based on an analysis perspective and its nature. In this context a financial perspective is followed. The analysis is developed on three main fronts:

– analysis of the economic and financial dynamics over time. The analysis can start from each economic and financial figure. In this context, it can be useful to start the analysis from capital sources and their investment in the company's activities and to measure their returns in terms of earnings and dividends;
– the analysis of the main financial ratios. There are many well-known ratios in literature. In this context the most commonly used financial ratios used in the financial community are considered. They are able to complete the analysis because they are in line with the analytical schemes regarding Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity;
– the analysis of the growth rate. The fundamental company analysis leads to investigate into the expected consistency of future economic and financial dynamics. Consequently, one of the most relevant key of the analysis is the estimate of the company's future growth rate with regards mainly to both Net Income and Operating Income.

2.1 Analysis of Economic and Financial Dynamics

The analytical schemes showed in the previous chapter focus on the definition of the Operating and Net Income, Capital Invested and Capital Structure, and Free Cash-flow from Operations and Free Cash-flow to Equity.

The analysis can start from each of them. In this context, it may be useful to start the analysis from the sources of capital and their investment in company's activities and to measure their returns in terms of earnings and dividends.

Equity and debt are the two sources of capital to finance company activities.

Equity refers to the shareholders source of capital. It can be increased through new capital or self-financing.

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_2

Equity must be repaid by dividends. Therefore, the equity remuneration is residual because it is achieved after the repayment of all production factors only in the case of specific conditions.

Debt in capital structure refers to the financial debt only. Usually, more specifically it refers to the Net Financial Position (NFP) that it is equal to the difference between financial debts (both long and short) and liquidity (equal to the sum between financial credits, marketable securities and cash) as proposed in the analytical scheme in the previous chapter. Therefore, the increase (decrease) of financial debt increases (decreases) the Net Financial Position.

Financial debt generates financial costs mainly in terms of interest on debt. It represents the debt holders' remuneration and then the cost of debt. It is located between EBIT and EBT by reducing it. It also generates a cash-out with negative effects on the Free Cash-flow to Equity (FCFE).

It is important to point out that if financial costs are not paid, they generate new financial debt to be added to the original one.

Finally, it is important to point out that the increase in financial debt increases the Net Financial Position and it increases the Free Cash-flow to Equity (FCFE).

The sum between equity and financial debt defines the Capital Structure of the company. Capital sources are invested in assets, both operating and non-operating, Net Working Capital and Liquidity in order to achieve company activities.

Investments in Operating Assets

They can be distinguished in operating capital expenditure (Capex) and in Net Working Capital (NWC).

Their total defines the Net Operating Capital Invested (NOCI). Consequently, the increase in operating assets increase investments in NOCI and then it increases the amount of Capital Invested (CI).

Investments in Operating Capital Expenditure (*Capex*) refer to investments in operating assets strictly such as operating tangible, intangible, financial assets, and inventory stable over time.

The main aim of these investments is to increase the EBITDA over time. It can be achieved by increasing operating revenues (through the increase of production capacity) and by decreasing operating costs (through the increase of technology efficiency) over time.

The positive effect on EBITDA is contrasted by the negative effect on EBIT due to the amortization and depreciation related to investments in tangible and intangible assets. Therefore, the increase of investments increases the distance between EBITDA and EBIT

Finally, the investments in Capex generate cash-out with negative effects on Free Cash-flow from Operations (FCFO). However, it is important to consider that the increase in Capex should increase the EBITDA by pushing up the FCFO over time. The main problem is related to the time of these two movements: while the cash-outs are immediate, the increase in EBITDA should be achieved in the near future.

Note that investments in operating financial assets can generate operating financial revenues with a positive effect on EBT-Operating.

Financial assets can be subject to the impairment test. The main problem concerns the difference between their market value and book value: if the first is greater than the second, the company can increase asset value through its revaluation; otherwise if the first is lower than the second, the company must decrease the asset value. In both cases there are no cash movements and they affect the EBIT. Note that the same effects concern the change in value of tangible and intangible assets.

Investments in Net Working Capital (*NWC*) refer to the investments in operating receivables and payables and inventory. The operating receivables and payables refer to the operating revenues and costs that they do not generate cash-in and cash-out yet. The inventory refers to the products created but not yet sold.

Specifically, sales revenues produce trade receivables while costs of goods sold generate trade payables. The difference between them defines the Trade Working Capital (TWC): if the first is higher than the second, the TWC increases; otherwise, if the second is higher than the first, the TWC decreases. The TWC is the main part of the NWC. Therefore, the increase (decrease) in TWC increases (decreases) the NWC.

On the other hand, other operating revenues and costs produce other operating receivables and payables. These operating receivables and payables plus inventory define the second part of the NWC. Therefore, if the other operating receivables plus inventory are greater than the other operating payables, the NWC increases; otherwise, if other operating payables are greater than other operating receivables plus inventory, the NWC decreases.

The increase in NWC (due to the increase in TWC, other operating receivables and inventory and decrease of other operating payables) increases the Net Operating Capital Invested (NOCI) and subsequently the amount of Capital Invested (CI); otherwise, the decrease of NWC (due to the decrease in TWC, other operating receivables and inventory and increase of other operating payables) decreases the Net Operating Capital Invested (NOCI) and subsequently the amount of Capital Invested (CI).

The increase and decrease in NWC over time have direct effects on cash-flows. In cash-flow terms the increase (decrease) of receivables (both trade and other operating) and inventory are considered cash-out (cash-in) while the increase (decrease) of payables (both trade and other operating) are considered cash-in (cash-out). Therefore, the increase in NWC can be considered a cash-out and it subsequently reduces the Free Cash-flow from Operations (FCFO); otherwise, the decrease in NWC can be considered as a cash-in and then it increases the Free Cash-flow from Operations (FCFO).

Investments in Non-operating Assets

They refer to the investments in Surplus Assets. It is possible to distinguish between investments in tangible, intangible and financial Surplus Assets, and investments in non-operating and non-current operating receivables and payables.

The investments in Surplus Assets are not considered in Net Operating Capital Invested (NOCI). Therefore, they increase the amount of Capital Invested (CI) only.

Investments in tangible and intangible and financial Surplus Assets refer to the investments in non-core business activities. The amortization and depreciation of intangible and tangible Surplus Asset are located between EBT-Operating and EBT. The main reasoning for this positioning is not to damage the EBIT. In this way, the EBIT can be considered strictly operating.

Investments in tangible and intangible Surplus Assets can affect the non-operating revenues and costs that they affect the EBT.

Investments in financial Surplus Assets can generate non-operating financial revenues. They have a positive effect on EBT.

Similarly, the investments in financial assets computed in Capex, are subject to the impairment test. In this case the change in value has direct effects on the EBT. The same effect on EBT is due to the change in value of tangible and intangible Surplus Assets.

In cash-flow terms, investments in Surplus Assets (tangible, intangible, financial) generates cash-out with negative effect on Free Cash-flow to Equity (FCFE). Indeed, they are not considered in Free Cash-flow from Operations (FCFO) because they are non-operating.

Investments in non-operating and non-current operating receivables and payables refer to the investments in non-operating and non-current operating revenues and costs that have not yet been translated into cash-in and cash-out.

They are not considered in NWC because they refer to the revenues and costs not related to the core-business of the company.

If the non-operating and non-current operating receivables are higher than the non-operating and non-current operating payables, the Surplus Asset increases; otherwise, if the non-operating and non-current operating receivables are lower than the non-operating and non-current operating payables, the Surplus Asset decreases.

In cash-flow terms, the increase (decrease) of non-operating and non-current operating receivables can be considered as cash-out (cash-in) and the increase (decrease) of non-operating and non-current operating payables can be considered cash-in (cash-out). Therefore, the increase in Surplus Assets is a cash-out and therefore it reduces the Free Cash-flow to Equity (FCFE); otherwise, the decrease in Surplus Assets is a cash-in and thus it increases the Free Cash-flow to Equity (FCFE). Also, in this case the Free Cash-flow from Operations (FCFO) are not affected because they are not related to the core-business.

Investment in Liquidity

Specific reasoning must be reserved to investments in Liquidity. It refers to the amount of cash, cash equivalent, financial credits and marketable securities that a company decides to maintain in the company. Therefore, it is not a capital source but capital invested because it is an investment decision. Indeed, liquidity is positioned in the Net Financial Position by reducing its value.

It is necessary to distinguish between: (i) the temporary movements of Liquidity, that they are due to the effect of cash-in and cash-out used to balance the movements of other items; (ii) the investment in Liquidity that defines the level of liquidity that company wants to maintain over time. Only the second type can be considered as a proper investment.

Liquidity can generate operating and non-operating financial revenues with positive effects on EBT and Free Cash-flow to Equity (FCFE).

In cash-flow terms, it can be considered as a cash-out. Therefore, the increase in Liquidity reduces the Free Cash-flow to Equity (FCFE).

Provisions

Finally, it is necessary to consider Provisions for risk and charges, taxes and employees. They reduce the Capital Invested (CI).

Accruals for provisions (employees, taxes, risk and charges) are located in the operating costs (non-cash). Therefore, they have direct effects on EBIT.

In cash-flow terms, they only have effects if they are really used with consequent cash-out by reducing the Free Cash-flow from Operations (FCFO).

2.2 Financial Ratios Analysis

The analysis of economic and financial dynamics can be improved by using some ratios. There are many ratios well-known in literature (Benninga 2014; Brealey et al. 2016; Berk and DeMarzo 2008; Damodaran 2012, 2015; Vernimmen et al. 2014; Hillier et al. 2016; Koller et al. 2015; Copeland et al. 2004; Fuller and Farrell 1987; Ross 2015). In this context the most commonly used Financial Ratios in the financial community are considered. They are able to complete the analysis because they are in line with the analytical schemes regarding Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity.

It is worth noting that the most significant problem about ratios is the possibility to use different sources. If it happens, use of the same formula is not enough to ensure comparable results. Thus, it is necessary to standardize the data before using the ratios.

Use of the same analytical schemes over the time together with the normalization process of past data and the rebalance between them and the expected future data, allows for data standardisation.

Also, it allows for outlining of the relationships among the economic and financial dynamics as defined. To this end, the Financial Ratios used in this context can be summarized as in Fig. 2.1.

Figure 2.1 can be read starting from one of three dimensions: income (Operating Income and Net Income), capital (Capital Invested and Capital Structure), cash-flows (Free Cash-flow from Operations and Free Cash-flow to Equity).

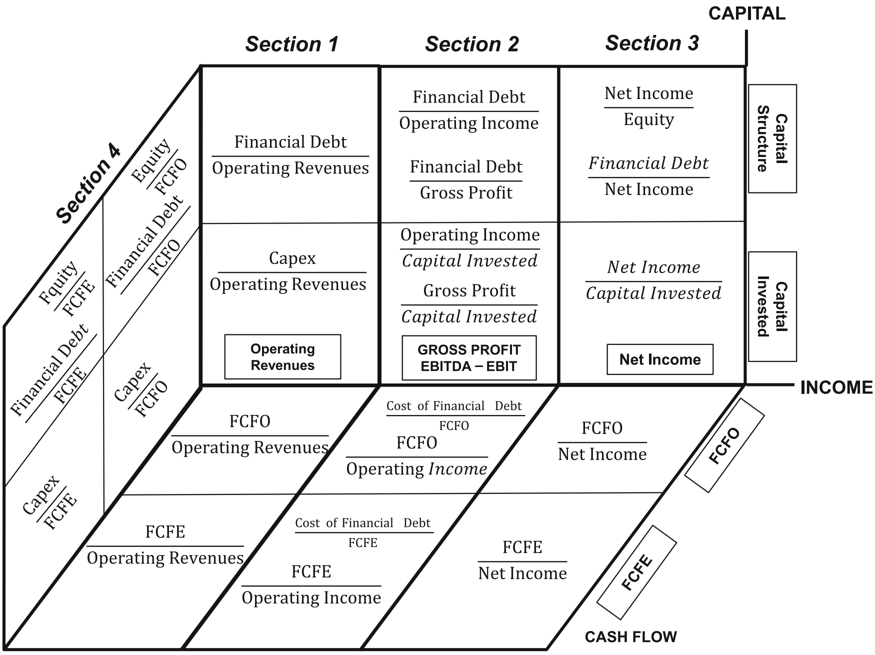


Fig. 2.1 Financial ratios

However, it is common use to start the analysis with the economic dynamic. In this sense, the following relationships can be analysed:

- *Section 1: Operating Revenues, Capital Invested and Capital Structure*, and *Free Cash-flow from Operations* and *Free Cash-flow to Equity*;
- *Section 2: Operating Income, Capital Invested and Capital Structure*, and *Free Cash-flow from Operations* and *Free Cash-flow to Equity;*
- *Section 3: Net Income, Capital Invested and Capital Structure*, and *Free Cash-flow from Operations* and *Free Cash-flow to Equity*;
- *Section 4: Capital Invested and Capital Structure,* and *Free Cash-flow from Operations* and *Free Cash-flow to Equity.*

The analysis must be completed by considering the Operating Leverage (OL) and Financial Risk Level (FRL) jointly of Sections 1–4. The Operating Leverage relates to the relationship between operating revenues and contribution margin, while the Financial Risk Level relates to the relationship between the EBIT and Net Income due to the weight of financial cost of debt.

Section 1

An analysis of the relationships between *Operating Revenues, Capital Invested and Structure*, and *Free Cash-flows from Operations and Free Cash-flow to Equity*, it can be achieved on the basis of several main ratios:

(a) *Capex on Operating Revenues (COR);*
(b) *Financial Debt on Operating Revenues (FDOR);*
(c) *Cash-flow on Operating Revenues (CFOR);*

Before starting the analysis it is important to point out that Operating Revenues are used rather than Revenues because they refer to the core-business of the company strictly. However, also if there are other Revenues (such as non-operating revenues, non-current operating revenues, and financial revenues) the Operating Revenues represent the greatest and most relevant part of Total Revenues.

(a) ***Capex on Operating Revenues (COR)***:

$$COR = \frac{Capex}{Operating\,Revenues\,(OR)} \tag{2.1}$$

The COR ratio measures the return of capital invested in Capex on the basis of Operating Revenues. Specifically a low value of ratio means a high return of capital invested in Capex in terms of Operating Revenues. In this case, the investments in Capex are able to push-up the Operating Revenues. On the contrary, a high value of ratio means a low return of capital invested in Capex in terms of Operating Revenues. In this case, the investments in Capex are not able to push-up the Operating Revenues.

(b) ***Financial Debt on Operating Revenues (FDOR)***:

$$FDOR = \frac{Financial\,Debt\,(FD)}{Operating\,Revenues\,(OR)} \leftrightarrow NFDOR = \frac{Net\,Financial\,Position\,(NFP)}{Operating\,Revenues\,(OR)} \tag{2.2}$$

The FDOR and NFDOR ratios measure the capability of the Operating Revenues to face Financial Debt and Net Financial Position in Capital Structure. In both cases the meaning is the same. Specifically a high value of ratio means a low level of ability of Operating Revenues to face Financial Debt (FD) and Net Financial Position (NFP). On the contrary, a low value of ratio means a high capability of Operating Revenues to face Financial Debt (FD) and Net Financial Position (NFP).

(c) ***Cash-flow on Operating Revenues (CFOR)***:

$$CFOR = \frac{Cash\text{-}flow\,(CF)}{Operating\,Revenues\,(OR)} \tag{2.3}$$

The CFOR ratio measures the relationship between Cash-flow and Operating Revenues. It can be considered as an indirect and approximate measure of the company's ability to transform Operating Revenues into Cash-flows. Specifically a low value of ratio means a bad relationship between Operating Revenues and Cash-flows. On the contrary, a high value of ratio means a good relationship between Operating Revenues and Cash-flows.

Considering two types of cash-flows, Free Cash-flow from Operations (FCFO) and the Free Cash-flow to Equity, the ratio can be applied in two versions:

$$CFOR = \frac{CF}{OR} \quad \rightarrow \quad \begin{array}{l} CFOOR = \frac{Free\,\text{Cash-flow from Operations}\,(FCFO)}{Operating\ Revenues\,(OR)} \\ CFEOR = \frac{Free\,\text{Cash-flow to Equity}\,(FCFE)}{Operating\ Revenues\,(OR)} \end{array} \tag{2.4}$$

Section 2

The analysis of the relationships between *Operating Income, Capital Invested and Capital Structure*, and *Free Cash-flows from Operations and Free Cash-flows to Equity* can focus on the following main ratios:

(a) *Return on Operating Capital Invested (ROIC);*
(b) *Financial Debt on Operating Income (FDOI);*
(c) *Cost of Financial Debt on Cash-flows (CDCF);*
(d) *Cash-flows on Operating Income (CFOI).*

(a) ***Return on Operating Capital Invested (ROIC)***:

$$ROIC = \frac{Operating\ Income\,(OI)}{Capital\ Invested\ in\ Operating\ Assets\,(CI)} \tag{2.5}$$

The ROIC ratio defines the relationship between Operating Income (OI) and Capital Invested in Operating Assets (CI) only. It is also called Return on Investment (ROI) if all company's investments are in core-business.

The ratio measures the return of capital invested in operating assets of the company. Specifically a high value of ratio means a high return of capital invested; on the contrary, a low value of ratio means a low return of capital invested.

The Capital Invested in Operating Assets is the capital sources invested to support the company's core-business activities. Therefore, investments in non-operating activities are not included. By following the analytical scheme proposed in the previous Chapter, the Capital Invested in Operating Assets is the Net Operating Capital Invested (NOCI). It is equal to the total of investments in Net Operating Capital Expenditures (Capex) and Net Working Capital (NWC).

The Operating Income is the difference between Operating Revenues and Operating Costs. Therefore, it measures the results of the Operating activities. It can be measured by three main figures: EBITDA, EBIT and NOPAT. Therefore, the ratio can be applied in three versions:

$$ROIC = \frac{OI}{CI} \rightarrow \begin{array}{l} ROIC = \frac{EBITDA}{Net\,Operating\,Capital\,Invested\,(NOCI)} \\ ROIC = \frac{EBIT}{Net\,Operating\,Capital\,Invested\,(NOCI)} \\ ROIC = \frac{NOPAT}{Net\,Operating\,Capital\,Invested\,(NOCI)} \end{array} \tag{2.6}$$

It is important to point out that if Capex is only considered as Capital Invested rather than the NOCI, it gets the Return on Assets (ROA) as follows:

$$ROA = \frac{OI}{Capex} \rightarrow \begin{array}{l} ROA = \frac{EBITDA}{Capex} \\ ROA = \frac{EBIT}{Capex} \\ ROA = \frac{NOPAT}{Capex} \end{array} \tag{2.7}$$

In order to acquire more information, it is also possible to measure the Return on Capital Invested in terms of Gross Profit (GP). In this case it takes on the following ratio (ROGP):

$$ROGP = \frac{Gross\,Profit\,(GP)}{Net\,Operating\,Capital\,Invested\,(NOCI)} \tag{2.8}$$

It is worth noting that in all cases, the increase of ratio is not necessarily good news. With equal Operating Income, the increase of the ratio is due to the reduction of total assets. If the reduction is due to the amortization and depreciation process, the increase of ratio is due to a reduction in investments. Most of all, the decrease of investments in Capex implies the reduction of future capabilities of the company to compete in the business and, subsequently a future reduction in Operating Income.

Otherwise, the increase of ratio due to the development of technologies, indicates an increase of the future capabilities of the company to compete in the business and, subsequently a future increase in the Operating Income.

Usually EBIT is preferred to EBITDA and NOPAT to measure Operating Income.

Using EBIT, the ratio can be decomposed on the basis of the Return on Sales (ROS) and the Turnover of Capital Invested (TCI).

Considering that the ROS is equal to the ratio between EBIT and Sales Revenues (SR), and the TCI is equal to the ratio between Sales Revenues (SR) and Net Operating Capital Invested (NOCI), the following is achieved:

$$ROI = ROS \cdot TCI \rightarrow \begin{array}{l} ROS = \frac{EBIT}{SR} \\ TCI = \frac{SR}{NOCI} \end{array} \rightarrow ROI = \frac{EBIT}{SR} \cdot \frac{SR}{NOCI} = \frac{EBIT}{NOCI} \tag{2.9}$$

The use of EBITDA highlights the effects of the amortization and depreciation process on the Operating Income. However, it is possible to move from the EBITDA to the EBIT by considering the Net Operating Capital Invested Gross of Amortization and Depreciation (NOCIG) as following (Silvi 2012):

$$ROIC = \frac{EBITDA}{NOCIG} \cdot \frac{NOCIG}{NOCI} \cdot \frac{EBIT}{EBITDA} = \frac{EBIT}{NOCI} \tag{2.10}$$

The use of NOPAT can only be calculated if taxes splitting is not used. Indeed, the use of tax splitting allows for a distinction between operating taxes and corporate taxes. In this case, the EBIT is calculated by also considering operating taxes. Therefore it is equal to NOPAT.

Differently, if tax splitting is not used in the tax area operating taxes and corporate taxes are considered jointly. In this case EBIT is the operating income before taxes while the NOPAT is the operating income after taxes.

Using NOPAT, the ROIC ratio measures the after-taxes operating profit divided by capital invested in core-business of the company (Koller et al. 2015).

(b) ***Financial Debt on Operating Income (FDOI)***:

$$FDOI = \frac{Financial\,Debt\,(FD)}{Operating\,Income\,(OI)} \tag{2.11}$$

The FDOI ratio measures the company's ability to face Financial Debt (FD) through the Operating Income (OI). Generally, the greater the distance between the amount of Operating Income and the amount of Financial Debt, the lower the financial risk.

Considering that the Operating Income can be measured by EBITDA, EBIT and NOPAT it is also important to consider that the Net Financial Position (NFP) can be used instead of Financial Debt (FD).

Therefore, the ratio can be applied in several versions as follows:

$$FDOI = \frac{FD}{OI} \rightarrow \begin{array}{l} FDOI = \frac{Financial\,Debt\,(FD)}{EBITDA} \leftrightarrow NFDOI = \frac{Net\,Financial\,Position\,(NFP)}{EBITDA} \\ FDOI = \frac{Financial\,Debt\,(FD)}{EBIT} \leftrightarrow NFDOI = \frac{Net\,Financial\,Position\,(NFP)}{EBIT} \\ FDOI = \frac{Financial\,Debt\,(FD)}{NOPAT} \leftrightarrow NFDOI = \frac{Net\,Financial\,Position\,(NFP)}{NOPAT} \end{array} \tag{2.12}$$

In order to acquire more information, the Gross Profit (GP) ability to face Financial Debt (FD) or Net Financial Position (NFP) can be measured. In this case, the FDGP ratio is the following:

$$FDGP = \frac{Financial\,Debt\,(FD)}{Gross\,Profit\,(GP)} \leftrightarrow NFDOI = \frac{Net\,Financial\,Position\,(NFP)}{Gross\,Profit\,(GP)} \tag{2.13}$$

(c) ***Cost of Financial Debt on Cash-flows (CDCF)***:

$$CDCF = \frac{Cost\ of\ Financial\ Debt\ (CD)}{Cash\ flows\ (CF)} \tag{2.14}$$

Despite the fact that the cost of financial debt has both economic (it is a cost) and financial dynamics (it is a cash-out), in this context an analysis of its impact on financial dynamics is preferred. Indeed, the effects of cost of debt on cash-flows is very relevant because if the company cannot face the relative cash-out, it is in a default condition.

The Cost of Financial Debt (CD) is equal to the Interest on Debt (ID) plus the other Financial Costs on Debt (FCD). Therefore:

$$CD = ID + FCD \tag{2.15}$$

The most relevant part is related to the Interest on Debt. Also, by considering that the cash-flows can be distinguished between Free Cash-flow from Operations (FCFO) and Free Cash-flow to Equity (FCFE), it is possible to apply the ratio in several versions as follows:

$$CDCF = \frac{CD}{CF} \rightarrow \begin{array}{l} CDCFO = \frac{CD}{FCFO} \leftrightarrow IDCFO = \frac{ID}{FCFO} \\ CDCFE = \frac{CD}{FCFE} \leftrightarrow IDCFE = \frac{ID}{FCFE} \end{array} \tag{2.16}$$

(d) ***Cash-flows on Operating Income (CFOI)***:

$$CFOI = \frac{Cash\ flows\ (CF)}{Operating\ Income\ (OI)} \tag{2.17}$$

The CFOI ratio measures the relationship between Operating Income and Cash-flows. It can be considered as an indirect and approximate measure of the company's ability to transform Operating Income in Cash-flows. Specifically, a high value of ratio means a good relationship between Operating Income and Cash-flows; otherwise, a low value of ratio means a bad relationship between Operating Income and Cash-flows.

Considering the two types of cash-flows, Free Cash-flow from Operations (FCFO) and the Free Cash-flow to Equity, the ratio can be applied in two versions:

$$CFOI = \frac{CF}{OI} \quad \rightarrow \quad \begin{array}{l} CFOOI = \frac{Free\ Cash\ flow\ from\ Operations\ (FCFO)}{Operating\ Income\ (OI)} \\ CFEOI = \frac{Free\ Cash\ flow\ to\ Equity\ (FCFE)}{Operating\ Income\ (OI)} \end{array} \tag{2.18}$$

Section 3

An analysis of the relationships between *Net Income, Capital Invested and Capital Structure*, and *Free Cash-flows from Operations* and *Free Cash-flow to Equity* can be focus on the following main ratios:

(a) *Return on Equity (ROE);*
(b) *Financial Debt on Net Income (FDNI);*
(c) *Net Income on Capital Invested (NICI);*
(d) *Cash-flows on Net Income (CFNI).*

(a) ***Return on Equity (ROE)***:

$$ROE = \frac{Net\,Income\,(NI)}{Equity\,(E)} \tag{2.19}$$

The ROE ratio is one of the most popular and most commonly used. It measures the return on Capital Invested in equity on the basis of Net Income.

In the financial approach the return of shareholders' investment is the aim of the company and thus the ROE is the true bottom-line measure of company performance.

The ROE can be defined on the basis of ROI and the Financial Leverage (L). Assuming that there are no non-operating activities, the ROI measures the return on Capital Invested in Operating Assets $(ROI \equiv ROIC)$. Also, the Net Income (NI) is equal to the EBIT less Interest on debt (ID) and Taxes (T):

$$NI = EBIT - ID - T \tag{2.20}$$

Considering that EBT is equal to EBIT less Interest in debt (ID), we can assume the calculation of Taxes (T) on the basis of a marginal tax rate (t_c) on EBT. Therefore:

$$EBT = EBIT - ID$$

$$T = EBT \cdot t_c \rightarrow T = (EBIT - ID) \cdot t_c$$

$$NI = EBIT - ID - [(EBIT - ID) \cdot t_c] = (EBIT - ID) \cdot (1 - t_c)$$

$$NI = (EBIT - ID) \cdot (1 - t_c) \tag{2.21}$$

Considering that:

$$ROE = \frac{NI}{E} \rightarrow NI = ROE \cdot E$$

$$ROI = \frac{EBIT}{CI} \rightarrow EBIT = ROI \cdot CI$$

and by considering that the interest on debt (ID) is equal to the Cost of Debt (K_D) multiplied by the amount of debt (D) in capital structure:

$$ID = D \cdot K_D$$

By replacing it, Eq. (2.21) can be re-written as follows:

$$ROE \cdot E = (ROI \cdot CI - D \cdot K_D) \cdot (1 - t_c) \tag{2.22}$$

We can assume that the entire capital, equal to Equity (E) plus Debt (D), is invested in operating assets only. In this case, the Capital Invested (CI) is equal to the Capital Structure (CS) that it is equal to the sum of Equity (E) and Debt (D):

$$CI = CS \rightarrow CI = E + D$$

By replacing it, Eq. (2.22) can be re-written as follows:

$$\begin{aligned} ROE \cdot E &= [ROI \cdot (E + D) - D \cdot K_D] \cdot (1 - t_c) \\ &= [ROI \cdot E + D \cdot (ROI - K_D)] \cdot (1 - t_c) \end{aligned}$$

Dividing first and second terms by Equity (E), we achieve:

$$ROE \cdot E \cdot \frac{1}{E} = [ROI \cdot E + D \cdot (ROI - K_D)] \cdot (1 - t_c) \cdot \frac{1}{E}$$

$$ROE = \left[ROI \cdot \frac{E}{E} + \frac{D}{E} \cdot (ROI - K_D)\right] \cdot (1 - t_c)$$

$$ROE = \left[ROI + \frac{D}{E} \cdot (ROI - K_D)\right] \cdot (1 - t_c)$$

The ratio between Debt (D) and Equity (E) defines the Leverage (L):

$$L = \frac{D}{E} \tag{2.23}$$

And subsequently:

$$ROE = [ROI + L \cdot (ROI - K_D)] \cdot (1 - t_c) \tag{2.24}$$

Equation (2.24) shows the relationship between ROE, ROI and Leverage (L). Specifically, it shows the multiple effects of debt on ROE. This effect can be positive or negative. It depends on the relationship between the ROI and the Cost of Debt (K_D) on the one hand, and the amount of Debt (D) in the Capital Structure on the other hand.

Specifically, the Financial Leverage (FL) can be defined as follows:

$$FL = L \cdot (ROI - K_D) \tag{2.25}$$

Therefore if:

- $ROI > K_D$: the Financial Leverage (FL) is positive. The Debt invested in the company generates a return on investment (ROI) greater than its costs (K_D); therefore, the increase in ROE is due to the ROI in an unlevered case plus the positive Leverage (L) effects. Consequently, other conditions equal, greater is the Leverage (L) and higher is the ROE. In this case, the Leverage (L) is a positive multiplier. The shareholders achieve earnings created from Debt. They benefit from the positive difference between ROI and K_D;
- $ROI < K_D$: the Financial Leverage (FL) is negative. The Debt invested in the company generates a return on investment (ROI) lower than its costs (K_D); therefore, the decrease in ROE is due to the ROI in unlevered case plus the negative Leverage (L) effects. Consequently, other conditions equal, greater is the Leverage (L) and lower is the ROE. In this case, the Leverage (L) is a negative multiplier. The shareholders claim losses arising from debt. Their loss is due to the negative difference between ROI and K_D.

Note that the ROE is an account measure. Usually, investors prefer the Total Return on Shareholders (TRS). It also takes into account market variations in stock price. Indeed, it combines the amount that shareholders gain through any increase in the share price over a given period with the sum of dividends paid to them over the period. Specifically, TRS is equal to the Percentage Change in Share Price (ΔSP) plus the Dividend Yield (DY) as follows:

$$TRS = \Delta SP + DY \tag{2.26}$$

The Dividend Yield (DY) measures the annual dividend per share paid by the company to its shareholders expressed as a percentage of its share price. Therefore, the Dividend Yield is equal to Dividend per Share (DPS) divided by Share Price (SP), as follows:

$$DY = \frac{DPS}{SP} \tag{2.27}$$

Note that Dividends are paid typically on a quarterly basis. Therefore, they must be annualized to calculate the Implied Dividend Yield (IDY) as follows:

$$IDY = \frac{Most\ Recent\ Quarterly\ DPS \times 4}{Current\ SP}$$

The TRS can be broken down (Koller et al. 2015). The analysis of the TRS's components can be useful for greater understanding of managers activities and for greater planning of future targets.

The Percentage Change in Share Price (ΔSP) can be expressed in terms of Percent Increase in Earnings (ΔEr) and the Percentage Change in a Company's Price-to-Earnings Ratio ($\Delta(\mathrm{P}/Er)$) as follows:

$$\Delta SP = \Delta Er + \Delta\left(\frac{P}{Er}\right) \tag{2.28}$$

where:

- the first term (ΔEr) is the Percentage Change in Share Price due to the change in the company's fundamentals as measured by the change in earnings;
- the second term $\Delta(P/Er)$ is the Percentage Change in Share Price due to the change in the company's fundamentals combined with the change in price due to the investors' expectations on company earnings.

On the basis of Eq. (2.28), Eq. (2.26) can be re-written as follows:

$$TRS = \Delta Er + \Delta\left(\frac{P}{Er}\right) + DY \tag{2.29}$$

Technically, there is an additional cross-term that reflects the interaction of the share price change and the P/Er change, but it is generally small and therefore it can be ignored in this context.

Three are the main problems of this traditional approach (Koller et al. 2015):

- first, the manager might assume that all forms of earnings can create an equal amount of value. But different sources of growth in earnings may create different amounts of value because they are associated with different returns on capital by generating different cash flows;
- second, it suggests that the dividend yield can be increased without affecting future earnings. But dividends are merely residual and they themselves are not able to create value;
- third, the impact of leverage is not considered.

However, it is possible to break down the TRS by overcoming these limits (Koller et al. 2015). Assuming that the company is financed only by equity (E). Also, assuming that the entire wealth generated is distributed in form of dividends.

The Percentage Increase in Earning (ΔEr) can be broken down into Percent Increase in Revenues (ΔR) and in the Percent Change in Profit Margin (ΔPM), as follows:

$$\Delta Er = \Delta R + \Delta PM \tag{2.30}$$

Note that the Profit Margin is equal to the Earnings divided by Revenues ($PM = Er/R$). Therefore, it measures the weight of costs indirectly.

Technically there is an additional cross-term that reflects the interaction of these two effects, but it is small and then it can be omitted in this context allowing to focus on the key point.

The Dividend Yield (DY) is equal to the Dividend amount (Div) divided by the Share Price (SP) as defined in Eq. (2.27).

In order to simplify the analysis without any loss of significance, it is possible to assume that all revenues immediately generate cash-in and all costs immediately generate cash-out. It implies that there is no Net Working Capital (NWC). It also assumes that there are no other cash movements other than investments. Under these assumptions, it is possible to define Dividend per Share (DPS) on the basis of Earning (Er), growth rate (g_n) and ROIC as follows:

$$DPS = Er \cdot \left(1 - \frac{g_n}{ROIC}\right) \tag{2.31}$$

The Share Price (SP) can be defined on the basis of Earnings (Er) and the Price-Earning ratio as follows:

$$SP = Er \cdot \left(\frac{P}{Er}\right) \tag{2.32}$$

On the basis of Eqs. (2.31) and (2.32), Eq. (2.27) can be re-written as follows:

$$\begin{aligned} DY &= \frac{DPS}{SP} = \frac{Er \cdot \left(1 - \frac{g_n}{ROIC}\right)}{Er \cdot \left(\frac{P}{Er}\right)} = \frac{1 - \frac{g_n}{ROIC}}{\frac{P}{Er}} = \left(1 - \frac{g_n}{ROIC}\right) \cdot \left(\frac{Er}{P}\right) \\ &= \left(\frac{Er}{P}\right) - \left(\frac{g_n}{ROIC} \cdot \frac{Er}{P}\right) \end{aligned}$$

and then:

$$DY = \left(\frac{Er}{P}\right) - \left(\frac{g_n}{ROIC} \cdot \frac{Er}{P}\right) \tag{2.33}$$

where:

- the first term $\left(\frac{Er}{P}\right)$ is the inverse of the Price-to-Earnings ratio and it is usually called the Earnings Yield (EY) or Zero Growth Return. It represents the return an investor would earn if the company did not grow or improve its profit margin and if it paid out all its earnings in dividends. Its share price would remain constant;
- the second term $\left(\frac{g_n}{ROIC} \cdot \frac{Er}{P}\right)$ represents the part of its earnings yield that the company must reinvest each year to achieve its growth at its level of ROIC.

On the basis of Eq. (2.30) it refers to the Percentage Increase in Earning (ΔEr) and Eq. (2.33) refers to the Dividend Yield (DY), Eq. (2.29) can be re-written as follows:

$$TRS = \Delta R + \Delta PM + \Delta\left(\frac{P}{Er}\right) + \left(\frac{Er}{P}\right) - \left(\frac{g_n}{ROIC} \cdot \frac{Er}{P}\right)$$

that can be rearranged as follows:

$$TRS = \Delta R - \left(\frac{g_n}{ROIC} \cdot \frac{Er}{P}\right) + \Delta PM + \left(\frac{Er}{P}\right) + \Delta\left(\frac{P}{Er}\right) \tag{2.34}$$

This decomposition leads to four key drivers of TRS:

- $\Delta R - \left(\frac{g_n}{ROIC} \cdot \frac{Er}{P}\right)$: it is the value generated from revenue growth net of the capital required to grow at the company's projected ROIC;
- ΔPM: it is the impact of the change in profit margin;
- $\left(\frac{Er}{P}\right)$: it is what TRS would have been, without any growth and profit margin improvements often called the Earning Yield or Zero Growth Return;
- $\Delta\left(\frac{P}{Er}\right)$: it is the changes in shareholders' expectations on company performance, measured by the change in its (P/Er) or other multiple earnings.

Equation (2.34) shows that investor's expectations have a big effect on TRS. It is relevant to note that a company whose TRS has consistently outperformed the market will reach a point in which it will no longer be able to satisfy expectations reflected in its share price. From that point onwards, TRS will be lower than it was in the past, even if the company may still be creating value.

(b) ***Financial Debt on Net Income (FDNI)***:

$$FDNI = \frac{Financial\,Debt\,(FD)}{Net\,Income\,(NI)} \leftrightarrow FDNI = \frac{Net\,Financial\,Position\,(NFP)}{Net\,Income\,(NI)} \tag{2.35}$$

The ratio FDNI can be applied by considering the Financial Debt (FD) or the Net Financial Position (NFPD). In both case, it measures the relationship between Net Income and Financial Debt. Generally the higher the ratio, the lower the company's ability to face Financial Debt.

(c) ***Net Income on Capital Invested (NICI)***:

$$NICI = \frac{Net\,Income\,(NI)}{Capital\,Invested\,(CI)} \tag{2.36}$$

The NICI ratio defines the relationship between Net Income (NI) and the Capital Invested (CI) in the company. Therefore, it measures the return of Capital Invested on Net Income.

It is possible to measure the return of Capital Invested in Operating Assets (NOCI) on Net Income (NI). It gets the NINOI ratio as follows:

$$NINOCI = \frac{Net\,Income\,(NI)}{Net\,Operating\,Capital\,Invested\,(NOCI)} \tag{2.37}$$

It is worth noting that in all cases, the increase of ratio is not necessarily good news. Net Income being equal, the increase of the ratio is due to a reduction in the capital invested. The decrease of investments implies the reduction of future company abilities to compete in the business and, therefore a future reduction in Operating and Net Income.

(d) ***Cash-flows on Net Income (CFNI)***:

$$CFNI = \frac{Cash\,flows\,(CF)}{Net\,Income\,(NI)} \tag{2.38}$$

The CFNI ratio measures the relationship between Net Income and Cash-flows. It can be considered as an indirect and approximate measure of the company's ability to transform Net Income in Cash-flows. Indeed a high value of ratio means a good relationship between Net Income and Cash-flows; otherwise, a low value of ratio means a bad relationship between Net Income and Cash-flows.

Considering two types of cash-flows, Free Cash-flow from Operations (FCFO) and the Free Cash-flow to Equity, the ratio can be applied in two versions:

$$CFNI = \frac{CF}{NI} \rightarrow \begin{array}{l} CFNI = \frac{Free\,Cash\,flow\,from\,Operations\,(FCFO)}{Net\,Income\,(NI)} \\ CFNI = \frac{Free\,Cash\,flow\,to\,Equity\,(FCFE)}{Net\,Income\,(NI)} \end{array} \tag{2.39}$$

Section 4

The analysis of the relationships between *Capital Invested and Capital Structure*, and *Free Cash-flows from Operations and Free Cash-flow to Equity* can be focused on the following main ratios:

(a) *Equity on Cash-flows (ECF);*
(b) *Financial Debt on Cash-flows (FDCF);*
(c) *Capex on Cash-flows (CCF).*

(a) ***Equity on Cash-flows (ECF)***:

$$ECF = \frac{Equity\,(E)}{Cash\,flows\,(CF)} \tag{2.40}$$

The ratio ECF defines the relationship between Equity (E) and the Cash-flows (CF). Considering two types of cash-flows, Free Cash-flow from Operations (FCFO) and the Free Cash-flow to Equity (FCFE), the ratio can be applied in two versions:

$$ECF = \frac{E}{CF} \rightarrow \begin{array}{l} ECF = \frac{Equity\,(E)}{Free\,Cash\,flow\,from\,Operations\,(FCFO)} \\ ECF = \frac{Equity\,(E)}{Free\,Cash\,flow\,to\,Equity\,(FCFE)} \end{array} \tag{2.41}$$

(b) ***Financial Debt on Cash-flows (FDCF)***:

$$FDCF = \frac{Financial\,Debt\,(FD)}{Cash\,flows\,(CF)} \tag{2.42}$$

The ratio ECF defines the relationship between Financial Debt (FD) and the Cash-flows (CF). Considering that it is possible to use the Net Financial Position (NFP) instead of Financial Debt (FD), and by considering that there are two types of cash-flows, Free Cash-flow from Operations (FCFO) and the Free Cash-flow to Equity, the ratio can be applied in two versions:

$$FDCF = \frac{FD}{CF} \rightarrow \begin{array}{l} FDCF = \frac{Financial\,Debt\,(FD)}{Free\,Cash\,flow\,from\,Operations\,(FCFO)} \\ FDCF = \frac{Net\,Financial\,Position\,(NFP)}{Free\,Cash\,flow\,from\,Operations\,(FCFO)} \\ FDCF = \frac{Financial\,Debt\,(FD)}{Free\,Cash\,flow\,to\,Equity\,(FCFE)} \\ FDCF = \frac{Net\,Financial\,Position\,(NFP)}{Free\,Cash\,flow\,to\,Equity\,(FCFE)} \end{array} \tag{2.43}$$

(c) ***Capex on Cash-flows (CCF)***:

$$CCF = \frac{Capex\,(C)}{Cash\,flows\,(CF)} \tag{2.44}$$

The ratio CCF defines the relationship between Capex (C) and the Cash-flows (CF). Considering two types of cash-flows, Free Cash-flow from Operations (FCFO) and the Free Cash-flow to Equity (FCFE), the ratio can be applied in two versions:

$$CCF = \frac{C}{CF} \rightarrow \begin{array}{l} CCF = \frac{Capex\,(c)}{Free\,Cash\,flow\,from\,Operations\,(FCFO)} \\ ECF = \frac{Capex\,(c)}{Free\,Cash\,flow\,to\,Equity\,(FCFE)} \end{array} \tag{2.45}$$

Operating Leverage and Financial Risk Level

The analysis of Sects. 2.1–2.4 must be completed by considering the Operating Leverage and Financial Risk Level (Anthony et al. 2011; Garrison et al. 2015).

The *Operating Leverage (OL)* of the company refers to the relationship between Operating Revenues and Contribution Margin (the complete analysis is in Chap. 3). Specifically, it measures the reaction of the EBIT to the variation in the Volume of goods sold.

Specifically, it is function of the costs structure: the greater the rigidity of the costs structure, the greater the negative effects on the EBIT of the negative variations of the Operating Revenues.

The Operating Leverage (OL) can be measured on the basis of EBIT and Volume of goods sold (V) in a lead-time t_0; t_1 as follows:

$$OL = \frac{\Delta(\%)EBIT}{\Delta(\%)V} = \frac{\frac{EBIT_1 - EBIT_0}{EBIT_0}}{\frac{V_1 - V_0}{V_0}} = \frac{EBIT_1 - EBIT_0}{EBIT_0} \cdot \frac{V_0}{V_1 - V_0} \tag{2.46}$$

The EBIT can be defined on the basis of Revenue per Unit (R_U), Variable Costs per Unit (VC_U), Total Fixed Costs (FC_T) and Volume of sales (V) as follows:

$$EBIT = (R_U - VC_U) \cdot V - FC_T \tag{2.47}$$

Note that the difference between Revenues (R) and Variable Cost (VC) defines the Contribution Margin (CM). Therefore, the difference between Revenue per Unit (R_U) and Variable Costs per Unit (VC_U) defines the Contribution Margin per Unit (CM_U):

$$CM = (R - VC) \leftrightarrow CM_U = (R_U - VC_U) \rightarrow CM = CM_U \cdot V = (R_U - VC_U) \cdot V$$

On the basis of Eq. (2.47), Eq. (2.46) can be rewritten as follows:

$$\begin{aligned} OL &= \frac{[(R_U - VC_U) \cdot V_1 - FC_T] - [(R_U - VC_U) \cdot V_0 - FC_T]}{(R_U - VC_U) \cdot V_0 - FC_T} \cdot \frac{V_0}{V_1 - V_0} \\ &= \frac{(R_U - VC_U) \cdot V_1 + (R_U - VC_U) \cdot V_0}{(R_U - VC_U) \cdot V_0 - FC_T} \cdot \frac{V_0}{V_1 - V_0} = \frac{(R_U - VC_U) \cdot (V_1 - V_0)}{(R_U - VC_U) \cdot V_0 - FC_T} \cdot \frac{V_0}{V_1 - V_0} \\ &= \frac{(R_U - VC_U) \cdot V_0}{(R_U - VC_U) \cdot V_0 - FC_T} \end{aligned}$$

and by considering that:

$$CM = (R_U - VC_U) \cdot V_0$$

$$EBIT = (R_U - VC_U) \cdot V - FC_T$$

The result is:

$$OL = \frac{CM}{EBIT} \tag{2.48}$$

Note that by expressing Eq. (2.46) for EBIT, it gets:

$$\Delta(\%)EBIT = OL \cdot \Delta(\%)V \tag{2.49}$$

Equation (2.49) shows that the higher the Operating Leverage (OL), the higher the variability of the EBIT to the variations in the Volume of goods sold (V).

This Operating Leverage (OL) is due to two main elements:

– the variation of the Volume of goods sold (V);
– the cost structure rigidity to be measured on the basis of relationship between fixed and variable costs.

Considering these two effects, subsequently:

– the higher fixed costs are than variable costs, the greater the rigidity of the costs structure. In this case, variations in the Volumes of goods sold generate reductions rather than proportional of the EBIT and, all other variables being equal, of the Net Income;
– the higher the variable costs over fixed costs, the greater the flexibility of the costs structure. In this case, variations in the volume of goods sold generate reductions less then proportional of the EBIT and, all other variables being equal, of the Net Income.

Two are the constraints to be kept in mind:

– first, this analysis is relevant in a specific time range (in terms of time or productivity capacity) only. Beyond this range, all costs are variable;
– second, this analysis assumes that the price per unit, the cost variable per unit and fixed costs are fixed in the time considered. Therefore, the volume of goods sold is the only variable that can be changed.

The *Financial Risk level (FRL)* of the company refers to the relationship between the EBIT and Net Income due to the weight of financial cost of debt.

Specifically, it measures the financial risk based on the relationship between EBIT and interest on debt. The financial risk is function of the company's ability to cover the interest on debt, due to the amount of debt in the capital structure, by the operating income: the greater the amount and the risk of debt in the capital structure, the higher the interest on debt and the lower is the company ability to cover it.

Assuming that the difference between EBIT and EBT is due to the interest on debt only, the financial risk can be measured by the ratio between EBIT and EBIT less Interest on Debt on the basis of a lead-time $t_0; t_1$ as follows:

$$FRL = \frac{\Delta(\%)EBT}{\Delta(\%)EBIT} = \frac{\frac{EBT_1 - EBT_0}{EBT_0}}{\frac{EBIT_1 - EBIT_0}{EBIT_0}} = \frac{EBT_1 - EBT_0}{EBT_0} \cdot \frac{EBIT_0}{EBIT_1 - EBIT_0} \tag{2.50}$$

By considering that:

$$EBT = EBIT - I$$

and by changing, it gets:

$$FRL = \frac{(EBIT_1 - I) - (EBIT_0 - I)}{EBIT_0 - I} \cdot \frac{EBIT_0}{EBIT_1 - EBIT_0}$$
$$= \frac{EBIT_1 - EBIT_0}{EBIT_0 - I} \cdot \frac{EBIT_0}{EBIT_1 - EBIT_0} = \frac{EBIT_1 - EBIT_0}{EBIT_0 - I} \cdot \frac{EBIT_0}{EBIT_1 - EBIT_0}$$

and then:

$$FRL = \frac{EBIT}{EBIT - I} \tag{2.51}$$

Note that by assuming that there are no non-operating revenues and costs or financial revenues (so that EBIT less Interest on Debt is equal to EBT) and by assuming equal taxation so that variations in EBT generate equal variations in Net Income (NI), it gets:

$$EBIT - I = EBT = NI$$

On the basis of these assumptions, Eq. (2.51) can be re-written as follows:

$$FRL = \frac{EBIT}{NI} \tag{2.52}$$

Note that by expressing Eq. (2.50) on the basis of EBIT, it gets:

$$\Delta(\%)EBT = FRL \cdot \Delta(\%)EBIT \tag{2.53}$$

And on the basis of assumptions related Eq. (2.52), it gets:

$$\Delta(\%)NI = FRL \cdot \Delta(\%)EBIT \tag{2.54}$$

The FRL measures the effects of interest on debt on Net Income dynamic: the higher the FRL, the greater the amount of interest on debt due to leverage, the lower is the ability of EBIT to cover it and the higher the probability of the company's default.

Therefore, the greater the FRL, the higher is the variability of the Net Income to the variances of the Volume of goods sold.

2.3 Growth Rate Analysis

The fundamental analysis of the firm leads to investigate into the expected consistency of the future economic and financial dynamics. Consequently, one of the most relevant keys to the analysis is an estimate of the company's future growth rate with regards to mainly both Net Income and Operating Income. Two are the mean approaches used to estimate the company's expected growth rate:

- the first is based on an analysis of the company's fundamentals;
- the second is based on an analysis of the historical trends.

An estimate of the growth rate based on an analysis of the company's fundamentals may be more specific, rigorous and reliable than the analysis based on historical trends. Unfortunately, analysts do not normally access the company's internal information. Therefore, they often use the analysis of the historical trends and integrate them with information that they can acquire about companies that they follow.

In this context, only the first approach is considered. Therefore, the growth rate estimate of both Operating and Net Income is based on the company's fundamentals (Damodaran 2012; Koller et al. 2015).

The estimate of the company's growth rate based on the analysis of the historical trends requires specific knowledge of advanced statistical models. For their analysis refer to the reference literature.

Growth Rate in Net Income

The expected Growth Rate in Net Income (g_{NI}) is linked to the equity reinvested in the company and to the return on equity (Damodaran 2012).

An estimate of the expected growth rate in Net Income (g_{NI}), can be created based on:

- *Retention Ratio* (RR), measuring the percentage of earnings retained by the company;
- *Return on Equity* (ROE), measuring measures the return on investment in equity.

Generally, the relationship between Retention Ratio (RR), Return on Equity (ROE) and expected Growth Rate in Net Income (g_{NI}), is characterized by a direct proportion: if the company has a high Retention Ratio and a high Return on Equity, it will have a high Growth Rate in Net Income.

Considering Net Income in time t (NI_t) and in previous time (NI_{t-1}), the Growth Rate in t -time (g_{NI_t}) can be measured in simple way as follows:

$$g_{NI_t} = \frac{NI_t - NI_{t-1}}{NI_{t-1}} = \frac{NI_t}{NI_{t-1}} - 1 \qquad (2.55)$$

It is worth noting that in order to measure the Growth Rate (g_{NI_t}) with regards to a specific period time, it is possible to use the CAGR (*Compound Annual Growth*

Rate). Denote with: NI_{t_n} is the ending value of Net Income and then its value at the end of the period considered; NI_{t_0} is the beginning value of Net Income and then its value at the start of the period considered; n is the number of years of the period considered. The CAGR is equal to:

$$g_{NI_t} = \left(\frac{NI_{t_n}}{NI_{t_0}}\right)^{\left(\frac{1}{n}\right)} - 1 \tag{2.56}$$

Note that CAGR is a much more accurate measure of true growth is the past earnings when year-to-year growth has been erratic.

However both equations are characterized by problems. Equation (2.55) considers percentage changes in earnings in each period but it ignores compounding effects in net income; on the other hand, Eq. (2.56) considers the compounding effects by considering the first (beginning value) and last (ending value) observations but it ignores what goes on between the start and the end of the period considered. Both approaches contain problems if the Net Income is negative. In both cases the equations are not meaningful.

Considering that the ROE in the t-time is equal to the ratio between Net Income in the t-time (NI_t) and the book value of Equity in the previous time (E_{t-1}) and by solving for Net Income, we have:

$$ROE_t = \frac{NI_t}{E_{t-1}} \quad \rightarrow \quad NI_t = E_{t-1} \cdot ROE_t \quad \rightarrow \quad \begin{array}{l} NI_{t-1} = E_{t-2} \cdot ROE_{t-1} \\ NI_{t-2} = E_{t-3} \cdot ROE_{t-2} \\ NI_{t-n} = E_{t-n+1} \cdot ROE_{t-n} \end{array}$$

Assuming that the company does not issue new shares and it retains part of the Net Income on the basis of Retained Earnings (RE). In t-time Equity (E_t) can be defined as follows:

$$E_t = E_{t-1} + RE_t \quad \rightarrow \quad \begin{array}{l} E_{t-1} = E_{t-2} + RE_{t-1} \\ E_{t-2} = E_{t-3} + RE_{t-2} \\ E_{t-n} = E_{t-n+1} + RE_{t-n} \end{array}$$

By replacing E_{t-1} and E_{t-2} with their equations respectively, and by assuming that the ROE is constant over time $(ROE \equiv ROE_t = ROE_{t-1} = ROE_{t-n})$, the equations of the Net Income in the period t and $t-1$ can be re-written as follows:

$$NI_t = E_{t-1} \cdot ROE_t = (E_{t-2} + RE_{t-1}) \cdot ROE_t = (E_{t-2} + RE_{t-1}) \cdot ROE$$

$$NI_{t-1} = E_{t-2} \cdot ROE_{t-1} = (E_{t-3} + RE_{t-2}) \cdot ROE_{t-1} = (E_{t-3} + RE_{t-2}) \cdot ROE$$

The difference between the Net Income in the two periods t and $t-1$ is equal to:

$$
\begin{aligned}
NI_t - NI_{t-1} &= [(E_{t-2} + RE_{t-1}) \cdot ROE] - [(E_{t-3} + RE_{t-2}) \cdot ROE] \\
&= (E_{t-2} + RE_{t-1} - E_{t-3} - RE_{t-2}) \cdot ROE
\end{aligned}
$$

and by considering that $E_{t-2} = E_{t-3} + RE_{t-2}$, we have:

$$NI_t - NI_{t-1} = RE_{t-1} \cdot ROE \tag{2.57}$$

By replacing Eq. (2.57), Eq. (2.55) can be re-written as follows:

$$g_{NI_t} = \frac{NI_t - NI_{t-1}}{NI_{t-1}} = \frac{RE_{t-1} \cdot ROE}{NI_{t-1}} = \frac{RE_{t-1}}{NI_{t-1}} \cdot ROE \tag{2.58}$$

The ratio between RE_{t-1} and NI_{t-1} is defined Retention Ratio on Net Income (RR). It measures the amount of equity reinvested back into the company to finance its investments in the business. By replacing this, we have:

$$g_{NI} = RR \cdot ROE \tag{2.59}$$

Therefore, an estimate of the growth rate on Net Income (g_{NI}) can be made by estimating the expected ROE and defining the amount of the retention on Net Income.

Obviously, Eq. (2.59) can be solved by Retention Ratio (RR_{NI}) as follows:

$$RR = \frac{g_{NI}}{ROE} \tag{2.60}$$

It is worth noting that by assuming a constant Equity over time, the growth rate of Net Income (g_{NI}) it is equal to the growth of Earnings per Share (g_{ES}) as follows:

$$g_{ES} \equiv g_{NI} = RR \cdot ROE$$

If this assumption is removed, the growth in Net Income can be different from the growth in earnings per share. If the company issues new equity to finance new projects, and if the new investments increase the Net Income, the increase in the earnings per share is not the same because the shares are changed. In this case, the relationship between Net Income and earnings per share must be redesigned.

By removing the assumption about the constant value of ROE over time, a new component of the growth must be considered. This additional growth is function of the changes in ROE over time. This amount has to be added to the growth rate as previously computed as follows:

$$g_{NI} = RR \cdot ROE + \frac{(ROE_t - ROE_{t-1})}{ROE_{t-1}} \tag{2.61}$$

Two main extensions of Eq. (2.59) can be considered.

The first refers to the ROE. The ROE can be substituted in the equation by considering the joint effects of ROI and Leverage as follows:

$$ROE = [ROI + L \cdot (ROI - K_D)] \cdot (1 - t_c)$$

$$g_{NI} = g_{NI} = RR \cdot ROE = RR \cdot [[ROI + L \cdot (ROI - K_D)] \cdot (1 - t_c)] \tag{2.62}$$

In this case, the effects on Growth Rate of ROI and Leverage (L) can be analysed.

The second refers to the Retention Rate (RR). It is possible to consider the Equity Reinvestment Rate (ER) rather than the Retention Rate on Net Income (RR). Specifically, the Equity Reinvestment Rate (ER) can be defined on the basis of equity reinvested in the business in form of Capex, Net Working Capital (NWC), debt reimbursement (DR), and Net Income (NI) as follows:

$$ER = \frac{Capex + NWC + DR}{NI} \tag{2.63}$$

In this case, the growth rate of Net income, is equal to:

$$g_{NI} = ER \cdot ROE = \left(\frac{\Delta Capex + \Delta NWC + DR}{NI} \right) \cdot ROE \tag{2.64}$$

These two extensions can be considered jointly as follows:

$$\begin{aligned} g_{NI} &= ER \cdot ROE \\ &= \left(\frac{\Delta Capex + \Delta NWC + DR}{NI} \right) \cdot [[ROI + L \cdot (ROI - K_D)] \cdot (1 - t_c)] \end{aligned} \tag{2.65}$$

Growth Rate in Operating Income

The expected Growth Rate in Operating Income (g_{OI}) is connected with the reinvestments in the company's operating assets and to the Return on Capital Invested (Damodaran 2012).

An estimate of the expected growth rate in Operating Income (g_{OI}) can be fulfilled based on:

- *Investment Rate* (IR): measuring the percentage of Operating Income reinvested in the operating assets;
- *Return on Capital Invested (ROCI)*: measuring the return on investment in operating assets. The Investment Rate (IR) can be defined on the basis of investments in Capex and Net Working Capital (NWC), and the EBIT, as follows:

$$IR = \frac{Capex + NWC}{EBIT}$$

By assuming a constant ROIC over time, the expected Growth Rate in Operating Income (g_{OI}) is equal to the product between the Return on Invested Capital (ROIC) and the Investment Rate (IR) as follows:

$$g_{OI} = IR \cdot ROIC \tag{2.66}$$

Equation (2.66) can be solved by Investment Rare (IR) as follows:

$$IR = \frac{g_{OI}}{ROIC} \tag{2.67}$$

It is worth noting that the relationship between the Growth Rate in Operating Income, Investment Rate and the ROIC can be summarized as follows:

– the higher the ROIC, the lower the Investment Rate (IR) must be to achieve the defined level of Growth Rate (g_{OI});
– the lower the ROIC, the higher the Investment Rate (IR) must be to achieve the defined level of Growth Rate (g_{OI}).

By removing the assumption of constant ROIC over time, a new component of the growth must be considered. This additional growth is function of the changes in ROIC over time. This amount has to be added to the growth rate as previously computed as follows:

$$g_{OI} = IR \cdot ROIC + \left(\frac{ROI_t - ROI_{t-1}}{ROI_t}\right) \tag{2.68}$$

Therefore, the Growth Rate will be increased if the additional part is positive, and will be decreased if the additional part is negative.

In this context it is important to focus the analysis on the relationship between ROIC and the Growth Rate in Operating Income (g_{OI}) (Koller et al. 2015).

The starting point is the baseline value equation: a company only creates value if the ROIC is greater than the cost of capital invested.

Therefore, based on the baseline equation, company growth increases its value only if the difference between the ROIC and the cost of capital is positive; otherwise, the growth of the company destroys its value.

This relationship can be clearly understood by considering Dividends (Div). In order to simplify the analysis without loss of significance, it is possible to assume that Earnings (Er) can be used as dividends less a part reinvested in the operating assets of the firm as follows:

$$Div = Er \cdot (1 - IR)$$

On the basis of Eq. (2.67), we have:

$$IR = \frac{g_{OI}}{ROIC}$$

and by substituting it, we have:

$$Div = Er \cdot \left(1 - \frac{g_{OI}}{ROIC}\right) \tag{2.69}$$

Equation (2.69) shows the relationship between Growth Rate in Operating Income (g_{OI}), ROIC, Investment Rate (IR) and Dividend (Div). Specifically, it shows how the key role is played by the ROIC. In fact:

- the higher the ROIC, the lower the Investment Rate (IR) must be to achieve the defined level of Growth Rate in Operating Income (g_{OI}) and, other conditions being equal, the higher Dividends will be;
- the lower the ROIC, the higher the Investment Rate (IR) must be to achieve the defined level of Growth Rate in Operating Income (g_{OI}) and, other conditions being equal, the lower the Dividends will be.

Three main aspects must be considered:

- first, with equal capital cost, function of the capital market, the higher the dividends and the higher company value as higher are the discounted dividends;
- second, as the value creation is the difference between the Return on Capital Invested and the cost of capital, the higher the ROIC than the capital cost, the higher the value creation in the event of faster growth;
- third, strictly related to the second point, when the ROIC is lower than capital cost, faster growth destroys company value. Specifically, if ROIC is lower than cost of capital, growing faster implies higher investments at a rate of return that destroys company value. Obviously, whenever the ROIC is equal to the cost of capital, the company value is neither created nor destroyed by the growth.

The second and the third points can be summarized as follows: with other equal conditions, a high ROIC is always positive; the same cannot be said of growth because it is only good if the ROIC is higher than the cost of capital. Consequently, a company with high ROIC and low Growth Rate may have a similar or even greater valuation than a company with higher Growth Rate but low ROIC. In other words, a company with a high ROIC can increase its value by increasing its Growth Rat rather than its ROIC. Otherwise, a company with a low ROIC can generate more value by focusing on the increase of its ROIC rather than on the increase of the Growth Rate. The concept can be summarized as follows: if a company has a high ROIC it should focus on improving Growth Rate; otherwise, if a company has a low ROIC it should focus on improving ROIC before improving Growth Rate (Koller et al. 2015).

As the higher the ROIC than the cost of capital (and the longer it can sustain a rate of return on that capital greater than its cost of capital), the greater is the creation of value, therefore it is critical to understand the drivers of the ROIC and assess impacts of every strategic and investment decision on them. The drivers of the ROIC are based in the Strategic Formula of the company.

2.4 Investment Analysis

The baseline concept of present value is that a dollar received in the future is less valuable than a dollar received today.

The main techniques for the investments analysis are the Net Present Value and the Internal Rate of Return (Benninga 2014; Brealey et al. 2016; Berk and DeMarzo 2008; Damodaran 2012, 2015; Vernimmen et al. 2014; Hillier et al. 2016; Koller et al. 2015; Copeland et al. 2004).

The present value (PV) is the baseline instrument to evaluate the profitability of an investment. The PV measures the present value of the future expected cash-flows (CF_t) discounted at interest rate (i), as follows:

$$PV = \sum_{t=1}^{n} \frac{CF_t}{(1+i)^t} \tag{2.70}$$

It is possible to assume that cash-flows are expected to grow at the same constant growth rate (g_n) in each period over time, as follows:

$$PV = \frac{CF_0}{(1+i)^1} + \frac{CF_0(1+g_n)^1}{(1+i)^2} + \frac{CF_0(1+g_n)^2}{(1+i)^3} + \cdots + \frac{CF_0(1+g_n)^{n-1}}{(1+i)^n}$$

That can be generalised as follows:

$$PV = \sum_{t=1}^{\infty} \frac{CF_0(1+g_n)^{t-1}}{(1+i)^t} \tag{2.71}$$

It is important to note that when $g_n > i$ cash-flows (CF) grow faster than they are discounted, the summation differences increase over time. The sum is infinite. It means that it is impossible to reproduce the cash-flows related to constant growth rate in perpetuity. In practice this type of eternity cannot exist. Consequently, the constant Growth Rate in perpetuity can be considered only if it is lower than the interest rate $g_n < i$ so that each successive term in the sum is less than previous term. In this case the sum is finite (Corelli 2016).

The baseline equation to measure the investment capability to create value is the Net Present Value (NPV).

The NPV measures the value as a difference between the present value of future expected cash-in (CFI) less the cash-out (CFO) as follows:

$$NPV = CFO_{(t_0)} + \sum_{t=1}^{n} \frac{CFI_t}{(1+i)^t} \tag{2.72}$$

Note that *CFO* is a negative value.

If the investment requires plus then one cash-out in different future years, Eq. (2.72) can be re-written as follows:

$$NPV = \sum_{t=0}^{n} \frac{CFO_t}{\left(1+R_f\right)^t} + \sum_{t=1}^{n} \frac{CFI_t}{(1+i)^t} \tag{2.73}$$

The Interest Rate (i) uses the time value of money through the risk-free rate $\left(R_f\right)$ to discount cash-flows measures, plus the premium risk related to the riskiness of investment (P_r) related to effective execution of the future expected cash-in.

However, it has a different meaning with regards to cash-out and cash-in:

- cash-out are future but they are not expected. Therefore, they are not estimated but certain in their amount and time. Consequently, the interest rate used to discount them only measures the time value of money $\left(i = R_f\right)$;
- cash-in are future and expected. Therefore, they are estimated and then uncertain in their effective amount. Consequently, the interest rate used to discount them measures the time value of money plus the premium risk $\left(i = R_f + P_r\right)$.

It is relevant to note that the risk related to future expected cash-in can be considered on the basis of two different and alternative approaches:

- the risk is considered in cash-in: in this case the expected future cash-in are "adjusted" on the basis of the probability of their realization and the interest rate measures only the time value of money $\left(i = R_f\right)$;
- the risk is considered in interest rate: in this case the expected future cash-in is not adjusted for the probability of their achievement and the interest rate measure the time value of money plus the premium-risk $\left(i = R_f + P_r\right)$.

Note that the NPV is an amount. Consequently if it is a positive number, the investment is profitable and its profitability is equal to the positive amount of NPV; otherwise, if it is a negative number, the investment is not profitable and its loss is equal to the negative amount of NPV. Obviously the higher the NPV, the higher the investment profitability.

The *Interest Rate of Return (IRR)*, also called Rate of Return (ROR) or Effective Interest Rate (EIR), is the rate at which it is assumed to discount the future cash-flows to obtain the initial cash-out. Therefore, the IRR measures the profitability of the investment on the basis of its initial costs (cash-out) and able to generate future cash-flows that are reinvested at the same rate. It implies the same interest rate for all periods considered.

Therefore, IRR does not incorporate environmental factors, but it only considers the cash-flows generated by the investment.

In general terms, the IRR of an investment measures the annualized effective compounded return rate that makes Net Present Value (NPV) of investment equal to zero: it is the discount rate that makes the Net Present Value of cash-out of the investment equal to the Net Present Value of the cash-in of the investment. Formally:

$$NPV = 0 \quad \rightarrow \quad \sum_{t=1}^{n} \frac{CFO_t}{(1+IRR)^t} = \sum_{t=1}^{n} \frac{CFI_t}{(1+IRR)^t} \tag{2.74}$$

Based on Eq. (2.74) the higher a IRR, the more profitable the investment.

Note that IRR is an indicator of efficiency, quality and yield of investment, as opposed to NPV, which refers more to the value and magnitude of an investment.

By considering the IRR and the cost of capital, the firm has to prefer available investment opportunities where the expected IRR exceeds the cost of capital. Therefore, if the IRR is greater than the cost of capital, the expected return of investment exceeds the investors' expectation.

Assuming an investment project that requires an initial cash-out equal to A and several cash-in C_n over time, we have:

$$A = \sum_{t=1}^{n} \frac{C_t}{(1+i)^t} \quad \rightarrow \quad A - \sum_{t=1}^{n} \frac{C_t}{(1+i)^t} = 0$$

The main problem is to define the value of Interest Rate i so that this equation is achieved. Therefore, the Interest Rate i is the IRR.

The problem can be solved by using the Newton's algorithm (Sydsaeter et al. 2012; Cesari 2012).

Assuming that the investment project requires a cash-out equal to A and it promises three cash-in C_1, C_2, C_3 respectively in t_1, t_2, t_3. In this case, we have:

$$A = \frac{C_1}{(1+i)^1} + \frac{C_2}{(1+i)^2} + \frac{C_3}{(1+i)^3} \quad \rightarrow \quad A - \frac{C_1}{(1+i)^1} + \frac{C_2}{(1+i)^2} + \frac{C_3}{(1+i)^3} = 0$$

Defining a new variable α as follows:

$$\alpha = \frac{1}{1+i}$$

so that:

$$A = C_1 \cdot \alpha + C_2 \cdot \alpha^2 + C_3 \cdot \alpha^3 \quad \rightarrow \quad A - C_1 \cdot \alpha + C_2 \cdot \alpha^2 + C_3 \cdot \alpha^3 = 0$$

It is possible to define the function $F(\alpha)$ as follows:

$$F(\alpha) = A - C_1 \cdot \alpha + C_2 \cdot \alpha^2 + C_3 \cdot \alpha^3 = 0 \tag{2.75}$$

Generally, a polynomial of degree n can have n solutions.

By using the Cartesio's theorem (if in a polynomial of degree n the coefficient signs change one time, than the polynomial has one solution real and positive also if it is of degree n), it is possible to say that IRR exists and is one.

Now the problem is to find this value of IRR. The Newton algorithm can be used as follows. The objective is to find the zero of the function $F(\alpha)$ and therefore to find the value of α so that the function is equal to zero.

Calculate the first derivative of the function $F(\alpha)$ respect to the variable α as following:

$$\frac{\partial F(\alpha)}{\partial \alpha} = -C_1 - 2 \cdot C_2 \cdot \alpha - 3 \cdot C_3 \cdot \alpha^2$$

or in equivalent form (by denoting the first derivative as $\frac{\partial F(\alpha)}{\partial \alpha} = F'(\alpha)$), we have:

$$F'(\alpha) = -C_1 - 2 \cdot C_2 \cdot \alpha - 3 \cdot C_3 \cdot \alpha^2 \tag{2.76}$$

The analytical first derivative in a generic point α_s can be approximated to the incremental ratio as follows:

$$\frac{\partial F(\alpha_s)}{\partial \alpha_s} \simeq \frac{F(\alpha_{s+1}) - F(\alpha_s)}{\alpha_{s+1} - \alpha_s}$$

and therefore:

$$\lim_{\alpha_{s+1} \rightarrow \alpha_s} \frac{F(\alpha_{s+1}) - F(\alpha_s)}{\alpha_{s+1} - \alpha_s} = \frac{F(\alpha_s)}{\alpha_s} = \frac{\partial F(\alpha_s)}{\partial \alpha_s}$$

We have to find the Interest Rate able to have the function $F(\alpha)$ equal to zero. Therefore, it is possible to assume $F(\alpha_{s+1}) = 0$ that is the objective value. In this case, it gets:

$$\frac{\partial F(\alpha_s)}{\partial \alpha_s} = \frac{-F(\alpha_s)}{\alpha_{s+1} - \alpha_s}$$

and by solving for α_{s+1} we have:

$$\alpha_{s+1} = \alpha_s - \frac{F(\alpha_s)}{\frac{\partial F(\alpha_s)}{\partial \alpha_s}}$$

by denoting the first derivative as $\frac{\partial F(\alpha_s)}{\partial \alpha_s} = F'(\alpha_s)$, we have:

$$\alpha_{s+1} = \alpha_s - \frac{F(\alpha_s)}{F'(\alpha_s)} \tag{2.77}$$

Equation (2.77) defines the relationship between α_{s+1} and α_s. It is an numerical algorithm able to define interactively the value of IRR which makes the function $F(\alpha) = 0$.

On the basis of the equations of $F(\alpha), F'(\alpha), \alpha_{s+1}$ it is possible to define the following procedure (Cesari 2012; Sydsaeter et al. 2012):

- *step 1*: assign to a α any arbitrary value so that $\alpha = \alpha_1$, so that:

$$\alpha_1 = \frac{1}{1 + i_1}$$

In this case it is a defined value (a number). Therefore, it is possible to calculate the Interest Rate i_1 that it gives the defined value of α_1:

$$i_1 = \frac{1}{\alpha_1} - 1$$

- *step 2*: introduce the value α_1 in the function $F(\alpha)$ and verify if it reaches a value near to zero $F(\alpha) \simeq 0$. In this sense, it is possible to introduce a convergence criterion. Specifically, if:

$$|F(\alpha)| < \varepsilon \quad \rightarrow \quad F(\alpha) \simeq 0$$

where ε is a positive value small enough to make that the $F(\alpha)$ is near to zero.

Therefore, by introducing the value of α_1 in the function, if $|F(\alpha_1)| < \varepsilon$ the value of α_1 satisfies the convergence criterion and therefore $F(\alpha_1) \simeq 0$. Therefore, the corresponding Interest Rate i_1 is the IRR sought and the procedure is closed. If this does not happen, the procedure moves into step 3.

- *step 3*: calculate the first derivative in the point α_1 as follows:

$$F'(\alpha_1) = -C_1 - 2 \cdot C_2 \cdot \alpha_1 - 3 \cdot C_3 \cdot \alpha_1^2$$

– *step 4*: calculate the new value α_2 on the basis of the α_1 as follows:

$$\alpha_2 = \alpha_1 - \frac{F(\alpha_1)}{F'(\alpha_1)}$$

– *step 5*: introduce the value α_2 in the function $F(\alpha)$ and if the convergence criterion is respected, so that:

$$|F(\alpha_2)| < \varepsilon \quad \rightarrow \quad F(\alpha_2) \simeq 0$$

the α_2 is correct and the corresponding Interest Rate i_2 is the IRR sought and the procedure is closed. If it does not happen, the procedure starts again from step 3 with the definition of the first derivative of the function with α_2 $F'(\alpha_2)$, and the new definition of the α_3 in step 4 and its validation in step 5. The procedure continues interactively until the convergence criterion is respected in step 5.

References

Anthony RN, Hawkins DF, Merchant KA (2011) Accounting: text and cases, 13th edn. McGraw-Hill, New York

Benninga S (2014) Financial modeling, 4th edn. MIT Press, Cambridge

Berk J, DeMarzo P (2008) Corporate finance. Pearson Education, Inc

Brealey RA, Myers SC, Allen F (2016) Principles of corporate finance, 12th edn. McGraw-Hill, New York

Cesari R (2012) Introduzione alla finanza matematica. Concetti di base, tassi e obbligazioni, 2nd edn. McGraw-Hill, New York

Copeland T, Weston F, Shastri K (2004) Financial theory and corporate policy, 4th edn. Addison-Wesley, Reading, MA

Corelli A (2016) Analytical corporate finance. Springer, Berlin

Damodaran A (2012) Investment valuation: tools and techniques for a determining the value of any assets, 3rd edn. Wiley, New Jersey

Damodaran A (2015) Applied corporate finance, 4th edn. Wiley, New Jersey

Fuller RJ, Farrell JL (1987) Modern investments and security analysis. McGraw-Hill, Inc

Garrison RH, Noreen EW, Brewer PC (2015) Managerial accounting, 15th edn. McGraw-Hill, New York

Hillier D, Ross S, Westerfield R, Jaffe J, Jordan B (2016) Corporate finance, 3rd edn. McGraw-Hill, New York

Koller T, Goedhart M, Wessels D (2015) Valuation, 6th edn. Wiley, New Jersey

Ross SA, Westerfield R, Jaffe J, Jordan BD (2015) Corporate finance, 11th edn. McGraw-Hill, New York

Silvi R (2012) Analisi di bilancio: la prospettiva manageriale. McGraw-Hill, New York

Sydsaeter K, Hammond P, Strom A (2012) Essential mathematics for economic analysis, 4th edn. Pearson, London

Vernimmen P, Quiry P, Dallocchio M, Le Fur Y, Salvi A (2014) Corporate finance: theory and practice. Wiley, New Jersey

Chapter 3
Product Profitability Analysis

Abstract In this third step of the fundamental company analysis, attention is placed on product profitability. Company ability to create profit over time is function of product profitability. An analysis of product capability to create profit is not easy because it requires good knowledge of product cost in every time of company life. In this context product cost is analysed on the basis of two main approaches:

– direct product cost: it is based on the difference between variable and fixed costs;
– full product cost: it is based on the difference between direct costs and indirect costs.

The definition of the standard product cost is a part of the problem. The other part is the analysis includes the actual product cost as function of the actual company cost. Therefore, the last part of the chapter focuses on the difference between budget and actual Net Income on the basis of the variance analysis.

3.1 Direct Cost of Product

Profitability of the company is function of product profitability. Therefore, to further understand the company's ability to create profit over time it is necessary to investigate into product profitability.

In each period of company life, the right definition and knowledge of the product cost is one of the most relevant success factors. It is fundamental for pricing of the product and subsequently for the right measurement of product profitability and, in general terms, for company profitability (Anthony et al. 2011; Garrison et al. 2014; Atrill and McLaney 2018; Banker et al. 2000; Bhimani et al. 2015; Drury 2016; Kaplan and Atkinson 1998; Sahaf 2013; Seal and Rohde 2014; Tennent 2014).

Therefore, with the correct measurement and knowledge in each period the product cost allows us to answer some fundamental questions such as: how much

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_3

did the product cost? What is the right price on the basis of production costs on the one hand, and market competition on the other? What is product profitability?

Consequently, the correct measurement of product cost is a fundamental for correct understanding of product profitability and company profitability over time.

The word "cost" is one of the most commonly used in accounting management. The most relevant problem is that the clear identification of cost requires a clear explanation. In fact, the word cost becomes more meaningful when it is preceded by a modifier in phrases such as direct cost, full cost, opportunity cost, differential cost, etc.

Before starting the analysis, it can be useful to define cost: cost is a measurement in monetary terms of the amount of resources used for a specific purpose (Anthony et al. 2011). Therefore, the cost requires three main structural elements:

- first, cost measures the use of resources in monetary terms. Cost then measures how many resources are used;
- second, cost is always expressed in monetary terms. Money represents the common denominator allowing the total amount of different resources, each of which is measured on the basis of its own scale. Therefore, thanks to money the total amount of the different resources used can be determined;
- third, cost must be defined always with regards to a well-defined cost object. The cost object can be anything for which monetary measurement of resources used is desired such as a product, a department, an activity, a project, a cost center, etc.

The correct determination of product margin requires the correct determination of product cost. It seems to be simple to define the product cost. Unfortunately, this simplicity is only apparent.

Usually, the wrong determination of product cost is one of the main elements of the wrong prediction of product margin and, on a general level, company profitability.

Before analysing the technique of product cost determination, it is necessary to introduce the different categories of costs.

It is worth noting that cost does not have a nature of its own with regards to its behaviour. Its classification with regards to its behaviour is function of the driver used for the analysis. In this sense, it is possible to define two main criteria of cost classification:

- Volume: costs and volume relationship;
- Cost Object: costs and product relationship.

Costs and Volume

In this case the driver used for the analysis of costs is volume. Therefore, the relationship between costs and volume is analysed.

The first question to be analysed is the measurement of volume. It must always be specified. It normally measures the level of activity. It can be defined on the basis of the *volume of production*, and therefore on the number of products produced, or *volume of sale* and therefore the number of products sold. Normally production volume is considered rather than sales volume.

Based on volume, costs can be distinguished as follows (Anthony et al. 2011):

- variable costs:
- fixed costs:
- semi-variable costs.

Variable Cost

The *Variable cost* is an item of cost that varies, in total, directly and proportionately with volume but the variable cost per unit of volume remains constant. Therefore, the total cost is variable because the cost per unit of volume remains constant (Anthony et al. 2011).

Denoting with VC_T the Total Variable Cost, VC_U the Variable Cost per Unit of Volume, V the Volume, it gets:

$$VC_T = VC_U \cdot V \tag{3.1}$$

Equation (3.1) shows that VC_T increases with the increase of V because VC_U remains constant over time. This relationship can be represented as in Fig. 3.1.

Variable costs are normally considered:

- the cost of materials, semi-products and products necessary for the products sold;
- the cost of services for external processing;
- cost of commercial services connected directly with volume.

Other costs can be considered variable only if there is a direct relationship with the volume of products created.

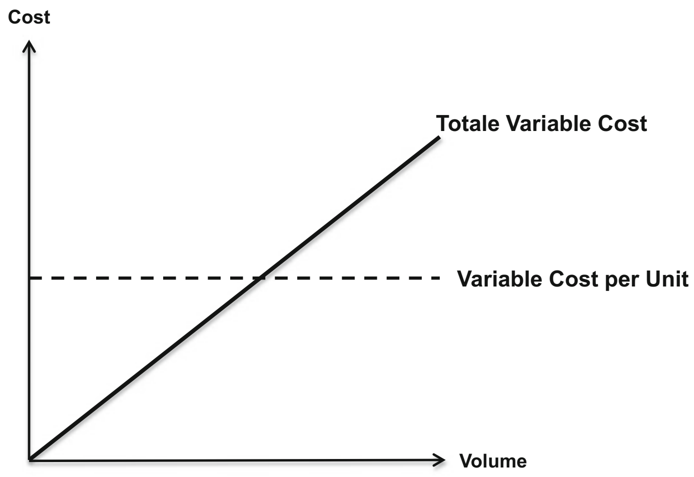


Fig. 3.1 Total variable cost and variable cost per unit

Sometimes, in specific cases, salaries of direct labour can be considered as variable costs. The direct relationship between the amount of salaries and the volume of products created is necessary. Based on these constraints, labour costs are usually considered as fixed costs.

Fixed Costs

The *Fixed cost* is an item of costs that, in total, does not change with volume. Consequently, the fixed cost per unit of volume decreases (increases) with the increase (decrease) of volume (Anthony et al. 2011).

There are two main constraints to be kept in mind:

- first, the cost is fixed with regards to volume only. It does not change with changes in volume. Otherwise, it changes for other reasons and specifically it changes in the case of management decisions on the item referenced;
- second, the cost is fixed within a defined production capacity and therefore within a defined volume of production. Out of the range, increases in volume require more production capacity, whose implementation increases fixed costs. Therefore, on the long term, assuming changes in the product capacity, all costs are variable.

Therefore, within a defined production capacity $(V \leq V^*)$ the total fixed cost remains constant while the fixed cost per unit decreases according to increases in volume.

Denoting with FC_T the total fixed cost, FC_U the fixed cost per unit of volume, V the volume, we have:

$$FC_U = \frac{FC_T}{V} \quad for\, V \leq V^* \tag{3.2}$$

Equation (3.2) shows how the fixed cost per unit of volume decreases (increase) to the increase (decrease) of volume. This relationship can be represented as in Fig. 3.2.

It is worth noting that by increasing the volume of product over a defined level (V^*) the total fixed cost varies. The volume V^* represents the maximum number of products that can be achieved with the assets in place. It subsequently represents the maximum level of production capacity. Consequently, if the company increases volume within the defined production capacity $(V \leq V^*)$, the total fixed cost remains constant and the fixed cost per unit decreases.

If the company wants to increase the volume of production over the level V^* $(V > V^*)$ it needs an increase in productivity capacity by the new assets and an increase in the amount of total fixed costs. Therefore, the difference between variable and fixed costs can only be achieved within a defined productivity capacity level as function of assets in place.

In this context fixed costs are considered all of those not related with product volume achieved or sales volumes such as administrative services expenses,

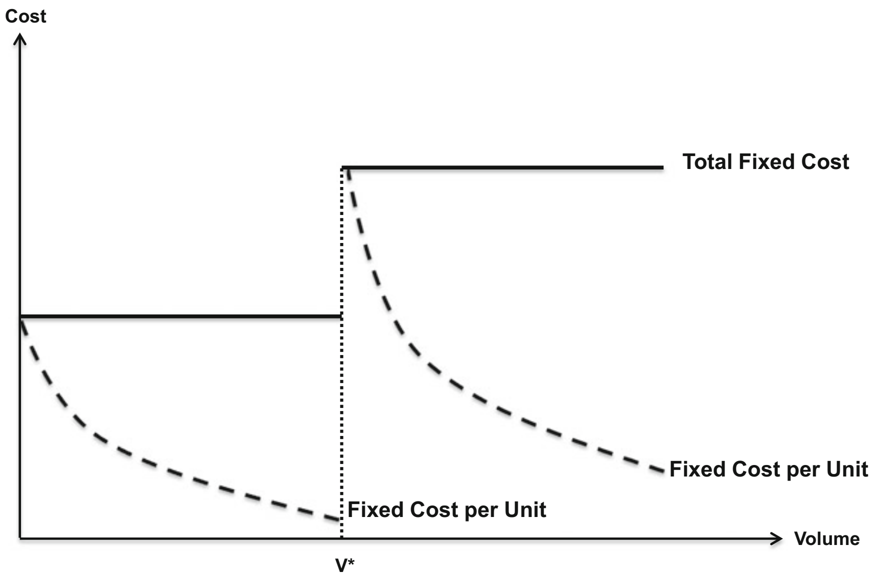


Fig. 3.2 Total fixed cost and fixed cost per unit

industrial and services expenses, leasing and rent expenses, all other costs for operating activities, amortizations and depreciations.

(c) **Semi-variable costs**

Unfortunately, it is not always easy to distinguish between variable and fixed costs. There are a lot of costs that can be defined as *Semi-variable costs* (or semi-fixed, partly variable, mixed costs). These costs are characterized by a combination of variable and fixed costs. It does not mean that they are divided exactly in half. The total amount of the semi-variable costs varies in the same direction, but less than proportionately with, changes in the volume of products achieved (Anthony et al. 2011). The dynamics of Semi-variable cost can be represented as shown in Fig. 3.3.

The relationship between volume and costs can be shown in a Cost-Volume Diagram. By considering all fixed or variable costs (by assuming that the semi-variable costs can be split into fixed and variable components) they can be illustrated approximately in two straight lines, with a linear approximation used to describe the relationship between volume and costs.

The general equation of a straight-line is the following:

$$y = a + bx$$

In our case: x, is the volume of products; y, is the total cost at a volume x; a, is the vertical intercept and then the fixed cost; b, is the slope and then the rate of cost change per unit of volume change. Therefore, it is the unit variable cost.

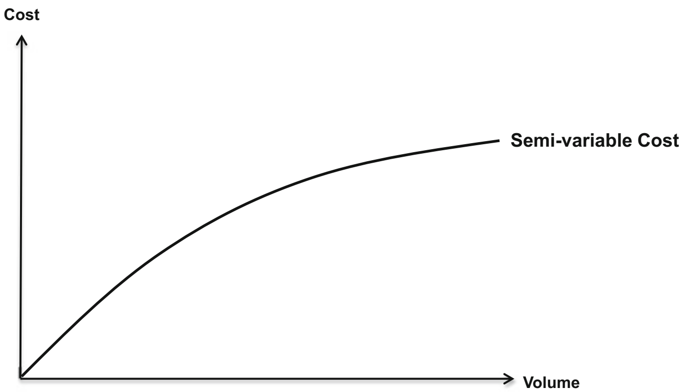


Fig. 3.3 Semi-variable costs

Denoting with: $x = V; y = TC_{(V)}; a = FC_T; b = VC_U$ and substituting, we have:

$$TC_{(V)} = FC_T + (VC_U \cdot V) \tag{3.3}$$

On an aggregate level, by considering that VC_T is equal to:

$$VC_T = VC_U \cdot V$$

Equation (3.3) can be rewritten as follows:

$$TC_{(V)} = FC_T + VC_T \tag{3.4}$$

Equation (3.4) shows how the total cost at volume V $(TC_{(V)})$ is equal to the total fixed costs (FC_T) plus the total variable costs (VC_T).

Equation (3.3) can be represented as in Fig. 3.4.

It is worth noting that the straight line of total costs in Fig. 3.4. starts from the vertical intercept defined by the total fixed cost and is parallel to the straight line of total variable costs because they have the same slope.

By considering that the variable cost per unit is constant and the fixed cost per unit decreases according to the increase in volume within a defined production capacity $(V \leq V^*)$, the total average cost per unit of volume is equal to the total cost divided per volume and therefore it decreases according to the increase in volume. Therefore, the total average cost per unit of volume behaves quite differently than total costs due to the different behaviour of fixed costs per unit and variable costs per unit (Anthony et al. 2011).

More specifically, when volume increases within a defined production capacity $(V \leq V^*)$, the total variable costs increase while the total fixed cost remains constant.

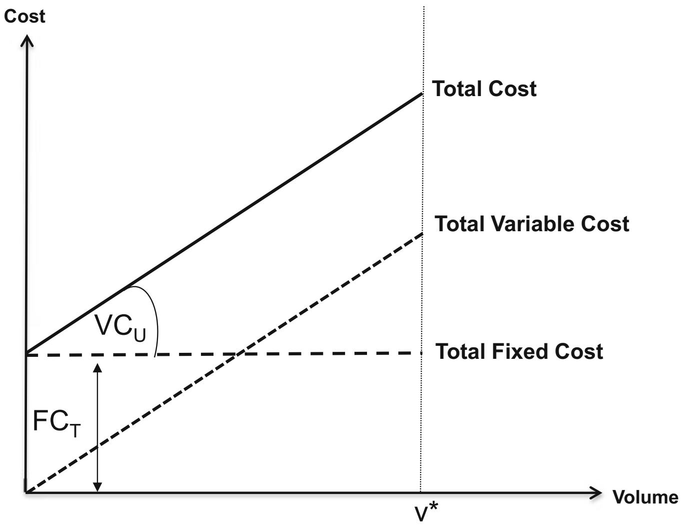


Fig. 3.4 Fixed cost, variable cost and total cost

Denoting TC_U the total average cost per unit of volume, FC_U the fixed cost per unit of volume, VC_U the variable cost per unit of volume, we have:

$$TC_U = FC_U + VC_U \tag{3.5}$$

The variable cost per unit of volume (VC_U) is constant. On the other hand, the total average cost per unit of volume (TC_U) and the fixed cost per unit of volume (FC_U) are function of Volume V. Specifically, TC_U is equal to the total cost at a volume V $(TC_{(V)})$ divided by volume V, as well as FC_U is equal to total fixed cost (FC_T) divided the volume V, as follows:

$$TC_U = \frac{TC_{(V)}}{V} \tag{3.6}$$

and

$$FC_U = \frac{FC_T}{V} \tag{3.7}$$

On the basis of Eqs. (3.6) and (3.7), the (3.5) can be re-written as follows:

$$\frac{TC_{(V)}}{V} = \frac{FC_T}{V} + VC_U \leftrightarrow TC_U = \frac{FC_T}{V} + VC_U \tag{3.8}$$

Therefore, with an increase of volume (V) within a defined prediction capacity $(V \leq V^*)$ the total cost per unit of volume (TC_U) decreases with the increase of

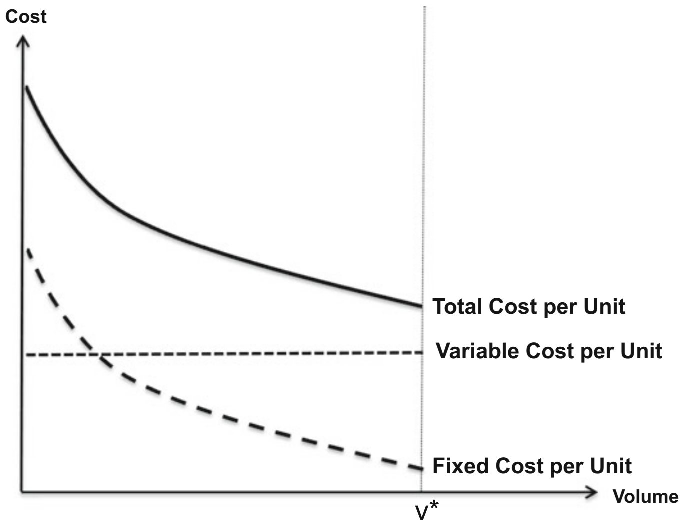


Fig. 3.5 Total cost per unit of volume

volume (V) because the fixed cost per unit of volume (FC_U) decreases with the increase of volume (V) while the variable cost per unit of volume (VC_U) remains constant. These relationships can be represented as in Fig. 3.5.

Figure 3.5 shows how the total average cost per unit of volume (TC_U) is parallel to the fixed cost per unit (FC_U). In fact, the reduction of the total cost per unit is due to the reduction of the fixed cost per unit.

The variable cost per unit (VC_U) represents a constant and therefore it is independent of the volume (V) and then increases the starting point of the total cost per unit of volume. By expanding the production to the volume V^* and assuming that the level of volume V^* is infinite, the fixed cost per unit is equal to zero, the total average cost per unit of volume is equal to the variable cost per unit of volume, as follows:

$$\lim_{V \to V^* = +\infty} TC_U = \lim_{V \to V^* = +\infty} \left[\frac{FC_T}{V} + VC_U \right] = VC_U \tag{3.9}$$

In a real world V^* cannot be infinite. Therefore, we have:

$$\lim_{V \to V^*} TC_U = \lim_{V \to V^*} \left[\frac{FC_T}{V} + VC_U \right] = \alpha + VC_U \tag{3.10}$$

where the value of α is the fixed cost per unit of volume for use of the maximum production capacity. Therefore, it is function of the quality and characteristics of the production system and managerial skills.

Therefore, as the volume increases within a defined production capacity $(V \leq V^*)$ there is an inverse relationship between total fixed costs and fixed costs per unit of volume, as well as between total variable costs and variable costs per unit of volume. Specifically, to the increase of volume we have:

– the total fixed cost remains constant, while the fixed cost per unit of volume reduces;
– the total variable cost increases, while the variable cost per unit of volume remains constant.

These movements imply that if the volume increases the total cost increases while the total average cost per unit of volume decreases.

The total average cost per unit of volume has relevant managerial implications. It generates the cost per unit of volume by distinguishing all fixed and variable costs. Therefore, it plays a key role in the product price decision. The difference between revenues and cost per unit of volume defines the margin per unit of volume. By multiplying the margin per unit of volume per volume of product sold, we have profit (loss) of the company.

Denoting with R_U the revenues per unit of volume, TC_U the total average cost per unit of volume, M_U the profit margin per unit of volume, V the volume of product, M_T the total margin of the company and therefore the profit (if it is positive) or loss (if it is negative) of the company, we have:

$$M_U = R_U - TC_U \tag{3.11}$$

and then:

$$M_U \cdot V = R_U \cdot V - TC_U \cdot V \leftrightarrow M_U \cdot V = (R_U - TC_U) \cdot V \tag{3.12}$$

Considering that the margin per unit of volume is the total margin (M_T), the revenues per unit of volume for volume are the total revenues (R_T) and the total average cost per unit of volume for volume is total costs (C_T), we have:

$$M_T = R_T - C_T \rightarrow \begin{array}{l} M_T = \mathit{Profit} \quad \mathit{if}\ R_T > C_T \rightarrow R_U > TC_U \\ M_T = \mathit{Loss} \quad \mathit{if}\ C_T > R_T \rightarrow\ R_U < TC_U \end{array} \tag{3.13}$$

The relationship between volume and costs as defined, is based on several conditions. They are truly restrictive and they must always be kept in mind to avoid misunderstandings and errors in decision processes. There are five main ones (Anthony et al. 2011):

(1) *Range of volume*: costs move along a straight-line only within a defined range of volume. Therefore, a single straight line gives a good approximation of the cost behaviour only within a defined and well specified range of volume that defines the Relevant Range. Only the relevant range must be considered in the analysis.

(2) *Length of the time period*: the amount of variable costs is function of the time period over which cost behaviour is being estimated. If the time period is only one day, few costs can be defined as variable. If the time period is one year, many more costs are variable. If the time period is infinite, all costs are variable. Therefore, the time period must be always defined and the relevant time period must be defined. Only the relevant time period must be considered in the analysis. The relevant time period is normally one year.
(3) *Stickiness of costs*: very few costs can be defined as truly variable costs. Usually they decrease or increase together with volume decreases and they increase more or less when volume increases. Also, they change more or less within a specific period of volume. These are defined as Sticky Costs (Anderson et al. 2003). Generally, the costs that are incurred at manager's discretion are stickier. The degree of stickiness varies across cost types and company types.
(4) *Environment*: it is assumed that cost varies with volume only. All other variables are considered constant. The many changes due to the economic environment are not considered.
(5) *Linear assumption*: the relationship between volume and costs assumes that costs change with volume along a straight line. In practice this assumption is often not valid because the costs move along a curve rather than a straight line. This assumption can be used if in the relevant range and relevant time period considered, the curve can be approximated by a straight line. Also step function costs are more common than these curvilinear costs.

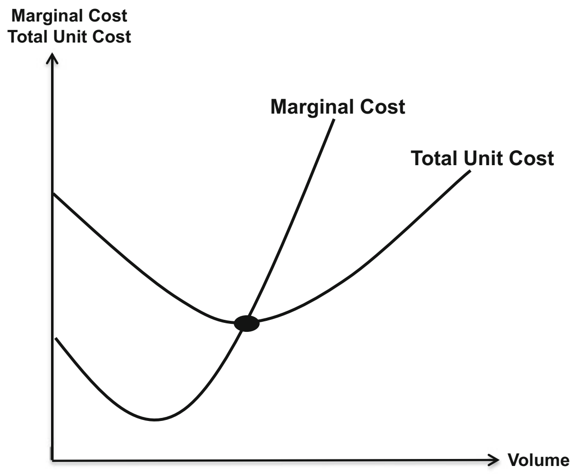
We have defined the total cost as the sum of total fixed costs and total variable costs. Similarly, the unit cost of a product is equal to the sum of unit variable costs and unit fixed costs. Since the unit variable cost remains constant to the increase in volume, while the unit fixed cost decreases to the increase of volume, to the increase of volume the unit cost per product decreases.

For a deep analysis, the marginal cost must be considered for the total cost of products and the unit cost of products: it is a baseline concept in economics and it measures the increase in total costs due to the increase in production of one extra unit. Therefore, the marginal cost is the cost of any additional unit of product to be produced.

A relevant relationship can be defined between the marginal cost and total unit cost:

- if the marginal cost for one additional unit is less than total unit cost, then the total unit cost decreases;
- if the marginal cost for one additional unit is higher than total unit cost, then the total unit cost increases;
- if the marginal cost for one additional unit is equal to the total unit cost, the curve in that point is flat.

Fig. 3.6 Relationship between marginal cost and total unit cost



On the long term, and therefore for different levels of production, the relationship between marginal costs (C_M) and total unit cost (TC_U) can be represented as in Fig. 3.6.

The most relevant indicator based on the relationship between volume and costs is the Contribution Margin $(\mathrm{CM\,or\,CM_T})$ and the Unit Contribution Margin $(\mathrm{CM_U})$.

It measures the rigidity/flexibility level of the company's cost structure. It can focus on the Operating Income or on Net Income. In the first case only operating revenues and costs are considered; in the second case, all revenues and costs are considered as indicated in Table 3.1.

The Contribution Margin is based on the difference between fixed and variable costs. The difference between revenues and total variable costs defines the Total Contribution Margin, while the difference between the revenue per unit of volume and the variable cost per unit of volume defines the Unit Contribution Margin.

The Contribution Margin measures the company's ability to cover the fixed costs.

The unit contribution margin $(\mathrm{CM_U})$ or contribution margin per unit of volume, measures the difference between the revenue per unit of volume (R_U) and variable cost per unit of volume (VC_U). Formally:

$$CM_U = R_U - VC_U \tag{3.14}$$

It is important to point out that the unit contribution margin remains constant per unit of volume (V) because the variable cost per unit of volume is constant.

Table 3.1 Contribution margin

Contribution margin analysis on net income	Contribution margin analysis on operating income
Net operating revenues	***Net operating revenues***
Financial revenues	
Non-current operating revenues	
Non-operating revenues	
Revenues	***Revenues***
(Variable costs of materials)	(Variable costs of materials)
(Variable costs of industrial services)	(Variable costs of industrial services)
(Variable costs of commercial and distribution services)	(Variable costs of commercial and distribution services)
Variable costs	***Variable costs***
Contribution margin (CM)	**Contribution margin (CM)**
(Fixed costs of employees)	(Fixed costs of employees)
(Fixed costs of administration)	(Fixed costs of administration)
(Fixed costs of commercial and distribution services)	(Fixed costs of commercial and distribution services)
(Fixed costs of industrial services)	(Fixed costs of industrial services)
(Fixed costs of leasing and rent)	(Fixed costs of leasing and rent)
(Other operating fixed costs)	(Other operating fixed costs)
Operating fixed costs (cash)	***Operating fixed costs (cash)***
(Amortization of intangible assets)	(Amortization of intangible assets)
(Depreciation of tangible assets)	(Depreciation of tangible assets)
(Accruals for employees)	(Accruals for employees)
(Accruals for provision for risks, charges and taxes)	(Accruals for provision for risks, charges and taxes)
(Impairment of assets)	(Impairment of assets)
Operating fixed costs (non-cash)	***Operating fixed costs (non-cash)***
(Financial fixed costs)	
(Non-current operating fixed costs)	
(Non-operating fixed costs)	
(Operating and corporate taxes)	
Net income	**EBIT**

Consequently, the contribution margin (CM) varies according to volume changes because the unit contribution margin (CM_U) remains constant per unit of volume. Formally:

$$CM_T = CM_U \cdot V \leftrightarrow CM_T = (R_U - VC_U) \cdot V \tag{3.15}$$

By distinguishing costs between variable and fixed costs, the marginal income (MI) can be computed based on the unit contribution margin as follows:

$$MI = CM_U \cdot V - FC_T \leftrightarrow MI = (R_U - VC_U) \cdot V - FC_T \tag{3.16}$$

It is worth noting that if only operating revenues and costs are considered, the marginal income is the EBIT; otherwise, if all revenues and costs are considered, the marginal income is the Net Income.

The analytical scheme of the Contribution Margin is commonly used by both internal and external analysts because it allows for greater understanding of the effects of the Operating Revenue changes on Operating Income and therefore on Net Income due to the rigidity level of the cost structure.

The information required is usually confidential and arising from management accounts. In any case, the external analyst can use a proxy based on the characteristics of cost categories.

The relationship between volume and costs can be used to define the break-even point. It defines the equilibrium-volume at which the total costs are equal to the total revenues. Therefore, for a volume higher than the equilibrium-volume, the company generates profit; otherwise, for a volume lower than equilibrium-volume, the company generate loss. The amount of profit or loss expected at any volume level is the vertical distance between the points on the total cost and total revenue lines at that volume (Anthony et al. 2011).

Denote with V the volume, R_T the total revenue at any volume, C_T the total cost at any volume, P_U the unit price, VC_U the variable cost per unit of volume, FC_T the total fixed costs.

The equilibrium-volume is the volume at which the total revenues are equal to the total costs, and therefore:

$$R_T = C_T \quad \rightarrow \begin{matrix} R_T = P_U \cdot V \\ C_T = FC_T + (VC_U \cdot V) \end{matrix} \rightarrow P_U \cdot V = FC_T + (VC_U \cdot V)$$

and consequently:

$$V_{(E)} = \frac{FC_T}{P_U - VC_U} \tag{3.17}$$

Equation (3.17) shows how the equilibrium-volume $(V_{(E)})$ is equal to the total fixed costs (FC_T) divided for the unit contribution margin $(P_U - VC_U)$.

On the basis of Eq. (3.17) the volume achieved (V) must be compared with the equilibrium-volume $(V_{(E)})$ as follows:

$$\begin{aligned} V = V_{(E)} &\rightarrow R_T = C_T \; \textit{Equilibrium} \\ V > V_{(E)} &\rightarrow R_T > C_T \; \textit{Profit} \\ V < V_{(E)} &\rightarrow R_T < C_T \; \textit{Loss} \end{aligned}$$

It is easy to extend the break-even analyses to the case in which a profit target $(Pr^{(*)})$ is defined. In this case, we have:

$$Pr^{(*)} = R_T - C_T$$
$$Pr^{(*)} = P_U \cdot V - [FC_T + (VC_U \cdot V)] = V \cdot (P_U - VC_U) - FC_T$$

The volume (V) is the volume that can achieve the profit target $(V_{P_{(T)}})$. Therefore, by solving for the volume $(V \equiv V_{P_{(T)}})$, it gets:

$$V_{P_{(T)}} = \frac{Pr^{(*)} + FC_T}{P_U - VC_U} \tag{3.18}$$

The volume able to achieve the profit target $(V_{P_{(T)}})$ is equal to the sum between the profit target defined $(Pr^{(*)})$ and the total fixed costs (FC_T) divided up into the unit contribution margin $(P_U - VC_U)$. The equation can be shown as in Fig. 3.7.

By using a unit contribution margin, a graph of contribution profit can be created, as shown in Fig. 3.8 (Anthony et al. 2011).

In Fig. 3.8 the vertical axis shows income. The main components for the analysis of the graph are the following:

- profit is equal to zero for a volume equal to $V_{(E)}$ representing the break-even volume (the equilibrium volume between revenues and costs);
- the slope is the unit contribution margin and then the contribution margin per unit of volume;
- at zero volume the loss is due to the total fixed costs.

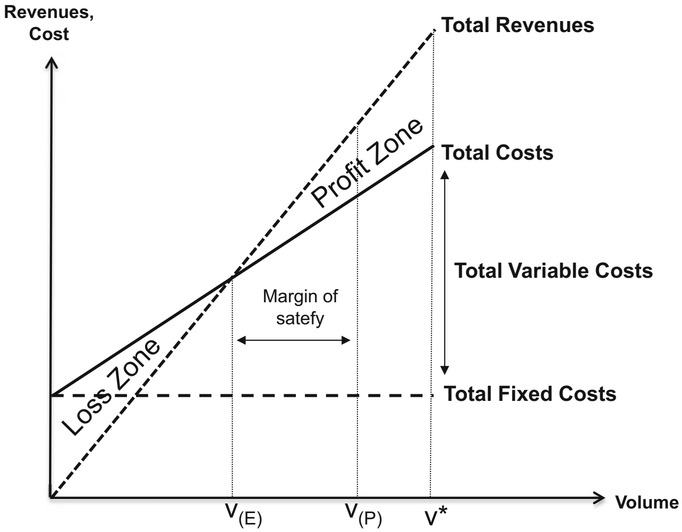
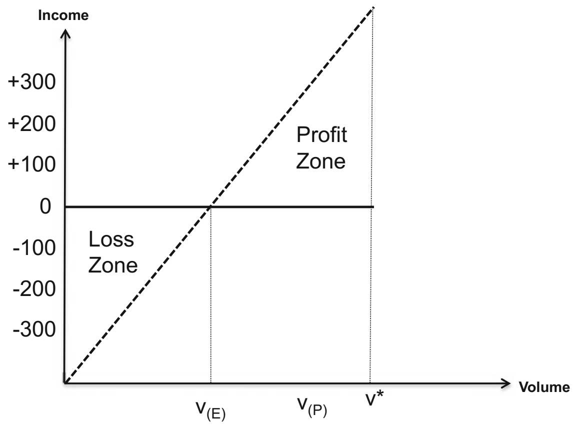


Fig. 3.7 Break-even analysis

Fig. 3.8 Contribution profit



The relationships previously analysed refers to company with a single product. If the company achieves more than one product with different contribution margins, the product mix must be calculated (Anthony et al. 2011).

If the proportion of each product's sales to the total remains constant (and therefore the product mix remains constant over time), it is possible to calculate the weighted-average unit contribution margin for all products (also called unit contribution margin of equivalent product) rather than the individual unit contribution margin of any product.

Considering two products: A and B. The company's Net Income (NI) is equal to the sum of contribution margin of product A $\left(CM_{(A)}\right)$ and contribution margin of product B $\left(CM_{(B)}\right)$ less the total fixed costs (FC_T).

The contribution margin is equal to the unit contribution margin (CM_U) (equal to the difference between price per unit (P_U) and unit variable cost (VC_U)) multiplied by the quantity of product (Q). Therefore, we have:

$$NI = \left\{\left[\left(P_{U(A)} \cdot Q_{(A)}\right) + \left(P_{U(B)} \cdot Q_{(B)}\right)\right] - \left[\left(VC_{U(A)} \cdot Q_{(A)}\right) + \left(VC_{U(B)} \cdot Q_{(B)}\right)\right]\right\} - FC_T$$
$$NI = \left[\left(P_{U(A)} - VC_{U(A)}\right) \cdot Q_{(A)} + \left(P_{U(B)} - VC_{U(B)}\right) \cdot Q_{(B)}\right] - FC_T$$

and then:

$$\begin{matrix} CM_{U(A)} = P_{U(A)} - VC_{U(A)} \\ CM_{U(B)} = P_{U(B)} - VC_{U(B)} \end{matrix} \rightarrow NI = \left(CM_{U(A)} \cdot Q_{(A)} + CM_{U(B)} \cdot Q_{(B)}\right) - FC_T$$

If the product mix is defined and well-known, the weighted percentage $(W(\%))$ of each product volume $\left(Q_{(A)}; Q_{(B)}\right)$ on total volume $\left(Q_{(T)}\right)$ can be calculated:

$$Q_{(T)} = Q_{(A)} + Q_{(B)} \rightarrow \begin{array}{l} W_A(\%) = \frac{Q_{(A)}}{Q_{(T)}} \\ W_B(\%) = \frac{Q_{(B)}}{Q_{(T)}} \end{array}$$

By replacing it, we have:

$$NI = \left(CM_{U(A)} \cdot W_A(\%) \cdot Q_{(T)} + CM_{U(B)} \cdot W_B(\%) \cdot Q_{(T)}\right) - FC_T$$
$$NI = \left[\left(CM_{U(A)} \cdot W_A(\%) + CM_{U(B)} \cdot W_B(\%)\right) \cdot Q_{(T)}\right] - FC_T$$

Denoting with $CM_{U(E)}$ the weighted-average unit contribution margin for all products (or unit contribution margin of equivalent product), we have:

$$CM_{U(E)} = CM_{U(A)} \cdot W_A(\%) + CM_{U(B)} \cdot W_B(\%)$$

and then:

$$NI = \left(CM_{U(E)} \cdot Q_{(T)}\right) - FC_T \tag{3.19}$$

3.2 Full Cost of Product

The full cost defines all resources used for a defined cost object that can be anything for which the cost is measured. With regards to any well defined cost object, the main difference is between direct cost and indirect cost (Anthony et al. 2011; Garrison et al. 2014).

Direct Cost
Direct cost is a cost that is specifically "traced to", or "caused by", that "cost object". Therefore, it is always possible to identify a clear and objective causal-relationship between the cost and the cost object. The direct cost is "attributed" to the cost object because they are "directly charged" to the cost object (Anthony et al. 2011). It is possible to identify three main types of direct cost:

(a) *Direct Material Cost*: it refers to the quantities of material that can be specifically identified with a cost object in an economically feasible manner, priced at the unit price of direct material. The measurement of the Direct Material Cost is characterized by two main aspects: (1) the quantity of material needed for each cost object and (2) the price per unit (cost per unit) of material. These materials are often called raw materials or simple materials;
(b) *Direct Labour Cost*: it refers to the labour quantities that can be specifically identified with a cost object in an economically feasible manner, priced at a unit price of direct labour. Specifically, the earnings of workers who assemble parts and transform them into finished goods or operate machines for production are direct labour costs of the product. The measurement of Direct Labour Costs is

characterized by two main aspects: (1) the quantity of labour time and (2) the price per unit (cost per unit) of labour time. It is worth noting that labour regulations are normally subject to the law. Therefore, the difference between Direct Labour and Indirect Labour tend to be blurred;

(c) *Other Direct Costs*: it refers to any cost due to goods and services traced to a single cost object. The costs must be identifiable with a single cost object. The measurement of each Other Direct Cost is based on two elements: (1) the quantity of services or goods required for each cost object and (2) the price per unit of goods and services (or cost per unit of goods and services). However, most companies only classify direct materials and direct labour costs as direct costs and they consider all other costs as indirect costs.

Indirect Cost

Indirect costs are costs "associated with", or "caused by", two or more "cost objects jointly". Therefore, unlike the direct costs, it is not directly traced to each of the cost objects. In this case, it is not possible, or at least not feasible, to define a clear and objective causal-relationship between the cost and each cost object separately. Consequently, it is not possible or feasible, to measure directly how much of the cost is attributable to a single cost object (Anthony et al. 2011).

Generally, there are three reasons for which cost is not traced directly to a cost object: (1) it is impossible to do; (2) it is not feasible to do; (3) management chooses not to do.

The Indirect Cost is "allocated" to the cost object on the basis of the allocation rate (or absorption rate). They include (Anthony et al. 2011):

(a) *Indirect Cost of Material*: it represents the cost of production materials not caused by a specific cost object. Therefore, this cost refers to materials used in the production process but not directly traced to individual cost objects;
(b) *Indirect Labour Cost*: it is a labour cost not related to a specific cost object. Therefore, unlike direct labour cost, employee efforts are related to the entire process of production;
(c) *Other costs*: they refer to all other costs of the company that are not caused by a specific cost object.

On the basis of these two definitions, it is clear how the terms "*Direct*" and "*Indirect*" are meaningful only in the context of a specified cost object.

In the business the most relevant cost object is the product. In fact, the correct definition of product cost allows for the definition of the product price on the basis of its costs on the one hand, and the market competition on the other hand. Therefore, the correct definition of the product cost is fundamental to understand product profitability and therefore company profitability.

The *product costing system* is the system that accumulates and reports product costs.

It is worth noting that the common confusion between direct costs and variable costs is very wrong as well as the confusion between indirect costs and fixed costs. This confusion occurs because if the cost object is a product, there are many costs directly connected with the product that are also variable with production volume of the product. Similarly, several costs that are indirect to the product are also fixed with the production volume of the product. Although it seems true, it is wrong to use direct costs and variable costs as synonyms as well as indirect costs and fixed costs. In fact, the direct/indirect cost dichotomy and variable/fixed cost dichotomy are included in a different concept: while the direct/indirect cost dichotomy relates to the traceability of costs to specific cost object and therefore it is an accountant's concept, the variable/fixed cost dichotomy relates to the behaviour of cost as volume fluctuates and therefore it is a economist's concept (Anthony et al. 2011).

In order to define the full cost of product, on the basis of the distinction between Direct and Indirect costs of product, it is possible to define several cost configurations (Anthony et al. 2011):

(a) *Conversion Cost*: it refers to the sum of Direct Labour Cost and Indirect Production Costs. Therefore, it includes all production costs needed to convert Direct Materials into finished products;
(b) *Full Production Cost*: it refers to the sum between the Direct Material Cost and Conversion Cost. It can also be defined as the sum between Direct Costs and Indirect Production Costs.
    Therefore, the Full Production Costs do not include distribution or selling costs, or those general and administrative costs that are unrelated to production operations.
(c) *Non-production Cost*: it refers to the all costs incurred in an organization other than production costs, such as selling costs, research and development costs general and administrative costs and interest costs.
(d) *Full Cost of Product*: it is the sum of Full Production Cost and Non-production Cost. Therefore, it is simply the sum of all the cost elements described above. One of the most common errors is to consider Full Cost of Product the Production Costs only or, even worse, the only Direct Material Costs. It has relevant negative implications because the Full Cost of Product is the basis on which the pricing is defined.

It is possible to summarize the Full Cost of Product as in Fig. 3.9.

On the basis of these cost configurations, it is possible to define the full cost of product as the sum of direct costs attributed to the product plus a fair-share of indirect costs allocated to the product.

This definition is only apparently simple. The most relevant problem is the right definition, in monetary terms, of the "*fair-share*" of the indirect costs to be allocated to each product.

The most common way to define the fair share is the proportion of indirect costs caused by each of the various cost objects. In this sense it is used the *allocation rate* (or absorption rate).

One of the most useful techniques is the *step-down order procedure*. It requires the definition of two different cost objects:

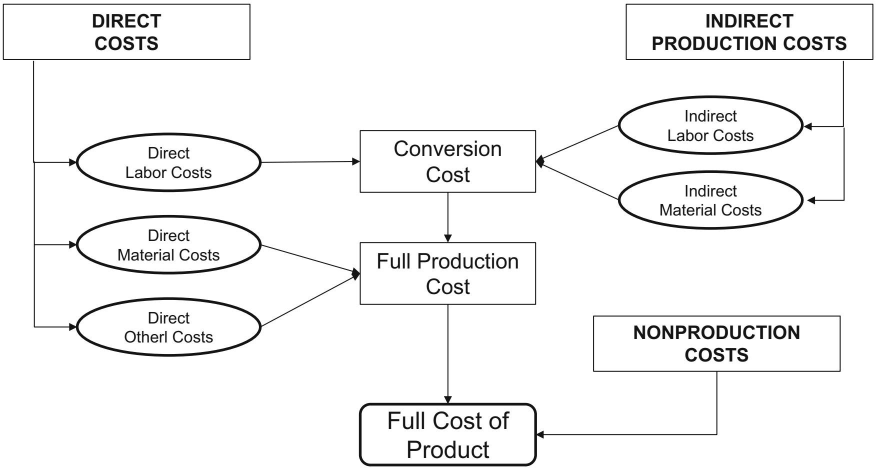


Fig. 3.9 Full cost of product

(a) *cost of product*: the final cost object is the product. Therefore, in this case it can be distinguished between direct and indirect cost of product;
(b) *cost of production and service centers*: they are intermediate cost objects and they are necessary for the application of the step-down order procedure. Therefore, in this case it can be distinguished between direct and indirect cost of cost centers.

The operative procedure to define the full cost of product, can be organised in six main steps.

Step 1: *Identification of Direct and Indirect Costs of Product*

The first step of the procedure requires the distinction between direct and indirect costs with regards to the product. Therefore, for each item of cost its Direct Cost or Indirect Cost must be defined.

Direct cost is specifically "traced to", or "caused by", that "Product". Therefore, it is always possible to identify a clear and objective causal relationship between the cost and the product. They are: Direct Material Costs, Direct Labour Costs and Other Direct Costs.

Direct Material Costs refer to the quantities of material required for each product priced at the unit price of direct material. Therefore, the Direct Material Cost for each product is obtained by multiplying (1) the quantity of material required for each product and (2) the price per unit of material (or cost per unit of material). These materials are often called *raw materials* or simple materials.

Direct Labour Costs refer to the quantities of labour required for each product priced at a unit price of direct labour. Therefore, Direct Labour Costs for each product are obtained by multiplying (1) the quantity of labour time needed for each product and (2) the price per unit of labour time (or cost per unit of labour time).

Other Direct Costs refer to the quantities of services or goods required for each product priced at the unit price. Therefore, every other direct cost for each product is obtained by multiplying (1) the quantity of goods and service required for each product and (2) the price per unit of good and service (or cost per unit of goods and services).

Direct Material Costs, Direct Labour Costs and Other Direct Costs are Direct Costs and they are "attributed" to the product.

Indirect costs are all of the costs that are not directly caused by the product for three main reasons: (1) it is impossible to do; (2) it is not feasible to do; (3) management chooses not to do.

These costs must be "allocated" to the product by using Production and Service Centers.

At the end of this first step of the process, all company costs are distinguished between direct and indirect cost of products.

Consequently, the sum of direct and indirect product costs must be equal to the sum of all company costs. This is the first checkpoint of the process.

Step 2: Assignment of Indirect Cost of Product to Production and Service Cost Centers

In this second step, it is necessary to "attribute" and to "allocate" the indirect cost of product to the Cost Centers.

A cost center is a cost object in which costs are accumulated. In a production cost system, items of cost are first accumulated in cost centers and then assigned to products. Therefore, in this step the cost object is not the product but the cost center. For this reason, they are often called an "intermediate cost object" to create distinction from the product that is the final cost object.

In other terms the allocation of indirect costs to product (the final cost object) requires that these costs are allocated in advance to the cost centers (the intermediate cost object). Therefore, all indirect costs of period must be assigned to the Production and Services Costs.

By considering cost centers as the cost object, two types of costs can be distinguished:

(1) *Direct Cost of Cost Center*: it refers to any indirect cost (with regards to the product object) item that can be uniquely associated with a cost center that is directly charged to that center. In this case, cost items are attributed to the cost center, one cost item at a time on the basis of a clear relationship between the cost and the cost center;
(2) *Indirect Cost of Cost Center*: it refers to any indirect cost (with regards to the product object) that benefit from several cost centers jointly are allocated to those centers. In this case, cost items that are allocated to the centers, different allocation rates may be used for different items. Therefore, every cost item is assigned to the cost center, one cost item at a time on the basis of a specific

allocation rate. Therefore, there must be a fair share of indirect costs on each cost center.

There are two main types of Cost Centers:

(1) *Production Cost Center*: the center (i) produces a product or a component of a product or (ii) performs a distinct step or task of such production;
(2) *Service Cost Center*: the center provides services to production cost centers or to other service cost centers or to general company activities. Note that not all service cost centers are identifiable organization units.

Therefore, each Production and Service Cost Center is characterized by a direct cost attributed plus a fair share of indirect costs.

At the end of this second step of the process, all indirect product costs are assigned to the Production Cost Centers and Service Cost Centers.

Consequently, the sum of the costs assigned to the Production Cost Centers and the costs assigned to the Service Cost Centers must be equal to the entire indirect costs of product. This is the second checkpoint of the process.

Step 3: Identification of the Primary and Secondary Service Cost

At this point in the process, all indirect costs assigned to Service Cost Centers must be assigned to the Product Cost Centers. Indeed, all Production Cost Centers can be attributed directly or allocated by an allocation rate to the product.

The reassignment of costs assigned to the Service Cost Centers to the Product Cost Centers, can be achieved by following a step-down order procedure. This procedure is complex but it is more accurate in the definition of the full product cost. On the basis of step-down order costs assigned to Service Cost Centers can be:

– attributed directly or allocated indirectly on the basis of allocation rates to the Product Cost Centers;
– attributed directly or allocated indirectly on the basis of allocation rates to the other Service Cost Centers and then from these to the Product Cost Centers through direct attribution or allocation.

Therefore, it is possible to distinguish between two different types of Service Cost Centers:

(1) *Primary Service Cost Center*: the costs assigned to the Service Cost Center are attributed directly or allocated through allocation rates to the Production Cost Centers that receive the service;
(2) *Secondary Service Cost Center*: the costs assigned to the Service Cost Center are attributed directly or allocated through allocation rates to the other Service Cost Centers that receive the service.

Therefore, only the costs included in the Primary Service Cost Centers can be assigned to the Production Cost Centers. Consequently, the Secondary Service Cost Centers are emptied.

Consequently, each Primary Service Cost Center is characterized by its original costs assigned plus a fair share of costs of the Secondary Service Cost Centers.

At the end of this third step of the process, all service costs are assigned to the Primary Service Cost Centers.

Consequently, the sum of the costs assigned to the Primary Service Cost Centers must be equal to the entire indirect service costs. This is the fourth checkpoint of the process;

Step 4: Reassignment of Service Cost Centers to the Production Cost Centers

At this fourth step of the process, the costs assigned to the Primary Service Cost Centers must be attributed directly or allocated through allocation rates to the Production Cost Centers. Consequently, the Primary Service Cost Centers are emptied.

Therefore, each Production Cost Center is characterized by its original costs assigned plus a fair share of costs of the Primary Service Cost Centers.

At the end of this fourth step of the process, all costs assigned to the Primary Service Cost Centers must be assigned to the Production Cost Centers.

Consequently, the sum of costs assigned to the Production Cost Centers must be equal to the sum of all indirect costs of product. This is the fifth checkpoint of the process.

Step 5: Identification of the Primary and Secondary Production Cost Centers

In this fifth step of the process, the costs assigned to the Production Cost Centers must be attributed or allocated through the allocation rates to the product.

Also in this case, it is possible to distinguish two types of Production Cost Centers:

(1) *Primary Production Cost Center*: the costs assigned to the Production Cost Centers are attributed directly or allocated through allocation rates to the product that receives the service;
(2) *Secondary Cost Center*: the costs assigned to the Production Cost Center are attributed directly or allocated through allocation rates to the other Production Cost Centers that receive the service.

Therefore, only the costs included in the Primary Production Cost Centers can be assigned to the Product. Consequently, the Secondary Production Cost Centers are emptied.

Therefore, each Primary Production Cost Center is characterized by its original costs assigned plus a fair share of costs of the Secondary Production Cost Centers.

At the end of this fifth step of process, all production costs are assigned to the Primary Production Cost Centers.

Consequently, the sum of the costs assigned to the Primary Production Cost Centers must be equal to the entire indirect production costs. This is the fifth checkpoint of the process.

Step 6: Assignment of the costs in Primary Production Cost Centers to product

In this sixth step of process, the costs assigned in the Primary Production Cost Centers must be attributed directly or allocated through allocation rates to the product.

Two different types of production system can be distinguished (Anthony et al. 2011):

(1) *Process cost system*: in this case, the costs in a defined period are considered. Therefore, all costs included in the Primary Production Cost Centers in the period are divided up into the quantity of products manufactured in the same period. Therefore, the unit cost per product is an *average unit cost per product*: total product costs divided by the quantity of products manufactured during the same period. Consequently, it is not possible to have different costs among a single unit of product. It is worth noting that since a time period is considered, it is possible that at the end of the period there are several products that have not yet been manufactured. In this case, it must be used a common unit for the products manufactured and the products not yet manufactured. This common unit is the *equivalent unit of production*: this unit is equivalent to the complete unit. In this case the inventory at the start of the period must be considered, plus the product manufactured in the period divided by the equivalent unit of production in the period.
(2) *Job order cost system*: in this case the cost of each job order is formed independently according to the period in which the activities are carried out. Each job order must border a share-part of the costs included in Primary Production Cost Centers according to their acquired services. The main problem is that each job order is characterized by a different use of each Primary Production Cost Center. In this case, different allocation rates must be used for one Primary Production Cost Center. Therefore, each Primary Production Cost Center uses one single allocation rate to allocate its costs to each order.

Each Primary Production Cost Center is characterized by its original costs assigned plus a fair share of costs of the Secondary Production Cost Centers.

At the end of this sixth step of the process, all costs assigned to the Primary Production Cost Centers must be assigned to the Product.

Consequently, the sum of costs assigned to the Production Cost Centers must be equal to the sum of all indirect costs of product. This is the sixth checkpoint of the process.

Step 7: Definition of the Full Cost of Product

In this last step of the process, Direct Costs with regards to Direct Material Costs, Labour Costs and Other Direct Costs should be added to the Indirect Costs assigned to each product.

The Full Product Cost is obtained by summing the Direct Cost attributed to the product and the Indirect Costs allocated to the product through the Primary Production Cost Centers. Therefore, the Unit Cost of Product is equal to the Direct Costs plus the share-part of Indirect Costs.

Consequently, by multiplying the Cost per Unit of Product for the quantity of product manufactured, total company costs are achieved. This is the seventh checkpoint of the process.

The entire process as described on the basis of these seven steps, can be summarized as in Fig. 3.10.

In the entire procedure as defined in seven steps, a key role is played by the *allocation bases* also called *cost drivers*.

The allocation bases should express, in as far as is possible, a causal relationship between the costs and the cost object that are the cost centers (intermediate cost object) and the product (final cost object). It is possible to use several alternative allocation bases to allocate indirect costs either to cost center and to product. They are usually defined on the basis of the specific characteristic of the company.

The main problem of the allocation rate is the time in which it is defined. Each allocation rate is defined in advance once a year before the beginning of the accounting year. Therefore, they are predetermined. Consequently, *predetermined allocation rates* are defined once a year before the start of the year and they are used throughout the year (Anthony et al. 2011).

It implies that amounts used to indicate the activity levels and costs estimated for the forthcoming year rather than their actual levels. Consequently, the value of the allocation rates is always expected and not actual. It implies that the procedure to compute the product cost through the procedure analysed previously (the seven steps) is based in the estimated value and not on actual values. Consequently, the product cost is estimated.

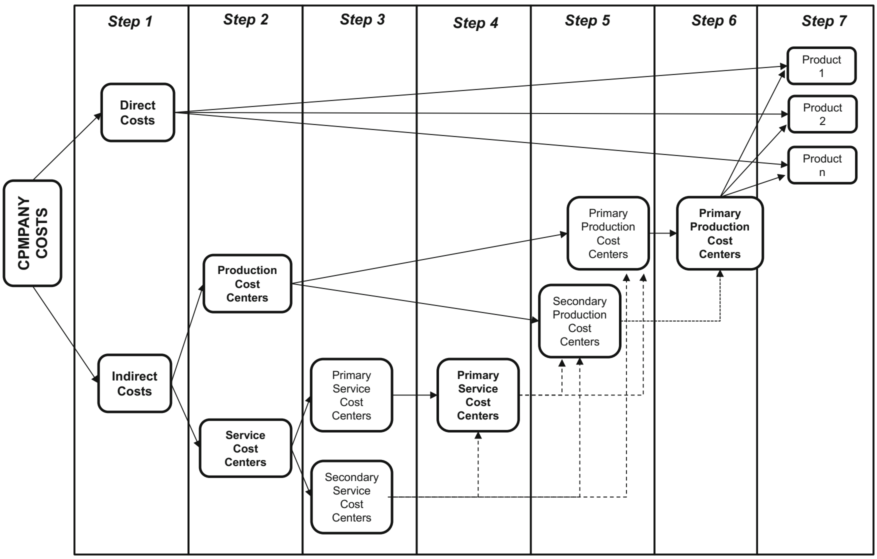


Fig. 3.10 Full cost of product procedure

The use of predetermined allocation rates rather than actual allocation rates computed during the year, has three main advantages (Anthony et al. 2011):

(1) first, allocation rates computed during the year (monthly for example) would be affected by conditions peculiar to that period. Variables on which the allocation rates are defined could change during the period by generating a fluctuation of the allocation rate. Consequently, the product cost would be misleading because allocation of the indirect costs is affected by these fluctuations;
(2) second, the predetermined allocation rate allows for allocation of indirect costs of the product in the same time in which the direct costs are attributed;
(3) third, the predetermined allocation rate allows for calculation of the product cost ex-ante in budget and with calculation of the variance during the year on the basis of the actual product costs. In fact, by using the predetermined allocation rates, expected costs and volumes are used and subsequently the product cost is estimated. During the year, the estimated costs are replaced by the actual costs, and then the product cost becomes actual. The difference between estimated product cost and actual cost of product defines the variance. The analysis of the variance is the most relevant activity in order to measure the difference between the estimate and the actual date for decision making during the year.

The most uncertain part of the definition of predetermined allocation rates is estimating whether or not the level of activity will be monthly in the forthcoming year. This amount is called the *standard volume*. Usually, the monthly standard volume is defined as equal to one-twelfth of the total volume estimated for the forthcoming year (Anthony et al. 2011).

The volume has a relevant value for allocation rates.

The main problem due to the relationship between volume and allocation rate is that the cost per unit of product changes on the basis of the volume estimated. Specifically, there are two main problems:

– first, the cost per unit of product is higher if the volume is estimated lower and it is lower if the volume is estimated higher;
– second, the cost per unit of product includes indirect costs for activities defined even if these activities have not been carried out.

The main problem due to the predetermined allocation rates is that the amount of indirect costs allocated to the product in a given month on the basis of the estimate is likely to differ from the amount of indirect costs allocated to the product on the basis of actual data. It occurs because the indirect costs assigned to the cost centers as well as the actual activity level for the month, are likely to be different from the estimates used when the predetermined allocation rates were calculated.

It generates the two phenomenon called (Anthony et al. 2011):

– *under-absorbed of indirect costs*: the amount of estimated indirect costs absorbed by the product are less than the amount of actual indirect costs absorbed by product;

– *over-absorbed of indirect costs*: the amount of estimated indirect costs absorbed by product exceeds the amount of actual indirect costs absorbed by the product.

The problem of cross subsidies is strictly connected with these two phenomenon. Generally, if the indirect cost structure is quite complex, use of the allocation rate leads to understatement of some product costs and overstatement of other product costs.

3.3 Variance Analysis

The procedure used to calculate the full cost as described previously, gives two types of outputs according to the definition of direct and indirect costs:

– the cost of product as it *should be*: if direct and indirect costs *estimated* is used, on the basis of the predetermined allocation rates, it results in the *standard cost of product*;
– the cost of product as it *actually is*: if direct and indirect costs *actual* is used, on the basis of the predetermined allocation rates, it results in the *actual cost of product*.

The standard costs (and in general the standard values) are the basis of the budget while the actual costs (and in general the actual values) are revealed during the year and they are the real values.

A key role in the standard system is played by the bill of materials: it includes all standards in terms of quantity, time, price and cost to obtain the standard costs and revenues in budget.

In this context, the variance is the simple difference between two numbers: (i) the standard performance; (ii) the actual performance.

Therefore, the variance analysis investigates into the difference between the standard and actual performance. Specifically, a variance analysis involves the decomposition of the variance into the individual factors that have originated the variance. There is no one way of carrying out the variance analysis because many types can be appropriated on the basis of company characteristics, its business and the decision-making process. Generally, a variance analysis must be implemented only if it allows for management to make decisions on the basis of the analysis between the estimated value in budget and actual value revealed.

A general overview of Variance Analysis scheme is summarized in Fig. 3.11 (adaptation from Anthony et al. 2011).

The aim of the variance analysis is to split up the total difference between Actual Net Income and Budget Net Income into the factors that have caused this difference and their specific relevance on total difference.

Based on such analysis management can understand relevant questions about the causes of the variances and take appropriate action based on the answers to these questions. Therefore, the variance analysis is focused on a deep understanding of

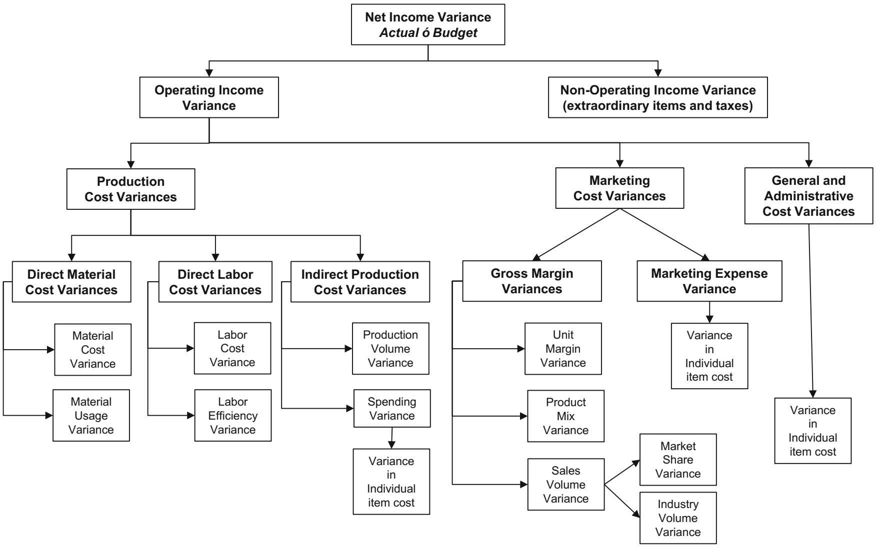


Fig. 3.11 Variance analysis scheme

the causes able to generate differences between budget and actual for the decision-making process at each level of the organization.

The variance analysis must be developed according to two different plans:

– first, is the correct calculation of all variances in order to define the right amount of the variances and further understand fragmentation of the variance between Budget Net Income and Actual Net Income;
– second, is the correct interpretation of these variances in order to make decisions to realign the budget with the actual.

In this sense, the first plan defines "what" the amounts of difference between actual and budget results were, and the second plan is more relevant because it defines "why" these variances occurred.

On the basis of this second plan, at a favourable or unfavourable variance does not imply good or bad performance. It requires the opinion of management according to the defined standard and the actual dynamics and their meaning in the decision-making process and strategies. Also a variance can get out of management control.

Despite the fact that there is no a standard variance analysis (because the depth of analysis and its complexity is function of the company characteristics and management's needs and judgments and it is possible to use different level of techniques) it is possible to identify a basic general rule: the variance analysis has to generate outputs whose relevance to increase company profit is higher than the cost of the analysis.

As shown in Fig. 3.11, the variance between Budget Net Income and Actual Net Income is due to the variance between Budget Operating Income and Actual Operating Income. Indeed, the variance in extraordinary items and taxes are not relevant in terms of decision-making process. The variance in Operating Income can be attributed to variances in three main areas (Anthony et al. 2011):

(a) Production Variances;
(b) Marketing Variances;
(c) General and Administrative Variances.

Before starting the variance analysis, it is important to note that it is not important if the variance is calculated by considering the standard (or budget) minus the actual or reverse. It changes the sign but not its meaning. By considering the cost variances, if:

- the actual cost is lower than the standard cost, it implies a negative cost and therefore the variance is *favourable*;
- the actual cost is higher than the standard cost, it implies a positive cost and therefore the variance is *unfavourable*.

Differently, by considering the revenue variances, if:

- the actual revenue is lower than the budget revenue, it implies a negative revenue and therefore the variance is *unfavourable*;
- the actual revenue is higher than the budget revenue, it implies a positive revenue and therefore the variance is *favourable*.

Therefore, it is possible to generalize by saying that *unfavourable variance* makes Actual Net Income lower than Budget Net Income. It occurs when:

- actual revenue is less than budget revenue;
- actual cost is higher than budget cost.

Differently, *favourable variance* makes Actual Net Income higher than Budget Net Income. It occurs when:

- actual revenue is higher than budget revenue;
- actual cost is less than budget cost.

Finally, note that words "*standard*" and "*budget*" refer both to the estimated data and therefore what costs and revenues should be. Usually, standard is used with per-unit cost and revenue amounts, while budget is used with total amounts of revenues and costs.

(A) **Production Cost Variances**

The production cost variances involve the analysis of the variance with regards to direct costs (direct material and direct labour) and indirect costs. Therefore, there are three main components (Anthony et al. 2011):

- Direct Material Costs;
- Direct Labour Costs;
- Indirect Production Costs.

Direct Material Variances

The standard direct material cost per unit of product (or one unit of output) ($MC_{U(S)}$) is equal to the standard quantity of material per unit of product ($QM_{U(S)}$) multiplied by the standard cost (or price) of material per unit of product ($C_{U(S)}$) (Anthony et al. 2011).

The total standard direct material cost ($MC_{T(S)}$) for an accounting period is equal to the standard direct material cost per unit of product ($MC_{U(S)}$) multiplied by the number of units of product (or output) produced actually ($Q_{(A)}$) and then effectively in that period (Anthony et al. 2011).

Formally:

$$MC_{U(S)} = QM_{U(S)} \cdot C_{U(S)} \Rightarrow MC_{T(S)} = MC_{U(S)} \cdot Q_{(A)} \tag{3.20}$$

Similarly, actual direct material cost per unit of product (or one unit of output) ($MC_{U(A)}$) is equal to the actual quantity of material per unit of product ($QM_{U(A)}$) multiplied by the actual cost of material per unit of product ($C_{U(A)}$).

The total actual direct material cost ($MC_{T(A)}$) for an accounting period is equal to the actual direct material cost per unit of product ($MC_{U(A)}$) multiplied by the number of units of product currently produced ($Q_{(A)}$), and then effectively, in that period (Anthony et al. 2011).

Formally:

$$MC_{U(A)} = QM_{U(A)} \cdot C_{U(A)} \Rightarrow MC_{T(A)} = MC_{U(A)} \cdot Q_{(A)} \tag{3.21}$$

The direct material cost variance (DMV) is equal to the difference between total standard direct material cost ($MC_{T(S)}$) and the total actual direct material cost ($MC_{T(A)}$) for an accounting period.

Since the number of units of product currently produced ($Q_{(A)}$) is the same in both cases, standard and actual, the difference is due to the standard quantity of material per unit of product ($QM_{U(S)}$) and the current direct material cost per unit of product ($MC_{U(A)}$) (Anthony et al. 2011).

Formally:

$$DMV = MC_{T(S)} - MC_{T(A)} \Rightarrow DMV = \left[MC_{U(S)} - MC_{U(A)}\right] \cdot Q_{(A)} \tag{3.22}$$

Therefore, direct material cost variance is based on the actual quantity of a period (and therefore standard quantity in the budget is not relevant in the analysis) and it is function of the difference between the standard direct material cost per unit of product and the actual direct material cost per unit of product.

The direct material cost variance (DMV) can be divided up into two components (Anthony et al. 2011):

(1) *Material usage variance* (also called yield variance or simply quantity component) $(DMV_{(Q)})$: it is equal to the difference between total standard quantity of material $(QM_{T(S)})$ and total current quantity of material $(QM_{T(A)})$, with each total quantity priced at the standard cost (or price) of material per unit of product $(C_{U(S)})$.
Both total quantities are based on the number of units of product actually $(Q_{(A)})$ produced. Therefore, total standard quantity of material $(QM_{T(S)})$ is equal to the standard quantity of material per unit of product $(QM_{U(S)})$ multiplied by the number of units of product actually $(Q_{(A)})$, and the total actual quantity of material $(QM_{T(A)})$ is equal to the actual quantity of material per unit of product $(QM_{U(A)})$ multiplied by the number of units of current product $(Q_{(A)})$. Formally:

$$\begin{array}{l} QM_{T(S)} = QM_{U(S)} \cdot Q_{(A)} \\ QM_{T(A)} = QM_{U(A)} \cdot Q_{(A)} \end{array} \Rightarrow QM_{T(S)} - QM_{T(A)} = \left[QM_{U(S)} - QM_{U(A)}\right] \cdot Q_{(A)} \tag{3.23}$$

and therefore:

$$DMV_{(Q)} = \left[QM_{T(S)} \cdot C_{U(S)} - QM_{T(A)} \cdot C_{U(S)}\right] = \left[QM_{T(S)} - QM_{T(A)}\right] \cdot C_{U(S)} \tag{3.24}$$

or in equivalent terms:

$$DMV_{(Q)} = \Delta\, QM_{T(S-A)} \cdot C_{U(S)} \tag{3.25}$$

Equation (3.25) shows how the material usage variance is equal to delta (Δ) between standard and actual quantity of material multiplied by standard material costs:

$$DMV_{(Q)} = \Delta\, Quantity \cdot Standard\, Cost$$

(2) *Material cost variance* (also called material price variance or price component) $(DMV_{(C)})$: it is equal to the difference between standard cost (or price) of material per unit of product $(C_{U(S)})$ and the actual cost (or price) of material per unit of product $(C_{U(A)})$ multiplied by the total actual quantity of material $(QM_{T(A)})$ used. Therefore, the material cost variance is equal to delta (Δ)

between standard and actual cost of material per unit of product multiplied by actual quantity of material:

$$DMV_{(C)} = \Delta\, Cost \cdot Actual\, Quantity$$

Formally:

$$DMV_{(C)} = \left[C_{U(S)} - C_{U(A)}\right] \cdot QM_{T(A)} \leftrightarrow DMV_{(P)} = \Delta C_{U(S-A)} \cdot QM_{T(A)} \quad (3.26)$$

The distinction of Direct Material Variances in its two components, Material Usage Variance and Material Price Variance, is fundamental for the management analysis for planning and control.

Direct Labour Variances

The standard direct labour cost per unit of product (or output) $(\mathrm{LC_{UP(S)}})$ is equal to the standard labour time per unit of product (or outputs) $(\mathrm{LT_{UP(S)}})$ usually expressed in hours, multiplied by a standard labour cost (or standard rate) per unit of time $(\mathrm{LC_{UT(S)}})$.

Note that if workers are paid on a price-rate basis, the standard labour cost per unit of product is the rates for producing that unit.

Total standard direct labour cost $(\mathrm{LC_{T(S)}})$ of an accounting period is equal to the standard direct labour cost per unit of product (or output) $(\mathrm{LC_{UP(S)}})$ multiplied by the number of units of product (or output) produced actually $(\mathrm{Q_{(A)}})$ and then effectively in that period (Anthony et al. 2011).

Formally:

$$LC_{UP(S)} = LT_{UP(S)} \cdot LC_{UT(S)} \Rightarrow LC_{T(S)} = LC_{UP(S)} \cdot Q_{(A)} \quad (3.27)$$

Similarly, actual direct labour cost per unit of product (or unit of output) $(\mathrm{LC_{UP(A)}})$ is equal to the actual labour time per unit of product (or output) $(\mathrm{LT_{UP(A)}})$ multiplied by actual labour cost (or standard rate) per unit of time $(\mathrm{LC_{UT(A)}})$.

The total actual direct labour cost $(\mathrm{LC_{T(A)}})$ for an accounting period is equal to the actual direct labour cost per unit of product (or output) $(\mathrm{LC_{UP(A)}})$ multiplied by the number of units of product (or output) produced actually $(\mathrm{Q_{(A)}})$ and then effectively in that period (Anthony et al. 2011).

Formally:

$$LC_{UP(A)} = LT_{UP(A)} \cdot LC_{UT(A)} \Rightarrow LC_{T(A)} = LC_{UP(A)} \cdot Q_{(A)} \quad (3.28)$$

The direct labour cost variance (DLV) is equal to the difference between total standard direct labour cost $(\mathrm{LC_{T(S)}})$ and the total actual direct labour cost $(\mathrm{LC_{UP(A)}})$ for an accounting period. Since the number of units of product (or output) produced actually $(Q_{(A)})$ is the same in both cases, standard and actual, the difference is due to the standard direct labour cost per unit of product (or output) $(LC_{UP(S)})$ and the actual labour time per unit of product (or output) $(LT_{UP(A)})$. Formally:

$$DLV = LC_{T(S)} - LC_{UP(A)} \Rightarrow DLV = \left[LC_{UP(S)} - LC_{UP(A)}\right] \cdot Q_{(A)} \tag{3.29}$$

Therefore, direct labour cost variance is based on the actual quantity of a period (and therefore standard quantity in the budget is not relevant in the analysis) and it is function of the difference between the standard direct labour cost per unit of product (or output) and the actual direct labour cost per unit of product.

The direct labour cost variance (DLV) can be divided up into two components (Anthony et al. 2011):

(1) *Labour efficiency labour* (also called the quantity variance or the usage variance) $(DLV_{(Q)})$: it is equal to the difference between standard labour time per unit of product (or outputs) $(LT_{UP(S)})$ and actual labour time per unit of product (or output) $(LT_{UP(A)})$ multiplied by standard labour cost (or standard rate) per unit of time $(LC_{UT(S)})$. Therefore, the labour efficiency variance is equal to delta (Δ) between standard and actual of labour time per unit of product multiplied by standard labour cost per unit of time:

$$DLV_{(Q)} = \Delta\, Time \cdot Stancard\, Cost$$

Formally:

$$DLV_{(Q)} = \left[LT_{UP(S)} - LT_{UP(A)}\right] \cdot LC_{UT(S)} \leftrightarrow DLV_{(Q)} = \Delta\, LT_{UP(S-A)} \cdot LC_{UT(S)} \tag{3.30}$$

(2) *Labour cost variance* (also called labour rate variance or labour price variance) $(DLV_{(C)})$: it is equal to the difference between standard labour cost (or standard rate) per unit of time $(LC_{UT(S)})$ and actual labour cost (or standard rate) per unit of time $(LC_{UT(A)})$ multiplied by the actual labour time per unit of product (or output) $(LT_{UP(A)})$. Therefore, the labour cost variance is equal to delta (Δ) between standard and actual of labour cost per unit of product multiplied by actual labour time per unit of product:

$$DLV_{(C)} = \Delta\, Cost \cdot Actual\, Time$$

Formally:

$$\begin{aligned} DLV_{(Q)} &= \left[LC_{UT(S)} - LC_{UT(A)}\right] \cdot LT_{UP(A)} \leftrightarrow DLV_{(Q)} \\ &= \Delta\, LC_{UT(S-A)} \cdot LT_{UP(A)} \end{aligned} \tag{3.31}$$

Indirect Production Cost Variances

Indirect Production Costs are allocated to the product through predetermined allocation rates. These rates are calculated by dividing the estimated product activity level (normal or standard volume) by the estimated total indirect costs to be incurred at that volume (Anthony et al. 2011).

By distinguishing between variable indirect costs and fixed indirect costs, total indirect costs (IC_T) are equal to total fixed indirect costs per period (IFC_T) plus variable indirect cost per unit of product (IVC_U) multiplied by quantity of product in volume (Q), so we have:

$$IC_T = IFC_T + (IVC_U \cdot Q) \tag{3.31}$$

Equation (3.31) represents the *flexible budget straight line*. It illustrates the amount of costs expected for each level of volume. Since some indirect production costs are variable, others are fixed and still others are semi-variable, total indirect production costs will be different at each volume level. Therefore, within an estimated volume range, total budget indirect production costs are expected to vary according to the equation, as shown in Fig. 3.12 (adaptation from Anthony et al. 2011).

The allocation rate (AR) defines the average indirect cost per unit of product at the standard quantity volume (Q_S). Therefore, it is found by dividing the total indirect costs at the standard volume by the standard volume and therefore the number of units represented by that volume number of units, as follows:

$$AR = \frac{IC_T}{Q_S} = \frac{IFC_T + (IVC_U \cdot Q_S)}{Q_S} = \frac{IFC_T}{Q_S} + \frac{IVC_U \cdot Q_S}{Q_S} = \frac{IFC_T}{Q_S} + IVC_U \tag{3.32}$$

Therefore, the allocation rate (AR) is equal to the sum of the variable indirect cost per unit ($ICVU$) and the average fixed indirect costs per unit at standard volume ($\frac{\mathrm{ICF}}{\mathrm{Q_S}}$). Figure 3.13 (adaptation from Anthony et al. 2011) illustrates the absorbed cost at standard volume by considering the predetermined allocation rate.

The Indirect Production Cost Variance (ICV) is the difference between actual indirect production costs and indirect production costs absorbed by production (and therefore allocated to products).

The Indirect Cost Variance can be divided up into two parts (Anthony et al. 2011):

(1) *Production Volume Variance* (PVV): it is due to the difference between standard volume of production in terms of units used in calculating the predetermined allocation rate and actual volume of production in terms of units.

Indirect production costs are allocated to each unit product produced, through the predetermined allocation rate. Therefore, each unit product absorbs a share-part of indirect production costs.

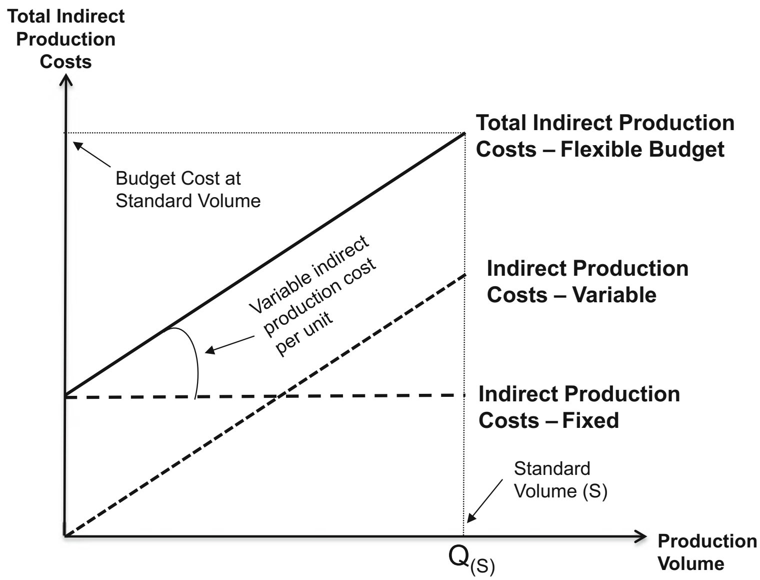


Fig. 3.12 Flexible budget

The amount of total indirect costs absorbed (AIC) by the units produced is equal to the predetermined allocation rate (AR) multiplied by the actual volume (in terms of number of units) produced ($Q_{(A)}$):

$$AIC = AR \cdot Q_{(A)} \tag{3.33}$$

It is important to note that budget indirect costs and absorbed indirect costs will only be equal at the standard volume ($Q_{(S)}$) in terms of number of units. Therefore (Anthony et al. 2011):

- actual volume is equal to standard volume ($Q_{(A)} = Q_{(S)}$): the amount of indirect production costs absorbed is equal to the budgeted indirect production costs;
- at any actual volume below standard volume ($Q_{(A)} < Q_{(S)}$): the amount of indirect production costs absorbed is less than the budgeted indirect production costs at that volume. At these volumes, budgeted indirect costs are *under-absorbed* (or unabsorbed). It produces unfavourable variance;
- at any actual volume above standard volume ($Q_{(A)} > Q_{(S)}$): the amount of indirect production costs absorbed is more than the budgeted indirect production costs at that volume. At these volumes, budgeted indirect costs are *over-absorbed*. It produces a favourable variance.

Therefore, *under-absorbed* and *over-absorbed* are function of the difference between budgeted and absorbed costs caused solely by the fact that actual volume is

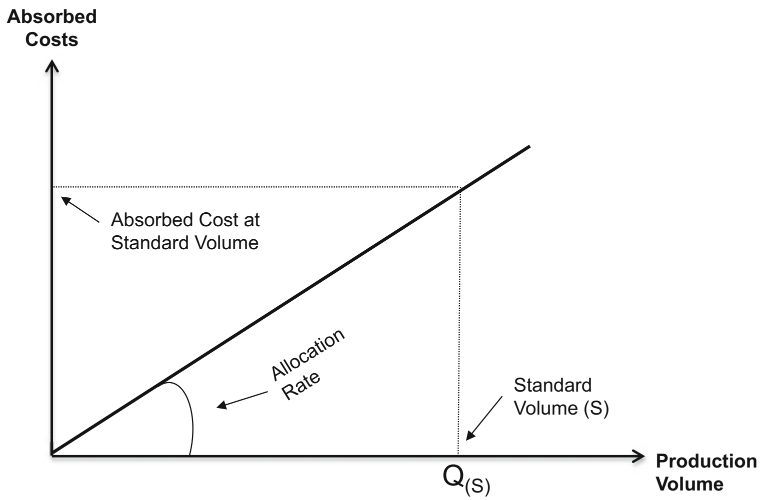


Fig. 3.13 Allocation rate

different from standard volume on the basis of the predetermined allocation rate calculated. For this reason, it is defined as *production volume variance*.

It is important to know that, the standard volume used in calculating the predetermined allocation rate is the volume expected during the forthcoming year (Anthony et al. 2011).

(2) *Spending Variance* (*SV*): is due to difference between the standard total indirect production costs in budget and the actual total indirect production costs achieved. Specifically, the spending variance is equal to the difference between the budgeted indirect production costs for the period's actual level of volume less the period's actual indirect costs. If:

- the actual indirect production costs are below the budgeted indirect production costs, the variance is favourable;
- the actual indirect production costs are above the budgeted indirect production costs, the variance is unfavourable.

Since the spending variance is based on the amount of indirect production costs estimated for the actual level of volume in the period, it implies that the spending variance is due to the difference between standard volume and actual volume.

It is worth noting that the spending variance for indirect production costs has the same meaning as the sum of usage and price (cost) variances for direct material cost and direct labour cost. Therefore, the spending variance can be divided up for each item cost between usage variance and price variance as described for direct cost variances (Anthony et al. 2011).

The *Indirect Production Cost Variance* (ICV) can be summarized as in Fig. 3.14 (adaptation from Anthony et al. 2011).

Therefore, the Indirect Production Cost Variance (ICV) is equal to the sum of Production Volume Variance (PVV) and Spending Variance (SV). Specifically, there are three main relationships (Anthony et al. 2011):

(a) Indirect Production Cost Variance (ICV): it is equal to absorbed costs less actual costs;
(b) Production Volume Variance (PVV): it is equal to absorbed costs less budgeted costs. It is always borne in mind that both the absorbed and budgeted cost amounts are based on the actual production volume of the period. Note that the production volume variance is also equal to the fixed portion of the allocation rate multiplied by the difference between actual volume and standard volume.
(c) Spending Variance (SV): it is equal to budgeted costs less actual costs.

(B) **Marketing Variances**

Usually, the marketing department focuses on two main goals: the gross margin to be achieved and the expenses limit in its budget to be respected. Based on these two goals, the marketing variance can be divided up into two parts: (i) Expenses Variances and (ii) Gross Margin Variances (Anthony et al. 2011).

Expenses Variances (EV) refers to each item of marketing expense and is equal to the difference between Actual Expenses (or Cost) $(\mathrm{E}_{(\mathrm{A})})$ and Budget Expense (or Cost) $(\mathrm{E}_{(\mathrm{B})})$:

$$EV = E_{(A)} - E_{(B)} \tag{3.34}$$

Gross Margin Variances (GMV) refers to the difference between Actual Gross Margin $(\mathrm{GMV}_{(\mathrm{A})})$ and Budget Gross Margin $(\mathrm{GMV}_{(\mathrm{B})})$ both calculated on the basis of the standard cost per unit of product:

$$GMV = GMV_{(A)} - GMV_{(B)} \tag{3.35}$$

The Gross Margin is equal to the difference between sales revenues (equal to the sum of the multiplications for each type of product of the price per unit of product and the volume in terms of units of product) and cost of sales (equal to the sum of the multiplications for each type of product of the cost per unit of product and the volume in terms of units of product) defines the Gross Margin. Generally, the marketing department is responsible for product sales volume and unit selling prices but not for the cost per unit of product. Therefore, in order to calculate the gross margin variance the cost per unit of product used should be the standard cost. The variance between the actual and standard cost per unit of product is calculated in production variance (Anthony et al. 2011).

The Gross Margin Variance can be divided up into three main components (Anthony et al. 2011):

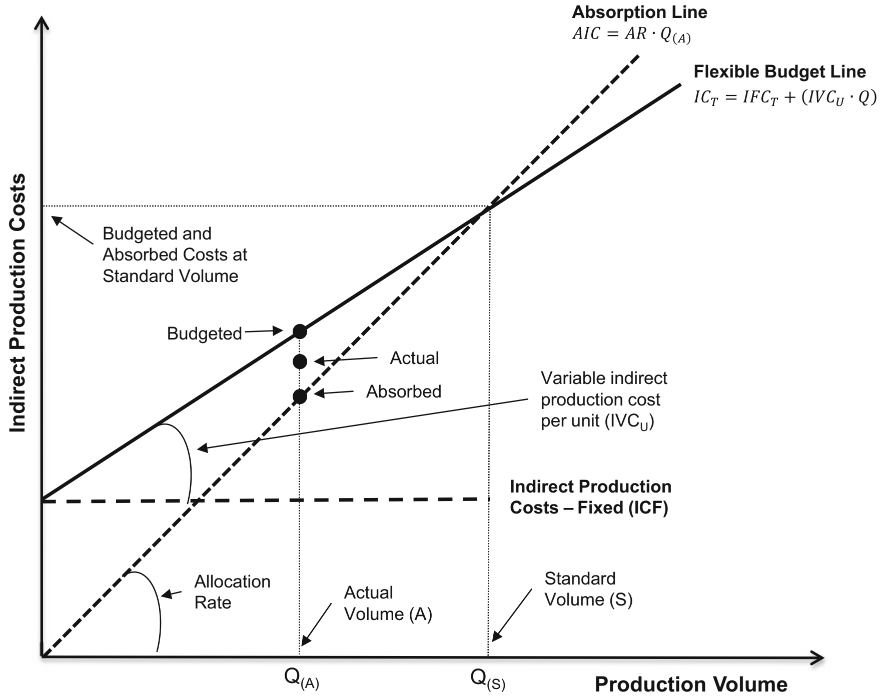


Fig. 3.14 Indirect production cost variance: budgeted, absorbed and actual indirect production costs

- Unit Margin Variance;
- Sales Volume Variance;
- Product Mix Variance.

Unit Margin Variance

It is due to the difference between budget gross margin per unit of product and actual gross margin per unit of product. Specifically, the Unit Margin Variance (UMV) is equal to the difference between actual unit margin $(MU_{(A)})$ and budget unit margin $(MU_{(B)})$ multiplied by actual volume $(Q_{(A)})$:

$$UMV = \left(MU_{(A)} - MU_{(B)}\right) \cdot Q_{(A)} \leftrightarrow MUV = \Delta MU_{(A-B)} \cdot Q_{(A)} \qquad (3.36)$$

It is worth noting that unit margin variance can be also be called the selling price variance if it is assumed that the actual unit cost turns out to be equal to the budget standard unit cost during the period (Anthony et al. 2011). Indeed, the unit margin variance (UMV) is equal to the difference between the actual unit margin $(MU_{(A)})$ and budget unit margin $(MU_{(B)})$. The actual unit margin $(MU_{(A)})$ is equal to the difference between actual unit selling price $(UP_{(A)})$ and actual unit cost $(UC_{(A)})$ as well as the budget unit margin $(MU_{(B)})$ is equal to the difference between budget unit selling price $(UP_{(B)})$ and budget standard unit cost $(UC_{(B)})$. Substituting, we have:

$$UMV = MU_{(A)} - MU_{(B)} \rightarrow \begin{matrix} MU_{(A)} = UP_{(A)} - UC_{(A)} \\ MU_{(B)} = UP_{(B)} - UC_{(B)} \end{matrix} \rightarrow UMV = \left(UP_{(A)} - UC_{(A)}\right) - \left(UP_{(B)} - UC_{(B)}\right) \tag{3.37}$$

If it is assumed that actual unit cost $(UC_{(A)})$ it is equal to the budget standard unit cost $(UC_{(B)})$ during the period, we have:

$$UC_{(A)} = UC_{(B)} \rightarrow UMV = UP_{(A)} - UP_{(B)} \tag{3.38}$$

Therefore, in this case the unit margin variance (UMV) is caused solely by the difference between actual unit selling price $(UP_{(A)})$ and budget unit selling price $(UP_{(B)})$.

Consequently, if the company uses the same standard unit costs during the year as were used in preparing that year's budget, the unit margin variance is called selling price variance because it is due to the difference between actual and budget selling unit price.

Sales Volume Variance

It is due to the difference between budget sales volume (in units of product) and actual sales volume (in units of product). Specifically, the sales volume variance (SVV) is equal to the difference between actual volume $(Q_{(A)})$ and budget volume $(Q_{(B)})$ multiplied by budget unit margin $(MU_{(B)})$:

$$SVV = \left(Q_{(A)} - Q_{(B)}\right) \cdot MU_{(B)} \leftrightarrow MUV = \Delta Q_{(A-B)} \cdot MU_{(B)} \tag{3.38}$$

If data is available on total sales of a product by all companies, the volume variance can be divided up into industry volume variance and markets share variance (Anthony et al. 2011). This division is useful because it is possible to distinguish the variance due to changes in total industry sales, that reflect the general economic conditions and therefore they are out of management control, and the variance due to changes in company's market share that are attributable to management strategies.

Specifically, industry volume variance (IVV) is equal to the difference between actual industry volume $(IV_{(A)})$ and budget industry volume $(IV_{(B)})$ multiplied by budget market share $(MS_{(B)})$ and budget unit margin $(MU_{(B)})$:

$$IVV = \left(IV_{(A)} - IV_{(B)}\right) \cdot MS_{(B)} \cdot MU_{(B)} \rightarrow IVV = \Delta IV_{(A-B)} \cdot MS_{(B)} \cdot MU_{(B)} \tag{3.39}$$

Market share variance (MSV) is equal to the difference between actual market share $(MS_{(A)})$ and budget market share $(MS_{(B)})$ multiplied by actual industry volume $(IV_{(A)})$ and budget unit margin $(MU_{(B)})$:

$$MSV = \left(MS_{(A)} - MS_{(B)}\right) \cdot IV_{(A)} \cdot MU_{(B)} \rightarrow MSV = \Delta\, MS_{(A-B)} \cdot IV_{(A)} \cdot MU_{(B)} \tag{3.40}$$

Product Mix Variance

It highlights if the company sells several products with different unit gross margins. If the company sells several products with different unit gross margins, the mix of high-margin and low-margin products have relevant effects on total gross margin.

Therefore, the product mix variance measures the variance in gross margin due to the difference between the budget mix that indicates the mix of products estimated to sell, and the actual mix that indicated the mix of products truly sold (Anthony et al. 2011).

The mix variance is the sum of the mix variances attributable to each product calculated on the difference between the Actual Volume (in units of product) Sold and Budget Volume (in units of product) Expected to Sell.

Note the difference use of volume for the calculation of production variances and marketing variances (Anthony et al. 2011):

- the analysis of marketing variances is based on sales volume and the difference between actual sales volume and budget sales volume. Indeed, the actual net income is based on actual sales volume as well as the budget net income is based on budget sales volume;
- the analysis of production variances is based on production volume and not on sales volumes. Also, the difference between actual and standard (in budget) production volume is relevant only with regards to indirect product costs volume variance. Indeed, in direct material and direct labour costs variances, the budget production volume is always irrelevant.

(C) **General and Administrative Costs Variances**

The general and administrative expenses (or Costs) include all others administrative costs, financial costs, non-operating costs and taxes.

For each item of cost, its variance may be divided up into volume and price. However, the variance for each cost is normally calculated as a simple difference between the actual cost and budget cost (Anthony et al. 2011). Therefore, the cost variance (CV) for each cost is equal to the difference between its actual cost $(C_{(A)})$ and its budget cost $(C_{(B)})$ as follows:

$$CV = C_{(A)} - C_{(B)} \tag{3.41}$$

References

Anthony RN, Hawkins DF, Merchant KA (2011) Accounting. Text and cases, 13th edn, International Edition. McGraw-Hill, New York

Anderson MC, Banker R, Janakiraman S (2003) Are selling, general and administrative costs "Sticky"? J Acc Res 47–63

Atrill P, McLaney E (2018) Management accounting for decision makers, 9th edn. Pearson

Banker RD, Kaplan RS, Young SM, Atkinson AA (2000) Management accounting. Pearson College, Boston

Bhimani A, Horngren CT, Datar SM, Rajan M (2015) Management and cost accounting, 6 edn. Pearson, New York

Drury C (2016) Management accounting for business, 6th edn. Thomson Learning, UK

Garrison RH, Noreen EW, Brewer PC (2014) Managerial accounting, 15th edn. McGraw-Hill, New York

Kaplan R, Atkinson AA (1998) Advanced management accounting, 3rd edn. Prentice Hall, Upper Saddle River

Sahaf MA (2013) Management accounting: principles and practice, 3rd edn. Paperback, Vikas Publishing House

Seal WB, Rohde C (2014) Management accounting. McGraw-Hill, London

Tennent J (2014) Guide to financial management: principles and practice, 2nd edn. The Economist

Part II
Risk and Return in the Capital Market

Chapter 4
Utility Function Approach

Abstract In the context of decisions under uncertainty investors try to maximize the expected return on investment and minimize investment risk. Unfortunately, there is a trade-off between these two aims. The theory of the choices under uncertainty leads the decision-making process in capital markets. The aim is to analyse the behaviour of the rational investor under uncertainty. Specifically, the aim of the theory is not to define a set of criteria for the investor's preference for general validity because all investors are different from one another. Otherwise, the aim of the theory is to define a set of criteria of the decision-making process based on a few principles characterized by generality, rationality, economic significance, consistency with individual criteria, and therefore able to have a normative function. In this regard, the theory defines the criteria by which the rational investor chooses between the real possible options, considering the restrictions, on the basis of the expected effects that could be achieved according to their nature and that can be sorted in consideration of the relative probability. The portfolio choices (or portfolio selection) is a problem related to wealth allocation between different investment assets. In this context, the portfolio choices will be analysed based on the two main criteria:

– utility functions criteria;
– mean-variance criteria.

This chapter analyses the first criteria, while the next chapter analyses the second criteria.

4.1 Decision Under Uncertainty

Any financial operation involves movements of money over time. There are two main variables: capital (in monetary dimensions) and time. On the basis of these two variables, investors try to maximize the expected return on investment and minimize the time period of the investment. There is a trade-off between these two aims.

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_4

In conditions of certainty, capital and time are known. There is a perfect and known relationship between decision-making and its effects in terms of both capital and time. Consequently, for each decision its effects in terms of capital and time are known, and the decision to be made can be presumed from the defined and expected effects in terms of capital and time.

On the contrary, in conditions of uncertainty, at least one of the two variables (capital and time) is unknown. The effects of the decision cannot be known ex-ante. Normally, once time has been defined, the capital value can only be expected. Indeed, it can take on different values, none of which are known at the moment of the decision.

Therefore, in the case of uncertainty between decisions and its effects in terms of capital and time, a third variable must be introduced: the "*states of nature*". It is uncontrollable and usually defined in terms of probability distribution associated with the possible events to be achieved.

Therefore, the effects on capital and time are a function of the decision made as well as the state of nature that could be determined in the future. Consequently, for each decision there could be N different effects on capital and time unknown ex-ante, and also once the expected effects in terms of capital and time have been defined, the decision to be made cannot be presumed.

Since the investors can be defined as risk averse, the variable states of nature can be defined in terms of risk (Saltari 2011; Castellani et al. 2005).

The theory of the choices under uncertainty leads the decision-making process in capital markets. The aim is to analyse the behaviour of the rational investor under uncertainty (Campbell 2015; Varian 1992; Kreps 2012; Gravelle and Rees 1992; Mankiw 2017; Perloff 2016). Specifically, the aim of the theory is not to define a set of criteria for the investor's preference for general validity because all investors are different from one another. On the contrary, the aim of the theory is to define a set of criteria of the decision-making process based on few principles characterized by generality, rationality, economic significance, consistency with individual criteria, and therefore able to have a normative function.

In this sense, the theory defines the criteria by which the rational investor chooses between the real possible options by considering the restrictions, on the basis of the expected effects that could be achieved according to the state of nature and that can be sorted by considering the associate probability (Castellani et al. 2005; Saltari 2011).

In conditions of uncertainty, by assuming the time defined, the relationship between decision, states of nature and effects in terms of capital, can be defined as follows:

$$\mathrm{y_{i,j}} = \mathrm{p}\left(\mathrm{a_i}; \mathrm{s_j}\right) \tag{4.1}$$

where:

- a_i: are the decisions (for $i = 1, 2, 3, \ldots, n$) and each of them belongs to the set of possible decisions (D) so that $a_i \in D$;

- s_j: are the states of nature (for $j = 1, 2, 3, \ldots, n$) and each of them belongs to the set of possible states of nature (S) so that $s_j \in S$;
- $y_{i,j}$: is the effect of the i-th decision (a_i) when the j-th state of nature (s_j) is achieved. It is a real number the real number $(y_{i,j})$ associated with each pair $(a_i; s_j)$ (noted $p(a_i; s_j)$).

This relationship can be represented by using the matrix Decisions-States of Nature as in Table 4.1 as follows:

It is worth noting that this relationship is based on the assumption of no correlation between the set of possible decisions (D) and the set of possible states of nature (S). Then, each decision (a_i) can never affect the state of nature (s_j).

The best way to solve the problem of the choices under uncertainty, is to assimilate, in some way, the problem of decisions under uncertainty to the problem of decisions under certainty. This would allow for the use of decisions under uncertainty of the analytical instrumentation usually used in decisions under certainty. In this sense, the reasoning can be structured as follows.

The uncertainty is due to the states of nature. A probability can be assigned $(\pi_s \, for \, s = 1, 2, 3, \ldots S)$ to each state of nature to be achieved. Equation (4.1) can be rewritten as follows:

$$\mathrm{y_{i,j}} = \mathrm{p}\left(\mathrm{a_i}; \pi_\mathrm{s}\left(\mathrm{s_j}\right)\right) \tag{4.2}$$

In this case, it can be defined a random variable $(\tilde{X}_a)$ to associate the real number $(y_{i,j})$ to the pair (a_i, s_j), as follows:

$$\tilde{\mathrm{X}}\left(\mathrm{a_i}, \mathrm{s_j}\right) = \mathrm{y_{i,j}} \tag{4.3}$$

Therefore, it is possible to substitute with decision (a_i) the random variable associated $(\tilde{\mathrm{X}}_\mathrm{a})$. Therefore, the probability distribution $(\mathrm{F_a})$, can be associated to each random variable, as follows:

$$\mathrm{F_a(s)} = \mathrm{y} \tag{4.4}$$

Table 4.1 Matrix decisions-states of nature

Matrix: decisions—states of nature					
		States of nature (s)			
		s_1	s_2	s_3	s_m
Decisions (a)	a_1	y_{11}	y_{12}	y_{13}	y_{1m}
	a_2	y_{21}	y_{22}	y_{23}	y_{2m}
	a_3	y_{31}	y_{32}	y_{33}	y_{3m}
	a_n	y_{n1}	y_{n2}	y_{n1}	y_{nm}

The effects are the results of the combination between decisions and the probability associated with the state of nature expected. Therefore it is possible to associate a probability distribution to each decision. This probability can be estimated based on the relative frequencies or purely subjective.

The choice between different decisions becomes the choice between different probability distributions. Investor's preferences are expressed according to the probability distribution associated with the decision. The choice refers to the probability distribution associated with the decision.

By assuming that investors have to choose among different probability distributions, they can be considered as goods and basket of goods in the theory of choice under certainty. By choosing among different probability distributions, it is implicitly assumed that they are characterized by certainty and not uncertainty. In other words, the investor choosing among different probability distributions assumes that they are certain (Cesari 2012a, b; Saltari 2011). Consequently, it is possible to define the problem of choice under uncertainty in the same way as the problem of choice under certainty (Hirshleifer and Riley 1992).

This consequence has a relevant corollary. If the problem of choice under uncertainty can be equated to the problem of choice under certainty, then it is possible to use the analytical instruments of the choices under certainty to the choices under uncertainty. Therefore, the theory of the utility function can be used and therefore the best choice is the one that allows for utility maximization.

In this context, the utility function is used to represent and order the investor's preferences. Decisions are evaluated based on their expected utility only. The preferred decision is the one that has the highest expected utility, or in other terms, maximizing expected utility.

Based on this process, the problem of the choice under uncertainty can be faced as a sorting problem of the decisions based on their associated probability distribution and by using the utility functions. Therefore, there are two variables to be defined (Saltari 2011):

- the probability distribution associated to each decision;
- the utility function to be used.

Formally, the utility of the i-th decision (a_i) can be defined as follows:

$$\mathrm{E}[\mathrm{U}(\mathrm{a_i})] = \pi_1 \mathrm{U}\left(\mathrm{y}_{\mathrm{a_i},1}\right) + \pi_2 \mathrm{U}\left(\mathrm{y}_{\mathrm{a_i},2}\right) + \cdots + \pi_S \mathrm{U}\left(\mathrm{y}_{\mathrm{a_i},S}\right) = \sum_{s=1}^{S} \pi_s \mathrm{U}\left(\mathrm{y}_{\mathrm{a_i},s}\right) \quad (4.5)$$

where:

- a_i: is the i-th decision (for $i = 1, 2, \ldots, n$) among the possible ones $(a \in D)$;
- $U(.)$: is the utility function defined on the effects of the decision;
- $y_{a_i,S}$: are the effects arising from the joint combination between the i-th decision (a_i) and the s-th state of nature (for $s = 1, 2, \ldots, S$);
- π_S: is the probability associated with execution of the s-th state of nature (for $s = 1, 2, \ldots, S$).

Based on Eq. (4.5) the decision *A* is preferred to the decision *B*, only if the expected utility of the decision *A* is greater than the decision *B*, as follows:

$$A \succ B \leftrightarrow \sum_{s=1}^{S} \pi_s U(y_{A,s}) > \sum_{s=1}^{S} \pi_s U(y_{B,s}) \tag{4.6}$$

The possibility to equate the choices under uncertainty to the choices under certainty, with the consequence of using the utility functions, requires the strict adhesion of the same postulates about the investor's behaviour. These postulates define the *axiomatic approach*. Generally, the axioms define the baseline properties of the rational behaviour of the investor. The investor's decisions must be aligned with the axioms. The investors behaviour can be considered as rational only if there is coherence between the investors behaviour and the axioms and his preferences can be classified according to the utility functions (Saltari 2011).

It is worth noting that it is the axiomatic approach that guarantees the coherence and the rigorous nature of the theory of decisions under uncertainty (Von Neumann and Morgenstern 1944; Arrow 1984; Hirshleifer and Riley 1992; Varian 1992; Kreps 1979, 1990; Fishburn and Kochenberger 1979; Heap et al. 1992; Saltari 2011). There are five axioms:

(1) completeness and consistency;
(2) mono-tonicity;
(3) continuity;
(4) independence or substitutability;
(5) reduction.

Among these, the most relevant axioms are those of *continuity* and *independence*. In fact the most relevant criticisms over time have been focused on these axioms.

Axiom 1: Completeness and Consistency
The individual preferences are complete and consistent.

Preferences are *complete* in the sense that, given two probability distributions, there is always the possibility for the investor to find the one he prefers or to express his indifference to the choice. This helps to avoid conditions of doubt between alternatives.

By assuming two probability distribution, *p* e *q*, we have:

$$p \succcurlyeq q \quad or \quad q \succcurlyeq p \quad or \quad p \sim q \tag{4.7}$$

The preferences are also consistent because they are transitive.

If the distribution of probability *p* is preferred, at least to the probability distribution of *q*, and if, in turn, the probability distribution *q* is preferred at least to the probability distribution *r*, then the probability distribution *p* is preferred at least to the probability distribution *r*. If the preferences are not transitive, it is not possible to identify an optimal probability distribution. Formally:

$$p \succcurlyeq q \quad and \quad q \succcurlyeq r \Rightarrow p \succcurlyeq r \tag{4.8}$$

Axiom 2: Mono-tonicity

Given a decision, if two distributions of probability are associated with the same effects, the decision that provides the best effects with the highest level of probability is preferred.

Given two degenerate distributions (δ), the first providing the best result (δ_m) with certainty and the second providing the worst result (δ_p) with certainty and using α and β to indicate the probability including between 0 and 1, and using the symbols o and $\oplus$ to indicate how the distributions are made, we have:

$$\alpha\, o\, \delta_m \oplus (1-\alpha)\, o\, \delta_p \succcurlyeq \beta\, o\, \delta_m \oplus (1-\beta)\, o\, \delta_p \Leftrightarrow \alpha \geq \beta \tag{4.9}$$

Therefore, given two degenerate distributions (or deterministic distribution) (δ_m and δ_p), the preference is function of the probability (α and β) of execution and their combination.

It is worth noting that a non-degenerate distribution can be obtained through their combination with the probability α and β from the two degenerate distributions (δ_m and δ_p). In this sense, $[\alpha\, o\, \delta_m \oplus (1-\alpha) o\, \delta_p]$ and $[\beta\, o\, \delta_m \oplus (1-\beta) o\, \delta_p]$ are two of the possible non-degenerate distributions. For each one there is a probability (α or β) to obtain δ_m and the complement to 1 of this probability ($(1-\alpha)$ or $(1-\beta)$) to obtain δ_p.

Axiom 3: Continuity

The preferences are continuous. Given two positions, it is always possible to construct an intermediate position in function of the investor's preference. Generally, this allows for construction of a sequence of positions that are closer and closer to the preferred position.

Denoted by r the degenerate distributions that assigns probability 1 to the result x so that $r = \delta_x$, with α the probability between 0 and 1, and with P the set of probability distribution. Assuming that the choices can be made between two alternatives:

– alternative 1: achieving a defined result with certainty that is assumed to be equal to δ_x;
– alternative 2: achieving a probability distribution $[\alpha\, o\, \delta_m \oplus (1-\alpha) o\, \delta_p]$ resulting in the degenerate distributions to achieve the best result (δ_m) and the worst result (δ_p) with certainty.

In this case, we have:

$$\forall r \in P\, \exists \alpha \in [0,1] : r \sim \alpha\, o\, \delta_m \oplus (1-\alpha) o\, \delta_p \tag{4.10}$$

The relationship shows that:

– if the probability α is close to 1, the investor prefers the second alternative: the probability distribution provides a higher probability of achieving the result equal to the best result of the degenerate distribution δ_m;

– if the probability α is close to 0, the investor prefers the first alternative: the probability distribution provides a higher probability of achieving the result equal to the worst result of the degenerate distribution δ_p.

Therefore, given the two positions it is always possible to construct an intermediate position based on a specific investor's preference operating through the "mixture". In fact, it is always possible to find a value of α in order to make the investor indifferent between the probability distribution and the certain consequence x.

Therefore, given two alternatives, one excellent and the other one bad, due to the effects of continuity, it is always possible to obtain an equi-preferred value based on their combination through the probability α.

It is worth noting that the value of α is the utility of x: $\alpha = U(x)$. Therefore, for this value of α, the certaint result x defines the certainty equivalent of the probability distribution $[\alpha\, o\, \delta_m \oplus (1-\alpha) o\, \delta_p]$. In other words, x is the result, that if obtained with certainty, it is equivalent to the probability distribution because it has the same utility. Consequently, the probability α can be considered as an investor's risk aversion.

Axiom 4: Independence or Substitutability
The probability distribution p is indifferent to the probability distribution q if any other probability distribution r is considered and the probability α between 0 and 1, we have:

$$\alpha\, o\, p \oplus (1-\alpha) o\, r \sim \alpha\, o\, q \oplus (1-\alpha) o\, r \tag{4.11}$$

Each of the two terms of the relationship can be considered as a particular probability distribution, where with probability α the probability distribution p is obtained (left side of the relationship) or the probability distribution q (right side of the relationship), and with residual probability $(1-\alpha)$ the probability distribution r for both terms is obtained.

The difference between the two sides of the relationship ($[\alpha\, o p \oplus (1-\alpha) o\, r]$ and $[\alpha\, o\, q \oplus (1-\alpha) o\, r]$) is due to the probability distributions p and q. Therefore, the indifference between the two terms of the relationship implies the indifference between p and q. Consequently, there is no change to the preference between two probability distributions when an additional random element is introduced if it is added in the same manner in both probability distributions.

This axiom is also called "axiom of substitutability", because if p and q are indifferent among them, the probability distribution p can be replaced by the probability distribution q, obtaining a new probability distribution that is equivalent.

Axiom 5: Reduction
For the decision-making process only the final probability associated to the effects is relevant. The way in which they are defined is not important.

Given the two degenerate distributions (δ), the first that generates the best result with certainty (δ_m) and the second that generates the worst result with certainty (δ_p), and using α and β to indicate the probability including between 0 and 1, we have:

$$\alpha \, o \left[\beta \, o \, \delta_m \oplus (1-\beta) \, o \, \delta_p\right] \oplus (1-\alpha) o \, \delta_p \sim \gamma \, o \, \delta_m \oplus (1-\gamma) o \, \delta_p \tag{4.12}$$

Therefore, $\gamma = \alpha\beta$ is the total probability to obtain δ_m. Consequently, for the investor it is indifferent to obtain δ_m at one time with probability γ, or twice with probability α and β.

Based on the axiomatic approach, if the kind of the investor's preferences respects these five axioms, then it is possible to use the utility function $(U(.))$. They are used to describe and order the investor's preferences only. Then, the value of the utility function, in absolute terms, is not relevant.

The probability distribution p can be preferred to the probability distribution q only if the expected utility of p is higher than the expected utility of q. Formally:

$$\begin{aligned} p \succ q &\Leftrightarrow \sum U(x)p(x) \geq \sum U(x)q(x) \\ p \succcurlyeq q &\Leftrightarrow \sum U(x)p(x) \geq \sum U(x)q(x) \\ p \sim q &\Leftrightarrow \sum U(x)p(x) \geq \sum U(x)q(x) \end{aligned} \tag{4.13}$$

4.2 Investor Behaviour and Risk Measurement

To further understand the investor's behaviour on risk, there are two basic principles that should be kept in mind: the first one refers to reducing margin utility and the second one is always preferred more money than less.

A fair lottery can be defined as one with an expected value equal to zero. Risky aversion implies that the individual would not accept the fair lottery. Specifically, consider the random payoff x where:

$$x = \begin{cases} x_1 \, with \, probability \, (p) \\ x_2 \, with \, probability \, (1-p) \end{cases}$$

A fair lottery must have an expected value of zero. Therefore, we have:

$$E(x) = x_1 p + x_2(1-p) = 0$$

It implies that:

$$E(x) = x_1 p + x_2(1-p) = 0 \rightarrow x_1 p + x_2(1-p) = 0 \rightarrow \begin{matrix} p = -\frac{x_2}{(x_1 - x_2)} \\ or \\ \frac{x_1}{x_2} = -\frac{(1-p)}{p} \end{matrix} \tag{4.14}$$

One of the useful concepts is the expected utility. Considering a random variable end-of-period wealth W can have n possible value W_i with probability p_i so that $\sum_{i=1}^{n} p_i = 1$. From any wealth outcome W_i the form $U(W_i)$ denotes the utility from any wealth outcome W_i while $E[U(W_i)]$ denotes the expected utility from any wealth outcome W_i.

The relation between the probability p_i associated to any possible wealth outcome W_i, its utility and expected utility, is the following:

$$E[U(W_i)] = \sum_{i=1}^{n} p_i U(W_i) \tag{4.15}$$

Now it is possible to analyse the relationship between the utility functions and the risk in the investor's perspective.

To do this, three baseline assumptions are necessary.

The first baseline assumption states that the decision is measured in terms of wealth as defined in its monetary value (w). Then money is the only argument of the utility function $(U(w))$. This assumption is restrictive and has three main consequences (Saltari 2011):

(1) the decisions have "*mono-dimensional consequences*": the consequences can be shown through a number on the real number line;
(2) the decisions have "*homogeneous consequences*": the heterogeneous behaviour of investors can be neglected;
(3) the function of the expected utility is "*monotonous*".

The second baseline assumption states that the utility function $U(w)$ must be *continuous* and *twice-differentiable*.

The third baseline assumption states that "*more is always preferred to less*". In mathematical terms it implies that the first derivative must always be strictly positive, as follows:

$$\mathrm{U}(\mathrm{w}_2) > U(\mathrm{w}_1) \quad \forall\, \mathrm{w}_2 > \mathrm{w}_1 \rightarrow U'(W) > 0 \tag{4.16}$$

It is true regardless of the behaviour of the investor to risk. It defines the sign of the second derivative and then the shape of their utility function. Specifically, the second derivative can be:

- $U''(w) < 0$: in this case the utility function is *concave* and the investor is *risk averse*;
- $U''(w) > 0$: in this case the utility function is *convex* and the investor is *risk lover*;
- $U''(w) = 0$: in this case the utility function is *linear* and the investor is *risk neutral*.

It is possible to summarize these different behaviours as in Fig. 4.1.

Risk Behaviour

There are three main types of investor's behaviour at risk: (a) risk aversion; (b) risk lover; (c) risk neutral (Arrow 1965, 1971; Campbell 2015; Kreps 2012; Varian 1992)

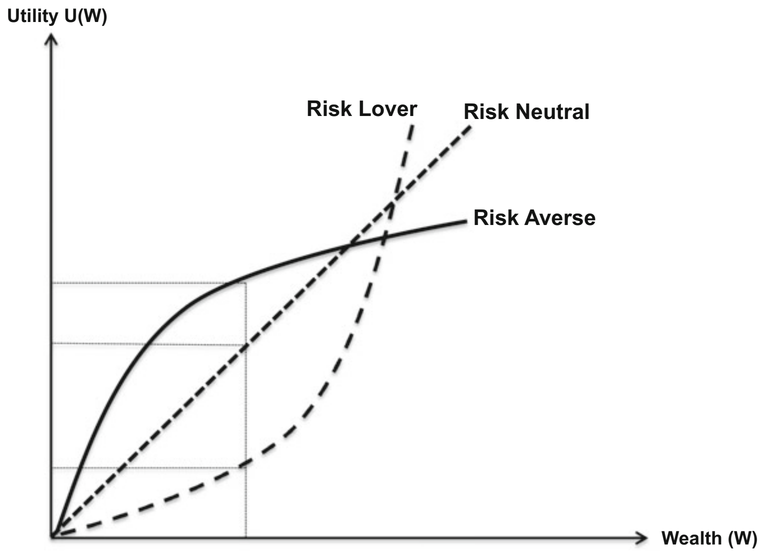


Fig. 4.1 Utility function and risk behaviour

(A) ***Risk Aversion***

In the case of risk aversion, the second derivative of the utility function is negative $(U''(w) < 0)$. It is possible to use the Jensen's inequality. Given a function $f(x)$ where x is a random variable, $f(E(x)) > E[f(x)]$ only if $f(x)$ is a function with a concave shape. If $f(x)$ has a concave shape no one segment is capable of joining two points on the graph in any point. Therefore, the segment $l(x) = a + bx$ lies always above the function: $l(x) = a + bx \geq f(x)$. The same values $l(E(x)) = f(E(x))$ can only be found in the tangent point $E(x)$. Since $l(x)$ is a linear function, it follows that: $l(E(x)) = E(l(x))$. Therefore $f(E(x)) = l(E(x)) = E(l(x)) \geq E(f(x))$. The inequality follows the condition that $l(x) - f(x) \geq 0$. Therefore, by considering the expectation, we have: $E[l(x) - f(x)] \geq 0$, and therefore $E(l(x)) \geq E(f(x))$. Therefore, on the basis of on the Jensen's inequality, we have:

$$\mathrm{E}[\mathrm{U(w)}] < U[\mathrm{E(w)}] \tag{4.17}$$

Assuming that the random variable (w) can assume two values, w_1 and w_2, with probability p_1 and $(1 - p_1)$ respectively. The $E[U(w)]$ and $U[E(w)]$ can be explicated as follows:

$$E[U(w)] = p_1 U(w_1) + (1 - p_1)U(w_2)$$

and

$$U[E(w)] = U[p_1 w_1 + (1 - p_1)w_2]$$

Substituting in Eq. (4.17), we have:

$$\mathrm{p_1 U(w_1) + (1 - p_1)U(w_2)} < U[\mathrm{p_1 w_1 + (1 - p_1)w_2}] \tag{4.18}$$

Equation (4.18) implies the concave shape of the function $U(w)$ in all of domain D, as shown in Fig. 4.2.

In (*Part A*) of the Fig. 4.2, the point A_1 coordinates are $(w_1, U(w_1))$, while the coordinates of the point A_2 are $(w_2, U(w_2))$. The coordinates of the point A^* are:

$$w^* = p_1 w_1 + (1 - p_1)w_2$$

$$U^* = p_1 U(w_1) + (1 - p_1)U(w_2)$$

Therefore, point A^* is positioned on the segment A_1A_2. If the utility function $U(w)$ is concave, for $w = w^*$ the function has an ordinate $(U(w^*))$ that is greater than the point A^* ordinate that is (U^*). It is because the curve is positioned above the segment A_1A_2, and therefore $p_1 U(w_1) + (1 - p_1)U(w_2) < U[p_1 w_1 + (1 - p_1)w_2]$.

In (*Part B*) of Fig. 4.2, the coordinates of the point A_1 are $(w_1; U(w_1))$, while the coordinates of the point A_2 are $(w_2; U(w_2))$.

Considering the probability distribution (p), and associating the portability p_1 to the realization of A_1 and the probability $(1 - p_1)$ to the realization of A_2, the coordinates of the point A^* on the segment A_1A_2 are:

$$A^* \equiv (E(w); E[U(w)])$$

The point $C(w)$ is the certainty equivalent. The distance π measures risk aversion in monetary terms rather than in terms of utility. Specifically, the amount of money π is the known maximum amount that the investor would be willing to pay to avoid the investment. If the investor pays π he receives the expected value of investment for certain.

Therefore, in the case of risk aversion, the utility function is:

(a) strictly increasing and therefore the first derivative is positive $(U'(w) > 0)$;
(b) concave and therefore the second derivative is negative $(U''(w) < 0)$.

It is worth noting that a risk-averse investor is also said to have a diminishing margin utility of wealth. Therefore, each additional unit of wealth adds less to

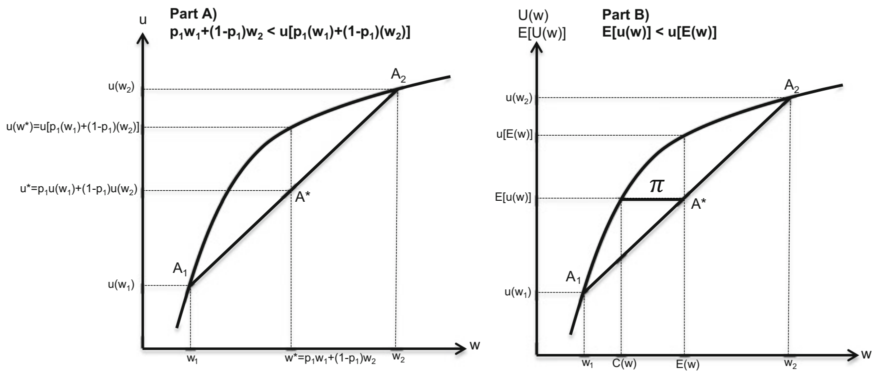


Fig. 4.2 Risk aversion and the concave shape of the utility function

utility, the higher the initial level of wealth and then in mathematical terms it implies that $U''(w) < 0$. The degree of risk aversion is given by the concavity of the utility function and it is equivalent to the absolute size of $U''(w)$. Also, the degree of risk aversion, even for a specific individual, may depend on initial wealth and the size of the investment.

(B) ***Risk Lover***

In the case of a risk lover (or risk seeker), the second derivative of the utility function is positive $(U''(w) > 0)$. In this case, we have:

$$\mathrm{E[U(w)]} > U[\mathrm{E(w)}] \tag{4.19}$$

Assuming that the random variable (w) can assume two values, w_1 and w_2, with probability p_1 and $(1 - p_1)$ respectively. The $E[U(w)]$ and $U[E(w)]$ can be explicated as follows:

$$E[U(w)] = p_1 U(w_1) + (1 - p_1) U(w_2)$$

and

$$U[E(w)] = U[p_1 w_1 + (1 - p_1) w_2]$$

Substituting in Eq. (4.19), we have:

$$\mathrm{p_1 U(w_1)} + (1 - \mathrm{p_1})\mathrm{U(w_2)} > U[\mathrm{p_1 w_1} + (1 - \mathrm{p_1})\mathrm{w_2}] \tag{4.20}$$

Equation (4.20) implies that the convex shape of the function $U(w)$ in all domain D, as shown in Fig. 4.3.

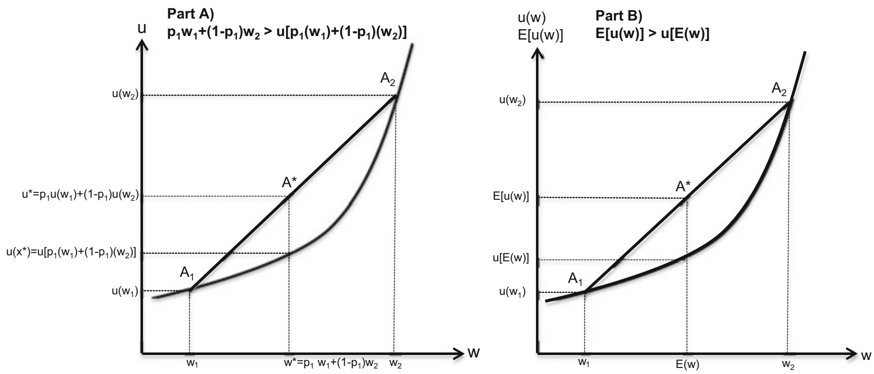


Fig. 4.3 Risk lover and the convex shape of the utility function

In (*Part A*) of Fig. 4.3, the coordinates of the point A_1 are $(w_1, u(w_1))$ while the coordinates of the point A_2 are $(w_2, u(w_2))$. The coordinates of point A^* are:

$$w^* = p_1 w_1 + (1 - p_1) w_2$$

$$U^* = p_1 U(w_1) + (1 - p_1) U(w_2)$$

Point A^* is positioned on the segment $A_1 A_2$.

If the utility function $U(w)$ is convex, for $w = w^*$ the function has an ordinate $(U(w^*))$ that is lower than the point A^* ordinate that is (U^*). This is because the curve is positioned below the segment $A_1 A_2$, and therefore $p_1 U(w_1) + (1 - p_1) U(w_2) > U[p_1 w_1 + (1 - p_1) w_2]$.

In (*Part B*) of Fig. 4.3, the coordinates of point A_1 are $(w_1; U(w_1))$, while the coordinates of point A_2 are $(w_2; U(w_2))$.

Considering the probability distribution (p), and associating probability p_1 to the realization of A_1 and the probability $(1 - p_1)$ to the realization of A_2, the coordinates of point A^* on the segment $A_1 A_2$ are:

$$A^* \equiv (E(w); E[U(w)])$$

Therefore, in the case of *risk seeking*, the utility function is:

(a) strictly increasing and therefore the first derivative is positive $(U'(w) > 0)$;
(b) convex and therefore the second derivative is negative $(U''(w) > 0)$.

(C) ***Risky Neutral***

In the case of risk neutral (or indifferent to risk), the second derivative of the utility function is equal to zero $(U''(w) = 0)$. In this case we have:

$$E[U(w)] = U[E(w)] \tag{4.21}$$

Assuming that the random variable (w) can assume two values, w_1 and w_2, with probability p_1 and $(1 - p_1)$ respectively. The $E[U(w)]$ and $U[E(w)]$ can be explicated as follows:

In this case, $U(w) = a + bw$ with $b > 0$. The utility function is unique unless it has a linear transformation. Therefore, we have $a = 0$ and $b = 1$. Therefore, $U(w) = w$ and consequently:

$$p_1 U(w_1) + (1 - p_1) U(w_2) = U[p_1 w_1 + (1 - p_1) w_2] \tag{4.22}$$

Equation (4.22) implies the linearity of the function $U(w)$ in all of the domain D, as shown in Fig. 4.4.

Therefore, in the case of risk neutral, the utility function is:

(a) strictly increasing (non satiety) and therefore the first derivative is positive $(U'(w) > 0)$;
(b) linear and therefore the second derivative is null $(U''(w) = 0)$.

Risk Aversion Measurement

Having defined the relationship between utility function $U(w)$ and the risk behaviour of the investor, the main problem is the measurement of risk aversion. There are two main instruments that can be used (Pratt 1964; Arrow 1965):

– the *Absolute Risk Aversion (ARA)*, and the corresponding *Absolute Risk Tolerance (ART)*;
– the *Relative Risk Aversion (RRA)*, and the corresponding *Relative Risk Tolerance (RRT)*.

(A) ***Absolute Risk Aversion***

The Absolute Risk Aversion *(ARA)* (λ) is defined by the ratio between the first and second derivative of the utility function $(U(.))$ as defined in the current monetary wealth (w_0) (it is a monetary amount) so that $U(w_0)$, as follows:

$$\lambda_{(w_0)} = -\frac{U''(w_0)}{U'(w_0)} \tag{4.23}$$

The coefficient $\lambda_{(w_0)}$ measures the concavity of the utility function $(U(w_0))$ in the point w_0. It is the Arrow (1970)-Pratt (1964) that measures the absolute (local) risk aversion. The measure of risk aversion is defined as "*local*" because it is a function of the initial level of wealth. The larger $\lambda_{(w_0)}$, the greater the degree of risk aversion.

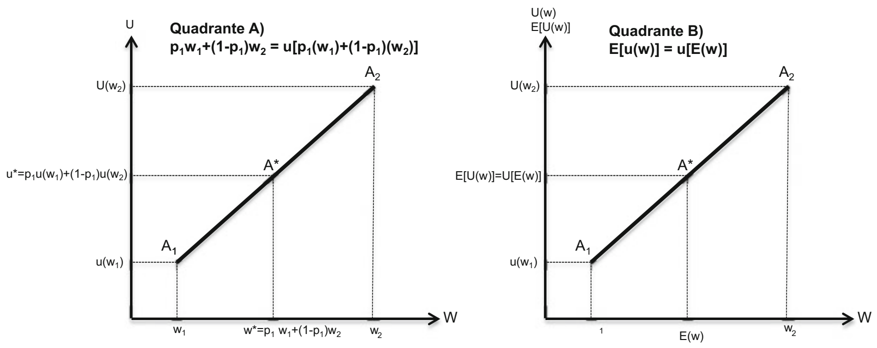


Fig. 4.4 Risk neutral and the linearity of the utility function

The first derivative is always positive (function increasing), while the second derivative is negative if the function is concave (risk aversion) and positive if the function is convex (risk seeking). Specifically:

– *risk aversion*: the utility function $(U(w_0))$ is concave. Consequently, the coefficient $\lambda_{(w_0)}$ is positive:

$$\lambda_{(w_0)} = -\frac{(-U''(w_0))}{U'(w_0)} = \frac{U''(w_0)}{U'(w_0)} > 0 \quad \forall w_0 \in D \tag{4.24}$$

The coefficient $\lambda_{(w_0)}$ is always positive. It increases as the second derivative increases and therefore with an increase in the absolute risk aversion of the investor;

– *risk lover*: the utility function $(U(w_0))$ is convex. Consequently, the coefficient $\lambda_{(w_0)}$ is negative:

$$\lambda_{(w_0)} = -\frac{(+U''(w_0))}{U'(w_0)} = -\frac{U''(w_0)}{U'(w_0)} < 0 \quad \forall w_0 \in D \tag{4.25}$$

Therefore, the higher the coefficient $\lambda_{(w_0)}$, the higher the absolute risk aversion of the investor.

In comparative terms, if the investor *A* is more absolute risk aversion than the investor *B*, the coefficient of investor *A* $(\lambda^{A}_{(w_0)})$ is higher than the coefficient of the investor B $(\lambda^{B}_{(w_0)})$ for each level of wealth (w_0), as follows:

$$\lambda^{A}_{(w_0)} > \lambda^{B}_{(w_0)}$$

The coefficient $\lambda_{(w_0)}$ does not change in the case of linear transformation of the utility function $(U(w_0))$. It is calculated according to the ratio between the first and the second derivatives. Therefore, by considering a utility function:

$$V(w_0) = aU(w_0) + b$$

We have:

$$\lambda^{V}_{(w_0)} = -\frac{V''(w_0)}{V'(w_0)} = -\frac{aU''(w_0)}{aU'(w_0)} = -\frac{U''(w_0)}{U'(w_0)} = \lambda^{U}_{(w_0)} \tag{4.26}$$

The inverse of the Absolute Risk Aversion is defined *Absolute Risk Tolerance (ART)* (τ). Formally, it is equal to:

$$\tau_{(w_0)} = \frac{1}{\lambda_{(w_0)}} = -\frac{U'(w_0)}{U''(w_0)} \tag{4.27}$$

Therefore, the higher the absolute risk aversion of the investor, the lower the absolute risk tolerance, and also, the lower the absolute risk aversion of the investor, the higher the absolute risk tolerance.

Therefore, the higher the coefficient $\tau_{(w_0)}$, the higher the absolute risk tolerance of the investor.

In comparative terms, if the investor A is more absolute risk tolerant than investor B, the coefficient of the investor A $(\tau^{A}_{(w_0)})$ is higher than the coefficient of the investor B $(\tau^{B}_{(w_0)})$, for each level of wealth (w_0), as follows:

$$\tau^{A}_{(w_0)} > \tau^{B}_{(w_0)}$$

(B) ***Relative Risk Aversion***

The Relative Risk Aversion (RRA) (ρ) based on the current monetary wealth (w_0) (it is monetary amount), is defined as follows:

$$\rho_{(w_0)} = -w_o \frac{U''(w_0)}{U'(w_0)} \tag{4.28}$$

The coefficient $\rho_{(w_0)}$ is expressed in the same unit of wealth measure (w_0). It also implies stabilization of the coefficient $\lambda_{(w_0)}$.

As in the case of the coefficient $\lambda_{(w_0)}$, the first derivative is always positive (function increasing), while the second derivative is negative if the function is concave (risk aversion) and positive if the function is convex (risk seeking), we have:

– *risk aversion*: the utility function $(U(w_0))$ is concave in all dominions (D). Consequently, the coefficient $\rho_{(w_0)}$ is positive:

$$\rho_{(w_0)} = -w_o \frac{(-U''(w_0))}{U'(w_0)} = w_o \frac{U''(w_0)}{U_{'}(w_0)} > 0 \quad \forall w_0 \in D \tag{4.29}$$

The coefficient $\rho_{(w_0)}$ is always positive. It increases with an increase in the second derivative and therefore with an increase in the relative risk aversion of the investor. Generally, $\lambda_{(w_0)}$ and $\rho_{(w_0)}$ are measures of how the investor's risk preferences change with changes in wealth around the initial "local" level of wealth.

– *risk lover*: the utility function $(U(w_0))$ is convex in all dominions (D). Consequently, the coefficient $\rho_{(w_0)}$ is negative:

$$\rho_{(w_0)} = -w_o \frac{(+U''(w_0))}{U'(w_0)} = -w_o \frac{U''(w_0)}{U'(w_0)} < 0 \quad \forall w_0 \in D \tag{4.30}$$

Therefore, the higher the coefficient $\rho_{(w_0)}$, the higher the relative risk aversion of the investor.

In comparative terms, if the investor *A* is more relative risk aversion than the investor *B*, the coefficient of the investor *A* $(\rho^A_{(w_0)})$ is higher than the coefficient of the investor B $(\rho^B_{(w_0)})$, for each level of wealth (w_0), as follows:

$$\rho^A_{(w_0)} > \rho^B_{(w_0)}$$

Also, in this case, the coefficient $\rho_{(w_0)}$ does not change in the case of linear transformation of the utility function $(U(w_0))$. It is calculated on the ratio between the first and the second derivatives. Therefore, by considering a utility function:

$$V(w_0) = aU(w_0) + b$$

We have:

$$\rho^V_{(w_0)} = -w_o \frac{V''(w_0)}{V'(w_0)} = -w_o \frac{aU''(w_0)}{aU'(w_0)} = -w_o \frac{U''(w_0)}{U'(w_0)} = \rho^U_{(w_0)} \tag{4.31}$$

The inverse of Relative Risk Aversion is defined *Relative Risk Tolerance (RRT)* $(\mathcal{T})$ as follows:

$$\mathcal{T}_{(w_0)} = \frac{1}{\rho_{(w_0)}} = -\frac{1}{w_o}\frac{U'(w_0)}{U''(w_0)} \tag{4.32}$$

Therefore, the higher the relative risk aversion of the investor, the lower the relative risk tolerance, and in the same way, the lower the relative risk aversion of the investor, the higher the relative risk tolerance. Therefore, the higher the coefficient $\mathcal{T}_{(w_0)}$, the higher the relative risk tolerance of the investor.

In comparative terms, if the investor *A* is more relative risk tolerant than the investor *B*, the coefficient of the investor *A* $(\mathcal{T}^{A}_{(w_0)})$ is higher than the coefficient of the investor *B* $(\mathcal{T}^{B}_{(w_0)})$, for each level of wealth (w_0), as follows:

$$\mathcal{T}^{A}_{(w_0)} > \mathcal{T}^{B}_{(w_0)}$$

4.3 Utility Functions

Different mathematical functions generate different implications for the form of risk aversion.

According to the dynamic of the coefficients of Absolute and Relative Risk Aversion, the utility functions can be classified into five main categories:

- *Constant Absolute Risk Aversion (CARA)*: the utility function is characterized by a constant Absolute Risk Aversion (ARA) (λ). In this class the most useful utility function is the negative exponential.
- *Constant Relative Risk Aversion (CRRA)*: the utility function is characterized by a constant Relative Risk Aversion (RRA) (ρ). In this class the most useful utility functions are power, logarithmic, and quadratic.
- *Hyperbolic Absolute Risk Aversion (HARA)*: the utility function is characterized by a hyperbolic Absolute Risk Aversion (ARA) (λ).
- *Hyperbolic Relative Risk Aversion (HRRA)*: the utility function is characterized by a hyperbolic Relative Risk Aversion (RRA) (ρ).
- *Decreasing Absolute Risk Aversion (DARA)*: it generalizes the class of utility functions HARA.

Among all of the utility functions that can be used, some of them can be defined as “standard” because they are the most common ones used in literature.

Linear
The function can be defined as follows:

$$U(w) = a + bw \quad b > 0 \tag{4.33}$$

where *a* is an arbitrary constant.

The first derivative is always positive and therefore the function is always increasing. Indeed:

$$U'(w) = b$$

The second derivative is equal to zero. It implies that there is no concavity and then the investor is risk neutral.

Therefore, it implies that the risk neutral is equivalent to the linear utility of the function.

Power
The function can be defined as follows:

$$U(w) = \frac{1}{a} w^{a} \quad a\langle 1; a \neq 0; w\rangle 0 \tag{4.34}$$

The first and second derivatives are the following:

$$U'(w) = \frac{1}{a} a w^{a-1} = w^{a-1}$$

$$U''(w) = (a - 1) w^{a-2}$$

The absolute risk aversion $(\lambda_{(w)})$ and the relative risk aversion $(\rho_{(w)})$ are the following:

$$\lambda_{(w)} = -\frac{U''(w)}{U'(w)} = -\left[\frac{(a-1)w^{a-2}}{w^{a-1}}\right] = -\left[\frac{(a-1)w^{a-1}w^{-1}}{w^{a-1}}\right] = -\left[\frac{a-1}{w}\right] = \frac{1-a}{w}$$

$$\rho_{(w)} = -w\frac{U''(w)}{U'(w)} = -w\left(\frac{a-1}{w}\right) = 1 - a$$

Therefore, the absolute risk aversion $(\lambda_{(w)})$ decreases as wealth (w) increases. The relative risk aversion $(\rho_{(w)})$ is independent of the level of wealth (w) and therefore it is constant. For this reason, this function is classified according to function groups *CRRA*.

Also by considering that:

$$\ln\left[\mathrm{U}'(\mathrm{w})\right] = \ln\left[w^{a-1}\right] = (a-1)\ln\left[w\right]$$

the $(a-1)$ can be considered as the elasticity of marginal utility with respect to wealth.

Logarithmic
The function can be defined as follows:

$$\mathrm{U}(\mathrm{w}) = \mathrm{a}\ \ln w + b \quad w > 0;\ a, b\ \textit{arbitrary constants} \tag{4.35}$$

The basic assumption is that the increase in utility is directly proportional with wealth (w) increases and it is inversely proportional to initial wealth.

The first and second derivatives are the following:

$$\mathrm{U}'(\mathrm{w}) = \mathrm{a}\frac{1}{w} = aw^{-1}$$

$$\mathrm{U}''(\mathrm{w}) = \mathrm{a}(-1)w^{-2} = -aw^{-2}$$

The absolute risk aversion $(\lambda_{(\mathrm{w})})$ and the relative risk aversion $(\rho_{(\mathrm{w})})$ are the following:

$$\lambda_{(\mathrm{w})} = -\frac{\mathrm{U}''(\mathrm{w})}{\mathrm{U}'(w)} = -\left[\frac{-aw^{-2}}{aw^{-1}}\right] = w^{-1} = \frac{1}{w}$$

$$\rho_{(\mathrm{w})} = -\mathrm{w}\frac{\mathrm{U}''(w)}{\mathrm{U}'(w)} = -w\left[\frac{-aw^{-2}}{aw^{-1}}\right] = -w\left[-w^{-1}\right] = -w\left[-\frac{1}{w}\right] = 1$$

Therefore, the absolute risk aversion $(\lambda_{(\mathrm{w})})$ decreases as wealth (w) increases. The Relative Risk Aversion $(\rho_{(\mathrm{w})})$ is independent on the level of wealth (w) and therefore it is constant. For this reason, this function is classified according to function groups *CRRA*.

Negative Exponential
The function can be defined as follows:

$$\mathrm{U}(\mathrm{w}) = \mathrm{a}\left(1 - e^{-\frac{w}{a}}\right) \quad a > 0 \tag{4.36}$$

This is a superiorly limited exponential function. The parameter $\underline{a}$ is the upper extremity and therefore it represents the maximum potentiality. Indeed, the function for $\mathrm{U}(\mathrm{w}) = a$ has a horizontal asymptote:

$$\lim_{w\to+\infty}[\mathrm{U}(\mathrm{w})] = \lim_{w\to+\infty}\left[\mathrm{a}\left(1 - e^{-\frac{w}{a}}\right)\right] = \mathrm{a}(1 - e^{-\infty})$$
$$= \mathrm{a}\left(1 - \frac{1}{e^{\infty}}\right) = \mathrm{a}(1-0) = a$$

The first and second derivatives are the following:

$$\mathrm{U}'(\mathrm{w}) = \mathrm{a}\left(-\left(-\frac{1}{a}\right)e^{-\frac{w}{a}}\right) = \mathrm{a}\left(\frac{1}{a}e^{-\frac{w}{a}}\right) = e^{-\frac{w}{a}}$$

$$\mathrm{U}''(\mathrm{w}) = \left(-\frac{1}{a}\right)e^{-\frac{w}{a}}$$

The absolute risk aversion $(\lambda_{(\mathrm{w})})$ and the relative risk aversion $(\rho_{(\mathrm{w})})$ are the following:

$$\lambda_{(\mathrm{w})} = -\frac{\mathrm{U}''(\mathrm{w})}{\mathrm{U}'(w)} = -\left[\frac{\left(-\frac{1}{a}\right)e^{-\frac{w}{a}}}{e^{-\frac{w}{a}}}\right] = \frac{1}{a}$$

$$\rho_{(\mathrm{w})} = -\mathrm{w}\frac{\mathrm{U}''(w)}{\mathrm{U}'(w)} = -w\left[\frac{\left(-\frac{1}{a}\right)e^{-\frac{w}{a}}}{e^{-\frac{w}{a}}}\right] = -w\left(-\frac{1}{a}\right) = w\frac{1}{a}$$

In this case, the relative risk aversion $(\rho_{(\mathrm{w})})$ increases as wealth (w) increases. Differently, the absolute risk aversion $(\lambda_{(\mathrm{w})})$ is independent on the wealth (w) and therefore it is constant. For this reason, this function is classified into the group of functions *CARA*.

Quadratic

The function can be defined as follows:

$$\mathrm{U}(\mathrm{w}) = \mathrm{w} - \frac{\mathrm{a}}{2}\mathrm{w}^2 \quad \mathrm{a} > 0 \tag{4.37}$$

The first and second derivatives are the following:

$$\mathrm{U}'(\mathrm{w}) = 1 - \frac{\mathrm{a}}{2}2\mathrm{w}^{2-1} = 1 - aw$$

$$\mathrm{U}''(\mathrm{w}) = -\mathrm{a}$$

The absolute risk aversion $(\lambda_{(\mathrm{w})})$ and the relative risk aversion $(\rho_{(\mathrm{w})})$ are the following:

$$\lambda_{(\mathrm{w})} = -\frac{\mathrm{U}''(\mathrm{w})}{\mathrm{U}'(w)} = -\left[\frac{-\mathrm{a}}{1 - aw}\right] = \frac{a}{1 - aw}$$

$$\rho_{(\mathrm{w})} = -\mathrm{w}\frac{\mathrm{U}''(w)}{\mathrm{U}'(w)} = -w\left[\frac{-\mathrm{a}}{1 - aw}\right] = w\left(\frac{a}{1 - aw}\right) = \frac{wa}{1 - aw}$$

It is relevant to know that the first derivative must be positive, and therefore the quadratic is only defined for a value of wealth (w) equal to:

$$1 - aw > 0 \rightarrow w < \frac{1}{a} \tag{4.38}$$

This is known as the "bliss point".

Marginal utility is linear in wealth and this can sometimes be a useful property.

It is worth noting that the absolute risk aversion $(\lambda_{(w)})$ and relative risk aversion $(\rho_{(w)})$ are not constant and they are both function of wealth (w). Specifically, the absolute and relative risk aversion increases with the increase in wealth (w). Undoubtedly, it seems a counter-intuitive result. However, this utility function has two main advantages that justify its preference (Cesari 2012b; Castellani et al. 2005).

The first advantage of utility function is that the quadratic utility function can be considered as the approximation to the second order of any utility function based on the Taylor's polynomial development. Therefore, it can be considered as a generalization of the specific utility functions used by the investor.

Assuming that the investor has a wealth equal to k and he is characterized by a utility function $U(w)$ derivable infinitely. Therefore, it can be developed in a Taylor series around the point w_0 as follows:

$$\mathrm{U(w_0)} = \sum_{\mathrm{n=0}}^{\infty} \frac{\mathrm{U^{(n)}(w_0)}}{\mathrm{n!}} (\mathrm{w - w_0})^{\mathrm{n}} \rightarrow (\mathrm{k + k_1}) = \sum_{\mathrm{n=0}}^{\infty} \frac{\mathrm{U^{(n)}(k)}}{\mathrm{n!}} \mathrm{k_1^n}$$

where k_1 is the increase of initial wealth (k).

Since the terms of degree are higher than the second, it represents an infinitesimal of a higher order than k_1^2, for small increments in k_1 their contribution can be neglected.

It generates an equality to be considered as an approximation of the second order, as follows (Castellani et al. 2005):

For $n = 2$, we have:

$$U(k + k_1) = \sum_{n=0}^{2} \frac{U^{(n)}(k)}{n!} k_1^n$$

and therefore:

$$U(k + k_1) = \frac{U^{(0)}(k)}{0!} k_1^0 + \frac{U'(k)}{1!} k_1^1 + \frac{U''(k)}{2!} k_1^2$$

by convention:

$$0! = 1; 1! = 1; 2! = 2 \cdot 1 = 2$$

We have:

$$\mathrm{U}(\mathrm{k}+\mathrm{k}_1) = \mathrm{U}(\mathrm{k}) + \mathrm{U}'(\mathrm{k})k_1^1 + \frac{\mathrm{U}''(\mathrm{k})}{2}\mathrm{k}_1^2$$

Subtracting $U(k)$ and dividing by $U'(k)$ both terms, we have:

$$U(k+k_1) - U(k) = U(k) + U'(k)k_1 + \frac{U''(k)}{2}k_1^2 - U(k)$$

and therefore:

$$\frac{U(k+k_1)}{U'(k)} - \frac{U(k)}{U'(k)} = \frac{U'(k)}{U'(k)}k_1 + \frac{1}{2}\frac{U''(k)}{U'(k)}k_1^2$$

$$\frac{\mathrm{U}(\mathrm{k}+\mathrm{k}_1)}{\mathrm{U}'(\mathrm{k})} - \frac{\mathrm{U}(\mathrm{k})}{\mathrm{U}'(\mathrm{k})} = \mathrm{k}_1 + \frac{1}{2}\frac{\mathrm{U}''(\mathrm{k})}{\mathrm{U}'(\mathrm{k})}\mathrm{k}_1^2$$

Placing:

$$\mathrm{a} = \frac{1}{\mathrm{U}'(\mathrm{k})} \rightarrow \mathrm{U}'(\mathrm{k}) > 0 \rightarrow a > 0$$

$$\mathrm{b} = -\frac{\mathrm{U}(\mathrm{k})}{\mathrm{U}'(\mathrm{k})}$$

The first term of equation is equal to:

$$\frac{\mathrm{U}(\mathrm{k}+\mathrm{k}_1)}{\mathrm{U}'(\mathrm{k})} - \frac{\mathrm{U}(\mathrm{k})}{\mathrm{U}'(\mathrm{k})} = \mathrm{aU}(\mathrm{k}+\mathrm{k}_1) + \mathrm{b}$$

It is a positive linear transformation of the equation $U(k+k_1)$. Therefore, it is equivalent as follows:

$$\frac{\mathrm{U}(\mathrm{k}+\mathrm{k}_1)}{\mathrm{U}'(\mathrm{k})} - \frac{\mathrm{U}(\mathrm{k})}{\mathrm{U}'(\mathrm{k})} = \mathrm{aU}(\mathrm{k}+\mathrm{k}_1) + \mathrm{b} \sim \mathrm{U}(\mathrm{k}+\mathrm{k}_1)$$

The second term of equation can be rewritten on the basis of the absolute risk aversion coefficient as follows:

$$\lambda_{(k)} = -\frac{U''(k)}{U'(k)} \rightarrow k_1 + \frac{1}{2}\frac{U''(k)}{U'(k)}k_1^2 = k_1 - \frac{1}{2}\lambda_{(k)}k_1^2$$

Replacing the first and second terms in the equation, we have:

$$U(k + k_1) = k_1 - \frac{1}{2}\lambda_{(k)}k_1^2$$

In order to represent the absolute risk aversion coefficient in the same unit measures of k, the absolute risk tolerance coefficient $\left(\tau_{(w_0)} = \frac{1}{\lambda_{(w_0)}}\right)$ can be used. In this case Eq. (4.44) can be rewritten as follows:

$$U(k + k_1) = k_1 - \frac{1}{2\tau_{(w_0)}}k_1^2 \tag{4.39}$$

Equation (4.39) can be considered as a generic quadratic utility function.

The absolute risk tolerance coefficient $\left(\tau_{(w_0)}\right)$ is greater, the higher the value of k around which the Taylor series is developed.

In order to be certain of the goodness of the approximation carried out, we should assume that the absolute risk tolerance $\tau_{(w_0)}$ is much greater than k_1 (Castellani et al. 2005):

$$\tau_{(w_0)} \gg k_1$$

The second advantage of utility function is that the quadratic utility function is coherent with the mean-variance criteria. Consequently, the quadratic utility function allows for compatibility between the utility function approach and the mean-variance approach.

The investor with a quadratic utility function chooses according to criteria coherent with mean and variance. He chooses based on mean $E[u(w)]$ and variance $Var(w)$.

Specifically, if the investor maximizes expected utility of end-of-period portfolio wealth, we can see that this is equivalent to maximising a function of expected portfolio returns and portfolio variance providing:

(a) either utility is quadratic,
(b) or portfolio returns are normally distributed and utility is concave.

Assuming that an initial wealth equal to w_0 and the stochastic portfolio return is equal to R_p. At the end of the period, we have:

$$w = w_0\left(1 + R_p\right) \leftrightarrow U(w) = U\left[w_0\left(1 + R_p\right)\right]$$

Expanding $U(R_p)$ in a Taylor series around the mean of $R_p \approx \mu_p$, we have:

$$U(R_p) = U(\mu_p) + (R_p - \mu_p)U'(\mu_p) + \frac{1}{2}(R_p - \mu_p)^2 U''(\mu_p) + \textit{higher order terms}$$

Considering that $E(R_p - \mu_p) = 0$ and $E(R_p - \mu_p)^2 = \sigma_P^2$. Therefore, we have:

$$U(R_p) = U(\mu_p) + \frac{1}{2}\sigma_P^2 U''(\mu_p) + E[\textit{higher} - \textit{order terms}]$$

If the utility function is quadratic, then the derivative greater than the second are equal to zero.

If the return are normally distributed, we have:

(a) $E[(R_p - \mu_p)^n] = 0$ for n odd, and
(b) $E[(R_p - \mu_p)^n]$ for n even is a function only of the variance σ_P^2.

Therefore, in both cases (utility is quadratic or portfolio returns are normally distributed) so $E[U(R_p)]$ is a function of only the mean μ_p and the variance σ_P^2.

It is worth noting that until a specific utility function is specified, the functional relationship between $E[U(R_p)]$ and (μ_p, σ_P^2) is not known and hence it is impossible to determine whether or not there is an analytic closed-form solution for asset demands (Cuthbertson and Nitzsche 2014).

This result can be verified through a simple reasoning (Cesari 2012b). Assuming a random variables $\tilde{W}$. Assuming that investor uses the quadratic utility function as defined in Eq. (4.37) as follows:

$$\mathrm{U(w) = w - \frac{a}{2}w^2 \quad a > 0}$$

The expected utility for the variable can be computed as follows:

$$\mathrm{E}\left[\mathrm{U}(\widetilde{W})\right] = \mathrm{E}\left[\widetilde{\mathrm{W}} - \frac{\mathrm{a}}{2}\widetilde{\mathrm{W}}^2\right] = \mathrm{E}\left(\widetilde{\mathrm{W}}\right) - \frac{\mathrm{a}}{2}\mathrm{E}\left(\widetilde{\mathrm{W}}^2\right)$$

Based on the following property of variance:

$$Var(x) = E(X^2) - E(X)^2 \quad \rightarrow \quad E(X^2) = Var(x) + E(X)^2$$

the equation can be rewritten as follows:

$$\mathrm{E}\left(\widetilde{\mathrm{W}}^2\right) = \mathrm{Var}\left(\widetilde{\mathrm{W}}\right) + \mathrm{E}\left(\widetilde{\mathrm{W}}\right)^2$$

and by substituting we have:

$$E\left[U\left(\widetilde{W}\right)\right] = E\left(\widetilde{W}\right) - \frac{a}{2}\left[Var\left(\widetilde{W}\right) + E\left(\widetilde{W}\right)^2\right]$$

This equation shows how the investor chooses on the basis of two variables: mean $E(\widetilde{W})$ and variance $Var(\widetilde{W})$. The relationship can be defined as follows:

$$E\left[U\left(\widetilde{W}\right)\right] = \Psi\left[\left(\overset{+}{E\left(\widetilde{W}\right)}\right); \left(\overset{-}{Var\left(\widetilde{W}\right)}\right)\right] \quad (4.40)$$

The function Ψ is called *mean-variance indirect utility (or indirect mean-variance utility function*), and it shows that:

- the mean has a positive effect: the mean increases, increase the level of welfare;
- the variance has a negative effect: the variance increases, decrease the level of welfare.

Also, the function Ψ shows that as the means are equal, the investor's choices are based on risk.

By considering two random variables $(\widetilde{W_1}; \widetilde{W_2})$,

Therefore, if there are two random variables $(\widetilde{W_1}; \widetilde{W_2})$ the investor choose on the basis of Eq. (4.40) as follows:

$$\begin{aligned} E\left(\widetilde{W_1}\right) = E\left(\widetilde{W_2}\right) \quad &\text{and} \quad Var\left(\widetilde{W_1}\right) < Var\left(\widetilde{W_2}\right) \Rightarrow \widetilde{W_1} \gtrsim \widetilde{W_2} \\ E\left(\widetilde{W_1}\right) > E\left(\widetilde{W_2}\right) \quad &\text{and} \quad Var\left(\widetilde{W_1}\right) = Var\left(\widetilde{W_2}\right) \Rightarrow \widetilde{W_1} \gtrsim \widetilde{W_2} \end{aligned}$$

Hyperbolic Absolute Risk Aversion (HARA) and Decreasing Absolute Risk Aversion (DARA)

Generalisation of the utility functions that can be classified in the *Hyperbolic Absolute Risk Aversion (HARA*) class, is the following:

$$U(w) = \frac{1}{a_2 - 1}(a_1 + a_2 w)^{1-\frac{1}{a_1}} \quad w > -\frac{a_1}{a_2}; a_2 \neq 0 \quad (4.41)$$

where a_1 and a_2 are arbitrary constants able to guarantee positive value of absolute risk aversion $(\lambda_{(w)})$.

Utility functions grouped in this class have the following absolute risk aversion $(\lambda_{(w)})$:

$$\lambda_{(w)} = \frac{1}{a_1 + a_2 w} \quad (4.42)$$

It is worth noting that the utility function power, negative exponential and quadratic can be grouped in the HARA class. Indeed they use a hyperbolic absolute risk aversion as follows:

- the absolute risk aversion of the power utility function is obtained by assuming $a_1 = 0$ and $a_2 = \frac{1}{1-a}$ as follows:

$$\lambda_{(w)} = \frac{1}{a_1 + a_2 w} \rightarrow \lambda_{(w)} = \frac{1}{\frac{w}{1-a}} = \frac{1-a}{w}$$

- the absolute risk aversion of the negative exponential utility function is obtained by assuming $a_1 = a$ and $a_2 = 0$ as follows:

$$\lambda_{(w)} = \frac{1}{a_1 + a_2 w} \rightarrow \lambda_{(w)} = \frac{1}{a}$$

- the absolute risk aversion of the quadratic utility function is obtained by assuming $a_1 = \frac{1}{a}$ and $a_2 = -1$ as follows:

$$\lambda_{(w)} = \frac{1}{a_1 + a_2 w} \rightarrow \lambda_{(w)} = \frac{1}{\frac{1}{a} - w} = \frac{1}{\frac{1-aw}{a}} = \frac{\mathrm{a}}{1 - aw}$$

The utility functions classified in the HARA group can be generalized in the class of *Decreasing Absolute Risk Aversion (DARA).* Generally, the HARA class uses a hyperbolic absolute risk aversion.

The generalization of the utility functions able to be grouped in DARA class, is the following:

$$\mathrm{U(w)} = \left(w + \frac{H}{G}\right)^{1-\frac{1}{G}} \tag{4.43}$$

where the absolute risk aversion is the following:

$$\lambda_{(w)} = \frac{1}{\mathrm{G}w + H} = \frac{1}{\mathrm{G}} \cdot \frac{1}{\mathrm{w} + \frac{H}{G}} > 0 \tag{4.44}$$

It is worth noting that by changing the value of G and H in the absolute risk aversion $(\lambda_{(w)})$ it is possible to obtain other classes. Indeed:

- for $H = 0; G > 0$: utility functions grouped in CRRA class are obtained;
- for $G = 0; H > 0$: utility functions grouped in CARA class are obtained.

The DARA generalization is the following:

$$\frac{\partial \lambda_{(w)}}{\partial (w)} = -\frac{U'''(w) \cdot U'(w) - [U''(w)]^2}{[U'(w)]^2} < 0 \tag{4.45}$$

In this case the third derivative is positive $(U'''(w) > 0)$. It implies that as wealth increases, there is less absolute risk aversion.

In order to choose the utility function to be implemented, the following criteria can be used (Litner 1970):

- short-term decisions: a neutral risk can be assumed. Therefore, a zero absolute risk aversion is assumed $(\lambda_{(w)} = 0)$ and therefore the linear utility function can be used;
- medium-term decision: a constant risk can be assumed. Therefore, a constant absolute risk aversion can be assumed $(\lambda_{(w)} = a)$ and therefore the negative exponential utility function can be used;
- long-term period decision: a utility function in the DARA class can be assumed.

4.4 Utility Functions and Portfolio Choices

The portfolio choices (or portfolio selection) is a problem of the wealth allocation between different investment assets (Ingersoll 1987; Gravelle and Rees 1992; Markowitz 1952, 1956, 1959; Tobin 1958; Litner 1965a, b).

In this context, the portfolio choices will be analysed based on the two main criteria:

- the utility functions criteria;
- the mean-variance criteria.

This paragraph will focus on the first, while the next paragraph will focus on the second.

The aim of investors is to define the "optimum portfolio" capable of achieving maximisation of the expected utility.

The definition of this aim is simplification because the aim is based on a single parameter. However, it presents a disadvantage because it represents an over-simplification that ends up losing several key elements. In a context of risk aversion, in order to reduce the effects of the over-simplification, the objective of maximization of the expected utility can be declined in two sub-objectives:

- profit maximization;
- risk minimization.

The utility of the expected value of wealth $(U[E(w)])$ is equal to the sum of the expected value of the utility function $(E[U(w)])$ and the risk $(\Phi(w))$ (Castellani et al. 2005), as follows:

$$U[E(w)] = E[U(w)] + \Phi(w) \tag{4.46}$$

In a condition of risk aversion, the risk is always positive, while the risk is null in a condition of risk neutral or in the use of a degenerate variable.

Equation (4.46) can be re-written in terms of risk $(\Phi(w))$ or in terms of the expected value of the utility function $(E[U(w)])$ as follows:

$$U[E(w)] = E[U(w)] + \Phi(w) \rightarrow \begin{array}{c} \Phi(w) = U[E(w)] - E[U(w)] \\ E[U(w)] = U[E(w)] - \Phi(w) \end{array} \tag{4.47}$$

Note that maximization of the expected value of the utility function $(E[U(w)])$ is due to maximization of the expected value of the wealth $(U[E(w)])$, as well as minimization of the risk $(\Phi(w))$. Consequently, the expected value of the utility is function of the trade-off between risk and return.

Specifically, for a defined level of wealth (w), using m for the expected value of the wealth $(m = E(w))$, φ the risk level $(\varphi = \Phi(w))$, and U the expected utility $(U = E[U(w)])$, each financial position can be defined on the basis of the trade-off between risk (φ) and return (m) as follows (Castellani et al. 2005);

$$U(\varphi, m) = U(m) - \varphi \tag{4.48}$$

There is a subset of the plane (φ, m) where the set of investment opportunities can be seen. If the utility function is continuous and derivable at least twice in the plane (φ, m), the two partial derivatives are the following:

$$\begin{array}{c} \frac{\partial U(\varphi,m)}{\partial m} = U'(m) \\ \frac{\partial U(\varphi,m)}{\partial \varphi} = -1 \end{array} \tag{4.49}$$

Therefore, given two points in the plane (φ, m) they represent two financial positions:

- at the same risk (φ) level: the point with the higher expected return (m) is preferred. Therefore, for any value of the abscissa (φ), the point with the higher ordinate (m) is preferred because the utility function (U) is increasing based on m for any given φ;
- at the same expected return (m) level: the point with the lower risk (φ) is preferred. Therefore, for any value of the ordinate (m), the point with the lower abscissa (φ) is preferred because the utility function (U) is concave based on φ for any given m.

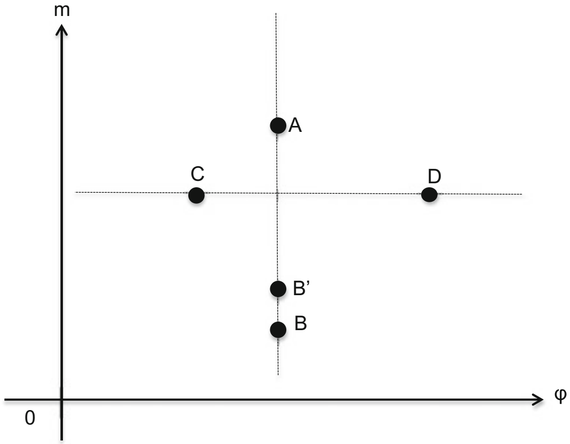


Fig. 4.5 Dominance of financial position in the plane (φ, m)

Each point defines a financial position and it can be represented in the plane (φ, m) as in Fig. 4.5 (adapted from Castellani et al. 2005).

Figure 4.5 shows that:

- point A is preferred to points B and B' because they have the same risk (φ), but the expected return (m) of point A is higher than that of points B and B'; similarly, point B' is preferred to point B: $A \succ B; A \succ B'; B' \succ B$.
- point C is preferred to point D because they have the same expected return (m), but the risk of point C is lower than that of point D: $C \succ D$.

At this stage of the analysis the main problem is related to the impossibility of defining relationships between points with different ordinates or abscissa (such as the preferences between point A and points C and D and also between points B, B' and C and D) while it is possible to define relationships between points with the same abscissa or ordinate (such as the preferences between point A and point B' and B or between C and D).

The solution of the problem requires consideration of the objectives of φ and m jointly. For this objective the function $U(\varphi, m)$ can be used. Therefore, it is necessary to derive the level set (level curve) of the space $U(\varphi, m)$, and then the shape of the points in the plane (φ, m) corresponding to the same level of the expected utility $(\bar{U})$, as follows (Castellani et al. 2005):

$$\mathrm{U}(\varphi, \mathrm{m}) = \bar{\mathrm{U}} \rightarrow \bar{U} = U(m) - \varphi \tag{4.50}$$

and therefore:

$$\mathrm{U}(\mathrm{m}) = \varphi + \bar{\mathrm{U}} \tag{4.51}$$

The utility function is always increasing (the first derivative is always positive). Then it is an injective function (one-to-one function) and its inverse is equal to:

$$\mathrm{m}(\varphi) = \mathrm{U}^{-1}\left(\varphi + \bar{\mathrm{U}}\right) \tag{4.52}$$

Note that by placing the risk equal to zero $(\varphi = 0)$, we have:

$$\mathrm{U}(\mathrm{m}) = \bar{\mathrm{U}}; \quad \bar{\bar{\mathrm{m}}} = \mathrm{U}^{-1}(\bar{\mathrm{U}})$$

Representing the intersection between the curve level $U(\varphi, m) = \bar{U}$ and the ordinate, and it represents the certainty equivalent of all points on the curve.

Based on the implicit function, the two first partial derivatives are the following:

$$\begin{aligned}\frac{\partial \mathrm{m}}{\partial \varphi} &= -\frac{\frac{\partial \mathrm{U}}{\partial \varphi}}{\frac{\partial \mathrm{U}}{\partial \mathrm{m}}} = -\frac{-1}{\mathrm{U}'(\mathrm{m})} = \frac{1}{\mathrm{U}'(\mathrm{m})} \\ \frac{\partial \varphi}{\partial \mathrm{m}} &= -\frac{\frac{\partial \mathrm{U}}{\partial \varphi}}{\frac{\partial \mathrm{U}}{\partial \mathrm{m}}} = -\frac{-1}{\mathrm{U}'(\mathrm{m})} = \frac{1}{\mathrm{U}'(\mathrm{m})}\end{aligned} \tag{4.53}$$

The first derivative is always positive because the function is increasing, and therefore:

$$\frac{1}{\mathrm{U}'(\mathrm{m})} > 0 \tag{4.54}$$

The second partial derivatives are the following:

$$\begin{aligned}\frac{\partial^2 \mathrm{m}}{\partial \varphi^2} &= \frac{\partial\left(\frac{1}{\mathrm{U}'[\mathrm{m}(\varphi)]}\right)}{\partial \varphi} = -\frac{\mathrm{U}''(\mathrm{m})}{[\mathrm{m}(\varphi)]^2}\frac{\mathrm{dm}}{\mathrm{d}\varphi} = -\frac{\mathrm{U}''(\mathrm{m})}{[\mathrm{U}'(\mathrm{m})]^3} \\ \frac{\mathrm{d}^2 \varphi}{\mathrm{dm}^2} &= \frac{\mathrm{d}}{\mathrm{d}\varphi}\frac{1}{\mathrm{U}'[\mathrm{m}(\varphi)]} = -\frac{\mathrm{U}''(\mathrm{m})}{[\mathrm{m}(\varphi)]^2}\frac{\mathrm{dm}}{\mathrm{d}\varphi} = -\frac{\mathrm{U}''(\mathrm{m})}{[\mathrm{U}'(\mathrm{m})]^3}\end{aligned} \tag{4.55}$$

The second derivative is negative because the risk aversion is assumed and therefore:

$$-\frac{\mathrm{U}''(\mathrm{m})}{[\mathrm{U}'(\mathrm{m})]^3} > 0 \tag{4.56}$$

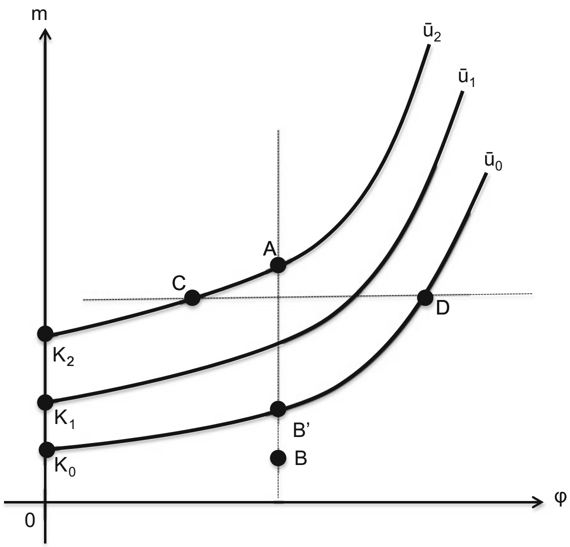
On the basis of the signs of the first and second derivatives, the curves in the plane (φ, m) are increasing function (for *m*) and convex (for *φ)*, as shown in Fig. 4.6 (adapted from Castellani et al. 2005).

Figure 4.6 registers the indifference curves with regards to the values $\bar{u}_0, \bar{u}_1, \bar{u}_2$ of the expected utility where $\bar{u}_0 < \bar{u}_1 < \bar{u}_2$.

The points on the curve represent the financial positions of the investors.

Points *A* and *C* are indifferent among them because they have the same expected utility $(\bar{u}_2)$.

Fig. 4.6 Dominance of financial position in the plane (φ, m) based on indifference curves



Point K_2 represents the certainty equivalent of points A and C, and also the points on the same curve.

Point D is indifferent to point B' that, in turn, is preferred to point B. Therefore, point D is preferred to point B.

Point K_0 is indifferent to points B' and D, and all other points of the curve.

Having defined the single financial positions on the indifference curves, the problem of the portfolio choices can be solved based on the utility function criteria. There are two main phases (Castellani et al. 2005; Saltari 2011):

(1) *optimization phase*;
(2) *maximization phase*.

Optimization Phase

It is quite clear that in the absence of constraints on risk (φ) and expected return (m) the optimal solution is given by: $\varphi = 0; m = \infty$. The presence of constraints generate a *set of opportunities* (W) that always defines a subset of the plane (φ, m). It is the presence of constraints on the two variables that results in the need for the Optimization phase. Its aim is to analyse separately the different partial objectives of the financial positions.

Within the set of opportunities (W) a key role is played by the *opportunity of frontier*: it can be defined as the opportunity that minimizes the risk for a given level of the expected value. The *constrained optimization problem* can be defined as follows (Castellani et al. 2005):

$$\begin{cases} \min_{w \in W} \Phi(w) \\ E(w) = m_0 \end{cases} \tag{4.57}$$

where the first is the function to minimize and the second is the constraint.

In Fig. 4.7 (adapted from Castellani et al. 2005) the set of opportunities (W) is represented by the area bounded by the curve between points S_5 and S_4. It implies that the set of opportunities never touches the ordinate. Consequently, the free-risk positions are not included in the set of opportunities (W) (Fig. 4.7).

Having defined a level of expected return equal to m_0, point S_0 (where $S_0 \in W$) defines the *opportunity frontier*. Point S_0 dominates all points to its right (S_i). They are all characterized by greater risk for the same expected return. Therefore, point S_0 is preferred to all of the others on its right (S_i), so that: $S_0 \succ S_i$.

By solving the constrained optimization of Eq. (4.57) for all values of m_0 the frontier opportunities are defined, defining the *frontier of opportunities* $(\mathcal{B})$: it is a subset of opportunities of frontier (W) defined by the curve between points S_5 and S_2, and between points S_3 and S_4. On the frontier of the opportunity $(\mathcal{B})$ there may be opportunities that presents the same risk but with a different expected value (such as points S_0 and S_0').

The *efficient opportunity* can be defined for every opportunity on the frontier of the opportunities $(\mathcal{B})$ which has a maximum expected value (m) for the same risk level (φ). Therefore, the efficient opportunity is the solution of the constrained optimization problem, as follows (Castellani et al. 2005):

$$\begin{cases} \max_{w \in \mathcal{B}} E(w) \\ \Phi(w) = \varphi_0 \end{cases} \tag{4.58}$$

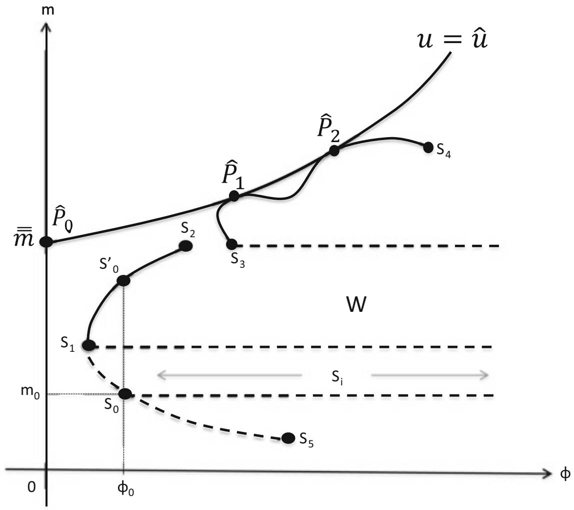


Fig. 4.7 Opportunities frontier and the maximization points

where the first is the function to be maximized and the second is the constraint.

The set of the efficient opportunities for the different level of risk (φ_0) defines the *efficient frontier* (ε): it is a subset of the *frontier of the opportunities* $(\mathcal{B})$ that, in turn, is a subset of the *set of opportunities* (W).

In Fig. 4.7 the *efficient frontier* is given by the curve portions (indicated by a continuous line) between points S_1 and S_2, and between points S_3 and S_4.

The curve portion (indicated by a dotted line) between points S_1 and S_5 is the portion of the frontier of opportunity $(\mathcal{B})$ that is dominated, and then is not efficient: for each point of the curve, it is possible to identify a point with same level of risk that offers higher expected value. The point S_1 has the minimum risk level.

The efficient frontier is given by an increasing function by definition because the first derivative is always positive. All points on the efficient frontier are characterized to be Pareto optimality: it is impossible to improve any one point without worse at least one other point.

The definition of the efficient frontier, allows for the separate analysis of the partial objectives of the financial positions. Therefore, with the definition of the efficient frontier, the Optimization phase is completed.

Maximization Phase

The aim of the *maximization phase* is to identify the points that maximize expected utility.

Therefore, after identification of the points of optimum of the set of opportunities (W), the maximum point is chosen. This point is positioned on the indifference curve with the maximum expected utility.

Therefore, it is necessary to identify the points $\hat{P}$ on the efficient frontier which are positioned on the indifference curve $U(\varphi, m) = \hat{u}$ with the highest expected utility $\hat{u}$.

In Fig. 4.7, points $\hat{P}_1$ and $\hat{P}_2$ represent the position of maximum expected utility and they are indifferent among them. Also point $\hat{P}_0$ is indifferent to points $\hat{P}_1$ and $\hat{P}_2$. Specifically, the ordinate $\overline{\overline{m}}$ of $\hat{P}_0$ is the certainty equivalent of the risky positions represented by the points $\hat{P}_1$ and $\hat{P}_2$. But $\hat{P}_0$ is outside of the set of the opportunities (W) and therefore it is not a real accessible position. However, its meaning is relevant: the investor is indifferent to obtaining $\overline{\overline{m}}$ with certainty $(\hat{P}_0)$ or the set of opportunities (W). Therefore, $\overline{\overline{m}}$ can be defined as the *indifference price* of the set of opportunities (W). It is worth noting, that point $\hat{P}_0$, and therefore the value $\overline{\overline{m}}$, can be identified at the end of the optimization phase only (Castellani et al. 2005).

For greater understanding of the two phases of Optimization and Maximization for the portfolio choices under the utility function approach, three main cases can be analysed (Saltari 2011):

– (*Case 1*) two assets in the portfolio;
– (*Case 2*) more than two assets in the portfolio;
– (*Case 3*) more than two assets in the portfolio and one of them free-risk.

(Case 1) Two Assets in the Portfolio

Assuming two assets in the portfolio. Therefore, the investor has to share his wealth between asset A_1 and asset A_2. The current wealth of the investor to be invested, is the constraint: it does not change but it can only be shared between the two assets. Formally (Saltari 2011):

$$P_1^A A_1 + P_2^A A_2 = P_1^A \overline{A_1} + P_2^A \overline{A_2} = w \tag{4.59}$$

where:

– w: is the current wealth held by the investor;
– P_1^A and P_2^A: are the prices of Asset 1 and Asset 2 respectively;
– A_1 and A_2: are the amounts (number) of Asset 1 and Asset 2 respectively, purchased by the investor;
– $\overline{A_1}$ e $\overline{A_2}$: are the amounts (number) of Asset 1 and Asset 2 respectively, held by investor.

The current portion of wealth invested in the i-th asset (α_i) is equal to:

$$\alpha_i = \frac{P_i^A A_i}{w} \quad \text{or} \quad \alpha_i = \frac{P_i^A \bar{A}_i}{w} \tag{4.60}$$

where, the first is defined according to the amount of the i-th Asset purchased by the investor, while the second according to the amount of the i-th Asset held by the investor.

Considering that there are only two assets, the constraint can be defined on the bases of the sum of the portions of wealth invested in Asset 1 (α_1) and Asset 2 (α_2) as follows:

$$P_1^A A_1 + P_2^A A_2 = w$$

by dividing each term by wealth (w), we have:

$$\frac{P_1^A A_1}{w} + \frac{P_2^A A_2}{w} = \frac{w}{w}$$

and on the basis of Eq. (4.60), we have:

$$\alpha_1 + \alpha_2 = 1 \tag{4.61}$$

The expected return of each Asset is function of the states of nature (s). For simplicity, assuming that only two states are possible: $s = 1; s = 2$. The matrix expected return—states of nature, of the two assets can be defined as follows: (Table 4.2).

The $z_{i,s}$ is the expected return of the i-th asset (for $i = 1, 2$) when the s-th states of nature (for $s = 1, 2$) is achieved.

Therefore, the portfolio's expected return is function of the portion of the current wealth invested in each of the two assets, as well as the expected return of the two assets as function of the states of nature. Formally (Saltari 2011):

$$\begin{aligned} y_1 &= A_1 z_{1,1} + A_2 z_{2,1} \quad \text{for } s = 1 (y_s = y_1) \\ y_2 &= A_1 z_{1,2} + A_2 z_{2,2} \quad \text{for } s = 2 (y_s = y_2) \end{aligned} \tag{4.62}$$

where:

- the first equation is the portfolio's expected return when the state of nature 1 $(s = 1 \rightarrow y_s = y_1)$ is achieved and is equal to the sum of the amount (units) of the two assets in portfolio (A_1, A_2) multiplied by their respective expected return when the state of nature 1 $(z_{1,1}, z_{2,1} \text{ for } i = 1, 2 \text{ and } s = 1)$ is achieved;
- the second equation is the portfolio's expected return when the state of nature 2 $(s = 2 \rightarrow y_s = y_2)$ is achieved and is equal to the sum of the amount (units) of the two assets in portfolio (A_1, A_2) multiplied by their respective expected return when the state of nature 2 $(z_{1,2}, z_{2,2} \text{ for } i = 1, 2 \text{ and } s = 2)$ is achieved.

Based on Eq. (4.60), we have:

$$\alpha_i = \frac{P_i^A A_i}{w} \rightarrow A_i = \frac{\alpha_i w}{P_i^A}$$

and by substituting in Eq. (4.62) we have:

$$\begin{aligned} y_1 &= \alpha_1 w \frac{z_{1,1}}{P_1^A} + \alpha_2 w \frac{z_{2,1}}{P_2^A} \quad \text{for } s = 1 (y_s = y_1) \\ y_2 &= \alpha_1 w \frac{z_{1,2}}{P_1^A} + \alpha_2 w \frac{z_{2,2}}{P_2^A} \quad \text{for } s = 2 (y_s = y_2) \end{aligned} \tag{4.63}$$

The rate of return of the i-th asset when the state of nature s $(r_{i,s})$ is achieved can be defined on the basis of its expected return in the state of nature s $(z_{i,s})$ and its purchase price (P_i^A), as follows:

Table 4.2 Matrix expected return—state of nature

Matrix expected return—states of nature			
		States of nature	
		State of nature 1 $(s = 1)$	State of nature 1 $(s = 1)$
Expected return	Asset 1 $(i = 1)$	$z_{i,s} = z_{1,1}$	$z_{i,s} = z_{1,2}$
	Asset 2 $(i = 2)$	$z_{i,s} = z_{2,1}$	$z_{i,s} = z_{2,2}$

$$r_{i,s} = \frac{z_{i,s} - P_i^A}{P_i^A} = \frac{z_{i,s}}{P_i^A} - 1 \tag{4.64}$$

The total return of the i-th asset when the state of nature s $(R_{i,s})$ is achieved, is equal to:

$$1 + r_{i,s} = \frac{z_{i,s}}{P_i^A} \equiv R_{i,s} \rightarrow R_{i,s} \equiv 1 + r_{i,s} \tag{4.65}$$

Therefore, Eq. (4.63) can be rewritten as follows (Saltari 2011):

$$\begin{aligned} y_1 &= \left[\alpha_1 R_{1,1} + \alpha_2 R_{2,1}\right] w \quad \text{for } s = 1 (y_s = y_1) \\ y_2 &= \left[\alpha_1 R_{1,2} + \alpha_2 R_{2,2}\right] w \quad \text{for } s = 2 (y_s = y_2) \end{aligned} \tag{4.66}$$

where:

- the first equation is the portfolio's expected return when the state of nature 1 $(s = 1 \rightarrow y_s = y_1)$ is achieved and is equal to the sum of wealth portions invested in the two assets $(\alpha_1 w; \alpha_2 w)$ multiplied by their expected total return respectively when the state of the nature 1 $(R_{1,1}, R_{2,1} \text{ for } i = 1, 2 \text{ and } s = 1)$ is achieved;
- the second equation is the portfolio's expected return when the state of nature 2 $(s = 2 \rightarrow y_s = y_2)$ is achieved and is equal to the sum of wealth portions invested in the two assets $(\alpha_1 w; \alpha_2 w)$ multiplied by their expected total return respectively when the state of the nature 2 $(R_{1,2}, R_{2,2} \text{ for } i = 1, 2 \text{ and } s = 1)$ is achieved.

In general, we have:

$$y_s = \alpha_i R_{i,s} w \quad i = 1, 2; \ s = 1, 2 \tag{4.67}$$

Therefore, the portfolio's return is function of the state of nature $(s = 1, 2)$ that it is achieved and of the total return of the two assets $(R_{1,s}; R_{2,s})$ on the basis of the wealth portion invested in each of them $(\alpha_1 w; \alpha_2 w)$.

The equations can be schematically represented as in Fig. 4.8 (adapted from Saltari 2011).

In Fig. 4.8, the portfolio's return is registered on the abscissa when the state of nature 1 $(s = 1 \rightarrow y_s = y_1)$ is achieved, while the portfolio's return is registered on the ordinate when the state of nature 2 $(s = 2 \rightarrow y_s = y_2)$ is achieved. Point *A* identifies the portfolio consisting of Asset 1 only $(\alpha_1 = 1; \alpha_2 = 0)$ while point *B* identifies the portfolio consisting of Asset 2 only $(\alpha_1 = 0; \alpha_2 = 1)$. Therefore, each point between *A* and *B* can be obtained by changing the amount of the two assets in the portfolio $(\Delta\alpha_1; \Delta\alpha_2)$.

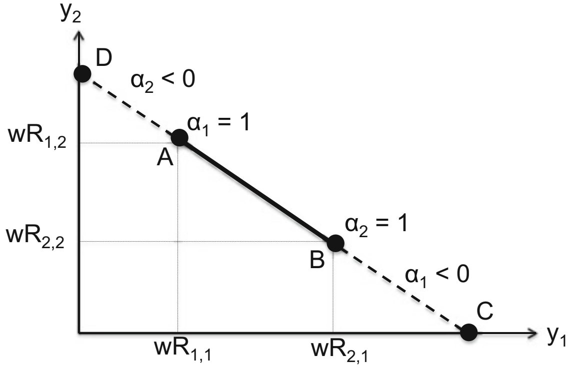


Fig. 4.8 The portfolio's return

If there is no short selling, and therefore for a positive value of α_i $(\alpha_i > 0)$, segment *AB* defines the constraint. The negative slope of the segment is due to the non-dominance of each of the two assets on the other one:

- for the state of nature 1, the return of asset 2 is higher than that of asset 1:

$$s = 1 \rightarrow R_{2,1} > R_{1,1}$$

- for the state of nature 2, the return of asset 1 is higher than that of asset 2:

$$s = 2 \rightarrow R_{1,2} > R_{2,2}$$

If there is short selling, and therefore also for negative value of α_i $(\alpha_i < 0)$, the constraint goes beyond points *A* and *B* up to points *C* and *D*. Specifically:

- point *D* represents a portfolio achieved through the short selling of Asset 2. Point A defines a portfolio that includes Asset 1 only. Therefore, if the revenues due to short selling of Asset 2 are used to purchase new amounts of Asset 1, it can go beyond point A up to point D. In this case, if the state of nature 1 $(s = 1)$ is achieved, the portfolio's return is null and the wealth held by the investor is just enough to repay debt due to short selling;
- point *C* represents a portfolio achieved through the short selling of Asset 1. Point B defines a portfolio that includes Asset 2 only. Therefore, if the revenues due to short selling of Asset 1 are used to purchase new amounts of Asset 2, it

can go beyond point B up to point C. In this case, if the state of nature 2 $(s = 2)$ is achieved, the portfolio's return is null and the wealth held by the investor is just enough to repay debt due to short selling.

It is worth noting that to assume that points C and D cannot be exceeded, it implies the assumption that the investor cannot fail. Beyond these points, the income is negative.

By assuming the absence of short selling, and therefore by considering the segment AB only, the sum of the portions of wealth invested in the two assets must be equal to the total wealth of the investor, as follows:

$$\alpha_1 + \alpha_2 = 1 \tag{4.68}$$

The constraint can be expressed based on one of the two assets. Therefore, based on the Eq. (4.68) the wealth portion invested in the asset 2 (α_2) can be expressed as function of the wealth portion invested in the asset 1 (α_1) as the follows:

$$\alpha_2 = 1 - \alpha_1 \tag{4.69}$$

By substituting Eq. (4.66) can be rewritten as follows:

$$\begin{aligned} y_1 &= \left[\alpha_1 R_{1,1} + \alpha_2 R_{2,1}\right] w = \left[\alpha_1 R_{1,1} + (1-\alpha_1) R_{2,1}\right] \\ w &= \left[\alpha_1 R_{1,1} + R_{2,1} - \alpha_1 R_{2,1}\right] w = \left[\alpha_1\left(R_{1,1} - R_{2,1}\right) + R_{2,1}\right] w \\ y_2 &= \left[\alpha_1 R_{1,2} + \alpha_2 R_{2,2}\right] w = \left[\alpha_1 R_{1,2} + (1-\alpha_1) R_{2,2}\right] \\ w &= \left[\alpha_1 R_{1,2} + R_{2,2} - \alpha_1 R_{2,2}\right] w = \left[\alpha_1\left(R_{1,2} - R_{2,2}\right) + R_{2,2}\right] w \end{aligned}$$

and then:

$$\begin{aligned} \mathrm{y}_1 &= \left[\alpha_1\left(\mathrm{R}_{1,1} - \mathrm{R}_{2,1}\right) + \mathrm{R}_{2,1}\right]\mathrm{w} \quad \text{for } \mathrm{s} = 1(\mathrm{y}_\mathrm{s} = \mathrm{y}_1) \\ \mathrm{y}_2 &= \left[\alpha_1\left(\mathrm{R}_{1,2} - \mathrm{R}_{2,2}\right) + \mathrm{R}_{2,2}\right]\mathrm{w} \quad \text{for } \mathrm{s} = 2(\mathrm{y}_\mathrm{s} = \mathrm{y}_2) \end{aligned} \tag{4.70}$$

It is possible to define the relationship between y_1 and y_2 on the basis of the constraint. By solving the first equation for α_1, we have:

$$\begin{aligned} y_1 &= \left[\alpha_1\left(R_{1,1} - R_{2,1}\right) + R_{2,1}\right] w \\ \frac{y_1}{w} &= \alpha_1\left(R_{1,1} - R_{2,1}\right) + R_{2,1} \\ \frac{y_1}{w} - R_{2,1} &= \alpha_1\left(R_{1,1} - R_{2,1}\right) \\ \alpha_1 &= \frac{\left(\frac{\mathrm{Y}_1}{\mathrm{W}}\right) - \mathrm{R}_{2,1}}{\mathrm{R}_{1,1} - \mathrm{R}_{2,1}} \end{aligned}$$

By substituting in the second equation, we have:

$$\mathrm{y}_2 = \left[\alpha_1\left(\mathrm{R}_{1,2} - \mathrm{R}_{2,2}\right) + \mathrm{R}_{2,2}\right]\mathrm{w}$$

$$\mathrm{y}_2 = \left[\frac{\left(\frac{\mathrm{y}_1}{\mathrm{W}}\right) - \mathrm{R}_{1,2}}{\mathrm{R}_{1,1} - \mathrm{R}_{2,1}}\left(\mathrm{R}_{1,2} - \mathrm{R}_{2,2}\right) + \mathrm{R}_{2,2}\right]\mathrm{w}$$

$$\mathrm{y}_2 = \frac{\mathrm{R}_{1,2} - \mathrm{R}_{2,2}}{\mathrm{R}_{1,1} - \mathrm{R}_{2,1}}\left[\left(\frac{\mathrm{y}_1}{\mathrm{W}}\right) - \mathrm{R}_{2,1}\right]\mathrm{w} + \mathrm{R}_{2,2}\mathrm{w}$$

$$\mathrm{y}_2 = \frac{\mathrm{R}_{1,2} - \mathrm{R}_{2,2}}{\mathrm{R}_{1,1} - \mathrm{R}_{2,1}}\left[\frac{\mathrm{y}_1 - \mathrm{R}_{2,1} \cdot \mathrm{W}}{\mathrm{W}}\right]\mathrm{w} + \mathrm{R}_{2,2}\mathrm{w}$$

and then:

$$\mathrm{y}_2 = \frac{\mathrm{R}_{1,2} - \mathrm{R}_{2,2}}{\mathrm{R}_{1,1} - \mathrm{R}_{2,1}}\left[\mathrm{y}_1 - \mathrm{wR}_{2,1}\right] + \mathrm{wR}_{2,2} \tag{4.71}$$

Equation (4.72) defines the relationship between y_1 and y_2 based on the constraint. It can be defined as the *constraint's equation* (Saltari 2011). The ratio $(R_{1,2} - R_{2,2})/(R_{1,1} - R_{2,1})$ defines the slope of the constraint.

At this stage of the analysis, the main problem is the *optimal allocation* of the investor's wealth among the assets, and therefore the definition of the *optimal portfolio*. Based on the utility function criteria, this aim can be redefined in terms of the *utility function maximization*. Then, the optimal portfolio is the one that maximizes the investor's utility function, and consequently maximizes the expected return of the portfolio (y_s).

It is worth noting that the portfolio's return is function of two variables:

- the share of wealth $\alpha_i (i = 1, 2)$ to be invested in each of the two assets;
- the state of nature (s) which could be achieved in the future and its associate probability $\pi_s (s = 1, 2)$.

Formally, a problem of constrained optimization can be defined, as follows (Saltari 2011):

$$\begin{cases} \max\limits_{\alpha_1, \alpha_2} E(U) = \pi_1 U(y_1) + \pi_2 U(y_2) \\ \alpha_1 + \alpha_2 = 1 \end{cases} \tag{4.72}$$

where the first is the equation to be maximized and the second is the constraint.

The Lagrangian $(\mathcal{L})$ function is the following:

$$\mathcal{L} = \pi_1 U(y_1) + \pi_2 U(y_2) + \lambda[\alpha_1 + \alpha_2 - 1]$$

By substituting Eq. (4.66) the equation to be maximized can be rewritten as follows:

$$\begin{aligned} E(U) &= \pi_1 U(y_1) + \pi_2 U(y_2) \\ &= \pi_1 U\left[\left(\alpha_1 R_{1,1} + \alpha_2 R_{2,1}\right)w\right] + \pi_2 U\left[\left(\alpha_1 R_{1,2} + \alpha_2 R_{2,2}\right)w\right] \end{aligned}$$

The first order conditions are the following:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \alpha_1} = \pi_1 U'(y_1) w R_{1,1} + \pi_2 U'(y_2) w R_{1,2} - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \alpha_2} = \pi_1 U'(y_1) w R_{2,1} + \pi_2 U'(y_2) w R_{2,2} - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 1 - (\alpha_1 + \alpha_2) = 0 \end{cases} \tag{4.73}$$

By solving the first and the second equation for λ, we have:

$$\lambda = \pi_1 U'(y_1) w R_{1,1} + \pi_2 U'(y_2) w R_{1,2}$$
$$\lambda = \pi_1 U'(y_1) w R_{2,1} + \pi_2 U'(y_2) w R_{2,2}$$

and then

$$\pi_1 U'(y_1) w R_{1,1} + \pi_2 U'(y_2) w R_{1,2} = \pi_1 U'(y_1) w R_{2,1} + \pi_2 U'(y_2) w R_{2,2}$$

By dividing first and second terms for the wealth (w), we have:

$$\pi_1 U'(y_1) R_{1,1} + \pi_2 U'(y_2) R_{1,2} = \pi_1 U'(y_1) R_{2,1} + \pi_2 U'(y_2) R_{2,2}$$

Considering that the return of the i-th asset in the state of nature s $(R_{i,s})$ is equal to:

$$R_{i,s} = \frac{z_{i,s}}{P_i^A}$$

and by substituting we have:

$$\frac{\pi_1 U'(y_1) z_{1,1} + \pi_2 U'(y_2) z_{1,2}}{P_1^A} = \frac{\pi_1 U'(y_1) z_{2,1} + \pi_2 U'(y_2) z_{2,2}}{P_2^A} \tag{4.74}$$

The equality (4.74) shows that for the optimal allocation, the investor shares his wealth between the two assets so that the last euro invested in asset 1 has the same expected marginal utility of the last euro invested in asset 2 (Saltari 2011).

(Case 2) More Than Two Assets in the Portfolio

Assuming that there are more than two assets in portfolio. In this case, the investor has to share his wealth (w) among n assets (n > 2). The budget constrain can be formalized as follows (Saltari 2011):

$$\sum_{i=1}^{n} P_i^A A_i = w \tag{4.75}$$

By dividing the first and second terms for wealth (w), and considering the Eq. (4.60), the Eq. (4.75) can be expressed in terms of the share of wealth invested in each assets, as follows:

$$\sum_{i=1}^{n} \frac{P_i^A A_i}{w} = \frac{w}{w} \rightarrow \frac{P_i^A A_i}{w} = \alpha_i \rightarrow \sum_{i=1}^{n} \alpha_i = 1$$

By considering Eq. (2.67), we have:

$$y_s = \alpha_i R_{i,s} w \rightarrow y_s = \sum_{i=1}^{N} \alpha_i R_{i,s} w \rightarrow y_s = w \sum_{i=1}^{n} \alpha_i R_{i,s}$$

and by considering Eq. (2.65), we have:

$$R_{i,s} \equiv 1 + r_{i,s}$$

$$y_s = w \sum_{i=1}^{n} \alpha_i \left(1 + r_{i,s}\right) \tag{4.76}$$

In this case, the investor has to choose the share of wealth (α_i) to be invested in each asset in the portfolio with the objective of maximizing the expected utility of the portfolio's return (y_s). The optimization problem, can be formalized as follows (Saltari 2011):

$$\begin{cases} \max\limits_{\alpha_i} E(U) = \sum\limits_{s=1}^{S} \pi_s U(y_s) \\ \sum\limits_{i=1}^{n} \alpha_i = 1 \end{cases} \tag{4.77}$$

where the first is the equation to be maximized and the second is the budget constraint.

The Lagrangian function can be defined as follows:

$$\mathcal{L} = \sum_{s=1}^{S} \pi_s U(y_s) + \lambda \left(\sum_{i=1}^{n} \alpha_i - 1 \right)$$

and by considering that

$$y_s = w \sum_{i=1}^{N} \alpha_i R_{i,s}$$

the $E(U)$ function can be rewritten as follows:

$$E(U) = \sum_{s=1}^{S} \pi_s U(y_s) = \sum_{s=1}^{S} \pi_s U\left(w \sum_{i=1}^{N} \alpha_i R_{i,s} \right)$$

The first order conditions are the following:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \alpha_i} = \sum_{s=1}^{S} \pi_s U'(y_s) w R_{i,s} - \lambda = 0 \; i = 1, 2, \ldots, n \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_{i=1}^{n} \alpha_i = 0 \end{cases} \tag{4.78}$$

The solution of the system with regards to the value of α_i and λ are the optimal choice (Saltari 2011).

(Case 3) More Than Two Assets and one of Them Free-Risk in the Portfolio Assuming that there are more than two assets and one of them free-risk in the portfolio.

It is worth noting that, in this context, the asset is free-risk if its expected return is independent from the state of nature (s) that will be achieved.

Therefore, the portfolio consists of n risky assets and one asset free-risk $(n + 1)$. Therefore, the budget constraint can be defined as follows:

$$\sum_{i=0}^{n} \alpha_i = 1 \tag{4.79}$$

Also in this case, as in the previous cases 1 and 2, the investor has to choose the share of wealth (α_i) to be invested in each asset in the portfolio with the objective of maximizing the expected utility of the portfolio's return (y_s). The optimization problem can be formalized as follows (Saltari 2011):

$$\begin{cases} \max_{\alpha_i, \alpha_0} E(U) = \sum_{s=1}^{S} \pi_s U(y_s) \\ \sum_{i=0}^{n} \alpha_i = 1 \end{cases} \tag{4.80}$$

Denoting with r_0 the rate of return of the free-risk asset (so that its return is equal to $R_0 = 1 + r_0$) and considering that is independent from the state of nature, the index s can be omitted and with α_0 the part of the wealth invested in the risk-free asset, the Lagrangian $(\mathcal{L})$ can be formalized as follows:

$$\mathcal{L} = \mathrm{E(U)} + \lambda\left(1 - \sum_{i=0}^{N} \alpha_i\right)$$

By deriving partially with respect to α_i, α_0 and λ, the first order condition is achieved. By placing them equal to zero, we have:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \alpha_i} = \sum_{s=1}^{S} \pi_s U'(y_s) w R_{i,s} - \lambda = 0 \; i = 1, 2, \ldots, n \\ \frac{\partial \mathcal{L}}{\partial \alpha_0} = \sum_{s=1}^{S} \pi_s U'(y_s) w R_0 - \lambda = 0 \; i = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_{i=0}^{N} \alpha_i = 0 \end{cases} \tag{4.81}$$

It is worth noting, that in this case for $i = 0$ we have α_0 that it is independent of the state of nature, so that: $y_s = w\alpha_0 R_0$. Therefore, in this case, unlike the previous case, a new condition is introduced in which $i = 0$ with regards to the risk-free rate.

From the first equation we have:

$$\lambda = \sum_{s=1}^{S} \pi_s U'(y_s) w R_{i,s}$$

and from the second equation we have:

$$\lambda = \sum_{s=1}^{S} \pi_s U'(y_s) w R_0$$

Consequently, by considering the first and the second equation we have:

$$\sum_{s=1}^{S} \pi_s U'(y_s) w R_{i,s} = \sum_{s=1}^{S} \pi_s U'(y_s) w R_0$$

and then:

$$\sum_{s=1}^{S} \pi_s U'(y_s) w \left(R_{i,s} - R_0\right) = 0 \leftrightarrow \mathrm{E}\left[U'(y_s) w \left(R_{i,s} - R_0\right)\right] = 0 \quad i = 1, 2, \ldots, n \tag{4.82}$$

where the two forms are equivalent.

If the number of the risky assets is high, the solution of the optimal portfolio by defining the share of wealth to be invested in each asset (α_i), is very hard. The *mutual fund theorem* (Tobin 1958; Cass and Stiglitz 1970) can be used to solve the problem of the optimal portfolio.

On the basis of some conditions, the portfolio can be assumed as consisting only of two assets:

- the first is the asset risk-free;
- the second is a "common fund" obtained from the optimal combination of all risky assets. The weight of each risky asset in the common fund is independent from the wealth level.

Therefore, the investor chooses the share of wealth to invest in the asset free-risk and in the common fund.

Use of the separation theorem requires several conditions. Specifically, it is necessary to introduce some constraints on the probability distribution of the expected return, or alternatively, on the utility function used by investor (Saltari 2011). Therefore, constraints can be used on:

- probability distribution of the expected return, it is necessary to assume that the expected return follows a normal distribution;
- the utility function, is necessary to assume that it is classified in the HARA (Hyperbolic Absolute Risk Aversion) group. In this case, the absolute risk aversion coefficient is a hyperbolic function of y_s while the absolute risk tolerance has a linear form.

Specifically, the absolute risk aversion coefficient for the utility function in the class HARA can be defined as follows:

$$\lambda_{(y_s)} = -\frac{U''(y_s)}{U'(y_s)} = \frac{1}{(c + dy_s)} \quad \text{with } c \text{ and } d \text{ constant}$$

and the absolute risk tolerance coefficient is equal to (Saltari 2011):

$$\tau_{(y_s)} = \frac{1}{\lambda_{(y_s)}} = -\frac{U'(y_s)}{U''(y_s)} = c + dy_s$$

References

Arrow K (1965) "The theory of risk aversion", aspects of the theory of risk-bearing (Lecture 2). Yrjs Jansson Foundation, Helsinki

Arrow K (1971) Essays in the theory of risk-bearing. Markham, Chicago

Arrow K (1984) Individual choice under certainty and uncertainty, Collected Papers of Kenneth J. Arrow, Vol. 3, Belknap Press, Cambridge MA

Campbell RM (2015) Microeconomics: principles, problems, and policies, McGraw-Hill

Cass D, Stiglitz J (1970) The structure of investor preferences and asset returns, and separability in portfolio allocation: a contribution to the pure theory of mutual funds. Journal of Economic Theory 2(2):122–160

Castellani G, De Felice M, Moriconi F (2005) *Manuale di Finanza 2. Teoria del portafoglio e mercato azionario*, Il Mulino
Cesari R (2012a) Introduzione alla finanza matematica. Concetti di base, tassi e obbligazioni, 2nd edn. McGraw-Hill, New York
Cesari R (2012b) Introduzione alla finanza matematica. Mercati azionari, rischi, e portafogli, 2nd edn. McGraw-Hill, New York
Cuthbertson K, Nitzsche D (2014) Quantitative financial economics: stocks, bonds and foreign exhange, 2nd ed., Wiley
Fishburn PC, Kochenberger GA (1979) Two-piece Von Neumann-Morgenstern utility functions. Decis Sci 10(4):503–518
Gravelle H, Rees R (1992) Microeconomics, 2nd edn. Longman, London
Heap S, Lyons B, Hollis M, Sugden R, Weale A (1992) The theory of choice: a critical guide. Blackwell
Hirshleifer J, Riley JG (1992) The analytics of uncertainty and information. Cambridge University Press
Ingersoll JE (1987) Theory of financial decision making. Rowman&Littlefield Pub Inc., Baltimore
Kreps DM (1979) A representation theorem for “preference for flexibility”. Econometrica 47(3):565–577
Kreps DM (1990) A course in microeconomic theory. Princeton University Press
Kreps DM (2012) Microeconomic foundations I: choice and competitive markets. Princeton University Press
Litner J (1965a) Security prices, risk, and maximal gains from diversification. J Finance 20(4):587–615
Litner J (1965b) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Rev Econ Stat 47(1):13–37
Litner J (1970) The market price of risk, size of market and investor’s risk aversion. The Review of Economics and Statistics 52(1):87–99
Mankiw GN (2017) Principles of microeconomics. 8th ed., Cengage Learning
Markowitz H (1952) Portfolio selection. The Journal of Finance 7(1):77–91
Markowitz H (1956) The optimization of a quadratic function subject to linear constraints. Nav Res Logist 3(1–2):111–133
Markowitz H (1959) Portfolio selection: efficient diversification of investment. Wiley, New York
Perloff JM (2016) Microeconomics: theory and applications with calculus. 4th ed., Pearson
Pratt JW (1964) Risk aversion in the small and in the large. Econometrica 32(1/2):122–136
Saltari E (2011) Appunti di Economia Finanziaria. Esculapio
Tobin J (1958) Liquidity preference as behavior towards risk. The Review of Economic Studies 25(2):65–86
Varian HR (1992) Microeconomic analysis, 3rd edn. W.W. Norton & Company, New York
Von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press

Chapter 5
Mean-Variance Approach

Abstract The mean-variance approach is the most widely used in the portfolio selections. The portfolio selection is based on two variables: (i) expected value of the portfolio return; (ii) variance of the expected portfolio return measuring the portfolio risk. An efficient portfolio must satisfy the Pareto optimal condition. Therefore, the investor prefers the portfolio that is capable of maximising its expected return to an equal variance or the portfolio capable of minimizing its variance to an equal expected return. This approach simplifies the problem of portfolio selection. There are two main advantages: first, it does not require specification about probability distribution; second, it is simple and intuitive because it is only based on the mean and variance. However, it is also true that this approach neglects a lot of relevant information about distribution probability. The entire portfolio selection process can be simplified on the basis of two main phases of the portfolio selection process:

(1) *optimization phase*: the aim is to define the diversified portfolio and the efficient frontier. The definition of the diversified portfolio is based on the statistical characteristics of the assets. Specifically, the expected return of the portfolio is equal to the weighted average of the expected returns of the assets, while the portfolio variance is the function of the covariance between the assets' expected returns. The assumption refers to the investors' homogeneous expectations about the statistical characteristics of the assets implying that all investors define the same efficient frontier.
(2) *maximization phase*: the aim is to choose the optimal portfolio among the efficient portfolios defined on the efficient frontier. None of the efficient portfolios on the efficient frontiers can be preferred over the others by definition. The choice of the optimal portfolio among the efficient portfolios requires a clear definition of the investor's preferences about risk.

While the *optimization phase* is characterized by objectivity because it is valid for the entire market and not for the single investor, the *maximization phase* is characterized by subjectivity because it is the function of the investor's risk preferences. An analysis of the entire portfolio selection process based on the optimization and maximization phases can be carried out according to four main steps:

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_5

– (*step 1*) construction of the diversified portfolio;
– (*step 2*) construction of the efficient frontier;
– (*step 3*) definition of the efficient portfolios;
– (*step 4*) choice of the optimal portfolio.

The first three steps (1, 2, 3) define the *optimization phase* while the last step (4) defines the *maximization phase*.

5.1 Diversified Portfolio

The mean-variance approach is the most widely used in the portfolio selections (Markowitz 1952, 1956, 1959, 1976, 2014, 2016; Tobin 1958). The portfolio selection is based on two variables:

(1) expected value of the portfolio return (μ_P);
(2) variance of the expected portfolio returns (σ_P^2) measuring the portfolio's risk.

The efficient portfolio must satisfy the Pareto optimal condition: it is impossible to improve one objective without worsening the other. Therefore, the investor prefers the portfolio that is capable of maximising its expected return to an equal variance or the portfolio capable of minimising its variance to an equal expected return. Therefore, the portfolio is efficient only if:

– for a given level of expected return, it minimizes the variance (and therefore the portfolio risk);
– for a given level of variance, it maximizes the portfolio expected return.

By considering a portfolio (A) with mean μ_A and variance σ_A^2 and a portfolio (B) with mean μ_B and variance σ_B^2, the portfolio (A) dominates (it is strictly preferred) the portfolio (B) if one of the following two conditions is achieved:

$$A \succ B \quad \text{if} \begin{cases} \mu_A > \mu_B; \quad \sigma_A^2 \leq \sigma_B^2 \\ \qquad\quad or \\ \sigma_A^2 < \sigma_B^2; \quad \mu_A \geq \mu_B \end{cases}$$

This approach simplifies the problem of portfolio selection. There are two main advantages: first, it does not require specification about probability distribution; second, it is simple and intuitive because it is only based on the mean and variance. However, it is also true that this approach neglects a lot of relevant information about the probability distribution.

The portfolio selection in the mean-variance approach can be considered solely as a problem of definition of the weights to be assigned to each asset in the portfolio. Indeed, the statistical characteristics of each asset, with regards to the first and second order (mean, variance, and covariance) are known.

Similarly to the utility function approach, also in this case there are two main phases of the portfolio selection process (among the others: Elton et al. 2013; Elton and Gruber 1977; Ledoit and Wolf 2003; Epps 1981; Jennings 1971; Johnson and Shannon 1974; Rubinstein 1973; Statman 1987; Wagner and Lau 1971; Cesari 2012a, b; Castellani et al. 2005; Brennan and Kraus 1976; Brown and Barry 1985; Brumelle 1974; Latane et al. 1971; Meyers 1973; Canner 1997; Cass and Stiglitz 1970; Dalal 1983; Edwards and Goetzmann 1994; Robichek and Cohn 1974; Elton and Gruber 1971, 1974; Hakansson 1970; Merton 1972; Mossin 1968; Ohlson 1975; Pye 1973; Smith 1968; Sunder 1980; Sharpe 1971; Schafer et al. 1976; Russ and Rosenberg 1980; Pogue and Solnik 1974; Officer 1973; Martin and Klemkosky 1975; Kryzanowski and To 1983; Fama 1968, 1981; Fama and MacBeth 1973; Farrell 1974; Francis 1975):

(1) *optimization phase*: the aim is to define the diversified portfolio and the efficient frontier.
The definition of the diversified portfolio is based on the statistical characteristics of the assets. Specifically, the portfolio expected return is equal to the weighted average of the expected returns of the assets, while the portfolio variance is the function of the covariance between the assets' expected returns.
The definition of the efficient frontier is based on two main steps. The first step requires consideration of the *N* risky assets in the portfolio. In this case, the efficient frontier is characterized by a hyperbolic form. The second step requires consideration of the *N* risky assets and one risk-free asset in the portfolio. In this case, the efficient frontier is characterized by a linear form. Specifically, it is a straight line born on the level of risk-free rate on ordinate and it is tangent to the curve.
The assumption refers to the investors' homogeneous expectations about the statistical characteristics of the assets implying that all investors define the same efficient frontier, both hyperbolic and linear.
(2) *maximization phase*: the aim is to choose the optimal portfolio among the efficient portfolios defined on the efficient frontier. None of the efficient portfolios on efficient frontier can be preferred over all of the others by definition. The choice of the optimal portfolio among the efficient portfolios requires the clear definition of the investor's preferences about risk. Therefore, in the maximization phase a key role is played by the investor's preferences mainly regarding risk. In this context, the utility function can be used to choose the optimal portfolio among the efficient portfolios defined on the basis of the efficient frontier.

It is worth noting that while the *optimization phase* is characterized by objectivity, because it is valid for the entire market and not for the single investor, the *maximization phase* is characterized by subjectivity, because it is the function of the investor's preferences about risk.

Based on these two phases, the portfolio selection process can be divided up into four main steps (Cesari 2012b):

– (*step 1*) *construction of the diversified portfolio*;
– (*step 2*) *construction of the efficient frontier*;
– (*step 3*) *definition of the efficient portfolios*;
– (*step 4*) *choice of the optimal portfolio.*

While the first three steps (n. 1, 2, 3) define the *optimization phase*, the last step (n. 4) defines the *maximization phase*.

Note that in this context "asset" is considered stock, bond or another portfolio.

This paragraph focuses on the construction of the diversified portfolio (step 1), while the other three paragraphs are focused respectively on the construction of the efficient frontier (step 2), definition of the efficient portfolios (step 3) and choice of the optimal portfolio (step 4).

Therefore, the first step of the process is the construction of the diversified portfolio. It requires the definition of:

(a) portfolio expected return
(b) portfolio variance.

Portfolio Expected Return

The problem of the estimate of the portfolio expected return has a simple solution. Indeed, the portfolio expected return is equal to the weighted average of the expected return of its assets. The weight of the assets is equal to the part of wealth invested in it.

Assuming a portfolio of two assets denoted as *Asset 1* and *Asset 2* and assuming that the wealth invested in the portfolio is equal to w and that the weights of the two assets are α_1 and α_2 respectively, the result is:

$$w_1 + w_2 = w \rightarrow \alpha_1 = \frac{w_1}{w}; \quad \alpha_2 = \frac{w_2}{w} \rightarrow \alpha_1 + \alpha_2 = 1 \tag{5.1}$$

Assuming that the expected return of the two assets in the state of nature s are equal to $R_{1,s}$ for Asset 1 and $R_{2,s}$ for the Asset 2, the expected return of the portfolio (y_s) can be defined as follows (Saltari 2011):

$$\mathrm{y_s} = \mathrm{w}\left(\alpha_1 \mathrm{R}_{1,s} + \alpha_2 \mathrm{R}_{2,s}\right) \tag{5.2}$$

In Eq. (5.2) the weights assigned to the two assets $(\alpha_1; \alpha_2)$ are the only unknown variables. The choice of the couple $(\alpha_1; \alpha_2)$ allows for definition of the optimal portfolio.

In order to simplify, assuming that the wealth invested in the portfolio is equal to 1 $(\mathrm{w} = 1)$, Eq. (5.2) can be re-written as follows (Saltari 2011):

$$\mathrm{E}(\mathrm{y_s}) = \alpha_1 \mathrm{E}\left(\mathrm{R}_{1,s}\right) + \alpha_2 \mathrm{E}\left(\mathrm{R}_{2,s}\right)$$

and by considering that $E(.) \equiv \mu(.)$, we have:

$$\mu_P = \alpha_1\mu_1 + \alpha_2\mu_2 \tag{5.3}$$

Equation (5.3) shows that the portfolio expected return (μ_P) is equal to the linear combination of the expected returns of the two assets $(\mu_1; \mu_2)$ weighted on the basis of their weights $(\alpha_1; \alpha_2)$ in the portfolio.

To further understand the relevance of Eq. (5.3) is possible to define the problem on the basis of the following two conditions:

(1) definition of the constraint in terms of α and $(1-\alpha)$ and defining the wealth invested in the Asset 2 (α_2) equal to α (so that $\alpha_2 = \alpha$) and the part of wealth invested in the Asset 1 (α_1) equal to $(1-\alpha)$ (so that $\alpha_1 = 1-\alpha$);
(2) assuming that Asset 2 is riskier than Asset 1, on the basis of the direct relationship between risk and return (the higher risk, the higher return), the variance and mean of the Asset 2 are greater than Asset 1.

These due conditions can be formalized as follows:

$$\begin{gathered}\alpha + (1-\alpha) = 1 \rightarrow \begin{cases}\alpha_2 = \alpha \\ \alpha_1 = 1-\alpha\end{cases} \quad 0<\alpha<1 \\ \sigma_2^2 > \sigma_1^2\,(\sigma_2 > \sigma_1) \leftrightarrow \mu_2 > \mu_1\end{gathered} \tag{5.4}$$

On the basis of the conditions (5.4), Eq. (5.3) can be re-written as follows:

$$\mu_P = \mu_1 + \alpha(\mu_2 - \mu_1) \tag{5.5}$$

Equation (5.5) shows three main results:

(1) the relationship between the portfolio expected return (μ_P) and the assets' expected returns $(\mu_1; \mu_2)$ is linear;
(2) the slope of the straight line $(\mu_2 - \mu_1)$ is always positive because Asset 2 is riskier than Asset 1 and therefore $\mu_2 > \mu_1$. The distance between the expected returns of the two assets $(\mu_2; \mu_1)$ is a constant because the statistical characteristics of the two assets are known;
(3) the portfolio expected return (μ_P) is the function of the wealth invested in the riskier asset (Asset 2) (α); consequently, the higher the wealth invested in Asset 2, the higher the portfolio expected return.

Therefore, Eq. (5.5) draws a straight line in the plane $(\alpha; \mu_P)$ as in Fig. 5.1.

If short selling is not permitted, the value of α changes between zero and one $(0 \leq \alpha \leq 1)$. Specifically, if:

- $\alpha = 0$: wealth is invested only in the less risky asset (Asset 1);
- $\alpha = 1$: wealth is invested only in the riskier asset (Asset 2).

Otherwise, if short selling is permitted the value of α changes beyond zero and over one $(\alpha < 0;\ \alpha > 1)$, as indicated by the dotted line in Fig. 5.1. Specifically, if:

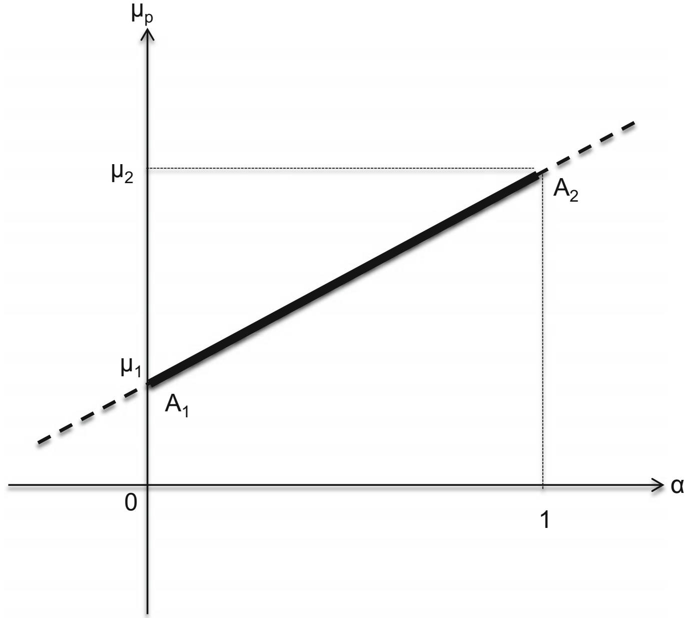


Fig. 5.1 Portfolio expected return as function of wealth invested in the riskier asset (α)

– $\alpha < 0$: there are short selling operations of Asset 2 and buying of Asset 1: this is a risk hedging operation because Asset 2 is riskier than Asset 1;
– $\alpha > 1$: there are short selling operations of Asset 1 and buying of Asset 2: this is a speculative operation because Asset 2 is riskier than Asset 1.

In general, for N assets in the portfolio, the portfolio expected return $(E(R_P) \equiv \mu_P)$ is equal to the weighted average of the expected returns of the assets in portfolio $(E(R_k) \equiv \mu_k)$. Their weights are equal to the share of wealth invested in each one $(\alpha_k = w_k/w)$ as follows:

$$E(R_P) = \sum_{k=1}^{n} \alpha_k E(R_k) \leftrightarrow \mu_P = \sum_{k=1}^{n} \alpha_k \mu_k \quad \text{with } \alpha_k = \frac{w_k}{w} \tag{5.6}$$

Portfolio Variance

The portfolio variance measures the variance of the expected portfolio return. It is not equal to the weighted average of the variance of the assets in the portfolio. Indeed, it depends on the covariance, and therefore correlations between the expected return of all assets in the portfolio considered in a pair.

Generally, the portfolio variance $\left(\sigma_P^2\right)$ is equal to the expected value of the squared difference between the expected return of the portfolio and its average expected return, as follows:

$$\sigma_P^2 = E\left(R_p - \mu_p\right)^2$$

Considering that:

$$R_P = \sum_{k=1}^{n} \alpha_k R_k; \quad \mu_P = \sum_{k=1}^{n} \alpha_k \mu_k$$

and by substituting, we have:

$$\sigma_P^2 = E\left(R_p - \mu_p\right)^2 = E\left(\sum_{k=1}^{n} \alpha_k R_k - \sum_{k=1}^{n} \alpha_k \mu_k\right)^2 \tag{5.7}$$

The analysis of the portfolio's variance can be performed on the basis of two main cases:

- (*Case 1*) two assets in the portfolio;
- (*Case 2*) more than two assets in the portfolio.

(*Case 1*) *Two Assets in the Portfolio*

Assuming a portfolio built on two assets: Asset 1 and Asset 2. Their weights are α_1 and α_2 respectively. On the basis of Eq. (5.7), the portfolio's variance $\left(\sigma_P^2\right)$ is equal to:

$$\sigma_P^2 = E[\alpha_1 R_1 + \alpha_2 R_2 - (\alpha_1 \mu_1 + \alpha_2 \mu_2)]^2$$

and then:

$$\sigma_P^2 = E[\alpha_1(R_1 - \mu_1) + \alpha_2(R_2 - \mu_2)]^2$$
$$\sigma_P^2 = E\left[(\alpha_1(R_1 - \mu_1))^2 + (\alpha_2(R_2 - \mu_2))^2 + 2(\alpha_1(R_1 - \mu_1))(\alpha_2(R_2 - \mu_2))\right]$$
$$\sigma_P^2 = E\left[\alpha_1^2(R_1 - \mu_1)^2 + \alpha_2^2(R_2 - \mu_2)^2 + 2\alpha_1\alpha_2(R_1 - \mu_1)(R_2 - \mu_2)\right]$$

Considering that: (i) the expected value of the sum of the returns is equal to the sum of the expected value of each return; and (ii) the expected value of the return for a constant is equal to the constant for the expected value of the return, the results are:

$$\sigma_P^2 = \alpha_1^2 E\left[(R_1 - \mu_1)^2\right] + \alpha_2^2 E\left[(R_2 - \mu_2)^2\right] + 2\alpha_1\alpha_2 E[(R_1 - \mu_1)(R_2 - \mu_2)]$$

and by considering that:

$$E\left[(R_1 - \mu_1)^2\right] = \sigma_1^2; \quad E\left[(R_2 - \mu_2)^2\right] = \sigma_2^2; \quad E[(R_1 - \mu_1)(R_2 - \mu_2)] = \sigma_{1,2}$$

We have:

$$\sigma_P^2 = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\alpha_1\alpha_2\sigma_{1,2} \tag{5.8}$$

Considering that the correlation coefficient $(\rho_{1,2})$ is equal to the ratio between the covariance of the assets' expected returns $(\sigma_{1,2})$ and the product of their standard deviations $(\sigma_1\sigma_2)$, we have:

$$\sigma_P^2 = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\alpha_1\alpha_2\rho_{1,2}\sigma_1\sigma_2 \tag{5.9}$$

$$\rho_{1,2} = \frac{\sigma_{1,2}}{\sigma_1\sigma_2} \rightarrow \sigma_{1,2} = \rho_{1,2}\sigma_1\sigma_2$$

On the basis of the correlation coefficient $(\rho_{1,2})$, the portfolio variance can be defined in terms of correlation rather than in terms of covariance. In this case, Eq. (5.8) can be re-written as follows:

$$\sigma_P^2 = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\alpha_1\alpha_2\rho_{1,2}\sigma_1\sigma_2 \tag{5.10}$$

As shown in Eq. (5.10) a key role is played by the covariance (and therefore by correlation) between the assets' expected returns. There are three basic cases:

- *positive correlation* $(\rho_{1,2} = +1)$: in this case the correlation is perfectly positive between the assets' expected returns. They move in the same direction;
- *negative correlation* $(\rho_{1,2} = -1)$: in this case the correlation is perfectly negative between the assets' expected returns. They move in the opposite direction;
- *no correlation* $(\rho_{1,2} = 0)$: in this case there is no correlation between the assets' expected return. They move independently between them.

Equation (5.10) can be rewritten on the basis of the conditions (5.4) as follows:

$$\sigma_P^2 = \alpha^2\sigma_2^2 + (1-\alpha)^2\sigma_1^2 + 2\alpha(1-\alpha)\rho_{1,2}\sigma_1\sigma_2 \tag{5.11}$$

and by explicating, we have:

$$\sigma_P^2 = \alpha^2\sigma_2^2 + \sigma_1^2 + \alpha^2\sigma_1^2 - 2\alpha\sigma_1^2 + 2\alpha\rho_{1,2}\sigma_1\sigma_2 - 2\alpha^2\rho_{1,2}\sigma_1\sigma_2$$

Equation (5.11) draws a curve. On the basis of the first order condition, and then by pointing the first derivative equal to zero, the stationary point is achieved. If the second derivative is positive, the stationary point is a minimum of the curve.

The first derivative of the portfolio variance (σ_P^2), compared with part of wealth (α) invested in the riskier asset (Asset 2), is equal to:

$$\frac{\partial\sigma_P^2}{\partial\alpha} = 2\alpha\sigma_2^2 + 2\alpha\sigma_1^2 - 2\sigma_1^2 + 2\rho_{1,2}\sigma_1\sigma_2 - 4\alpha\rho_{1,2}\sigma_1\sigma_2$$

and then the first derivative is equal to:

$$\frac{\partial \sigma_P^2}{\partial \alpha} = \alpha\sigma_2^2 + \alpha\sigma_1^2 - \sigma_1^2 + \rho_{1,2}\sigma_1\sigma_2 - 2\alpha\rho_{1,2}\sigma_1\sigma_2 \tag{5.12}$$

Placing the first derivative equal to zero, and by solving for α, the stationary point is achieved, as follows:

$$\frac{\partial \sigma_P^2}{\partial \alpha} = 0 \rightarrow \alpha\sigma_2^2 + \alpha\sigma_1^2 - \sigma_1^2 + \rho_{1,2}\sigma_1\sigma_2 - 2\alpha\rho_{1,2}\sigma_1\sigma_2 = 0$$

$$\alpha = \frac{\sigma_1^2 - \rho_{1,2}\sigma_1\sigma_2}{\sigma_2^2 + \sigma_1^2 - 2\rho_{1,2}\sigma_1\sigma_2} \tag{5.13}$$

The second derivative is equal to:

$$\frac{\partial^2 \sigma_P^2}{\partial^2 \alpha} = \sigma_2^2 + \sigma_1^2 - 2\rho_{1,2}\sigma_1\sigma_2 \tag{5.14}$$

The sign of the second derivative is the function of the correlation coefficient $(\rho_{1,2})$. Consider the three extreme cases of the correlation $(\rho_{1,2} = +1; -1; 0)$. If:

$$\begin{aligned}
&\rho_{1,2} = +1 \rightarrow \frac{\partial^2 \sigma_P^2}{\partial^2 \alpha} = \sigma_2^2 + \sigma_1^2 - 2\sigma_1\sigma_2 = (\sigma_2 - \sigma_1)^2 \rightarrow \frac{\partial^2 \sigma_P^2}{\partial^2 \alpha} > 0 \\
&\rho_{1,2} = -1 \rightarrow \frac{\partial^2 \sigma_P^2}{\partial^2 \alpha} = \sigma_2^2 + \sigma_1^2 + 2\sigma_1\sigma_2 = (\sigma_2 + \sigma_1)^2 \rightarrow \frac{\partial^2 \sigma_P^2}{\partial^2 \alpha} > 0 \\
&\rho_{1,2} = 0 \rightarrow \frac{\partial^2 \sigma_P^2}{\partial^2 \alpha} = \sigma_2^2 + \sigma_1^2 \rightarrow \frac{\partial^2 \sigma_P^2}{\partial^2 \alpha} > 0
\end{aligned} \tag{5.15}$$

In the first case $(\rho_{1,2} = +1)$ the second derivative is positive because $\sigma_2 > \sigma_1$ (Asset 2 is riskier than Asset 1). Also in the first case, as in the second case $(\rho_{1,2} = -1)$, it is a square of polynomial and therefore the second derivative is positive by definition. In the last case $(\rho_{1,2} = 0)$, the second derivative is positive by definition because the variances, and therefore the standard deviation, are positive.

Therefore, the sign of the second derivative is always positive. Consequently, the curve is convex and the stationary point defined by Eq. (5.13) is a minimum point of the curve. Specifically, α is the part of wealth to be invested in the riskier asset (Asset 2) to minimize the portfolio variance and therefore portfolio risk.

Now the problem is to define the correct position of α with regards to its abscissa. Its sign is the function of the sign of the correlation $(\rho_{1,2})$ and therefore: $-1 \leq \alpha \leq +1$.

Also in this case, by considering the tree extreme cases $(\rho_{1,2} = +1; -1; 0)$, and substituting in Eq. (5.13) we have:

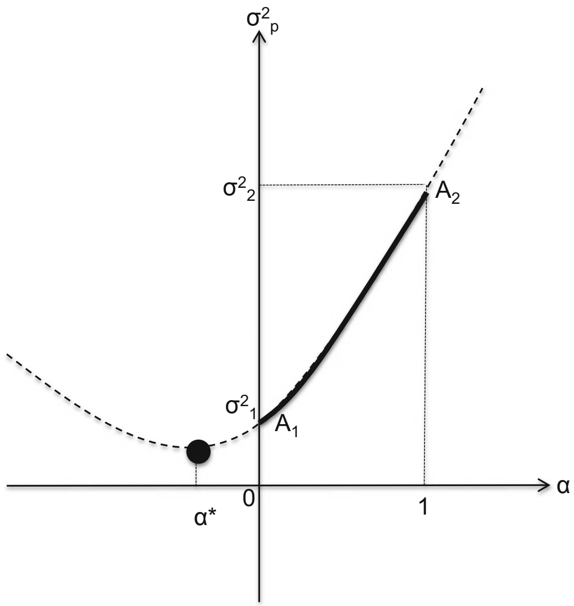
$$
\begin{aligned}
\rho_{1,2} &= +1 \rightarrow \alpha = \frac{\sigma_1^2 - \sigma_1\sigma_2}{\sigma_2^2 + \sigma_1^2 - 2\sigma_1\sigma_2} = \frac{\sigma_1(\sigma_1 - \sigma_2)}{(\sigma_2 - \sigma_1)^2} \\
&= \frac{\sigma_1}{(\sigma_1 - \sigma_2)} \rightarrow \sigma_2 > \sigma_1 \rightarrow \alpha < 0 \rightarrow -1 \leq \alpha \\
\rho_{1,2} &= -1 \rightarrow \alpha = \frac{\sigma_1^2 + \sigma_1\sigma_2}{\sigma_2^2 + \sigma_1^2 + 2\sigma_1\sigma_2} = \frac{\sigma_1(\sigma_1 + \sigma_2)}{(\sigma_2 + \sigma_1)^2} \\
&= \frac{\sigma_1}{(\sigma_1 + \sigma_2)} \rightarrow \alpha > 0 \rightarrow 0 \leq \alpha \leq +1 \\
\rho_{1,2} &= 0 \rightarrow \alpha = \frac{\sigma_1^2}{\sigma_2^2 + \sigma_1^2} = \frac{\sigma_1}{(\sigma_1 + \sigma_2)} \rightarrow \alpha > 0 \rightarrow 0 \leq \alpha \leq +1
\end{aligned}
\tag{5.16}
$$

Therefore, in the presence of a perfect positive correlation $(\rho_{1,2} = +1)$ between the expected returns of the two assets, the portfolio variance is minimized by a negative value of investment in the riskier asset $(\alpha < 0)$. Therefore, this point can be achieved only by short selling as shown in Fig. 5.2 (adapted from Castellani et al. 2005).

Otherwise, in both cases of the presence of perfect negative correlation $(\rho_{1,2} = -1)$ and absence of correlation $(\rho_{1,2} = 0)$ between the expected returns of the two assets, the portfolio variance is minimized by a positive value of investment in the riskier asset $(\alpha > 0)$ as shown in Fig. 5.3 (adapted from Castellani et al. 2005).

At this point of the analysis, it is relevant to analyse shifting from the variance to the standard deviation of the portfolio.

Fig. 5.2 The portfolio variance in the case of positive correlation



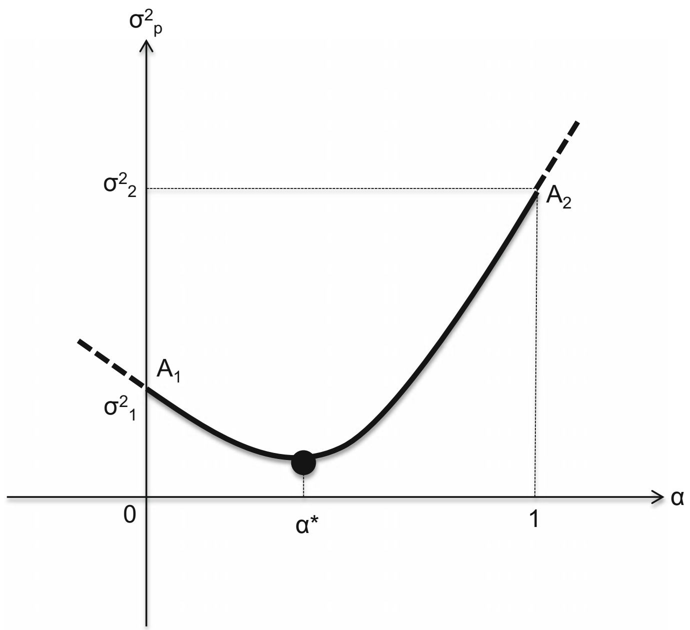


Fig. 5.3 The portfolio variance in the case of negative correlation

By starting from the portfolio's variance as defined by Eq. (5.11), the three extreme cases of the correlation $\left(\rho_{1,2} = +1; -1; 0\right)$ can be considered.

If the correlation between the expected returns of the two assets is positive $\left(\rho_{1,2} = +1\right)$, Eq. (5.11) can be rewritten as follows:

$$\sigma_P^2 = \alpha^2\sigma_2^2 + (1-\alpha)^2\sigma_1^2 + 2\alpha(1-\alpha)(1)\sigma_1\sigma_2$$

and then:

$$\sigma_P^2 = [\alpha\sigma_2 + (1-\alpha)\sigma_1]^2 = [\sigma_1 + \alpha(\sigma_2 - \sigma_1)]^2 \tag{5.17}$$

Asset 2 is riskier than Asset 1. Therefore, $\sigma_2 > \sigma_1$ and then the argument of the square is always positive. Consequently, it is possible to move directly from the portfolio variance $\left(\sigma_P^2\right)$ to its standard deviation $\left(\sigma_p\right)$ as follows:

$$\sigma_P = \sigma_1 + \alpha(\sigma_2 - \sigma_1)$$

and solving for part of wealth α invested in the riskier asset (Asset 2), we have:

$$\alpha = \frac{\sigma_P - \sigma_1}{\sigma_2 - \sigma_1} \tag{5.18}$$

Therefore, the portfolio with zero standard deviation is obtained by investing a part of wealth in the riskier asset equal to:

$$\sigma_P = 0 \to \sigma_1 + \alpha(\sigma_2 - \sigma_1) = 0 \to \alpha = \frac{\sigma_1}{\sigma_1 - \sigma_2} \equiv \frac{-\sigma_1}{\sigma_2 - \sigma_1} \to \sigma_2 > \sigma_1 \to \alpha < 0 \tag{5.19}$$

As shown in the case of the portfolio variance, also in this case, clearly to minimize the portfolio's standard deviation and therefore its risk, the part of wealth to be invested in the riskier asset (Asset 2) is negative. Consequently, the minimum risk of the portfolio can only be achieved by short selling if the correlation between the expected returns of the two assets is positive.

On the basis of Eq. (5.5) the portfolio expected return in terms of its standard deviation can be defined, as follows (Cesari 2012b):

$$\mu_P = \mu_1 + \left(\frac{\sigma_P - \sigma_1}{\sigma_2 - \sigma_1}\right)(\mu_2 - \mu_1) \leftrightarrow \mu_P = \mu_1 - \sigma_1\left(\frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}\right) + \sigma_P\left(\frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}\right) \tag{5.20}$$

Equation (5.20) shows that, since there is a perfect positive linear correlation between the expected returns of the two assets, the risk and the expected returns of the portfolio are a linear combination of the risk and the expected returns of assets.

Considering that the statistical characteristics of the two assets $(\mu_1; \sigma_1; \mu_2; \sigma_2)$ are known, and Asset 2 is riskier than Asset 1 (so that: $\sigma_2 > \sigma_1; \mu_2 > \mu_1 \to \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1} > 0$), the portfolio's expected return is linear function of σ_P. Specifically, Eq. (5.20) draws a straight line with intercept equal to $\left(\mu_1 - \sigma_1\left(\frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}\right)\right)$ and the positive slope equal to $\left(\frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}\right)$. Consequently, all possible combinations between the two assets expected returns must be positioned on a straight line as in Fig. 5.4.

Therefore, in this case the portfolio's volatility is linear function of the α and so the part of wealth invested in Asset 2 (the riskier asset). Since the standard deviation of Asset 2 (σ_2) is function of α, a line with interception equal to σ_1 is achieved and with an angular coefficient equal to $(\sigma_2 - \sigma_1)$. Then, with an increase in part of the wealth invested in Asset 2 (the riskier asset), and then an increase in α, the portfolio's risk increases linearly. All possible assets combinations cannot present a risk higher than the risk represented by the line between the two assets.

If the correlation between the expected returns of the two assets is negative $(\rho_{1,2} = -1)$, Eq. (5.11) can be rewritten as follows:

$$\sigma_P^2 = \alpha^2\sigma_2^2 + (1 - \alpha)^2\sigma_1^2 + 2\alpha(1 - \alpha)(-1)\sigma_1\sigma_2$$

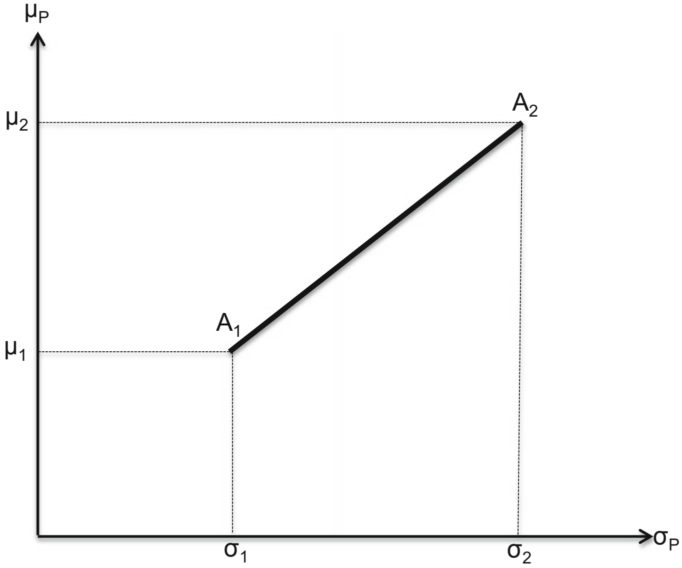


Fig. 5.4 Standard deviation and return of the portfolio in the case of perfect positive correlation $\left(\rho_{1,2} = +1\right)$ between expected returns of the two assets

and then:

$$\sigma_P^2 = [\alpha\sigma_2 - (1-\alpha)\sigma_1]^2 = [\alpha(\sigma_1 + \sigma_2) - \sigma_1]^2 \tag{5.21}$$

In this case, it is not possible to move from the portfolio variance $\left(\sigma_P^2\right)$ to its standard deviation $\left(\sigma_p\right)$ directly, as in the previous case. The argument under the square is function of the part of wealth invested in Asset 2 (the riskier asset) and of the distance between the standard deviation of Asset 2 and Asset 1. Consequently:

$$\sigma_p = |\alpha(\sigma_1 + \sigma_2) - \sigma_1| \rightarrow \sigma_P = \begin{cases} \alpha(\sigma_1 + \sigma_2) - \sigma_1 & \text{if} \quad \alpha(\sigma_1 + \sigma_2) - \sigma_1 > 0 \\ \sigma_1 - \alpha(\sigma_1 + \sigma_2) & \text{if} \quad \alpha(\sigma_1 + \sigma_2) - \sigma_1 < 0 \end{cases} \tag{5.22}$$

There are two linear relationships: one increasing (if positive) and the other decreasing (if negative). These two lines define the different level of the portfolio's risk (σ_P) to the α changes and therefore to the changes of part of wealth invested in the riskier asset (Asset 2).

Since one is always positive when the other is negative, there is always a single solution for calculation of the portfolio risk and return.

Therefore, from the first linear relationship, it follows that:

$$\sigma_P = \alpha(\sigma_1 + \sigma_2) - \sigma_1 \rightarrow \alpha = \frac{\sigma_P + \sigma_1}{\sigma_1 + \sigma_2} \tag{5.23}$$

Also in this case it is possible to define the portfolio's expected return in terms of standard deviation. By substituting α in Eq. (5.5) the portfolio's expected return is equal to:

$$\mu_P = \mu_1 + \left(\frac{\sigma_P + \sigma_1}{\sigma_1 + \sigma_2}\right)(\mu_2 - \mu_1) \leftrightarrow \mu_P = \mu_1 + \sigma_1\left(\frac{\mu_2 - \mu_1}{\sigma_1 + \sigma_2}\right) + \sigma_P\left(\frac{\mu_2 - \mu_1}{\sigma_1 + \sigma_2}\right) \tag{5.24}$$

Therefore, the portfolio's expected return (μ_P) is linear function of its standard deviation (σ_P). Equation (5.24) draws a straight line with interception equal to $\left(\mu_1 + \sigma_1\left(\frac{\mu_2-\mu_1}{\sigma_1+\sigma_2}\right)\right)$ and slope equal to $\left(\frac{\mu_2-\mu_1}{\sigma_1+\sigma_2}\right)$. Since Asset 2 is riskier than Asset 1 (so that: $\sigma_2 > \sigma_1; \mu_2 > \mu_1 \rightarrow \frac{\mu_2-\mu_1}{\sigma_1+\sigma_2} > 0$), the slope is positive.

Similarly, from the second linear equation, we have:

$$\sigma_P = \sigma_1 - \alpha(\sigma_1 + \sigma_2) \rightarrow \alpha = \frac{\sigma_1 - \sigma_P}{\sigma_1 + \sigma_2} \tag{5.25}$$

In this case, the portfolio's expected return in terms of standard deviation is equal to:

$$\mu_P = \mu_1 + \left(\frac{\sigma_1 - \sigma_P}{\sigma_1 + \sigma_2}\right)(\mu_2 - \mu_1) \leftrightarrow \mu_P = \mu_1 + \sigma_1\left(\frac{\mu_2 - \mu_1}{\sigma_1 + \sigma_2}\right) + \sigma_P\left(\frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2}\right) \tag{5.26}$$

The portfolio's expected return (μ_P) is linear function of its standard deviation (σ_P). Furthermore, also in this case Eq. (5.26) draws a straight line with interception equal to $\left(\mu_1 + \sigma_1\left(\frac{\mu_2-\mu_1}{\sigma_1+\sigma_2}\right)\right)$ and slope equal to $\left(\frac{\mu_2-\mu_1}{\sigma_1+\sigma_2}\right)$. Since the Asset 2 is riskier than Asset 1 (so that $\sigma_2 > \sigma_1; \mu_2 > \mu_1 \rightarrow \frac{\mu_1-\mu_2}{\sigma_1+\sigma_2} < 0$) the slope is negative.

Therefore, based on Eqs. (5.24) and (5.26), the relationship between risk and return of the portfolio with perfect negative correlation $(\rho_{1,2} = -1)$ among the expected returns of the two assets can be represented as in Fig. 5.5.

In the case of negative correlation between the assets' expected returns, their systemic movements in opposite directions allows for cancellation of the risk. It is achieved in the point of intersection of the two lines on the ordinate.

Specifically, by assuming a portfolio with zero standard deviation, the part of wealth invested in the riskier asset is equal to:

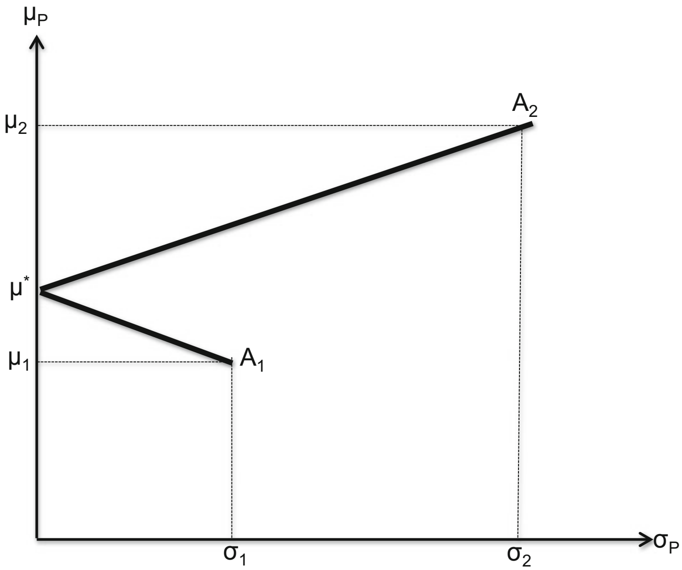


Fig. 5.5 Standard deviation and return of the portfolio in the case of perfect negative correlation $(\rho_{1,2} = -1)$ between the expected returns of the two assets

$$\sigma_P = 0 \rightarrow \begin{matrix} \alpha(\sigma_1 + \sigma_2) - \sigma_1 = 0 \rightarrow \alpha = \frac{\sigma_1}{\sigma_1 + \sigma_2} \\ \sigma_1 - \alpha(\sigma_1 + \sigma_2) = 0 \rightarrow \alpha = \frac{\sigma_1}{\sigma_1 + \sigma_2} \end{matrix} \rightarrow \alpha > 0 \tag{5.27}$$

Equation (5.27) shows two main elements:

- first, the point in which the two linear equations have the same value, the standard deviation of the portfolio is equal to zero and therefore the portfolio has no risk;
- second, the minimization of the portfolio's standard deviation can be achieved for a positive value of the investment in the riskier asset (Asset 2): $0 < \alpha^* < 1$.

Therefore, in the case of a perfect positive correlation, the portfolio with minimum risk can be achieved for negative value of investment $(\alpha^* < 0)$ in the riskier asset (Asset 2) by short selling, while in the perfect negative correlation it is achieved for positive value of investment $(0 < \alpha^* < 1)$ in the riskier asset.

The analysis of the two cases of perfect positive and negative correlation between the expected returns of the two assets $(\rho_{1,2} = +1; \rho_{1,2} = -1)$ shows that the portfolio's risk in terms of standard deviation (σ_P) is higher for the positive correlation than the negative one as indicated in Fig. 5.6.

In the plan $(\sigma_P; \mu_P)$, the portfolio's standard deviation is lower in the case of a perfect negative correlation $(\rho_{1,2} = -1)$ and it is higher in the case of a perfect positive correlation $(\rho_{1,2} = +1)$. The space defined by the lines classifies all

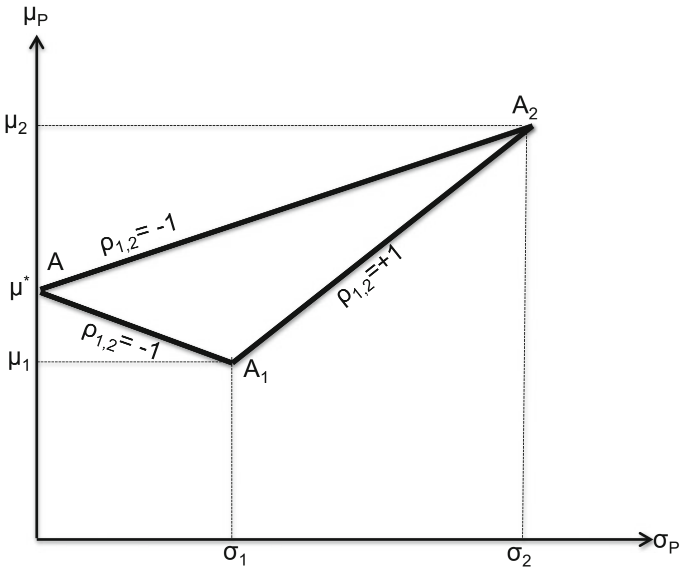


Fig. 5.6 Relationship between risk and return for $\rho_{1,2} = +1$ and $\rho_{1,2} = -1$

possible portfolios obtained by the changes in the correlations between the assets' expected returns $\left(-1 < \rho_{1,2} < +1\right)$.

If there is no correlation between the expected returns of the two assets is negative $\left(\rho_{1,2} = 0\right)$, Eq. (5.11) can be rewritten as follows:

$$\sigma_P^2 = \alpha^2\sigma_2^2 + (1-\alpha)^2\sigma_1^2 + 2\alpha(1-\alpha)(0)\sigma_1\sigma_2$$

and then:

$$\sigma_P^2 = \alpha^2\sigma_2^2 + (1-\alpha)^2\sigma_1^2 = \alpha^2\sigma_2^2 + \sigma_1^2 + \alpha^2\sigma_1^2 - 2\alpha\sigma_1^2$$

$$\sigma_P^2 = \alpha^2\left(\sigma_2^2 + \sigma_1^2\right) - 2\alpha\sigma_1^2 + \sigma_1^2 \tag{5.28}$$

In this case, the vertex of the parabola can be used. In this case the minimum of the parabola is the minimum point of the portfolio's variance and standard deviation. Therefore, the part of wealth to invest in the riskier asset (Asset 2) α is equal to the abscissa of the parabola's vertex as follows:

$$\alpha = -\frac{b}{2a} = -\frac{-2\sigma_1^2}{2\left(\sigma_2^2 + \sigma_1^2\right)} = \frac{\sigma_1^2}{\sigma_2^2 + \sigma_1^2} \tag{5.29}$$

Considering that σ_1^2 and σ_2^2 are variances and therefore they are positive by definition, the value of α is positive $(\alpha > 0)$ and it is possible to move from the variance to standard deviation as follows:

$$\alpha = \frac{\sigma_1}{\sigma_1 + \sigma_2} \rightarrow \alpha^* > 0 \tag{5.30}$$

Considering that α is positive, the concave nature of the parabola is upwards and therefore α is a minimum point. This is the same result of the negative correlation.

Consequently, by investing a part of wealth in the riskier asset (Asset 2) equal to α, the portfolio's variance and standard deviation is at a minimum point. Indeed, in this case the portfolio's variance is lower than the case in which the total wealth is invested in Asset 1 (the less risky asset). Therefore, the *minimum portfolio risk* (as measured by variance or standard deviation) is obtained by investing a part of wealth equal to α in Asset 2 (the riskier asset) and the remaining wealth in Asset 1, as shown in Fig. 5.7.

The portfolio *A* is defined as the *minimum portfolio risk*. Therefore, it is necessary to define the value of α that allows for a reduction to the minimum level of the portfolio's risk (variance or standard deviation) and therefore the portfolio's risk. In other words, it is necessary to define the part of wealth to be invested in the riskier asset (Asset 2) in order to obtain the portfolio with the minimum risk.

It is possible to represent Fig. 5.7 with regards to α by moving from the plan $(\sigma_P; \mu_P)$ to the plan $(\sigma_P; \alpha)$ as in Fig. 5.8 (adapted from Castellani et al. 2005).

It is important to note that generally, correlations between the assets' expected returns tend to be different from zero, because all assets are affected by economic dynamics and they tend to be more positive than negative.

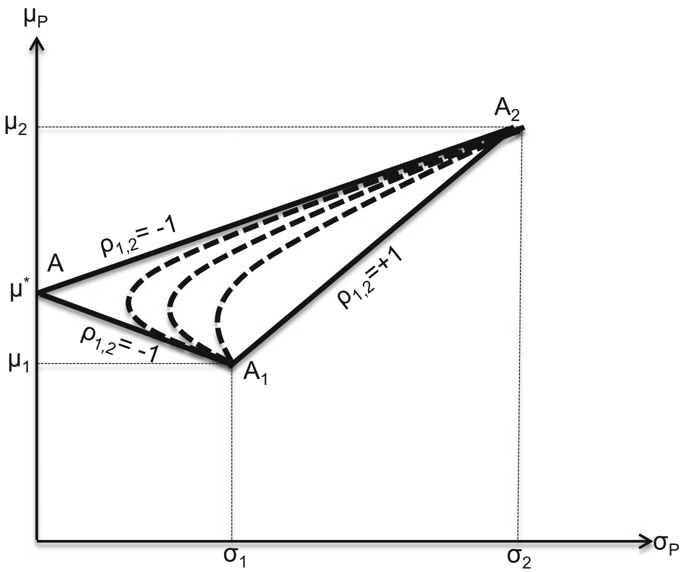


Fig. 5.7 Portfolios for the $-1 \leq \rho_{1,2} \leq +1$

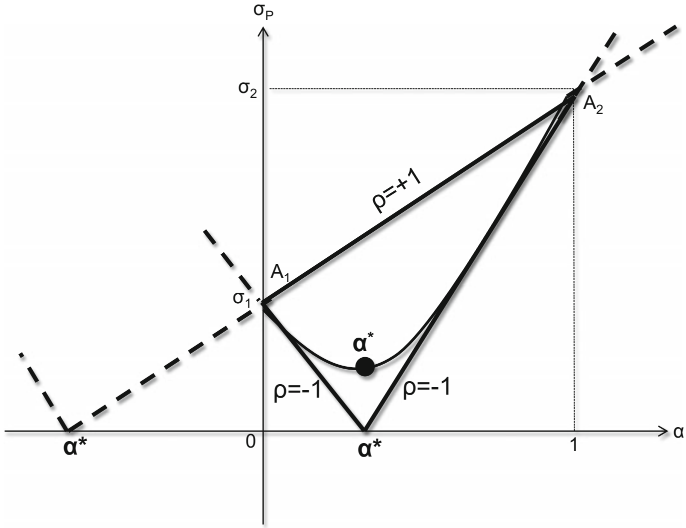


Fig. 5.8 Portfolio's standard deviation function of α

Specifically, the correlation coefficient changes between the range: $-1 \leq \rho_{1,2} \leq +1$. Theoretically the effects of diversification start for a $\rho_{1,2} < +1$ and they increase gradually until $\rho_{1,2} = -1$. The smaller the correlation, the more the curve bends; the greater inclination of the curve is achieved for $\rho_{1,2} = -1$, as shown in Figs. 5.7 and 5.8.

Unfortunately, in capital markets, the cases $\rho_{1,2} = -1$ is only theoretical (Ross et al. 2015). Generally, the cases of $\rho_{1,2} = +1; 0; -1$ represent a theoretical hypothesis and they create a space in which all portfolios can be found as function of the correlations between the assets' expected returns according to the changes of α, as shown in Figs. 5.7 and 5.8.

In capital markets the correlation between the assets' expected returns tend to be higher than zero and lower than 1 $\left(0 < \rho_{1,2} < +1\right)$. Consequently, in the absence of short selling (and therefore for positive value of assets in the portfolio), the diversification effect is greater as the lower the level of correlation between the asset's expected returns $(\rho \approx 0)$. Specifically, we have:

$$0 < \rho_{1,2} \leq \frac{\sigma_1}{\sigma_2} \tag{5.31}$$

with $\frac{\sigma_1}{\sigma_2}$ very small. In this case, the increase in wealth invested in Asset 2 (riskier asset) reduces the portfolio's variance more than the variance of the portfolio consisting of Asset 1 (the less risky asset) only. The part of wealth to be invested in Asset 2 to minimize the portfolio's variance is equal to:

$$\alpha = \frac{\sigma_1}{\sigma_1 + \sigma_2} \tag{5.32}$$

For investments in Asset 2 greater than α, the portfolio's variance increases.

The most important problem is the dimension of the ratio between the two standard deviations: the higher it is, and therefore the closer it gets to 1, the lower the effects of diversification.

It is worth noting that diversification does not have a value. It is a general concept that acquires value only by defining the target of risk and portfolio return. The trade-off between risk and return is the optimality mean-variance problem.

(Case 2) More than Two Assets in the Portfolio

The analysis developed for two assets in the portfolio can be generalized by considering a portfolio of N assets.

In this case the covariance, and therefore the correlation, between the expected returns of the assets are calculated by considering the assets in pairs.

To calculate the portfolio's variance $\left(\sigma_P^2\right)$ it is necessary to use the Covariance Matrix (C), as follows:

$$C = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n,1} & \sigma_{n,2} & \dots & \sigma_{n,n} \end{bmatrix} \tag{5.33}$$

In order to avoid the determinant of the matrix being equal to zero, the assets' returns in the rows and columns are not characterized by: (i) elements that are null; (ii) parallels rows or columns that are equal or proportional between them; (iii) parallels rows or columns that are linear combinations of each other.

It is worth noting that because correlation between pairs of assets (i-th and j-th) is calculated, the first index (i) can assume N value (one of each of the assets) while the second index (j) can assume a value of $(N-1)$ because $i \neq j$. Consequently, there are $N(N-1)$ correlation coefficients. Also, because the correlation coefficient between i and j is equal to the correlation coefficient between j and i, it is necessary to consider $[N(N-1)]/2$ correlation coefficients (for example, for 250 assets, there are 31.125 correlation coefficients to calculate). Formally:

$$\rho_{i,j} = [N(N-1)]/2 \tag{5.34}$$

The general term $\sigma_{k,j}$ (note that $\sigma_{k,j} = \sigma_{j,k}$ for symmetry) indicates the covariance between the expected returns of the k-th asset (I_k) and the j-th asset $\left(I_j\right)$ in portfolio, as follows:

$$\sigma_{k,j} = E\left[(I_k - \mu_k)\left(I_j - \mu_j\right)\right] \quad \text{per} \quad k, j = 1, 2, \dots, n \tag{5.35}$$

and $k = j$ we have a $\left(\sigma_k^2\right)$, and therefore the standard deviation (σ_k), as follows:

$$\sigma_{k,k} = E[(I_k - \mu_k)(I_k - \mu_k)] = E\left[(I_k - \mu_k)^2\right] = Var(I_k) = \sigma_k^2 \rightarrow \sigma_k = \sqrt{\sigma_k^2} \tag{5.36}$$

For positive values of the variance (and therefore not considering the risk-free asset that represents a specific case because its return is sure resulting in variance being null) we have:

$$\sigma_{k,j} = \sigma_k \sigma_j \rightarrow Positive\ Correlation$$
$$\sigma_{k,j} = -\sigma_k \sigma_j \rightarrow Negative\ Correlation$$
$$\sigma_{k,j} = 0 \rightarrow No\ Correlation$$

On the basis of the weight of each asset in the portfolio $\left(\alpha_k = \frac{W_k}{W} \text{ for } k = 1, 2, \ldots, n\right)$, it is possible to move from the Covariance Matrix (C) to the Covariance Weighted Matrix (C_α), as follows:

$$C_\alpha = \begin{bmatrix} \alpha_1\alpha_1\sigma_{1,1} & \alpha_1\alpha_2\sigma_{1,2} & \ldots & \alpha_1\alpha_n\sigma_{1,n} \\ \alpha_2\alpha_1\sigma_{2,1} & \alpha_2\alpha_2\sigma_{2,2} & \ldots & \alpha_2\alpha_n\sigma_{2,n} \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_n\alpha_1\sigma_{n,1} & \alpha_n\alpha_2\sigma_{n,2} & \ldots & \alpha_n\alpha_n\sigma_{n,n} \end{bmatrix} \tag{5.37}$$

Based on Eq. (5.7), the portfolio's variance $\left(\sigma_P^2\right)$ is equal to:

$$\sigma_P^2 = E\left[(R_P - \mu_P)^2\right] = \mathrm{E}\left[\left(\sum_{k=1}^{n} \alpha_k R_k - \sum_{k=1}^{n} \alpha_k \mu_k\right)^2\right]$$
$$= E\left[\left(\sum_{k=1}^{n} \alpha_k (R_k - \mu_k)\right)^2\right]$$

and therefore by adding all terms of the Covariance Weighted Matrix (C_α), the portfolio's variance $\left(\sigma_P^2\right)$ and standard deviation (σ_P) are equal to:

$$\sigma_P^2 = \sum_{k=1}^{n} \sum_{j=1}^{n} \alpha_k \alpha_j \sigma_{k,j} \rightarrow \sigma_P = \sqrt{\sum_{k=1}^{n} \sum_{j=1}^{n} \alpha_k \alpha_j \sigma_{k,j}} \tag{5.38}$$

Normally the linear correlation coefficient $\left(\rho_{k,j}\right)$ is used instead of the covariance: while the second only indicates the presence or absence of a linear relationship between the variables, the first measures the intensity of this relationship. It is equal to the ratio between the covariance among the returns of the k-th and j-th assets (σ_{kj}) and the product among their standard deviations $(\sigma_k; \sigma_j)$ as follows:

$$\rho_{k,j} = \frac{\sigma_{kj}}{\sigma_k \sigma_j} \rightarrow \sigma_{kj} = \rho_{k,j}\sigma_k\sigma_j$$

Using the correlation coefficient, the Correlation Matrix (C_ρ) can be used, as follows:

$$C_\rho = \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,n} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{n,1} & \rho_{n,2} & \cdots & 1 \end{bmatrix} \tag{5.39}$$

On the basis of the Correlation Matrix the portfolio's variance (σ_P^2) is computed by modifying Eq. (5.38) as follows:

$$\sigma_P^2 = \sum_{k=1}^{n}\sum_{j=1}^{n} \alpha_k \alpha_j \rho_{k,j} \sigma_k \sigma_j \rightarrow \sigma_P = \sqrt{\sum_{k=1}^{n}\sum_{j=1}^{n} \alpha_k \alpha_j \rho_{k,j} \sigma_k \sigma_j} \tag{5.40}$$

For greater understanding of the diversification effects due to the covariance (and therefore correlations) between the expected returns of the N assets in portfolio, a breakdown of the double summation as derived from the Covariance Weighted Matrix (C_α) may be useful, as follows (Elton et al. 2013):

$$\sigma_P^2 = \sum_{k=1}^{n} \alpha_k^2 \sigma_k^2 + \sum_{k=1}^{n} \sum_{\substack{j=1 \\ j \neq k}}^{n} \alpha_k \sigma_j \sigma_{k,j} \tag{5.41}$$

In the second part of Eq. (5.41) the double summation implies that the $j \neq k$. Indeed, for $j = k$ it results in the variance as for the matrix symmetry we have: $\alpha_{i,j} = \alpha_{j,i} \quad \forall i,j; i \neq j$.

Therefore, because $j = k$ and $k = j$, it is necessary to consider the square. Specifically:

- the first part refers to the terms of variance of each asset multiplied for the square of the investment. Therefore, it refers to the sum of the n terms square of the matrix;
- the second part refers to the terms of covariance between the expected return of the assets in portfolio considering them as a pair. Each covariance is multiplied by twice the product of the weights of the two assets according to the symmetry of the matrix. Therefore, it refers to the sum of the $N(N-1)$ triangle of the matrix (while k can assume N value, since $j \neq k$, j can assume $(N-1)$ value).

Equation (5.41) shows that the key term is the second. The most relevant effect on the portfolio's variance is represented by the covariance between the expected returns of the assets considered in a pair rather than the variance of the expected return of each asset. Consequently, the risk diversification is function of the covariance between the expected returns of the n assets in portfolio by considering them in a pair rather than their specific variance.

Specifically, if the:

- covariance is near to zero: the assets are linearly independent between them, and therefore the portfolio variance is lower than the variance of each asset;
- covariance is positive (negative): the assets are linearly dependent between them, and therefore the variance of the assets moves in the same direction (positive or negative). Therefore, the portfolio variance is higher the closer it gets to 1 (−1). In this case, the portfolio variance is higher than the variance of each asset.

Therefore, the risk of each asset is not important, but the change that it creates in the portfolio variance due to its introduction as function of the covariance between its expected returns and the expected returns of the other assets in portfolio.

It is worth noting that the concept of diversification is not defined in itself. It is not possible to define a general role capable of indicating the reduction of the portfolio's risk to the increase of assets.

For greater understanding of the diversification and its effects, we can assume a portfolio (P) composed of N risky assets. Assuming that the wealth (w) invested in the portfolio is equal to 1 $(w = 1)$, and the weight of each asset in the portfolio is equal to $\alpha_k = 1/n\,(k = 1, 2, \ldots, n)$ (Elton et al. 2013).

Based on Eq. (5.41) the portfolio variance $\left(\sigma_P^2\right)$ is equal to:

$$\sigma_P^2 = \sum_{k=1}^{n}\left(\frac{1}{n}\right)^2 \sigma_k^2 + \sum_{k=1}^{n}\sum_{\substack{j=1 \\ j \neq k}}^{n}\left(\frac{1}{n}\right)\left(\frac{1}{n}\right)\sigma_{k,j} \tag{5.42}$$

where:

- the first term is the variance (VT);
- the second term is the covariance (CT).

Equation (5.42) can be re-arranged as follows:

$$\sigma_P^2 = \left(\frac{1}{n}\right)\sum_{k=1}^{n}\left(\frac{\sigma_k^2}{n}\right) + \left(\frac{n-1}{n}\right)\sum_{k=1}^{n}\sum_{\substack{j=1 \\ j \neq k}}^{n}\left(\frac{\sigma_{kj}}{n(n-1)}\right)$$

The terms in brackets in the summation are mean: the first is the *mean variance* $\left(\bar{\sigma}_k^2\right)$ of the expected returns of assets in portfolio, while the second is the *mean* *covariance* $\left(\bar{\sigma}_{k,j}\right)$ between the expected returns of the assets in portfolio considering them in a pair.

By replacing them, the portfolio variance can be defined as follows (Elton et al. 2013):

$$\sigma_P^2 = \left(\frac{1}{n}\right)\bar{\sigma}_k^2 + \left(1 - \frac{1}{n}\right)\bar{\sigma}_{k,j} \leftrightarrow \sigma_P^2 = \left(\frac{1}{n}\right)\left(\bar{\sigma}_k^2 - \bar{\sigma}_{k,j}\right) + \bar{\sigma}_{k,j} \tag{5.43}$$

Equation (5.43) provides understanding of the contribution of the variance and covariance on the portfolio variance according to the increase in the number (n) of the assets. Indeed, the limits to infinity for the number of the asset (n), are the following:

$$\begin{gathered} \lim_{n\to\infty} \tfrac{1}{n}\bar{\sigma}_k^2 = 0; \quad \lim_{n\to\infty}\left(1 - \tfrac{1}{n}\right)\bar{\sigma}_{k,j} = \bar{\sigma}_{k,j} \\ \lim_{n\to\infty}\left(\tfrac{1}{n}\right)\left(\bar{\sigma}_k^2 - \bar{\sigma}_{k,j}\right) + \bar{\sigma}_{k,j} = \bar{\sigma}_{k,j} \end{gathered} \tag{5.44}$$

The limits to infinity of Eq. (5.43) show that the contribution to the portfolio variance of the variance of each asset goes to zero while the contribution of the covariance goes to the mean covariance. Consequently, the specific risk of each asset is diversified and the only relevant term for the portfolio's variance is the covariance between the expected returns of the assets considered in a pair.

Therefore, for the k-th asset in portfolio, only its marginal contribution to the portfolio risk is relevant. It can be called the *asset's marginal risk*. Consequently, introduction of the k-th asset in the portfolio requires the calculation of (Castellani et al. 2005; Elton et al. 2013):

- the k-th asset's *marginal contribution to portfolio's returns* $\left(PR_{MC(k)}\right)$;
- the k-th asset's *marginal contribution to portfolio's variance* $\left(PV_{MC(k)}\right)$;
- the k-th asset's *marginal contribution to portfolio's standard deviation* $\left(PSD_{MC(k)}\right)$.

To do this, the partial derivatives of expected returns should be calculated, variance and standard deviation of the portfolio compared with the k-th asset's weight (α_k) as defined on the basis of wealth invested on the total wealth invested in the portfolio.

The *k-th asset's marginal contribution to the portfolio returns* $\left(PR_{MC(k)}\right)$ is calculated based on the partial derivative of the portfolio expected returns (μ_P) with regards to the k-th asset's weight (α_k) (Castellani et al. 2005).

Consider that:

$$\mu_P = \sum_{i=1}^{n} \alpha_i \mu_i \quad i = 1, 2, 3, \ldots, n$$

It involves a lot of terms that do not contain an α_k and only one term including α_k. The derivatives of all terms that do not include α_k, and therefore for each

$\alpha_i \neq \alpha_k$, are equal to zero because they are constants as far as α_k is concerned. In this case, the term including α_k is $\alpha_k \mu_k$. Therefore, the first derivative is equal to:

$$\frac{\partial \mu_P}{\partial \alpha_k} = \frac{\partial}{\partial \alpha_k}\left[\sum_{i=1}^{n} \alpha_i \mu_i\right] = \frac{\partial}{\partial \alpha_k}[\alpha_k \mu_k] = 1\mu_k = \mu_k$$

Consequently, the marginal contribution of the k-th asset to the portfolio's returns, is equal to its expected returns:

$$PR_{MC(k)} = \frac{\partial \mu_P}{\partial \alpha_k} = \mu_k \tag{5.45}$$

The *k-th asset's marginal contribution to the portfolio variance* $\left(PV_{MC(k)}\right)$ is calculated based on the partial derivative of the portfolio variance $\left(\sigma_P^2\right)$ with respect to the k-th asset's weight (α_k). It measures the k-th asset's marginal risk (in terms of variance).

Considering that:

$$\sigma_P^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \alpha_i \alpha_j \sigma_{i,j}$$

We have:

$$\frac{\partial \sigma_P^2}{\partial \alpha_k} = \frac{\partial}{\partial \alpha_k}\left[\sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \alpha_i \alpha_j \sigma_{i,j}\right]$$

The partial derivative of the first term follows the same principles as discussed earlier. All of the terms that do not include α_k are constant as far as α_k is concerned. Therefore, their derivative is zero. Consequently, the derivative of each term α_i different from α_k, and therefore for each $\alpha_i \neq \alpha_k$, is equal to zero. In this case, the term involving α_k is:

$$\alpha_k^2 \sigma_k^2$$

and therefore its derivative is equal to:

$$\frac{\partial}{\partial \alpha_k}\left[\sum_{i=1}^{n} \alpha_i^2 \sigma_i^2\right] = \frac{\partial}{\partial \alpha_k}\left[\alpha_k^2 \sigma_k^2\right] = 2\alpha_k \sigma_k^2$$

The partial derivative of the second term is calculated by considering that α_k is calculated twice: once when $i = k$ and once when $j = k$.

In the first case, when $i = k$, we have:

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \alpha_i\alpha_j\sigma_{i,j} = \sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_k\alpha_j\sigma_{k,j} = \alpha_k \sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j}$$

and the partial derivative is equal to:

$$\frac{\partial}{\partial\alpha_k}\left[\alpha_k \sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j}\right] = 1\sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j} = \sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j}$$

In the second case, when $j = k$, we have:

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \alpha_i\alpha_j\sigma_{i,j} = \sum_{\substack{i=1\\ i\neq k}}^{n} \alpha_i\alpha_k\sigma_{i,k} = \alpha_k \sum_{\substack{i=1\\ i\neq k}}^{n} \alpha_i\sigma_{i,k}$$

and its derivatives are equal to:

$$\frac{\partial}{\partial\alpha_k}\left[\alpha_k \sum_{\substack{i=1\\ i\neq k}}^{n} \alpha_i\sigma_{i,k}\right] = 1\sum_{\substack{i=1\\ i\neq k}}^{n} \alpha_i\sigma_{i,k} = \sum_{\substack{i=1\\ i\neq k}}^{n} \alpha_i\sigma_{i,k}$$

Since i and j are simply summation indicators and therefore it does not matter which one is used, and since the variance-covariance matrix is symmetric, we have:

$$\sigma_{i,j} = \sigma_{j,i} \rightarrow \sigma_{i,k} = \sigma_{k,i} \leftrightarrow \sigma_{k,j} = \sigma_{j,k}$$

and therefore:

$$\sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j} = \sum_{\substack{i=1\\ i\neq k}}^{n} \alpha_i\sigma_{i,k}$$

Consequently, the derivative of the second term is equal to:

$$\frac{\partial}{\partial\alpha_k}\left[\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \alpha_i\alpha_j\sigma_{i,j}\right] = \sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j} + \sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j} = 2\sum_{\substack{j=1\\ j\neq k}}^{n} \alpha_j\sigma_{k,j}$$

Summing the first and the second derivatives, the portfolio variance derivative is achieved, as follows:

$$\frac{\partial}{\partial \alpha_k}\left[\sum_{i=1}^{n}\alpha_i^2\sigma_i^2+\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\alpha_i\alpha_j\sigma_{i,j}\right]=2\alpha_k\sigma_k^2+2\sum_{\substack{j=1\\ j\neq k}}^{n}\alpha_j\sigma_{k,j}$$

Summing the terms, and therefore by introducing the first term in the summation dropping the constraint $(j \neq k)$, we have:

$$2\alpha_k\sigma_k^2+2\sum_{\substack{j=1\\ j\neq k}}^{N}\alpha_j\sigma_{k,j}=2\sum_{j=1}^{n}\alpha_j\sigma_{k,j}$$

Therefore, the marginal contribution of the k-th asset to the portfolio variance is equal to:

$$PV_{MC(k)}=\frac{\partial \sigma_P^2}{\partial \alpha_k}=2\sum_{j=1}^{n}\alpha_j\sigma_{k,j} \tag{5.46}$$

It is relevant to note that by considering the covariance between the portfolio returns and the k-th asset's returns $(\sigma_{k,P})$, we have:

$$\sigma_{k,P}=E[(I_K-\mu_k)(I_P-\mu_P)]$$

and considering that:

$$\begin{aligned}\sigma_{k,P}=E[(I_K-\mu_k)(I_P-\mu_P)]&=E\left[(I_K-\mu_k)\sum_{j=1}^{n}\alpha_j\left(I_j-\mu_j\right)\right]\\&=E\left[\sum_{j=1}^{n}\alpha_j(I_k-\mu_k)\left(I_j-\mu_j\right)\right]\\&=\sum_{j=1}^{n}\alpha_j E\left[(I_k-\mu_k)\left(I_j-\mu_j\right)\right]\end{aligned}$$

and considering that:

$$E\left[(I_k-\mu_k)\left(I_j-\mu_j\right)\right]=\sigma_{k,j}$$

we have:

$$\sigma_{k,P} = \sum_{j=1}^{n} \alpha_j \sigma_{k,j}$$

and by replacing, Eq. (5.46) can be rewritten as follows:

$$\mathrm{PV}_{\mathrm{MC(k)}} = \frac{\partial \sigma_{\mathrm{P}}^2}{\partial \alpha_{\mathrm{k}}} = 2 \sum_{\mathrm{j}=1}^{\mathrm{n}} \alpha_{\mathrm{j}} \sigma_{\mathrm{k,j}} = 2\sigma_{\mathrm{k,P}} \tag{5.47}$$

Finally, the *k-th asset's marginal contribution to the portfolio standard deviation* $\left(PSD_{MC(k)}\right)$, is calculated based on the partial derivative of the portfolio standard deviation (σ_P) compared with the k-th asset's weight (α_k). It measures the k-th *asset's marginal risk* (in terms of standard deviation).

By considering that:

$$\sigma_P = \sqrt{\sigma_{\mathrm{P}}^2} = \sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}}^2 \sigma_{\mathrm{i}}^2 + \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{i}}}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \sigma_{\mathrm{i,j}}}$$

$$= \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}}^2 \sigma_{\mathrm{i}}^2 + \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{i}}}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \sigma_{\mathrm{i,j}} \right)^{\frac{1}{2}}$$

$$\frac{\partial \sigma_P}{\partial \alpha_{\mathrm{k}}} = \frac{\partial}{\partial \alpha_{\mathrm{k}}} \left[\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}}^2 \sigma_{\mathrm{i}}^2 + \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{i}}}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \sigma_{\mathrm{i,j}} \right)^{\frac{1}{2}} \right]$$

$$= \left(\frac{1}{2}\right) \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}}^2 \sigma_{\mathrm{i}}^2 + \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{i}}}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \sigma_{\mathrm{i,j}} \right)^{\frac{1}{2}-1} \left(2\alpha_{\mathrm{k}} \sigma_{\mathrm{k}}^2 + 2 \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{k}}}^{\mathrm{n}} \alpha_{\mathrm{j}} \sigma_{\mathrm{k,j}} \right)$$

$$= \left(\frac{1}{2}\right) \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}}^2 \sigma_{\mathrm{i}}^2 + \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{i}}}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \sigma_{\mathrm{i,j}} \right)^{-\frac{1}{2}} \left(2\alpha_{\mathrm{k}} \sigma_{\mathrm{k}}^2 + 2 \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{k}}}^{\mathrm{n}} \alpha_{\mathrm{j}} \sigma_{\mathrm{k,j}} \right)$$

$$= \frac{\left(\frac{1}{2}\right) \left(2\alpha_{\mathrm{k}} \sigma_{\mathrm{k}}^2 + 2 \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{k}}}^{\mathrm{n}} \alpha_{\mathrm{j}} \sigma_{\mathrm{k,j}} \right)}{\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}}^2 \sigma_{\mathrm{i}}^2 + \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\substack{\mathrm{j}=1 \\ \mathrm{j} \neq \mathrm{i}}}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \sigma_{\mathrm{i,j}} \right)^{\frac{1}{2}}}$$

$$= \frac{\left(\frac{1}{2}\right) \left(2 \sum_{\mathrm{j}=1}^{\mathrm{n}} \alpha_{\mathrm{j}} \sigma_{\mathrm{k,j}} \right)}{\sigma_P} = \frac{\left(\frac{1}{2}\right) \left(2\sigma_{\mathrm{k,P}} \right)}{\sigma_P} = \frac{\sigma_{\mathrm{k,P}}}{\sigma_P}$$

Therefore, the marginal contribution of the k-th asset to the portfolio's standard deviation is equal to:

$$\mathrm{PSD}_{\mathrm{MC(k)}} = \frac{\partial \sigma_P}{\partial \alpha_k} = \frac{\sigma_{k,P}}{\sigma_P} \tag{5.48}$$

It is worth noting that the constraint $\sum_{k=1}^{n} w_k = 1$ is not considered in the partial derivatives. Consequently, the portfolio variance $\left(\sigma_P^2\right)$ and the standard deviation (σ_P) are function of n and not $(n-1)$ assets, and therefore are all considered as variance and covariance between them. This is because only the change in the portfolio risk is relevant due to the introduction of the k-th asset (function of the covariance between its expected returns and the expected returns of all other assets considered in couples and the k-th weights in the portfolio), and not the diversification measure (Castellani et al. 2005).

Therefore, $\frac{\sigma_{k,P}}{\sigma_P}$ is the relevant measurement of risk for the k-th asset. It measures the marginal contribution to the portfolio risk due to the introduction of the k-th asset.

5.2 Efficient Frontier

The second step of the process is construction of the efficient frontier (Lintner 1965; Chen et al. 1971, 1975; Hill 1976; Chen 1977; Alexander 1976, 1977, 1978; Bawa et al. 1979; Bertsekas 1974; Buser 1977; Jones-Lee 1971; Dybving 1984; Jacob 1974; Lewis 1988). It has a different form as function of short selling. However, an analysis of the efficient frontier when there is short selling, can be interpreted as an extension of the analysis when there is no short selling (Elton et al. 2013).

For further understanding, the analysis can be developed on the basis of three main cases (Castellani et al. 2005):

- (*Case 1*) two risky assets in portfolio;
- (*Case 2*) N risky assets in portfolio;
- (*Case 3*) N risky assets and one free-risk asset in portfolio.

(Case 1) Two Risky Assets in Portfolio

Assuming a portfolio of two risky assets:

- *Asset 1* (A_1): its weight in portfolio (α_1) is equal to the part of wealth invested in it (w_1) on total wealth invested in the portfolio (w) so that $\alpha_1 = w_1/w$;
- *Asset 2* (A_2): its weight in portfolio (α_2) is equal to the part of wealth invested in it (w_2) on total wealth invested in the portfolio (w) so that $\alpha_2 = w_2/w$.

Assuming that Asset 2 is riskier than Asset 1. On the basis of Eq. (5.4) we have:

- $\mu_2 > \mu_1$;
- $\sigma_2^2 > \sigma_1^2 \rightarrow \sigma_2 > \sigma_1$.

and by assuming that the part of wealth invested in Asset 2 is equal to α and consequently the part of wealth invested in Asset 1 is equal to the $1 - \alpha$, we have:

$$\alpha_1 + \alpha_2 = 1 \leftrightarrow \begin{cases} \alpha_2 = \alpha \\ \alpha_1 = 1 - \alpha \end{cases} \quad (0 < \alpha < 1)$$

Changing α, and therefore the part of wealth invested in Asset 2 (riskier asset), all portfolios can be obtained on the basis of all possible combinations between the two assets.

It is worth noting that:

- if there is no short selling, α can change in the range $[0; 1]$ so that: $0 \leq \alpha \leq 1$;
- if there is short selling, α may exceed 1 $(a > 1)$ by short selling Asset 1 (short position on Asset 1) and acquiring Asset 2; similarly, α may exceed 0 $(a < 0)$ by short selling Asset 2 (short position on Asset 2) and acquiring the Asset.

In the plan $\left(\sigma_P^2; \mu_P\right)$ the *opportunities frontier* $(\mathcal{B})$ is defined by all portfolios on the basis of the following coordinates:

$$\begin{cases} \mu_P = (1 - \alpha)\mu_1 + \alpha\mu_2 \\ \sigma_P^2 = \alpha^2\sigma_2^2 + (1 - \alpha)^2\sigma_1^2 + 2\alpha(1 - \alpha)\rho_{1,2}\sigma_1\sigma_2 \end{cases} \tag{5.49}$$

These Eq. (5.49) define the system of parametric equations of the opportunity frontier $(\mathcal{B})$ and they allow for definition of the coordinates of each portfolio by changing α (Castellani et al. 2005).

It is important to note that the opportunities frontier $(\mathcal{B})$ can be obtained by explicating the portfolio variance $\left(\sigma_P^2\right)$ as function of its return (μ_P). To do this, it is necessary to explicate the first equation (portfolio's return) of Eq. (5.49) for α and substituting it in the second equation (portfolio variance).

Therefore, the first equation can be explicated by α as follows:

$$\begin{aligned} \mu_P &= (1 - \alpha)\mu_1 + \alpha\mu_2 \\ \mu_P &= \mu_1 - \alpha\mu_1 + \alpha\mu_2 \\ \mu_P - \mu_1 &= \alpha(\mu_2 - \mu_1) \end{aligned} \rightarrow \alpha = \frac{\mu_P - \mu_1}{\mu_2 - \mu_1}$$

By substituting α in Eq. (5.11), the portfolio variance can be rewritten as follows:

$$\sigma_P^2 = \alpha^2\sigma_2^2 + (1-\alpha)^2\sigma_1^2 + 2\alpha(1-\alpha)\rho_{1,2}\sigma_1\sigma_2$$
$$\sigma_P^2 = \left(\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)^2\sigma_2^2 + \left(1-\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)^2\sigma_1^2 + 2\left(\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)\left(1-\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)\rho_{1,2}\sigma_1\sigma_2$$
$$\sigma_P^2 = \left(\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)^2\sigma_2^2 + \left(\frac{\mu_2-\mu_1-\mu_P+\mu_1}{\mu_2-\mu_1}\right)^2\sigma_1^2 + 2\left(\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)\left(\frac{\mu_2-\mu_1-\mu_P+\mu_1}{\mu_2-\mu_1}\right)\rho_{1,2}\sigma_1\sigma_2$$
$$\sigma_P^2 = \left(\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)^2\sigma_2^2 + \left(\frac{\mu_2-\mu_P}{\mu_2-\mu_1}\right)^2\sigma_1^2 + 2\left(\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)\left(\frac{\mu_2-\mu_P}{\mu_2-\mu_1}\right)\rho_{1,2}\sigma_1\sigma_2$$
$$\sigma_P^2 = \left(\frac{\mu_P-\mu_1}{\mu_2-\mu_1}\right)^2\sigma_2^2 + \left(\frac{\mu_2-\mu_P}{\mu_2-\mu_1}\right)^2\sigma_1^2 + 2\left(\frac{\mu_P\mu_2-\mu_P^2-\mu_1\mu_2+\mu_1\mu_P}{(\mu_2-\mu_1)^2}\right)\rho_{1,2}\sigma_1\sigma_2$$
$$\sigma_P^2 = \frac{1}{(\mu_2-\mu_1)^2}\left[(\mu_P-\mu_1)^2\sigma_2^2 + (\mu_2-\mu_P)^2\sigma_1^2 + 2(\mu_P\mu_2-\mu_P^2-\mu_1\mu_2+\mu_1\mu_P)\rho_{1,2}\sigma_1\sigma_2\right]$$
$$\sigma_P^2 = \frac{1}{(\mu_2-\mu_1)^2}\left[\mu_P^2\sigma_2^2 + \mu_1^2\sigma_2^2 - 2\mu_P\mu_1\sigma_2^2 + \mu_2^2\sigma_1^2 + \mu_P^2\sigma_1^2 - 2\mu_P\mu_2\sigma_1^2 + 2\mu_P\mu_2\rho_{1,2}\sigma_1\sigma_2 - 2\mu_P^2\rho_{1,2}\sigma_1\sigma_2 - 2\mu_1\mu_2\rho_{1,2}\sigma_1\sigma_2 + 2\mu_1\mu_P\rho_{1,2}\sigma_1\sigma_2\right]$$
$$\sigma_P^2 = \frac{1}{(\mu_2-\mu_1)^2}\left[\mu_P^2(\sigma_2^2+\sigma_1^2-2\rho_{1,2}\sigma_1\sigma_2) - 2\mu_P\left[\mu_1\sigma_2^2+\mu_2\sigma_1^2-\rho_{1,2}\sigma_1\sigma_2(\mu_2+\mu_1)\right] + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 - 2\mu_1\mu_2\rho_{1,2}\sigma_1\sigma_2\right]$$

By placing:

$$A = \sigma_2^2 + \sigma_1^2 - 2\rho_{1,2}\sigma_1\sigma_2$$
$$B = 2\left[\mu_1\sigma_2^2 + \mu_2\sigma_1^2 - \rho_{1,2}\sigma_1\sigma_2(\mu_2+\mu_1)\right]$$
$$C = \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 - 2\mu_1\mu_2\rho_{1,2}\sigma_1\sigma_2$$

We have:

$$\sigma_P^2 = \frac{1}{(\mu_2-\mu_1)^2}\left[A\mu_P^2 - B\mu_P + C\right] \tag{5.50}$$

and in terms of standard deviation, we have:

$$\sigma_P = \frac{1}{(\mu_2-\mu_1)}\sqrt{A\mu_P^2 - B\mu_P + C} \tag{5.51}$$

In the plane $(\mu_P; \sigma_P^2)$ Eq. (5.50) draws a parabola with its axis parallel to the axis of ordinates and concave upwards (Saltari 2011).

Note, by considering the function linearity, explicating the portfolio's variance (σ_P^2) as a function of its return (μ_P) (so that $\sigma_P^2 = A\mu_P^2 + B\mu_P + C$) instead of as

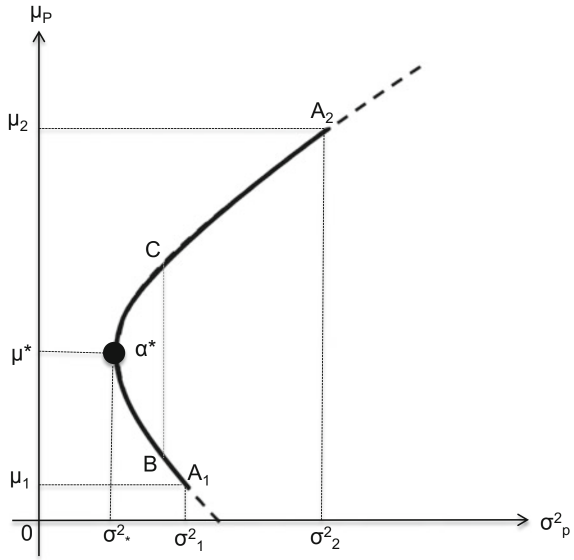


Fig. 5.9 The opportunity frontier $(\mathcal{B})$ in the plane $(\sigma_P^2; \mu_P)$

function of α (so that $\sigma_P^2 = A\alpha^2 + B\alpha + C$) it is simply equivalent to create a linear transformation of the abscissas.

Reversing the axes, and therefore in the plane $(\sigma_P^2; \mu_P)$, Eq. (5.50) draws a parabola with its axis parallel to the axis abscissa and concavity towards the right. This parabola defines the opportunity frontier $(\mathcal{B})$.

Assuming a well-diversified portfolio $\left(0 < \rho_{1,2} < \frac{\sigma_1}{\sigma_2}\right)$ the parabola can be represented as in Fig. 5.9 (adapted from Castellani et al. 2005).

The parabola in Fig. 5.9 represents the opportunities frontier $(\mathcal{B})$ and it contains all efficient portfolios obtained on the basis of a combination of the two assets by changing the α.

It is worth noting that, assuming that there are no shorts selling, the curve is delimited by the portfolios composed of Asset 1 only (portfolio A_1 created by $\alpha = 0$) and only by Asset 2 (portfolio A_2 achieved by $\alpha = 1$). All others portfolios, based on the combination of the two assets, move along the curve.

Otherwise, assuming that there are shorts selling, for the negative value of α and therefore for $(\alpha < 0)$, the portfolio A_1 is exceeded through the short selling of Asset 2 and acquiring of Asset 1. At the same time, for the positive value of α and therefore for $(\alpha > 1)$, portfolio A_2 is exceeded through the short selling of Asset 1 and acquisition of Asset 2.

Assuming that there is no short selling and therefore for $(0 \leq \alpha \leq 1)$, all efficient portfolios are positioned on the parabola. Therefore, the curve represents the frontier of minimum variance portfolio for each defined level of portfolio expected return (μ_P). Consequently, having defined the level of the portfolio's expected return, by moving along the curve and by changing the composition of the portfolio as function of α, it is possible to minimize the portfolio's variance $\left(\sigma_P^2\right)$.

Note all portfolios on the opportunity frontier $(\mathcal{B})$, and therefore on the frontier of minimum variance portfolios, can be considered as Pareto optimality. All portfolios positioned on the parabola lower branch are dominated by the portfolios positioned on the parabola upper branch: they are defined as dominant portfolios because they are characterized by a higher expected return for equal risk level.

The point of transition on the curve from dominant portfolio to dominated portfolios is represented by the minimum variance portfolio (α^*).

The α^* defines the part of wealth to be invested in Asset 2 (riskier asset) to minimize the portfolio's variance. Therefore, α^* the weight of Asset 2 and therefore the weight of Asset 1, in the portfolio in order to obtain the minimum variance portfolio.

Therefore, from the minimum variance portfolio, the parabola upper branch is concave while the parabola lower branch is convex. Only the parabola upper branch defines the frontier of efficient portfolios, because only these portfolios are really efficient. In Fig. 5.9 the frontier of efficient portfolios is defined by part of the curve $(\alpha^*; A_2)$.

Therefore, the subset of the minimum variance portfolios is larger than the efficient portfolios and therefore the frontier of minimum variance portfolios contains the frontier of efficient portfolios. In this sense, it is sufficient to analyse the position of portfolio C and of portfolio B. Both portfolios are positioned on the parabola and therefore on the frontier of minimum variance portfolios. But only portfolio C can be defined as an efficient portfolio because for the same variance $\left(\sigma_C^2 = \sigma_B^2\right)$ its expected return is higher than the portfolio B $\left(\sigma_C^2 = \sigma_B^2\right)$; therefore, portfolio C dominates portfolio B. Consequently, both portfolios are positioned on the frontier of minimum variance portfolios but only portfolio C is positioned on the frontier of efficient portfolios. Only the portfolios on the frontier of the efficient portfolios can be defined as Pareto optimality.

Therefore, having defined the frontier of minimum variance portfolios it is necessary to focus the analysis on the frontier of efficient portfolios where the efficient portfolios are positioned, starting from the minimum variance portfolio. All points:

- *above the frontier of efficient portfolios*: represent portfolios that cannot be obtained due to the statistical characteristics of the assets in portfolios;
- *below the frontier of efficient portfolios*: represent portfolios that can be obtained on the basis of the statistical characteristics of the assets in portfolios, but they are not efficient and therefore they are dominated by the portfolios positioned on the efficient frontier.

The frontier of efficient portfolios represents the solution of the optimization problem in the mean-variance criteria. The efficient portfolios show the lower variance for the same expected return level or, equivalently, the higher return for the same variance level.

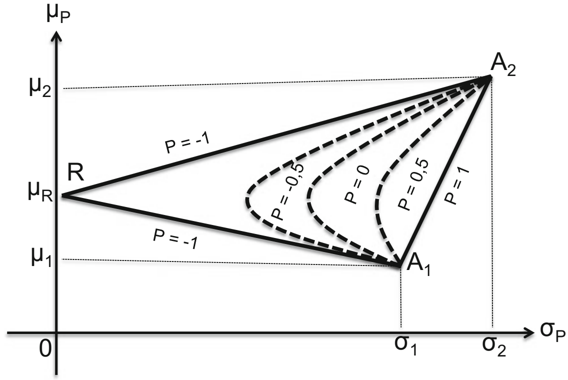


Fig. 5.10 The opportunity frontier $(\mathcal{B})$ in the plane $(\sigma_P; \mu_P)$

It is worth noting that the construction of the frontier of efficient portfolios is function of the investor's expectations about the statistical characteristics (mean, variance and covariance) of the assets in portfolios. Consequently, each investor works on his frontier of efficient portfolios.

Moving from the portfolio's variance $\left(\sigma_P^2\right)$ to its standard deviation (σ_P), in the plane $(\sigma_P; \mu_P)$ the opportunity frontier $(\mathcal{B})$ can be represented as the form of a hyperbola branch with concavity towards the right which increases with the decrease of the correlation coefficient, as shown in Fig. 5.10.

Note that the portfolio R can be obtained as follows:

$$\alpha_R = \frac{\sigma_1}{\sigma_1 + \sigma_2} \tag{5.52}$$

On the basis of the α_R the expected return of the portfolio R can be achieved, as follows:

$$\mu_R = (1 - \alpha_R)\mu_1 + \alpha_R \mu_2$$

$$\mu_R = \left(1 - \frac{\sigma_1}{\sigma_1 + \sigma_2}\right)\mu_1 + \frac{\sigma_1}{\sigma_1 + \sigma_2}\mu_2$$

$$\mu_R = \left(\frac{\sigma_1 + \sigma_2 - \sigma_1}{\sigma_1 + \sigma_2}\right)\mu_1 + \frac{\sigma_1}{\sigma_1 + \sigma_2}\mu_2$$

$$\mu_R = \frac{\sigma_2}{\sigma_1 + \sigma_2}\mu_1 + \frac{\sigma_1}{\sigma_1 + \sigma_2}\mu_2 \tag{5.53}$$

(*Case 2*) *N Risky Assets in Portfolio*

At this point, it is possible to analyse (*Case 2*) by assuming N risky assets in portfolio. Also, if the problem is more complex, the result is similar to the one of (*Case 1*).

In order to define the optimal portfolio, it is necessary to distinguish between two cases:

– short selling is permitted;
– short selling is not permitted.

If the *short selling is permitted* (it is the most relevant situation) there are no restraints on α considered; therefore, it can be positive or negative due to the short selling. Since the correlations between the assets' expected returns tend to be different from zero, and the positive correlation tens to be greater than negative correlation, the short selling is used according to portfolio optimization.

In this context, definition of the efficient portfolio can be defined as a constrained optimization problem, as follows (Castellani et al. 2005):

$$\begin{cases} \min\limits_{\boldsymbol{\alpha}\in\mathbb{R}^n} \sigma_P^2(\boldsymbol{\alpha}) \\ \mu_P(\boldsymbol{\alpha}) = \mu_0 \\ \sum\limits_{k=1}^{n} \alpha_k = 1 \end{cases} \tag{5.54}$$

Or alternatively:

$$\begin{cases} \max\limits_{\boldsymbol{\alpha}\in\mathbb{R}^n} \mu_P(\boldsymbol{\alpha}) \\ \sigma_P^2(\boldsymbol{\alpha}) = \sigma_0^2 \\ \sum\limits_{k=1}^{n} \alpha_k = 1 \end{cases} \tag{5.55}$$

where in Eq. (5.54) the first is the equation to be minimized and the second and the third are the constraints, and also in Eq. (5.55) the first is the equation to be maximized and the second and third are the constraints. In both cases, $\boldsymbol{\alpha}$ is a vector: $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$.

Considering that investors are characterized by the risk aversion, the constrained optimization problem is usually formalized on the basis of Eq. (5.54) rather than on the basis of Eq. (5.55).

In the extensive form, Eq. (5.54) can be rewritten as follows:

$$\begin{cases} \min\limits_{\boldsymbol{\alpha}\in\mathbb{R}^n} \sum\limits_{k=1}^{n} \sum\limits_{j=1}^{n} \alpha_k \alpha_j \rho_{k,j} \sigma_k \sigma_j \\ \sum\limits_{k=1}^{n} \alpha_k \mu_k = \mu_0 \\ \sum\limits_{k=1}^{n} \alpha_k = 1 \end{cases} \tag{5.56}$$

Equation (5.56) can be solved by using the Lagrangian $(\mathcal{L})$ as follows (Castellani et al. 2005):

$$\mathcal{L} = \sum_{k=1}^{n}\sum_{j=1}^{n}\alpha_k\alpha_j\rho_{k,j}\sigma_k\sigma_j - \lambda'\left(\sum_{k=1}^{n}\alpha_k\mu_k - \mu_0\right) - \lambda''\left(\sum_{k=1}^{n}\alpha_k - 1\right) \tag{5.57}$$

Remembering that in Eqs. (5.45), (5.46) and (5.48), we have:

$$\frac{\partial\mu_P}{\partial\alpha_k} = \frac{\partial}{\partial\alpha_k}\sum_{j=1}^{n}\alpha_j\mu_j = \mu_k$$

$$\frac{\partial\sigma_P^2}{\partial\alpha_k} = \frac{\partial}{\partial\alpha_k}\sum_{j=1}^{n}\sum_{h=1}^{n}\alpha_j\alpha_h\sigma_{j,h} = 2\sum_{j=1}^{n}\alpha_j\sigma_{k,j}$$

$$\frac{\partial\sigma_P}{\partial\alpha_k} = \frac{1}{2\sigma_P}\frac{\partial\sigma_P^2}{\partial\alpha_k} = \frac{\sum_{j=1}^{n}\alpha_j\sigma_{kj}}{\sigma_P}$$

and by considering the Legrangian first derivative:

$$\frac{\partial\mathcal{L}}{\partial\alpha_k} = 2\sum_{j=1}^{n}\alpha_j\rho_{k,j}\sigma_j - \lambda'\mu_k - \lambda'' = \sum_{j=1}^{n}\alpha_j\rho_{k,j}\sigma_j - \frac{1}{2}\lambda'\mu_k - \frac{1}{2}\lambda'' \tag{5.58}$$

and placing first derivatives first, we have:

$$\begin{cases} \sum_{j=1}^{n}\alpha_j\rho_{k,j}\sigma_j - \frac{1}{2}\lambda'\mu_k - \frac{1}{2}\lambda'' = 0 \quad per \quad k = 1, 2, \ldots, n \\ \sum_{k=1}^{n}\alpha_k\mu_k = \mu_0 \\ \sum_{k=1}^{n}w_k = 1 \end{cases} \tag{5.59}$$

Equation (5.59) are linear. Denoting by $\alpha_1(\mu_0), \alpha_2(\mu_0), \ldots, \alpha_n(\mu_0)$ the solutions obtained at the fixed value of the portfolio's expected return (μ_0), we have the coordinates of the minimum variance portfolio as follows (Castellani et al. 2005):

$$\mu_0 = \sum_{k=1}^{n}\alpha_k(\mu_0)\mu_k \tag{5.60}$$

$$\sigma_0^2 = \sum_{k=1}^{n}\sum_{j=1}^{n}\alpha_k(\mu_0)\alpha_j(\mu_0)\rho_{k,j}\sigma_k\sigma_j \tag{5.61}$$

The coordinates of the minimum variance portfolio are obtained by solving the constrained optimization problem through the Lagrangian without any constraint on the expected return.

In the plane $(\mu_P; \sigma_P^2)$, we have a parabola with axes parallel to the abscissa and concavity to the right as shown in Fig. 5.11. All assets are positioned in the space

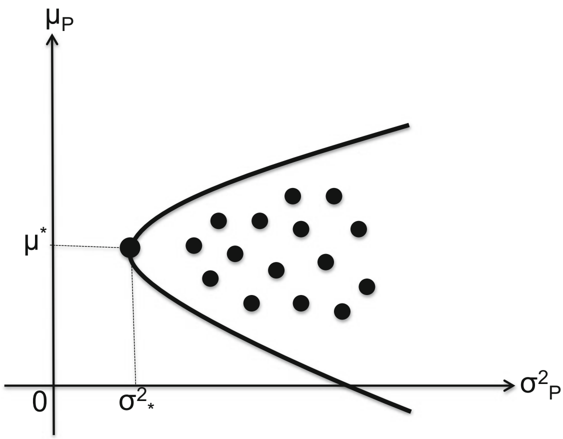


Fig. 5.11 The opportunity frontier for n risky assets in portfolio and short selling in the plane $\left(\sigma_{\mathrm{P}}^2, \mu_{\mathrm{P}}\right)$

delimited by the parabola. The curve represents the opportunity frontier $(\mathcal{B})$ in case of n risky assets and when short selling is permitted. The parabola's vertex indicates the minimum variance portfolio, as follows (Castellani et al. 2005):

$$\sigma_*^2 = \min_{w \in \mathcal{B}} \sigma_{\mathrm{p}}^2(\boldsymbol{\alpha}) \tag{5.62}$$

where $\boldsymbol{\alpha}$ is a vector as follows: $\boldsymbol{\alpha} = \{a_1, a_2, \ldots, a_n\}$

Similarly to the previous case, the efficient frontier is represented by the parabola's superior branch starting from the minimum variance portfolio.

Shifting the focus from the plane $\left(\sigma_P^2, \mu_P\right)$ to the plane (σ_P, μ_P), the opportunity frontier goes from hyperbolic to hyperbole with the axes parallel to the abscissa and the point of axe intersection on the ordinate, as shown in Fig. 5.12.

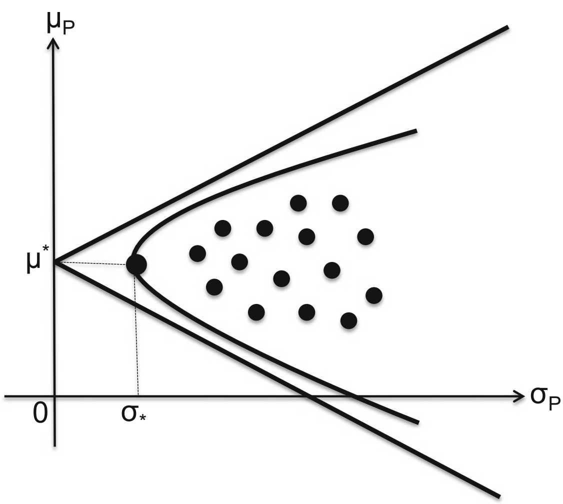


Fig. 5.12 The opportunity frontier for n risky assets in portfolio and short selling in the plane $(\sigma_{\mathrm{P}}, \mu_{\mathrm{P}})$

If *short selling is not permitted*, the opportunity frontier is defined only based on the assets owned. In this case, the constrained optimization problem can be defined as follows (Castellani et al. 2005):

$$\begin{cases} \min\limits_{w \in \mathbb{R}^n} \sigma_P^2(\boldsymbol{\alpha}) \\ \mu_P(\boldsymbol{\alpha}) = \mu_0 \\ \sum\limits_{k=1}^{n} w\alpha_k = 1 \\ \alpha_k \geq 0 \quad k = 1, 2, \ldots, n \end{cases} \tag{5.63}$$

where $\boldsymbol{\alpha}$ is a vector as follows: $\boldsymbol{\alpha} = \{a_1, a_2, \ldots, a_n\}$.

The problem does not have an analytical solution. The opportunity frontier is obtained through numerical procedures.

(Case 3) N Risky Assets in Portfolio and One Risk-Free Asset

Finally, the last case (*Case 3*) assumes *N* risky assets in portfolio and one risk-free asset.

Introduction of the asset free-risk in the portfolio, allows for simplification of the problem of the portfolio optimisation. Indeed, in this case the efficient frontier is represented by a straight-line and not by a curve.

Assuming a portfolio consisting of *N* risky assets $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ and one risk-free asset (α_0). Assuming that it is a zero-coupon bond issued in t-time and maturity in s-time. Denoting by Q_0 the price of the risk-free asset in t-time and by A_0 the price of the risk-free asset in s-time, it returns (R_F) (that it is sure and not expected), it is equal to:

$$R_F = \frac{A_0 - Q_0}{Q_0} = \frac{A_0}{Q_0} - 1 \tag{5.64}$$

Considering that the asset is risk-free, the price A_0 is known in t-time and therefore, the return R_f is sure and not expected.

The risky asset (a_k) requires an expected return (μ_k) higher than the risk-free asset return, so that:

$$\mu_k \gg R_F$$

The difference between the μ_k and R_F is the extra-return expected by investors for the risk burden, and therefore the *risk premium* (δ_k) required by investor for the risk:

$$\delta_k = \mu_k - R_F \tag{5.65}$$

Considering two assets in the portfolio: (i) risky asset (A_1) and (ii) risk-free asset (A_0). The expected return of the risky asset is equal to μ_1 and the sure return of the risk-free asset is equal to R_F. Denoting the part of wealth invested in the risky asset by α, and the part of wealth invested in the risk-free asset by $(1-\alpha)$, the portfolio's expected return (μ_P) can be calculated as follows:

$$\mu_P = \alpha\mu_1 + (1-\alpha)R_F$$

and therefore:

$$\mu_P = R_F + \alpha(\mu_1 - R_F) \tag{5.66}$$

Considering that the variance of the risk-free asset is equal to zero $\left(\sigma_i^2 = 0\right)$ by definition (its return is certain), the portfolio's variance $\left(\sigma_P^2\right)$ is equal to:

$$\sigma_P^2 = \alpha^2\sigma_1^2 + (1-\alpha)\sigma_2^2 + 2\alpha(1-\alpha)\sigma_{1,2}$$

$$\sigma_P^2 = \sigma_1^2\alpha^2 \tag{5.67}$$

The portfolio's variance is positive. Therefore, it is possible to move from the variance to the standard deviation directly as follows:

$$\sigma_P = \sigma_1\alpha \tag{5.68}$$

It is possible to define the portfolio's expected return (μ_P) as function of its standard deviation. By solving Eq. (5.68) by α and substituting in Eq. (5.66) we have:

$$\alpha = \frac{\sigma_P}{\sigma_1}$$

$$\mu_P = R_F + \frac{\sigma_P}{\sigma_1}(\mu_1 - R_F)$$

and therefore:

$$\mu_P = R_F + \sigma_P\left(\frac{\mu_1 - R_F}{\sigma_1}\right) \tag{5.69}$$

Equation (5.69) draws a straight-line in the space $(\sigma_P; \mu_P)$ where interception is equal to (R_F) and slope is equal to $\left(\frac{\mu_1 - R_F}{\sigma_1}\right)$, as shown in Fig. 5.13.

All possible combinations between the risky assets (and therefore the risky portfolio) and the risk-free asset are pointed on the straight-line. Therefore, it is the *efficient frontier* where:

- on the left of the point A_1, all combinations between the risk-free asset and the n-risky assets defining the risky portfolio are highlighted;

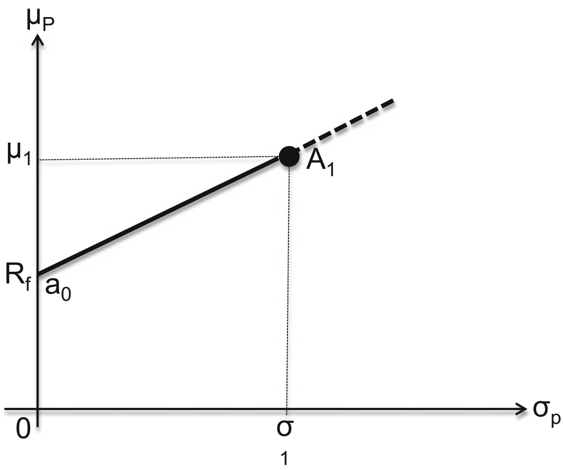


Fig. 5.13 The opportunity frontier for n risky assets and one risk-free asset in portfolio in the plane (σ_P, μ_P)

– on the right of the point A_1, all combinations created by borrowing at a risk-free rate and therefore through short selling of the risk-free asset are highlighted (Elton et al. 2013).

Note that on the basis of Eq. (5.65), Eq. (5.69) can be rewritten as follows:

$$\delta_1 = \mu_1 - R_F$$

$$\mu_P = R_F + \sigma_P\left(\frac{\delta_1}{\sigma_1}\right) \tag{5.70}$$

The ratio between the *risk premium* (δ_1) and the standard deviation of the risky asset (a_1) is the Sharpe index (π):

$$\frac{\mu_1 - R_f}{\sigma_1} = \frac{\delta_1}{\sigma_1} = \pi \tag{5.71}$$

In general terms, by considering the k-th risky asset, the Sharpe index can be expressed on the basis of the variance of the risky asset as follows:

$$\frac{\delta_k}{\sigma_k^2} = \pi \tag{5.72}$$

The Sharpe index can be considered as the *cost of capital for unit of risk* or, in equivalent term, *unit price for risk*.

Based on the Share index, Eq. (5.70) can be rewritten as follows:

$$\mu_P = R_f + \pi\sigma_P \tag{5.73}$$

Equation (5.73) defines the line of the efficient frontier.

The efficient frontier can be defined analytically (Saltari 2011). Assuming a portfolio P of two risky assets, A_1 and A_2, and one risk-free asset (A_0). The expected returns of the risky assets are equal to μ_1 and μ_2 respectively while the sure return of the risk-free asset is equal to R_F. The part of wealth invested in the risk-free asset is equal to α_0, while the part of wealth invested in the risky assets is equal to α_1 and α_2. The budget constraint can be defined as follows:

$$\sum_{i=1}^{n} \alpha_i = 1 \rightarrow \alpha_0 + \alpha_1 + \alpha_2 = 1 \tag{5.74}$$

The portfolio's expected return is equal to:

$$\mu_P = \alpha_0 R_F + \alpha_1 \mu_1 + \alpha_2 \mu_2 \tag{5.75}$$

Considering only two variables, α_1 and α_2, it is possible to define α_0 as function of α_1 and α_2 as follows:

$$\begin{aligned} \alpha_0 &= 1 - \alpha_1 \rightarrow \alpha_0 R_F + \alpha_1 \mu_1 = (1 - \alpha_1) R_F + \alpha_1 \mu_1 \\ &= R_F - \alpha_1 R_F + \alpha_1 \mu_1 = R_F + \alpha_1 (\mu_1 - R_F) \\ \alpha_0 &= 1 - \alpha_2 \rightarrow \alpha_0 R_F + \alpha_2 \mu_2 = (1 - \alpha_2) R_F + \alpha_2 \mu_2 \\ &= R_F - \alpha_2 R_F + \alpha_2 \mu_2 = R_F + \alpha_2 (\mu_2 - R_F) \end{aligned} \tag{5.76}$$

And therefore Eq. (5.75) can be rewritten as follows:

$$\mu_P = R_F + \alpha_1(\mu_1 - R_F) + \alpha_2(\mu_2 - R_F) \tag{5.77}$$

For a given expected return the minimum variance portfolio can be obtained by solving a optimization constrained problems as follows:

$$\begin{cases} \min\limits_{\alpha_1, \alpha_2} \frac{1}{2} \sigma_P^2 = \frac{1}{2} \left(\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + 2 \rho_{1,2} \sigma_1 \sigma_2 \alpha_1 \alpha_2 \right) \\ \mu_P = R_F + \alpha_1 (\mu_1 - R_F) + \alpha_2 (\mu_2 - R_F) \\ \alpha_1 + \alpha_2 = 1 \end{cases} \tag{5.78}$$

where the first equation is the equation to be minimized, while the second and the third equations are the constraints.

It is worth noting, that the use of half of the variance $\left(\frac{1}{2}\sigma_P^2\right)$ instead of the variance $\left(\sigma_P^2\right)$ is used only to simplify calculations (Saltari 2011).

By using the Lagrangia $(\mathcal{L})$, we have:

$$\mathcal{L} = \frac{1}{2}\sigma^2 + \lambda[\mu_P - (R_F + \alpha_1(\mu_1 - R_F) + \alpha_2(\mu_2 - R_F))]$$

and therefore:

$$\mathcal{L} = \frac{1}{2}\left(\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\rho_{1,2}\sigma_1\sigma_2\alpha_1\alpha_2\right) + \lambda[\mu_P - (R_F + \alpha_1(\mu_1 - R_F) + \alpha_2(\mu_2 - R_F))] \tag{5.79}$$

Placing the derivatives of the Lagrangian equal to zero with respect α_1, α_2 and λ we have:

$$\begin{cases} \frac{\partial\mathcal{L}}{\partial\alpha_1} = \alpha_1\sigma_1^2 + \rho\sigma_1\sigma_2\alpha_2 - \lambda(\mu_1 - R_f) = 0 \\ \frac{\partial\mathcal{L}}{\partial\alpha_2} = \alpha_2\sigma_2^2 + \rho\sigma_1\sigma_2\alpha_1 - \lambda(\mu_2 - R_f) = 0 \\ \frac{\partial\mathcal{L}}{\partial\lambda} = \mu - [R_f + \alpha_1(\mu_1 - R_f) + \alpha_2(\mu_2 - R_f)] = 0 \end{cases} \tag{5.80}$$

To solve the system of Eq. (5.80) it is possible to multiply the first equation for α_1, as well as the second equation for α_2, and to add the two equations together.

Therefore, by multiplying the first equation for α_1 and the second equation for α_2, we have:

$$\left[\alpha_1\sigma_1^2 + \rho\sigma_1\sigma_2\alpha_2 - \lambda(\mu_1 - R_F)\right]\alpha_1 = 0 \rightarrow \alpha_1^2\sigma_1^2 + \rho\sigma_1\sigma_2\alpha_2\alpha_1 - \lambda(\mu_1 - R_F)\alpha_1 = 0$$
$$\left[\alpha_2\sigma_2^2 + \rho\sigma_1\sigma_2\alpha_1 - \lambda(\mu_2 - R_F)\right]\alpha_2 = 0 \rightarrow \alpha_2^2\sigma_2^2 + \rho\sigma_1\sigma_2\alpha_1\alpha_2 - \lambda(\mu_2 - R_F)\alpha_2 = 0$$

By adding two equations, we have:

$$\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\rho_{1,2}\alpha_1\alpha_2\sigma_1\sigma_2 = \lambda[\alpha_1(\mu_1 - R_F) + \alpha_2(\mu_2 - R_F)]$$

The left side of the equation is the portfolio's variance $\left(\sigma_P^2\right)$ due to the risky assets.

Considering that:

$$\alpha_1 + \alpha_2 = 1 \leftrightarrow \mu_1 + \mu_2 = \mu$$

the right side of the equation is equal to the product between λ and the difference between the portfolio's expected return due to the risky assets (μ) and the sure return of the risk-free asset (R_F), as follows:

$$\sigma_P^2 = \lambda(\mu - R_F) \rightarrow \lambda = \frac{\sigma_P^2}{(\mu - R_F)}$$

Now it is possible to solve the first and the second equations of the system of Eq. (5.80) for α_1 and α_2. Substituting λ in the first equation, we have:

$$\alpha_1\sigma_1^2 + \rho\sigma_1\sigma_2\alpha_2 - \frac{\sigma_P^2}{(\mu - R_F)}(\mu_1 - R_F) = 0$$
$$\alpha_1\sigma_1^2(\mu - R_F) + \rho\sigma_1\sigma_2\alpha_2(\mu - R_F) - \sigma_P^2(\mu_1 - R_F) = 0$$
$$\alpha_1 = \frac{\sigma_P^2(\mu_1 - R_F) - \rho\sigma_1\sigma_2\alpha_2(\mu - R_F)}{\sigma_1^2(\mu - R_F)}$$
$$\alpha_1 = \frac{\sigma_P^2(\mu_1 - R_F)}{\sigma_1^2(\mu - R_F)} - \frac{\rho\sigma_1\sigma_2\alpha_2(\mu - R_F)}{\sigma_1^2(\mu - R_F)}$$
$$\alpha_1 = \frac{\sigma_P^2(\mu_1 - R_F)}{\sigma_1^2(\mu - R_F)} - \frac{\rho\sigma_1\sigma_2\alpha_2}{\sigma_1^2}$$

Substituting λ and α_1 in the second equation, we have:

$$\alpha_2\sigma_2^2 + \rho\sigma_1\sigma_2\left[\frac{\sigma_P^2(\mu_1 - R_F)}{\sigma_1^2(\mu - R_F)} - \frac{\rho\sigma_1\sigma_2\alpha_2}{\sigma_1^2}\right] - \frac{\sigma_P^2}{\mu - R_F}(\mu_2 - R_F) = 0$$
$$\alpha_2\sigma_2^2 + \frac{\rho\sigma_1\sigma_2\sigma_P^2(\mu_1 - R_F)}{\sigma_1^2(\mu - R_F)} - \frac{\rho^2\sigma_1^2\sigma_2^2\alpha_2}{\sigma_1^2} - \frac{\sigma_P^2}{\mu - R_F}(\mu_2 - R_F) = 0$$
$$\alpha_2\sigma_2^2\sigma_1^2(\mu - R_F) + \rho\sigma_1\sigma_2\sigma_P^2(\mu_1 - R_F) - \rho^2\sigma_1^2\sigma_2^2\alpha_2(\mu - R_F) - \sigma_P^2\sigma_1^2(\mu_2 - R_F) = 0$$
$$\alpha_2\sigma_2^2\sigma_1^2(\mu - R_F) - \rho^2\sigma_1^2\sigma_2^2\alpha_2(\mu - R_F) = \sigma_P^2\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2\sigma_P^2(\mu_1 - R_F)$$
$$\alpha_2\left(\mu - R_f\right)\left[\sigma_2^2\sigma_1^2 - \rho^2\sigma_1^2\sigma_2^2\right] = \sigma_P^2\sigma_1^2\left(\mu_2 - R_f\right) - \rho\sigma_1\sigma_2\sigma_P^2\left(\mu_1 - R_f\right)$$
$$\alpha_2 = \frac{\sigma_P^2\sigma_1^2(\mu_2 - R_F)}{(\mu - R_F)\left[\sigma_2^2\sigma_1^2 - \rho^2\sigma_1^2\sigma_2^2\right]} - \frac{\rho\sigma_1\sigma_2\sigma_P^2(\mu_1 - R_F)}{(\mu - R_F)\left[\sigma_2^2\sigma_1^2 - \rho^2\sigma_1^2\sigma_2^2\right]}$$
$$\alpha_2 = \frac{\sigma_P^2}{\mu - R_f}\cdot\frac{\sigma_1^2(\mu_2 - R_F)}{\sigma_2^2\sigma_1^2(1 - \rho^2)} - \frac{\sigma_P^2}{\mu - R_f}\cdot\frac{\rho\sigma_1\sigma_2(\mu_1 - R_F)}{\sigma_2^2\sigma_1^2(1 - \rho^2)}$$
$$\alpha_2 = \frac{\sigma_P^2}{\mu - R_F}\cdot\frac{1}{\sigma_2^2\sigma_1^2(1 - \rho^2)}\sigma_1^2(\mu_2 - R_F) - \frac{\sigma_P^2}{\mu - R_F}\cdot\frac{1}{\sigma_2^2\sigma_1^2(1 - \rho^2)}\cdot\rho\sigma_1\sigma_2(\mu_1 - R_F)$$

and remembering that:

$$\lambda = \frac{\sigma_P^2}{\mu - R_0}$$

and placing:

$$\Delta = \sigma_2^2\sigma_1^2\left(1 - \rho^2\right)$$

We have:

$$\alpha_2 = \frac{\lambda}{\Delta}\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]$$

Substituting λ and α_2 in the first equation, we have:

$$\alpha_1\sigma_1^2 + \rho\sigma_1\sigma_2\left\{\frac{\lambda}{\Delta}\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]\right\} - \lambda(\mu_1 - R_F) = 0$$

$$\alpha_1\sigma_1^2 + \frac{\lambda}{\Delta}\rho\sigma_1\sigma_2\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right] - \lambda(\mu_1 - R_F) = 0$$

$$\alpha_1\sigma_1^2 + \frac{\lambda}{\Delta}\left\{\rho\sigma_1\sigma_2\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]\right\} - \lambda(\mu_1 - R_F) = 0$$

$$\alpha_1 = \frac{\sigma_1^2 + \frac{\lambda}{\Delta}\left\{\rho\sigma_1\sigma_2\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]\right\} - \lambda(\mu_1 - R_F)}{\sigma_1^2}$$

$$\alpha_1 = \frac{\lambda}{\Delta}\frac{1}{\sigma_1^2}\left[(\mu_1 - R_F)\sigma_1^2\sigma_2^2\left(1-\rho^2\right) - \rho\sigma_1\sigma_2\sigma_1^2(\mu_2 - R_F) + \rho^2\sigma_1^2\sigma_2^2(\mu_1 - R_F)\right]$$

$$\alpha_1 = \frac{\lambda}{\Delta}\frac{1}{\sigma_1^2}\left[(\mu_1 - R_F)\left[\sigma_1^2\sigma_2^2\left(1-\rho^2\right) + \rho^2\sigma_1^2\sigma_2^2\right] - \rho\sigma_1\sigma_2\sigma_1^2(\mu_2 - R_F)\right]$$

$$\alpha_1 = \frac{\lambda}{\Delta}\frac{1}{\sigma_1^2}\left[(\mu_1 - R_F)\left[\sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 + \rho^2\sigma_1^2\sigma_2^2\right] - \rho\sigma_1\sigma_2\sigma_1^2(\mu_2 - R_F)\right]$$

$$\alpha_1 = \frac{\lambda}{\Delta}\frac{1}{\sigma_1^2}\left[(\mu_1 - R_F)\sigma_1^2\sigma_2^2 - \rho\sigma_1\sigma_2\sigma_1^2(\mu_2 - R_F)\right]$$

$$\alpha_1 = \frac{\lambda}{\Delta}\left[(\mu_1 - R_F)\sigma_1^2\sigma_2^2\frac{1}{\sigma_1^2} - \rho\sigma_1\sigma_2\sigma_1^2(\mu_2 - R_F)\frac{1}{\sigma_1^2}\right]$$

$$\alpha_1 = \frac{\lambda}{\Delta}\left[(\mu_1 - R_F)\sigma_2^2 - \rho\sigma_1\sigma_2(\mu_2 - R_F)\right]$$

$$\alpha_1 = \frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1 - R_F) - \rho\sigma_1\sigma_2(\mu_2 - R_F)\right]$$

$$\alpha_2 = \frac{\lambda}{\Delta}\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]$$

Defined the value of α_1 and α_2, and by substituting it in the budget constraint we have the efficient frontier. Specifically, the budget constraint is equal to:

$$\mu = R_F + \alpha_1(\mu_1 - R_F) + \alpha_2(\mu_2 - R_F)$$

By replacing it, we have:

$$\mu - R_F = \left\{\frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1 - R_F) - \rho\sigma_1\sigma_2(\mu_2 - R_F)\right]\right\}(\mu_1 - \mathrm{R_f})$$
$$+\left\{\frac{\lambda}{\Delta}\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]\right\}(\mu_2 - R_F)$$
$$\mu - R_F = \frac{\lambda}{\Delta}\left\{\left[\sigma_2^2(\mu_1 - R_F) - \rho\sigma_1\sigma_2(\mu_2 - R_F)\right](\mu_1 - R_F)\right.$$
$$\left.+\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right](\mu_2 - R_F)\right\}$$
$$\mu - R_F = \frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1 - R_F)^2 - \rho\sigma_1\sigma_2(\mu_2 - R_F)(\mu_1 - R_F) + \sigma_1^2(\mu_2 - R_F)^2\right.$$
$$\left.-\rho\sigma_1\sigma_2(\mu_1 - R_F)(\mu_2 - R_F)\right]$$
$$\mu - R_F = \frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1 - R_F)^2 + \sigma_1^2(\mu_2 - R_F)^2 - 2\rho\sigma_1\sigma_2(\mu_1 - R_F)(\mu_2 - R_F)\right]$$
$$\mu - R_F = \frac{\lambda}{\Delta}\left[\sigma_2^2\left(\mu_1^2 + \mathrm{R_F^2} - 2\mu_1 R_F\right) + \sigma_1^2\left(\mu_2^2 + \mathrm{R_F^2} - 2\mu_2 R_F\right)\right.$$
$$\left.-2\rho\sigma_1\sigma_2\left(\mu_1\mu_2 - \mu_1 R_F - \mu_2 R_F + \mathrm{R_F^2}\right)\right]$$
$$\mu - R_F = \frac{\lambda}{\Delta}\left[\sigma_2^2\mu_1^2 + \sigma_2^2\mathrm{R_F^2} - 2\mu_1 R_F\sigma_2^2 + \sigma_1^2\mu_2^2 + \sigma_1^2\mathrm{R_F^2} - 2\mu_2 R_F\sigma_1^2\right.$$
$$\left.-2\rho\sigma_1\sigma_2\mu_1\mu_2 + 2\rho\sigma_1\sigma_2\mu_1 R_F + 2\rho\sigma_1\sigma_2\mu_2 R_F - 2\rho\sigma_1\sigma_2\mathrm{R_F^2}\right]$$
$$\mu - R_F = \frac{\lambda}{\Delta}\left[\mathrm{R_F^2}\left(\sigma_2^2 + \sigma_1^2 - 2\rho\sigma_1\sigma_2\right) - 2\mathrm{R_f}\left(\sigma_2^2\mu_1 + \sigma_1^2\mu_2 - \rho\sigma_1\sigma_2\mu_1 - \rho\sigma_1\sigma_2\mu_2\right)\right.$$
$$\left.+\left(\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2 - 2\rho\sigma_1\sigma_2\mu_1\mu_2\right)\right]$$
$$\mu - R_F = \frac{\lambda}{\Delta}\left[\mathrm{R_F^2}\left(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\right) - 2R_F\left(\sigma_2^2\mu_1 + \sigma_1^2\mu_2 - \rho\sigma_1\sigma_2(\mu_1 + \mu_2)\right)\right.$$
$$\left.+\left(\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2 - 2\rho\sigma_1\sigma_2\mu_1\mu_2\right)\right]$$

By placing:

$$\mathrm{A} = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2;$$
$$\mathrm{B} = \sigma_1^2\mu_2 + \sigma_2^2\mu_1 - \rho\sigma_1\sigma_2(\mu_1 + \mu_2);$$
$$\mathrm{C} = \sigma_1^2\mu_2^2 + \sigma_2^2\mu_1^2 - 2\rho\sigma_1\sigma_2\mu_1\mu_2$$

And replacing it, we have:

$$\mu - R_F = \frac{\lambda}{\Delta}\left(\mathrm{AR_F^2} - 2\mathrm{B}R_F + \mathrm{C}\right)$$

and remembering that:

$$\lambda = \frac{\sigma^2}{\mu - R_F}$$

and substituting, we have:

$$\mu - R_F = \frac{\sigma^2}{(\mu - R_F)} \frac{1}{\Delta} \left(\mathrm{AR}_\mathrm{F}^2 - 2\mathrm{B}R_F + \mathrm{C} \right)$$

$$(\mu - R_F)^2 = \frac{\sigma^2}{\Delta} \left(\mathrm{AR}_\mathrm{F}^2 - 2\mathrm{B}R_F + \mathrm{C} \right)$$

$$(\mu - R_F)^2 = \sigma^2 \frac{\mathrm{AR}_\mathrm{F}^2 - 2\mathrm{B}R_F + \mathrm{C}}{\Delta}$$

$$\mu - R_F = \pm \sqrt{\sigma^2 \frac{\mathrm{AR}_\mathrm{F}^2 - 2\mathrm{B}R_F + \mathrm{C}}{\Delta}}$$

$$\mu = R_F \pm \sigma \sqrt{\frac{\mathrm{AR}_\mathrm{F}^2 - 2\mathrm{B}R_F + \mathrm{C}}{\Delta}} \tag{5.81}$$

Equation (5.81) is the equation of the efficient frontier (Saltari 2011).

It is important to note that the equation $\mu = R_f + \frac{\mu_T - R_f}{\sigma_T} \Delta\sigma$ is similar (not considering the negative sign) to the equation of the efficient frontier.

Therefore, by considering one risk-free asset in portfolio, we have two lines with a common interception equal to R_F and slope equal to σ_P in absolute terms, as shown in Fig. 5.14 (adapted from Saltari 2011).

In Fig. 5.14 only the line with a positive slope $(R_F T)$ defines the efficient frontier by considering N risky assets and one free-risk asset. Also, the line that originates from R_F and passes through the point T, defining the efficiency line.

The points between R_F and T represent all possible solutions of the trade-off between risk and return by changing the part of wealth invested in the portfolio of risky assets and in the risk-free asset.

The points over point T obtained by the fixed rate debt (or equivalently, by short selling the risk-free asset and acquiring the risky assets).

Generally, each portfolio on the lines TR_FE can be obtained by linear combination of the risk-free asset and the tangent portfolio (T) consisting of risky assets (Saltari 2011). Specifically:

- in the point T, all wealth is invested in the two risky assets, and therefore $\alpha_0 = 0$;
- in the point R_F, all wealth is invested in the risk-free, and therefore $\alpha_0 = 1$;
- to the right of the point T, short selling of the risk-free asset occurs and acquisition of the risky assets with an increase in the portfolio risk;

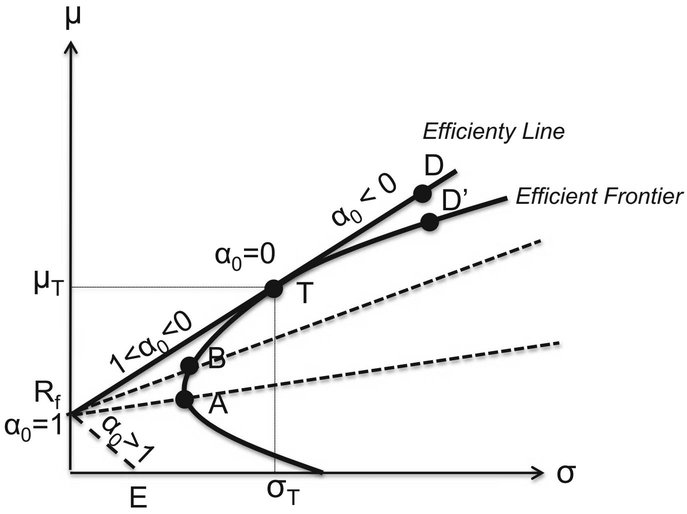


Fig. 5.14 The efficient frontier for a risk-free asset in portfolio

– in the section $R_F E$, there is short selling of the risky assets and acquisition of the risk-free asset, and therefore $\alpha_0 > 1$.

In point T, all wealth is invested in the risky assets. Therefore, there is no risk-free asset in the portfolio, and therefore $\alpha_0 = 0$ and consequently the budget constraint is equal to $\alpha_1 + \alpha_2 = 1$.

Based on this condition, it is possible to estimate the expected return and the variance of the tangent portfolio (T) (Saltari 2011).

Remembering that:

$$\alpha_1 = \frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1 - R_F) - \rho\sigma_1\sigma_2(\mu_2 - R_F)\right]$$

and

$$\alpha_2 = \frac{\lambda}{\Delta}\left[\sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]$$

and adding the two equations, we have:

$$\alpha_1 + \alpha_2 = \frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1 - R_F) + \sigma_1^2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_2 - R_F) - \rho\sigma_1\sigma_2(\mu_1 - R_F)\right]$$

and therefore:

$$\begin{aligned}
\alpha_1+\alpha_2 &= \frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1-R_F)+\sigma_1^2(\mu_2-R_F)-\rho\sigma_1\sigma_2(\mu_2-R_F+\mu_1-R_F)\right]\\
\alpha_1+\alpha_2 &= \frac{\lambda}{\Delta}\left[\sigma_2^2(\mu_1-R_F)+\sigma_1^2(\mu_2-R_F)-\rho\sigma_1\sigma_2(\mu_2+\mu_1-2R_F)\right]\\
\alpha_1+\alpha_2 &= \frac{\lambda}{\Delta}\left[\sigma_2^2\mu_1-\sigma_2^2R_F+\sigma_1^2\mu_2-\sigma_1^2R_F-\rho\sigma_1\sigma_2(\mu_1+\mu_2)+\rho\sigma_1\sigma_2 2R_F\right]\\
\alpha_1+\alpha_2 &= \frac{\lambda}{\Delta}\left[\sigma_2^2\mu_1+\sigma_1^2\mu_2-\rho\sigma_1\sigma_2(\mu_1+\mu_2)-R_f\left(\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2\right)\right]
\end{aligned}$$

Remembering that:

$$\begin{aligned}
A &= \sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2\\
B &= \sigma_1^2\mu_2+\sigma_2^2\mu_1-\rho\sigma_1\sigma_2(\mu_1+\mu_2)\\
\alpha_1+\alpha_2 &= 1
\end{aligned}$$

and replacing it, we have:

$$1=\frac{\lambda}{\Delta}(B-AR_F)$$

Therefore, it is possible to obtain the value of the Lagrangian multiplier (λ) for the portfolio T as follows:

$$\lambda=\frac{\Delta}{B-AR_F}$$

Substituting λ in the equation, we have:

$$\mu-R_F=\frac{\lambda}{\Delta}\left(AR_F^2-2BR_F+C\right)$$

and therefore the expected return of the tangent portfolio (μ_T) is equal to:

$$\mu_T-R_F=\frac{1}{\Delta}\frac{\Delta}{B-AR_F}\left(AR_F^2-2BR_F+C\right)$$

and therefore:

$$\mathrm{\mu_T}-\mathrm{R_F}=\frac{1}{\mathrm{B}-\mathrm{AR_F}}\left(\mathrm{AR_F^2}-2\mathrm{BR_F}+\mathrm{C}\right) \tag{5.82}$$

The variance of the tangent portfolio $\left(\sigma_T^2\right)$ is obtained by substituting the value $\left(\mu_T-R_f\right)$ in the equation:

$$(\mu - R_F)^2 = \sigma^2 \frac{AR_F^2 - 2BR_F + C}{\Delta}$$

and therefore:

$$\left[\frac{1}{B - AR_F}\left(AR_F^2 - 2BR_F + C\right)\right]^2 = \sigma^2 \frac{AR_F^2 - 2BR_F + C}{\Delta}$$

and therefore:

$$\begin{aligned}
\sigma_T^2 &= \frac{\left[\frac{1}{B-AR_F}\left(AR_F^2 - 2BR_F + C\right)\right]^2}{AR_F^2 - 2BR_F + C}\Delta \\
\sigma_T^2 &= \frac{\frac{\left(AR_F^2 - 2BR_F + C\right)^2}{(B-AR_F)^2}}{AR_F^2 - 2BR_F + C}\Delta \\
\sigma_T^2 &= \frac{\left(AR_F^2 - 2BR_F + C\right)^2}{(B - AR_F)^2} \cdot \frac{\Delta}{AR_F^2 - 2BR_F + C} \\
\sigma_T^2 &= \Delta \frac{AR_F^2 - 2BR_F + C}{(B - AR_F)^2}
\end{aligned} \tag{5.83}$$

Therefore, the investor chooses a portfolio consisting of a risk-free asset and the tangent portfolio defined by the risky assets (Saltari 2011). Therefore, the efficient portfolio consists of a mix between the risk-free asset and the risky asset. Consequently, they are positioned in the part of the line between R_F and the tangenty portfolio (T).

Finally, Fig. 5.14 shows that:

- the risky portfolio A is the minimum variance portfolio for all combinations between the risk-free asset and the risky assets on the line $R_F A$;
- the risky portfolio B, for all combinations between the risk-free asset and the risky assets on the line $R_F B$;
- the risky portfolio C, for all combinations between the risk-free asset and the risky assets on the line $R_F C$.

All combinations between the risk-free asset and the risky assets on the line $R_F B$ are preferred to the combinations on the line $R_F A$, because a defined risk level gives higher returns.

At the same time, the combinations between the risk-free asset and the risky assets on the line $R_F T$ are preferred to the combinations on the line $R_F B$, because a defined risk level gives higher returns.

Generally, with the slope increases of the line originating from $R_{F,}$ we have the best combination between the risk-free asset and the risky assets. The maximum slope possible is obtained from the tangent point between the line and the efficient frontier as in point (T). As there are no portfolios above the efficient frontier due to the statistical characteristics of the assets, it follows there are no portfolios above the line that originates from R_F and passing through (T) the *efficient line*.

Therefore, if the investor is risk adverse, the portfolio obtained by the combination between the risk-free asset and the risky asset (T) and therefore is positioned on the line $R_F T$. Otherwise, if the investor is inclined to risk, his portfolio is positioned on the line *TD* creating by borrowing at a risk-free rate and therefore by short selling the risk-free asset and acquiring the risky assets.

Consequently, the portfolio (T) is the obtimal portfolio. All investors with the same efficient frontier and the same risk-free rate, choose the same portfolio T.

It is worth noting that the optimal portfolio T can be defined without any information on investor characteristics (Elton et al. 2013).

If it is possible to invest in risk-free assets but it is not possible to borrow at a risk-free rate, the efficient frontier is equal to $R_F CD'$.

Otherwise, if it is possible to invest in a risk-free asset and borrow at a risk-free rate, the efficient frontier is equal to $R_F CD$.

Finally, it is important to point out that, with other conditions equal, the R_F reduction implies the downward movement of the tangent point on the curve. It implies a reduction in the expected return of the optimal portfolio. It leads investors to undertake more risks in an attempt to achieve a higher expected return.

5.3 Efficient Portfolios

The third step of the process concerns the definition of the efficient portfolios.

There are several techniques for the definition of the efficient portfolios. The most important one can be analysed on the basis of four main assumptions (Elton et al. 2013):

- *case 1:* short selling is permitted and it is possible to borrow and invest at a risk-free rate;
- *case 2:* short selling is permitted and it is not possible to borrow or invest at a risk-free rate;
- *case 3:* short selling is not permitted and it is possible to borrow and invest at a risk-free rate;
- *case 4:* short selling is not permitted and it is not possible to borrow or invest at a risk-free rate.

These four cases can be summarized as shown in Table 5.1.

(Case 1) Short Selling is permitted and it is possible to Borrow and Invest at a Risk-Free Rate

The risk-free rate (R_F) for borrowing and for investing implies the existence of one optimal portfolio of risky asset. It is positioned on the straight line originating from the R_F (interception point on the ordinate) and it is tangent to the opportunity

Table 5.1 Short selling and the opportunity to borrow and invest at a risk-free rate

		Borrowing and investing at a risk-free rate	
		Yes	Not
Short selling	Yes	*Case 1*	*Case 2*
	Not	*Case 3*	*Case 4*

frontier (the curve). The line between R_F and the tangent point with the opportunity frontier defines the efficient frontier. Indeed, this line has a greater slope possible based on the R_F value (defining the point of interception on the ordinate) and the statistical characteristics of the risky assets that define the risky portfolio.

The straight line between the risk-free asset and the risky assets portfolio is equal to the ratio between the portfolio's expected return (μ_P) less the risk-free rate (R_F) and the portfolio's standard deviation (σ_P) as follows (Elton et al. 2013):

$$\theta = \frac{\mu_P - R_f}{\sigma_P} \tag{5.84}$$

Note that in the tangent point the slope of the straight line and the curve is the same, as follows:

$$\frac{\mu_P - R_f}{\sigma_P} = \frac{\mu_1 - R_f}{\sigma_1}$$

The optimization constrained problem, can be formalized as follows:

$$\begin{cases} \max\limits_{\alpha_i} \theta = \frac{\mu_P - R_F}{\sigma_P} \\ \sum\limits_{i=1}^{N} \alpha_i = 1 \end{cases} \tag{5.85}$$

where the first is the equation to be maximized and the second is the constraint.

To avoid the Lagrangian the constraint can be replaced in the equation to be maximized (Elton et al. 2013).

It is possible to multiply both terms of the constraint for risk-free rate (R_F) as follows:

$$R_F \sum_{i=1}^{N} \alpha_i = R_F \rightarrow \sum_{i=1}^{N} (\alpha_i R_F) = R_F$$

Remember that the portfolio's expected return and its standard deviation are the following:

$$\mu_P = \sum_{i=1}^{N} \alpha_i \mu_i$$
$$\sigma_P = \sqrt{\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}} = \left(\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j} \right)^{\frac{1}{2}}$$

and by substituting the constraint in the equation to be maximized, we have:

$$\theta = \frac{\mu_P - R_F}{\sigma_P} = \frac{\sum_{i=1}^{N} \alpha_i \mu_i - \sum_{i=1}^{N} (\alpha_i R_F)}{\left(\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j} \right)^{\frac{1}{2}}}$$
$$= \frac{\sum_{i=1}^{N} \alpha_i (\mu_i - R_F)}{\left(\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j} \right)^{\frac{1}{2}}}$$
$$\theta = \frac{\sum_{i=1}^{N} \alpha_i (\mu_i - R_F)}{\left(\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j} \right)^{\frac{1}{2}}}$$

Therefore, the optimization constrained problem is moved into a simple maximization problem. It is necessary to calculate the first derivative with regards to each variable and placing equal to zero. The problem can be solved by solving this equation system (Elton et al. 2013):

$$\begin{cases} \frac{\partial \theta}{\partial \alpha_1} = 0 \\ \frac{\partial \theta}{\partial \alpha_2} = 0 \\ \cdots \\ \cdots \\ \frac{\partial \theta}{\partial \alpha_n} = 0 \end{cases} \tag{5.86}$$

Generally, the first derivative of θ respect to α_k, is equal to:

$$\frac{\partial \theta}{\partial \alpha_k} = -\left[\lambda \alpha_k \sigma_k^2 + \sum_{\substack{i=1 \\ j \neq k}}^{n} \lambda \alpha_j \sigma_{k,j} \right] + (\mu_k - R_F) = 0 \tag{5.87}$$

and in an extensive form, it is equal to:

$$\frac{\partial\theta}{\partial\alpha_k} = -\left(\lambda\alpha_1\sigma_{1,i} + \lambda\alpha_2\sigma_{2,i} + \lambda\alpha_3\sigma_{3,i}\ldots + \lambda\alpha_i\sigma_i^2 + \cdots + \lambda\alpha_{n-1}\sigma_{n-1,i} + \lambda\alpha_n\sigma_{n,i}\right) + \mu_k - R_F = 0$$

In the equation λ is a constant and it is equal to:

$$\lambda = \frac{\mu_P - R_f}{\sigma_P^2} = \frac{\sum_{i=1}^{N}\alpha_i\left(\mu_i - R_f\right)}{\sum_{i=1}^{N}\alpha_i^2\sigma_i^2 + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}} \tag{5.88}$$

Since each α_k (for $k = 1, 2, \ldots n$) is multiplied by the constant λ, it is possible to rewrite the first derivative on the basis of a new variable:

$$Z_k = \lambda\alpha_k \tag{5.89}$$

The λ is constant. Therefore, Z_k is proportional to α_k and therefore to the part of wealth invested in the k-th asset.

By replacing it we have:

$$\begin{aligned}
\frac{\partial\theta}{\partial\alpha_k} &= -\left(Z_1\sigma_{1,i} + Z_2\sigma_{2,i} + Z_3\sigma_{3,i}\ldots + Z_i\sigma_i^2 + \cdots + Z_{n-1}\sigma_{n-1,i} + Z_n\sigma_{n,i}\right) + \mu_k - R_f = 0\\
&- \left(Z_1\sigma_{1,i} + Z_2\sigma_{2,i} + Z_3\sigma_{3,i}\ldots + Z_i\sigma_i^2 + \cdots + Z_{n-1}\sigma_{n-1,i} + Z_n\sigma_{n,i}\right) + \mu_k - R_f = 0\\
&Z_1\sigma_{1,i} + Z_2\sigma_{2,i} + Z_3\sigma_{3,i}\ldots + Z_i\sigma_i^2 + \cdots + Z_{n-1}\sigma_{n-1,i} + Z_n\sigma_{n,i} = \mu_k - R_f
\end{aligned}$$

and therefore:

$$Z_1\sigma_{1,i} + Z_2\sigma_{2,i} + Z_3\sigma_{3,i}\ldots + Z_i\sigma_i^2 + \cdots + Z_{n-1}\sigma_{n-1,i} + Z_n\sigma_n = \mu_k - R_f$$

This is the equation for each value of i. The solution implies the following equation system (Elton et al. 2013):

$$\begin{cases}
\mu_1 - R_F = Z_1\sigma_1^2 + Z_2\sigma_{1,2} + Z_3\sigma_{1,3} + \cdots + Z_n\sigma_{1,n}\\
\mu_2 - R_F = Z_1\sigma_{1,2} + Z_2\sigma_2^2 + Z_3\sigma_{2,3} + \cdots + Z_n\sigma_{2,n}\\
\mu_3 - R_F = Z_1\sigma_{1,3} + Z_2\sigma_{2,3} + Z_3\sigma_3^2 + \cdots + Z_n\sigma_{3,n}\\
\ldots\\
\mu_n - R_F = Z_1\sigma_{1,n} + Z_2\sigma_{2,n} + Z_3\sigma_{3,n} + \cdots + Z_n\sigma_n^2
\end{cases} \tag{5.90}$$

The terms Z_k (for $k = 1, 2, \ldots, n$) are proportional to the optimal amount to be invested in each asset. To define the optimal amount to be invested, the equations for Z_s must be defined. There are N equations (one for each asset) and N terms unknown (the Z_k for each asset). Subsequently, the optimal amount (as a percentage) to be invested in the k-th asset is equal to (Elton et al. 2013):

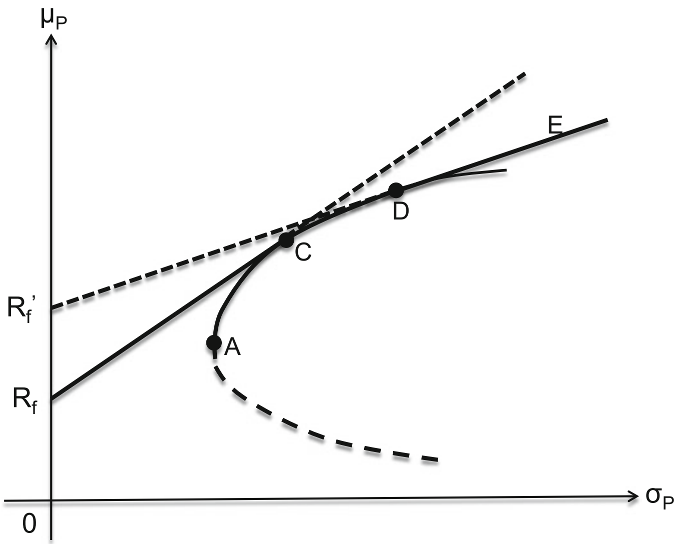


Fig. 5.15 The risk-free rate changes and optimal portfolio

$$\alpha_k = \frac{Z_k}{\sum_{i=1}^{N} Z_i} \tag{5.91}$$

(Case 2) Short Selling is permitted and it is not possible to Borrow or Invest at a Risk-Free Rate

In this case, the problem is related to the risk-free rate. By changing the risk-free rate, the tangent point on the curve changes, as shown in Fig. 5.15.

If the risk-free rate is R'_F, the optimal portfolio is equal to D, while if the risk-free rate is R_F the optimal protfolio is C.

The problem can be solved by considering that the optimal amount to be invested in each asset is linear function of the risk-free rate R_F. For each value of R_F the optimal portfolio can be defined on the basis of the efficient frontier as defined with regards to the statistical characteristics of the assets.

Formally, based on Eq. (5.89) it is possible to solve Z_k as function of R_F as follows (Elton et al. 2013):

$$Z_k = C_{ok} + C_{1k} R_F \tag{5.92}$$

where C_{ok} and C_{1k} are constant. They change with a change in k but they are constant compared with R_F. Consequently, by changing R_F, the amount to be invested in the k-th asset on the different point of the efficient frontier is changed.

(Case 3) Short Selling is not Permitted and it is Possible to Borrow and Invest at a Risk-Free Rate The difference between *Case 3* and *Case 1* is that the investors have to own a positive amount of the assets $(\alpha_i > 0)$.

Therefore, optimization constrained problem can be formalized as follows (Elton et al. 2013):

$$\begin{cases} \max\limits_{\alpha_i} \theta = \frac{\mu_P - R_f}{\sigma_P} \\ \sum\limits_{i=1}^{n} \alpha_i = 1 \\ \alpha_i \geq 0 \quad \forall i \end{cases} \tag{5.93}$$

where the first is the equation to be maximized while the second and third equations are the constraints.

It is important to note that the equation to be maximized is not linear but quadratic. It is due to the terms $\alpha_i^2, \alpha_i\alpha_j$ in the portfolio's variance (σ_P^2) and therefore in the standard deviation (σ_P). Therefore, the problem is a quadratic programming, and it requires the use of a specific algorithm that is usually structured on the basis of the Kuhn-Tucker conditions. They can be summarized as follows (Elton et al. 2013).

We must have a derivative of θ compared with each α_i and set it to zero in order to find the maximum value of θ. The problem regards the constraints. If there is no constraint, α_i can be positive or negative and the optimum can occur when α_i is positive or negative. Otherwise, if there is a constraint and α_i must not be negative $(\alpha_i \geq 0)$ the optimum could be achieved at a value of α_i which is not feasible, as shown in Fig. 5.16.

The variable θ as function of α_i may look like *Curve 1* or *Curve 2*.

If there is a constraint and α_i should not be negative $(\alpha_i \geq 0)$, in the case of *Curve 1* the maximum feasible value of θ occurs at point M. Otherwise, in the of *Curve 2*, the maximum feasible value of θ occurs at point M_1 rather than M_2.

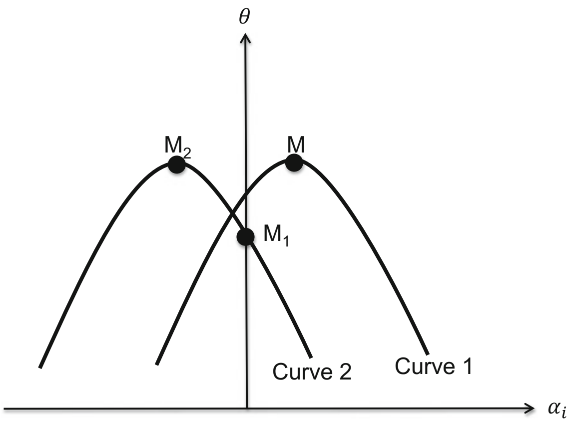


Fig. 5.16 Value of the function θ as α_i changes

If the maximum value for α_i occurs at M_1 (*Curve 2*) then the partial derivative of θ with regards to α_i is negative ($\frac{\partial\theta}{\partial\alpha_i} < 0$) at the maximum feasible value (for $\alpha_i = 0$).

Otherwise, if the maximum value for α_i occurs at M (*Curve 1*) then the partial derivative of θ with regards to α_i is equal to zero ($\frac{\partial\theta}{\partial\alpha_i} = 0$) at the maximum feasible value (for $\alpha_i \geq 0$).

Therefore, if there is a constraint based on which α_i should not be negative ($\alpha_i \geq 0$), we have:

$$\frac{\partial\theta}{\partial\alpha_i} \leq 0$$

This relationship can be re-written as follows:

$$\frac{\partial\theta}{\partial\alpha_i} + U_i = 0 \tag{5.94}$$

where U_i is the part to reach zero.

This is the first Kuhn-Tucker condition for maximization.

Therefore, if the optimum occurs when:

– α_i is positive (*Curve 1*): the derivative is equal to zero and then U_i it is equal to zero:

$$\alpha_i > 0 \rightarrow \frac{\partial\theta}{\partial\alpha_i} = 0 \rightarrow U_i = 0 \Leftrightarrow \alpha_i > 0, \quad U_i = 0$$

– the maximum occurs at α_i equal to zero (*Curve 2*): the derivative is negative and then U_i is positive:

$$\alpha_i = 0 \rightarrow \frac{\partial\theta}{\partial\alpha_i} \leq 0 \rightarrow U_i > 0 \Leftrightarrow \alpha_i = 0, \quad U_i > 0$$

In general terms, these relationships can be written in a compact form, as follows:

$$\begin{gathered} \alpha_i\ U_i = 0 \\ \alpha_i \geq 0 \\ U_i \geq 0 \end{gathered} \tag{5.95}$$

This is the second Kuhn-Tucker condition for maximization.

Therefore, the four Kuhn-Tucker conditions can be summarized as follows (Elton et al. 2013):

$$\begin{array}{ll} (1) & \frac{\partial\theta}{\partial\alpha_i} + U_i = 0 \\ (2) & \alpha_i \quad U_i = 0 \\ (3) & \alpha_i \geq 0 \\ (4) & U_i \geq 0 \end{array} \tag{5.96}$$

Therefore, if the solution respects these conditions, then it allows for definition of the optimal portfolio. In other words, if the solution suggested satisfies the Kuhn-Tucker conditions, then it gives the optimum portfolio.

(Case 4) Short Selling is not Permitted and it is not Possible to Borrow or Invest at a Risk-Free Rate In this case the efficient frontier is obtained by minimizing the portfolio's risk for a defined level of expected returns. Formally (Elton et al. 2013):

$$\begin{cases} \min\limits_{\alpha_i} \sigma_P^2 = \sum\limits_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum\limits_{i=1}^{N} \sum\limits_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j} \\ \sum\limits_{i=1}^{N} \alpha_i = 1 \\ \sum\limits_{i=1}^{N} \alpha_i \mu_i = \mu_P \\ \alpha_i \geq 0 \quad i = 1, 2, \ldots, N \end{cases} \tag{5.97}$$

The first is the equation to be minimized, while the second, third and fourth are the constraints. The first constraint $\left(\sum_{i=1}^{N} \alpha_i = 1\right)$ shows that short selling is not permitted; while the last $(\alpha_i \geq 0)$ shows that the amount to be invested in each asset is positive or null (it is not possible to borrow at a risk-free rate).

The equation to be minimized is quadratic due to the terms $\alpha_i^2, \alpha_i \alpha_j$ in the portfolio's variance. Therefore, it is necessary to use the four conditions of Kuhn-Tucker to solve the problem by defining the optimal portfolio.

5.4 Optimal Portfolio

The fourth step of the process is the choice of the optimal portfolio. It is function of the investor's preferences. A specific level of risk characterizes the investor. In this regard the investor solves the trade-off between risk and return.

The mean-variance criterion does not define which portfolio on the efficient frontier must be chosen. The choice is function of the investor's risk aversion level, and this variable is not considered in the model.

The investor's choices can be achieved by using the utility function. Therefore, the optimal portfolio for the investor is the efficient portfolio on the efficient frontier maximising the investor's utility function (Castellani et al. 2005; Cohen and Pogue

1967; Connor and Korajczyk 1993; Carhart 1997; Chan et al. 1999; Elton et al. 2006, 2013; Levy 1984).

It is important to note that the expected utility must be defined on the basis of mean and variance. The coherence between the expected utility criterion and the mean-variance criterion, the utility function must be quadratic and/or the returns follow a normal distribution.

Assuming that the investor invests in a time t in the portfolio P consisting of assets $a_0, a_1, a_2, \ldots, a_n$. Assuming that the investor is characterized by the utility function $\mu(x)$. Assuming that the value of the portfolio in a future period s is equal to A_P. The investor's choice is based on the maximization of the expected utility in a time s, as follows:

$$U = E[\mu(x)] = E[\mu(A_P)] \tag{5.98}$$

If the utility function is quadratic or the expected returns follow a normal distribution (so that there is coherence between the utility function criterion and the mean-variance criterion), the expected utility can be expressed as function of the portfolio's standard deviation (σ_P) and its expected return (μ_P), as follows:

$$U(\sigma_P, \mu_P) \tag{5.99}$$

Therefore, the expected utility is increasing function of the portfolio's expected return (μ_P) and it is decreasing function of its standard deviation (σ_P).

In the plane (σ_P, μ_P) it is possible to introduce lines corresponding to the same fixed level of the expected utility $\bar{\mu}$, as follows (Castellani et al. 2005):

$$U(\sigma_P, \mu_P) = \bar{\mu} \quad \sigma \geq 0 \tag{5.100}$$

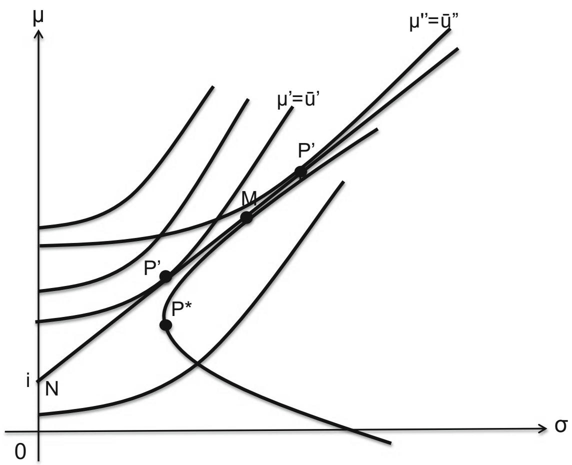


Fig. 5.17 Indifference curves and portfolio choices

All portfolios that satisfy this equation are indifferent for the investor with regards to his expected return. Consequently, the level lines of U (the indifference curves) identify the indifferent lines for the investor for each defined level $\bar{\mu}$.

The indifference curves in the plan (σ_P, μ_P) are convex (upward concavity). The higher the investor risk aversion, the higher the convexity; in the same way, the lower the investor risk aversion, the lower the convexity.

Moving towards the upper left in the plan, the curves correspond to the $\bar{\mu}$ levels of expected utility growth. Under this assumption, there is one efficient portfolio that maximizes U: this portfolio is represented by the point P in which the efficient frontier is tangent to the indifference curve corresponding to the $\bar{\mu}$ maximum value.

Investors with a different level of risk aversion, and therefore with a different convexity of the indifference curve, but defining the same efficient frontier, choose a different efficient portfolio on the efficient frontier as shown by the points P' and P'' in Fig. 5.17 (adapted from Castellani et al. 2005).

Using $\mu(\sigma;\bar{\mu})$ to denote the function that explicitly espresses the indifference curve according to the investor's expected utility $\bar{\mu}$, the abscissa of the point σ of the maximum utility is uniquely determined as the system solution in two equations in variables σ and $\bar{\mu}$, as follows (Castellani et al. 2005):

$$\begin{cases} \mu(\sigma;\bar{\mu}) = i + \pi_M \sigma \\ \frac{\partial \mu(\sigma;\bar{\mu})}{\partial \sigma} = \pi_M \end{cases} \tag{5.101}$$

The first equation defines the tangent condition between the indifference curve and the efficient line. The second equation indicates that the indifference curve has to have the same slope as the indifference line, in the tangent point.

The problem can be formalized in a matrix form as follows (Cesari 2012b):

$$\begin{cases} \max\limits_{w_1,\dots w_N} \Psi\left(\overset{+}{\mu_p};\overset{-}{\sigma_p^2}\right) \\ \mu_p = \boldsymbol{\alpha}'\mu \\ \sigma_p^2 = \boldsymbol{\alpha}'\Sigma\alpha \\ \boldsymbol{\alpha}'\mathbf{1} = 1 \end{cases} \tag{5.102}$$

The first equation is the function of indirect utility.

For calculation of the indifference curves, the differential of the utility function must be defined (Cesari 2012b):

$$\mathrm{d}\Psi = \frac{\partial \Psi}{\partial \mu_p} d\mu_p + \frac{\partial \Psi}{\partial \sigma_p} d\sigma_p = 0 \tag{5.103}$$

It is equal to zero, and therefore the utility level does not change, so:

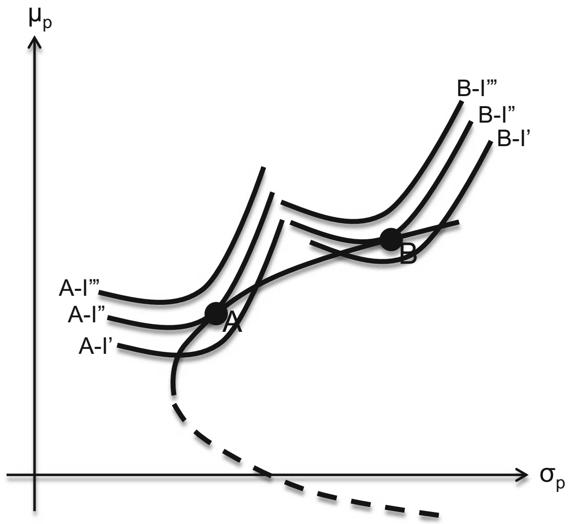


Fig. 5.18 The efficient frontier and the indifference curve

$$\frac{d\mu_p}{d\sigma_p} = \frac{-\frac{\partial \Psi}{\partial \sigma_p}}{\frac{\partial \Psi}{\partial \mu_p}} > 0 \tag{5.104}$$

Positivity of the marginal rate of substitution, identifies the trade-off between the portfolio's expected return and its standard deviation for a defined investor with preference Ψ. Consequently, if the portfolio risk increases, it can only be accepted with an increase in the expected return (μ_p).

Graphically, the indifference curves allow for definition of the optimal portfolio. They are perfectly increasing in the plan $(\sigma_p; \mu_p)$ and they represent constant levels of wealth on the same curve and increasing in step upwards (from the north-west) by a curve.

The growth in the indifference curve shows that the higher-level risk must be compensated by higher expected return. The slope of the indifference curves represents the subjective risk premium of the investor and therefore his aversion to risk.

The match between the information arising from capital markets as summarized in the efficient frontier on the one hand, and the information arising from the investor's preferences as summarized by the indifference curve on the other hand, allows for solving of the problem of portfolio choices. The optimal portfolio is positioned on the efficient frontier and it is tangent to the higher indifference curve, as shown in Fig. 5.18 (adapted from Cesari 2012b).

Figure 5.18 shows with regards to investor *A*, that the optimal portfolio is obtained by the tangent between curve $A - I''$ and the efficient frontier and also, as far as investor B is concerned, it is obtained by the tangent between curve $B - I''$ and the efficient frontier.

5.5 Single Index Model

One of the most important problems of the portfolio choices refers to the correct application of the theoretical approach due to the high number and quality of the input data required: the correct application of the theoretical model requires an estimate of the expected returns of each asset that may potentially be introduced into the portfolio as well as an assessment of all of the covariance between the assets' expected returns in a couple. Specifically, by considering N assets in portfolio, the correlation between the assets in a couple are equal to $[N(N-1)]/2$ as discussed previously.

Several models are developed to reduce the complexity by reducing the amount of data input to calculate the correlation between the assets' expected returns in the portfolio.

To do this, they allow for application of the theory by introducing some constraints and assumptions. These constraints and assumptions reduce the level of accuracy of the model but they allow for simplification. Among these models, the most useful model is the *Single Index Model* (Elton et al. 2013).

The Single Index Model is the most simple model of the portfolio's theory. It can be considered as a special case of the *Market Model* as illustrated at a later stage.

The model assumes one common factor, the single index, on which the movement of all assets depend.

The model is based on empirical evidence: usually, when the market increases, as indicated by the positivity of some indexes, the majority of the assets increase; on the contrary, when the market decreases, as indicated by the negativity of some indexes, the majority of the assets decrease. It leads us to believe that these assets are connected between them, because they respond to the same common factor. Consequently, a measure of the correlation between the assets can be obtained indirectly by considering the relationship between each asset's expected returns with the defined index's expected returns.

Formally, the i-th asset's expected return (R_i) can be defined as follows (Elton et al. 2013):

$$R_i = F_i + \beta_i R_m \tag{5.105}$$

where:

- F_i: is the part of the asset's expected return not related to the index. It is function of the company's specific characteristics. It is a random variable;
- R_m: is the defined index's expected return. It is a random variable;
- β_i: is a constant and it measures the sensibility of the asset's expected returns (R_i) to the change in the index's expected returns (R_m). It measures the expected change in the R_i due to the change in R_m.

Therefore, Eq. (5.105) divides the asset's expected return in two parts:

– the first (F_i), is independent of the market, as it is represented by the index defined, and it is based on the company's specific characteristics;
– the second $(\beta_i R_m)$, is dependent of the market, as it is represented by the index defined, on the basis of a sensibility coefficient.

The component F_i can be divided in two parts, as follows:

$$F_i = f_i + e_i \tag{5.106}$$

where:

– the first part (f_i) denotes the expected value of F_i and can be defined as "*anticipated term*";
– the second part (e_i) denotes the random (uncertain) element and therefore the residual term of F_i and it can only be known at the end of the period.

The asset's risk is function of the second part of Eq. (5.107), and therefore from the probability structure of e_i.

Expectations on the value of the assets in the starting time (t_0) is function of the information reflected in the price through the component f_i. Therefore, f_i is a component of the asset's price in t_0.

The component e_i is a random variable. The opinion on it usually reflects the investors risk behaviour. Its introduction in the price is mainly due to the investors risk aversion. Therefore, they require a risk premium to invest in risky assets. This premium risk is the largest return required by investors and it can be translated in a price discount.

On the basis of Eq. (5.106), Eq. (5.105) we have (Elton et al. 2013):

$$R_i = f_i + \beta_i R_m + e_i \tag{5.107}$$

where both e_i that R_m are random variables.

The Single Index Model is based on a set of parameters (Elton et al. 2013).

The standard equation is defined by Eq. (5.107) as follows:

$$R_i = f_i + e_i + \beta_i R_m \quad for\, i = 1, 2, 3, \ldots, N$$

By construction, the mean of e_i is equal to:

$$E(e_i) = 0 \quad for\, i = 1, 2, 3, \ldots, N \tag{5.108}$$

By definition, the variances of e_i and R_m are equal to:

$$E[e_i - E(e_i)]^2 = E(e_i - 0)^2 = E(e_i)^2 = \sigma^2_{e_i} \quad for\, i = 1, 2, 3, \ldots, N \tag{5.109}$$

$$E(R_m - \mu_m)^2 = \sigma^2_{R_m} \tag{5.110}$$

There are two main baseline assumptions:

Assumption 1 there is no correlation between e_i and R_m (for $i = 1, 2, 3, \ldots, N$). Therefore, we have:

$$\begin{aligned} Cov(e_i; R_m) &= E[(e_i - E(e_i))(R_m - E(R_m))] = E[e_i(R_m - E(R_m))] \\ &= E[e_i(R_m - \mu_m)] = 0 \end{aligned} \tag{5.111}$$

Therefore, the assumption ensures that the two variables (e_i and R_m) are independent among them. Specifically, the accuracy of the equation to describe the asset's return is independent from the index's return.

Assumption 2 each asset is only correlated to the market index representing the common factor. It implies non-correlation between the random elements (residual) of the assets (e_i and e_j for each *i*-th and *j*-th asset). Therefore, the covariance between the residuals of i-th asset and j-th asset (for $i = 1, 2, 3, \ldots, N$ and $j = 1, 2, 3, \ldots, N$) are equal to zero, as follows:

$$E\left[\left(e_i - \mu_{e_i}\right)\left(e_j - \mu_{e_j}\right)\right] = E\left(e_i e_j\right) = 0 \tag{5.112}$$

Therefore, the assumption ensures that the returns of the assets are connected with the market index's returns only, and not among them. Consequently, the returns of all assets move in the same direction because they move with the same common factor as represented by the market index, not because they are related among them. The assumption implies that the only reason assets vary together is because of common movement with the market index.

These two assumptions are very restrictive. They represent strong simplification and they represent a relevant approximation to reality. By combining these two assumptions, the Single Index Model is based on the simple concept that stock prices move together only because of common movement with the market as represented by market index. It assumes that the stock prices move together only because of a common co-movement with market index.

On the basis of these assumptions, it is possible to estimate the expected returns and the variance and the portfolio and the covariance between the assets in portfolio.

The i-th asset's *expected return* (R_i) is equal to (Elton et al. 2013):

$$\begin{aligned} R_i &= f_i + e_i + \beta_i R_m \rightarrow E(R_i) = E(f_i + e_i + \beta_i R_m) \rightarrow E(R_i) \\ &= E(f_i) + E(e_i) + E(\beta_i R_m) \end{aligned}$$

Considering that f_i and β_i are constant, and $E(e_i) = 0$, we have:

$$E(R_i) = f_i + \beta_i E(R_m)$$

$$E(R_i) = f_i + \beta_i E(R_m) \leftrightarrow \mu_i = f_i + \beta_i \mu_m \tag{5.113}$$

The i-th asset's *variance* $\left(\sigma_i^2\right)$ is equal to (Elton et al. 2013):

$$\sigma_i^2 = E(R_i - E(R_i))^2 \rightarrow \sigma_i^2 = E(R_i - \mu_i)^2$$

By substituting R_i and μ_i with relative Eqs. (5.107) and (5.113), we have:

$$\begin{aligned}\sigma_i^2 &= E[(f_i + e_i + \beta_i R_m) - (f_i + \beta_i \mu_m)]^2 = E[f_i + e_i + \beta_i R_m - f_i - \beta_i \mu_m]^2 \\ &= E[e_i + \beta_i (R_m - \mu_m)]^2 = E\left[(e_i)^2 + (\beta_i (R_m - \mu_m))^2 + 2e_i \beta_i (R_m - \mu_m)\right] \\ &= E(e_i)^2 + E[\beta_i (R_m - \mu_m)]^2 + E[2e_i \beta_i (R_m - \mu_m)]\end{aligned}$$

Considering that β_i it is constant, and by considering that $E(e_i)^2 = \sigma_{ei}^2$, $E(R_m - \mu_m)^2 = \sigma_m^2$ and $E[e_i(R_m - \mu_m)] = 0$, we have:

$$\sigma_i^2 = E(e_i)^2 + \beta_i^2 E(R_m - \mu_m)^2 + 2\beta_i E(e_i(R_m - \mu_m))$$

and then:

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma_{ei}^2 \tag{5.114}$$

The *covariance between the i-th asset and the j-th assets*, is equal to (Elton et al. 2013):

$$\sigma_{i,j} = E\left[(R_i - E(R_i))\left(R_j - E\left(R_j\right)\right)\right] \rightarrow \sigma_{i,j} = E\left[(R_i - \mu_i)\left(R_j - \mu_j\right)\right]$$

Substituting R_i, R_j, μ_i and μ_j with relative equations, we have:

$$\begin{aligned}\sigma_{i,j} &= E\left\{[(f_i + e_i + \beta_i R_m) - (f_i + \beta_i \mu_m)]\left[\left(f_j + e_j + \beta_j R_m\right) - \left(f_j + \beta_j \mu_m\right)\right]\right\} \\ &= E\left\{[f_i + e_i + \beta_i R_m - f_i - \beta_i \mu_m]\left[f_j + e_j + \beta_j R_m - f_j - \beta_j \mu_m\right]\right\} \\ &= E\left\{[e_i + \beta_i (R_m - \mu_m)]\left[e_j + \beta_j (R_m - \mu_m)\right]\right\} \\ &= E\left[e_i e_j + e_i \beta_j (R_m - \mu_m) + e_j \beta_i (R_m - \mu_m) + \beta_i \beta_j (R_m - \mu_m)^2\right] \\ &= E\left(e_i e_j\right) + E\left[e_i \beta_j (R_m - \mu_m)\right] + E\left[e_j \beta_i (R_m - \mu_m)\right] + E\left[\beta_i \beta_j (R_m - \mu_m)^2\right]\end{aligned}$$

and considering that β_j it is a constant, and by considering that $E(e_i) = 0$, $E(R_m - \mu_m)^2 = \sigma_m^2$, $E[e_i(R_m - \mu_m)] = 0$ and $E\left(e_i e_j\right) = 0$, we have:

$$\sigma_{i,j} = E\left(e_i e_j\right) + \beta_j E[e_i (R_m - \mu_m)] + \beta_i E\left[e_j (R_m - \mu_m)\right] + \beta_i \beta_j E\left[(R_m - \mu_m)^2\right]$$

and then:

$$\sigma_{i,j} = \beta_i \beta_j \sigma_m^2 \tag{5.115}$$

Based on Eqs. (5.113), (5.114) and (5.115) it is possible to estimate the expected return and the variance (and therefore standard deviation) of the portfolio.

The *portfolio's expected return* (μ_P) is equal to (Elton et al. 2013):

$$\mu_P = \sum_{i=1}^{N} \alpha_i \mu_i$$

and by considering that:

$$\mu_i = f_i + \beta_i \mu_m \rightarrow \mu_P = \sum_{i=1}^{N} \alpha_i (f_i + \beta_i \mu_m)$$

and therefore:

$$\mu_P = \sum_{i=1}^{N} \alpha_i f_i + \sum_{i=1}^{N} \alpha_i \beta_i \mu_m \tag{5.116}$$

The *portfolio's variance* $\left(\sigma_P^2\right)$ is equal to (Elton et al. 2013):

$$\sigma_P^2 = \sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}$$

by considering that:

$$\xrightarrow[\sigma_{i,j}=\beta_i\beta_j\sigma_m^2]{\sigma_i^2=\sigma_{e_i}^2+\beta_i^2\sigma_m^2} \sigma_P^2 = \sum_{i=1}^{N} \alpha_i^2 \left(\sigma_{e_i}^2 + \beta_i^2 \sigma_m^2\right) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \beta_i \beta_j \sigma_m^2$$

and then:

$$\sigma_P^2 = \sum_{i=1}^{N} \alpha_i^2 \beta_i^2 \sigma_m^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \beta_i \beta_j \sigma_m^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_{e_i}^2 \tag{5.117}$$

Therefore, in order to estimate the portfolio's expected return (μ_P) and its variance $\left(\sigma_P^2\right)$, it is necessary to estimate:

- three variables for each i-th asset: f_i, β_i and σ_i^2;
- two variables for the market index: μ_m and σ_m^2.

Consequently, $3N + 2$ terms rather than $[N(N - 1)]/2$ are required. A reduction in the estimates is significant.

The Single Index Model tries to reduce complexity due to the pure portfolio theory. To do this, the Single Index Model assumes that the stock prices move together only because of common movement with the market as represented by the market index. Therefore, it assumes that the stock prices move together only because of a common co-movement with market index. This restrictive assumption reduces the number of correlations to be estimated and therefore it allows for easy application of the portfolio theory.

It is interesting to re-write the portfolio's expected return and variance on the basis of the *portfolio's beta* (β_P) and its *anticipated term* (f_P). In both cases, they can be defined as a weighted average: the *portfolio's beta* (β_P) is equal to the weighted average betas (β_i) of the assets in portfolio, and the *portfolio's anticipated term* (f_P) is equal to the weighted average anticipated term (f_i) of the assets in portfolio (Elton et al. 2006). Formally:

$$\beta_P = \sum_{i=1}^{N} \alpha_i \beta_i; \quad f_P = \sum_{i=1}^{N} \alpha_i f_i \tag{5.118}$$

The *portfolio's expected return* (μ_P) can be re-written as follows:

$$\mu_P = \sum_{i=1}^{N} \alpha_i f_i + \sum_{i=1}^{N} \alpha_i \beta_i \mu_m \rightarrow \mu_P = f_P + \beta_P \mu_m \tag{5.119}$$

It is worth noting that if the portfolio (P) is equal to the market portfolio, their expected returns must be equal. It implies that $f_P = 0$ and $\beta_P = 1$:

$$\mu_P = \mu_m \rightarrow \begin{matrix} f_P = 0 \\ \beta_P = 1 \end{matrix} \tag{5.120}$$

It implies that the beta coefficient of the market must be equal to 1 and the i-th asset can be defined more or less risky than the market if the β_i is higher or lower than β_P respectively.

By considering that:

$$\sum_{i=1}^{N} \alpha_i \beta_i = \beta_P; \quad \sum_{j=1}^{N} \alpha_j \beta_j = \beta_P$$

the *portfolio's variance* $\left(\sigma_P^2\right)$ can be re-written as follows (Elton et al. 2013):

$$\sigma_P^2 = \sum_{i=1}^{N} \alpha_i^2 \beta_i^2 \sigma_m^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \beta_i \beta_j \sigma_m^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_{e_i}^2 \rightarrow \sigma_P^2$$
$$= \beta_{\mathrm{P}}^2 \sigma_{\mathrm{m}}^2 + \beta_{\mathrm{P}}^2 \sigma_{\mathrm{m}}^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_{e_i}^2$$

and by considering that the first and the second terms are the same, we have:

$$\sigma_P^2 = \beta_{\mathrm{P}}^2 \sigma_{\mathrm{m}}^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_{e_i}^2 \tag{5.121}$$

Note that the same result can be obtained if the condition $j \neq i$ is not considered in the portfolio's variance. In this case we have:

$$\sigma_P^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \beta_i \beta_j \sigma_m^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_{e_i}^2 = \left(\sum_{i=1}^{N} \alpha_i \beta_i \right) \left(\sum_{j=1}^{N} \alpha_j \beta_j \right) \sigma_m^2 + \sum_{i=1}^{N} \alpha_i^2 \sigma_{e_i}^2$$
$$= \beta_{\mathrm{P}}^2 \sigma_{\mathrm{m}}^2 + \sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}}^2 \sigma_{\mathrm{e_i}}^2$$

On the basis of the portfolio's variance, it is possible to analyse the effect of an increase in the portfolio assets. Assuming a portfolio of N assets and assuming an investment of the same part of wealth in each asset so that: $\alpha_i = \frac{1}{N}$. On the basis of Eq. (5.121), the portfolio's variance is equal to (Elton et al. 2013):

$$\sigma_P^2 = \beta_P^2 \sigma_m^2 + \sum_{i=1}^{N} \left(\frac{1}{N} \right)^2 \sigma_{e_i}^2 \rightarrow \sigma_P^2 = \beta_P^2 \sigma_m^2 + \frac{1}{N} \left(\sum_{i=1}^{N} \frac{1}{N} \sigma_{e_i}^2 \right) \rightarrow \sigma_P^2$$
$$= \beta_P^2 \sigma_m^2 + \frac{1}{N} \left(\sum_{i=1}^{N} \frac{\sigma_{e_i}^2}{N} \right)$$

It is important to note that the increase of the assets in portfolio (N) decreases the average residual risk $\left(\sigma_{e_i}^2 \right)$. In fact:

$$\lim_{N \to \infty} \frac{\sigma_{e_i}^2}{N} = 0 \rightarrow \lim_{N \to \infty} \frac{1}{N} \left(\sum_{i=1}^{N} \frac{\sigma_{e_i}^2}{N} \right) = 0$$

Therefore, by increasing the number of assets in the portfolio (N), the second part of the portfolio's risk is null and the only risk is associated with the portfolio's beta as follows:

$$\sigma_P^2 = \beta_P^2 \sigma_m^2 \rightarrow \sigma_P = \beta_P \sigma_m \tag{5.122}$$

By considering that the portfolio's beta (β_P) is equal to the weighted average betas (β_i) of the assets in portfolio, we have (Elton et al. 2013):

$$\beta_P = \sum_{i=1}^{N} \alpha_i \beta_i \rightarrow \sigma_P = \left(\sum_{i=1}^{N} \alpha_i \beta_i \right) \sigma_m$$

By considering that σ_m is the same for each i-th asset, the i-th asse's marginal contribution to the portfolio's risk is function of its coefficient beta (β_i).

Note that the i-th asset's variance, as defined by Eq. (5.114) can be split in two parts:

- the first part $\left(\sigma_{e_i}^2\right)$ measures *the specific risk of asset:* it is not relevant with regards to the marginal contribution to the portfolio's risk. Indeed, it tends to be zero to increase assets in the portfolio;
- the second part $\left(\beta_i^2 \sigma_m^2\right)$ measures the *systematic risk:* it is independent of the number of assets in the portfolio and therefore it does not change by changing N. Also, by considering that σ_m^2 is a constant for each of the N assets in portfolio, the contribution margin of the asset to the portfolio's risk is measured by β_i only. Therefore, this is the only element of risk to be considered and the coefficient beta (β_i) represents the risk measurement of the i-th asset in the portfolio.

It is worth noting that, by using the Single Index Model, the coefficient beta measures the systematic risk of the asset based on only the market index defined. In fact:

- the market portfolio is approximated by a specific market index and it is the same for each asset;
- the expected returns of each asset are not related among them but they are connected with the expected return of the market index only.

Therefore, the main problem is an estimate of the beta of the i-th assets that could be introduced in the portfolio. There are two main approaches:

- *first approach*: it is based on an estimate of the future betas on the correlations between the asset's expected returns and the index's expected returns. In this case, the beta estimate tends to be subjective as function of the subjective prevision of the analyst about the expected returns;
- *second approach*: it is based on an estimate of the future betas on the basis of historical betas. This second approach is more useful and there is proof that historical betas provide useful information on future betas (Blume 1970; Levy 1971). This approach is normally preferred by analysts rather than the first one.

By following the second approach, the i-th asset's beta is estimated on the basis of the historical betas.

Equation (5.107) defines the i-th asset's expected return (R_i). Based on past data, it is possible to observe the past returns of the market and the i-th asset. If f_i, β_i and $\sigma^2_{e_i}$ are assumed to be constant through time, it is possible to expect the same equation at each point in time. Therefore, the straightforward procedure exists for an estimate of the parameters f_i, β_i and $\sigma^2_{e_i}$ (Elton et al. 2013). To do this, it is possible to use the regression analysis.

Therefore, by considering the time (t), Eq. (5.107) can be rewritten as follows:

$$R_{it} = f_i + \beta_i R_{mt} + e_i \tag{5.123}$$

and applying the Ordinary Least Square (OLS), we have:

$$\beta_i = \frac{\sum_{t=1}^{n}[(R_{it} - \mu_{it})(R_{mt} - \mu_{mt})]}{\sum_{t=1}^{n}(R_{it} - \mu_{it})^2} = \frac{Cov(R_m; R_i)}{Var(R_m)} = \frac{\sigma_{(R_m;R_i)}}{\sigma^2_{R_m}} \tag{5.124}$$

and

$$f_i = R_{it} - \beta_i R_{mt} \tag{5.125}$$

Obviously, the estimate is subject to error and therefore, the estimated β_i and f_i could not be equal to their real values that existed in the period. Furthermore, β_i and f_i are not perfectly stationarity over time.

The Single Index Model, can be considered as a special case of the Market Model. The difference is due to an assumption: while in the Single Index Model the covariance of the assets to the same market index only is assumed $\left(E\left(e_i, e_j\right) = 0\right)$, in the Market Model a correlation between expected returns of the assets in portfolio $E\left(e_i, e_j\right) \neq 0$ is assumed. Therefore, the Market Model is characterized by a less restrictive form than the Single Index Model. In this sense, the Single Index Model can be considered a special case of the Market Model (Elton et al. 2013).

The standard Market Model can be defined as in Eq. (5.107). The change in assumption about the relationship among the expected returns of assets in portfolio, does not lead to the simple expressions of portfolio risk as in the Single Index Model. In other words, the Market Model reduces the complexity of the pure portfolio's theory less than the Single Index Model.

Therefore, while in the Single Index Model R_m is the expected returns of the market index choices as representative of the market portfolio, in the Market Model by removing the assumption $E\left(e_i e_j\right) = 0$, R_m represents the expected returns of the portfolio built by the investor. Specifically, the expected returns of assets in portfolio are related among them, and therefore the investor has to define the portfolio by choosing the assets to be held on the basis of their expected returns and the correlations with the other assets owned. Consequently, the beta of the i-th asset changes as function of the choice of the assets in portfolio.

In general and therefore by moving from the Single Index Model to the Market Model by removing the assumption $E(e_i e_j) = 0$, for a well-diversified portfolio, non-systematic risk approach to zero and the only relevant asset's risk is the systematic risk measured by coefficient beta.

It is function of the correlation between the i-th asset's expected returns and the expected returns of all of the assets in portfolio. Subsequently beta is a measure of the risk emerging arise from the relationship between the asset's expected returns and the portfolio's expected returns.

By assuming a portfolio of N assets and the i-th asset, the coefficient beta of the i-th asset can be changed if:

- *the i-th asset's expected returns change, with the assets in the portfolio equal as well as their expected returns.* In this case, the change in the i-th asset's beta is function of the change in the company's elements that have an effect on free cash-flows to equity and then on its expected returns. The changes in the i-th asset's expected return generate a change in the correlation between them and the portfolio's expected returns;
- *the assets in portfolio change, being equal the i-th asset's expected returns*. In this case, the redefinition of the portfolio by changing the assets owned, generates new correlations between the i-th asset expected returns and the portfolio's expected returns;
- *the assets in portfolio do not change but their expected returns change, with the i-th asset's expected returns equal*. In this case, the change in the assets' expected returns change the portfolio's expected returns and therefore the correlations between them and the i-th asset's expected returns;
- *the i-th asset's expected returns change together with a change in the portfolio assets or their expected returns*. In this case, the i-th asset's beta changes as function of the new correlations between the i-th asset's expected returns and the portfolio's expected returns.

Therefore, the i-th asset's beta, and therefore its systematic risk, is function of the company's elements according to their effects on the expected returns and the portfolio's characteristics. Consequently, it changes over time.

Finally, it is interesting to note that the covariance refers to the asset's expected returns. The asset's expected return is function of the change in market prices. The changes are not only due to the fundamental analysis of the asset but they include market movements. Consequently, the asset's fundamental analysis is a part of the price movements on the market.

References

Alexander G (1976) The derivation of efficient sets. J Financ Quant Anal XI(5):817–830

Alexander G (1977) Mixed security testing of alternative portfolio selection models. J Financ Quant Anal XII(4):817–832

Alexander G (1978) A reevaluation of alternative portfolio selection models applied to common stock. J Financ Quant Anal XIII(1):71–78

Bawa VS, Brown SJ, Klein RW (1979) Estimation risk and optimal portfolio choice. North Holland, Amsterdam

Bertsekas D (1974) Necessary and sufficient conditions for existence of an optimal portfolio. J Econ Theory 8(2):235–247

Blume M (1970) Portfolio theory: a step toward its practical application. J Bus 43(2):152–173

Brennan MJ, Kraus A (1976) The geometry of separation and myopia. J Financ Quant Anal XI (2):171–193

Brown S, Barry C (1985) Differential information and security market equilibrium. J Financ Quant Anal 20:407–422

Brumelle S (1974) When does diversification between two investments pay? J Financ Quant Anal IX(3):473–483

Buser S (1977) Mean-variance portfolio selection with either a singular or non-singular variance-covariance matrix. J Financ Quant Anal XII(3):436–461

Canner N (1997) An asset allocation puzzle. Am Econ Rev 87(1):181–193

Carhart M (1997) On persistence in mutual fund performance. J Financ 52:661–692

Cass D, Stiglitz J (1970) The structure of investor preferences and asset returns, and separability in portfolio allocation: a contribution to the pure theory of mutual funds. J Econ Theory 2(2):122–160

Castellani G, De Felice M, Moriconi F (2005) Manuale di Finanza 2. Teoria del portafoglio e mercato azionario, Il Mulino

Cesari R (2012a) Introduzione alla finanza matematica. Concetti di base, tassi e obbligazioni, 2nd edn. McGraw-Hill, New York

Cesari R (2012b) Introduzione alla finanza matematica. Mercati azionari, rischi, e portafogli, 2nd edn. McGraw-Hill, New York

Chan LKC, Karceski J, Lakonishok J (1999) On portfolio optimization: forecasting covariances and choosing the risk model. Rev Financ Stud 12:263–278

Chen A (1977) Portfolio selection with stochastic cash demand. J Financ Quant Anal XII(2):197–213

Chen A, Jen F, Zionts S (1971) The optimal portfolio revision policy. J Bus 44(1):51–61

Chen A, Kim H, Kon S (1975) Cash demands, liquidation costs and capital market equilibrium under uncertainty. J Financ Econ 2(3):293–308

Cohen K, Pogue J (1967) An empirical evaluation of alternative portfolio selection models. J Bus 46:166–193

Connor G, Korajczyk R (1993) A test for the number of factors in an approximate factor model. J Financ 48:1263–1291

Dalal AJ (1983) On the use of a covariance function in a portfolio model. J Financ Quant Anal XVIII(2):223–228

Dybving PH (1984) Short sales restrictions and kinks of the mean variance frontier. J Financ 39 (1):239–244

Edwards F, Goetzmann W (1994) Short horizon inputs and long horizon portfolio choice. J Portfolio Manage 20(4):76–81

Elton EJ, Gruber MJ (1971) Dynamic programming applications in finance. J Financ XXVI (2):473–505

Elton EJ, Gruber MJ (1974) Portfolio theory when investment relatives are lognormally distributed. J Financ XXIX(4):1265–1273

Elton EJ, Gruber MJ (1977) Risk reduction and portfolio size: an analytical solution. J Bus 50 (4):415–496

Elton EJ, Gruber MJ, Spitzer J (2006) Improved estimates of correlation coefficients and their impact on the optimum portfolios. Eur Financ Manage 12(3):303–318

Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN (2013) Modern portfolio theory and investment analysis, 9th edn. Wiley

Epps TW (1981) Necessary and sufficient conditions for the mean-variance portfolio model with constant risk aversion. J Financ Quant Anal XVI(2):169–176

Fama E (1968) Risk, return, and equilibrium: some clarifying comments. J Financ 23:29–40

Fama E (1981) Stock return, real activity, inflation and money. Am Econ Rev 71:545–565

Fama E, MacBeth J (1973) Risk, return, and equilibrium: empirical tests. J Polit Econ 38:607–636

Farrell J (1974) The multi-index model and practical portfolio analysis. In: The financial analysts research foundation occasional paper, vol 4

Francis JC (1975) Intertemporal differences in systemic stock price movements. J Financ Quant Anal 10(2):205–219

Hakansson N (1970) An induced theory of the firm under risk: the pure mutual fund. J Financ Quant Anal V(2):155–178

Hill R (1976) An algorithm for counting the number of possible portfolios given linear restrictions on the weights. J Financ Econ XI(3):479–487

Jacob N (1974) A limited-divesification portfolio selection model for the small investor. J Financ XXIX(3):847–856

Jennings E (1971) An empirical analysis of some aspects of common stock diversification. JFinanc Quant Anal VI(2):797–813

Johnson K, Shannon D (1974) A note of diversification and the reduction of dispersion. J Financ Econ 1(4):365–372

Jones-Lee MW (1971) Some portfolio adjustment theorems for the case of non-negativity conditions on security holdings. J Financ XXVI(3):763–775

Kryzanowski L, To MC (1983) General factor models and structure of security returns. J Financ Quant Anal 18:31–52

Latane H, Tuttle D, Young A (1971) How to choose a market index. Financ Anal J 27(4):75–85

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J Empir Financ 10(5):603–624

Levy H (1984) Measuring risk and performance over alternative investment horizons. Financ Anal J 40(2):61–67

Levy RA (1971) On the short-term, stationary of beta coefficients Financ Anal J 27(5):55–62

Lewis AL (1988) A simple algorithm for the portfolio selection problem. J Financ 43(1):71–82

Lintner J (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev Econ Stat XLVII:13–37

Markowitz H (1952) Portfolio selection. J Financ 7(1):77–91

Markowitz H (1956) The optimization of a quadratic function subject to linear constraints. Naval Res Log 3(1–2):111–133

Markowitz H (1959) Portfolio selection: efficient diversification of investment. Wiley, New York

Markowitz H (1976) Markowitz revisited. Financ Anal J 32(4):47–52

Markowitz H (2014) The theory and practice of rational investing. Risk-return analysis, vol I, McGraw-Hill

Markowitz H (2016) The theory and practice of rational investing. risk-return analysis, vol. II, McGraw-Hill

Martin J, Klemkosky R (1975) Evidence of heteroscedasticity in the market model. J Bus 48 (1):81–86

Merton R (1972) An analytic derivation of the efficient portfolio frontier. J Financ Quant Anal VII (4):1851–1872

Meyers S (1973) A re-examination of market and industry factors in stock price behavior. J Financ VIII(3):695–705

Mossin J (1968) Optimal multiperiod portfolio policies. J Bus 41(2):215–229

Officer RR (1973) The variability of the market factor of the New York stock exchange. J Bus 46 (3):434–453

Ohlson J (1975) Portfolio selection in a log-stable market. J Financ Quant Anal X(2):285–298

Pogue G, Solnik B (1974) The market model applied to European common stocks: some empirical results. J Financ Quant Anal IX(6):917–944

Pye G (1973) Lifetime portfolio selection in continuous time for a multiplicative class of utility functions. Am Econ Rev LXIII(5):1013–1020

Robichek A, Cohn R (1974) The economic determinants of systemic risk. J Financ XXIX:439–447

Ross SA, Westerfield R, Jaffe J, Jordan BD (2015) Corporate finance, 11th edn, McGraw-Hill

Rubinstein M (1973) The fundamental theorem of parameter-preference security valuation. J Financ Quant Anal VIII(1):61–69

Rudd A, Rosenberg B (1980) The market model in investment management. J Financ 35(2):597–606

Saltari E (2011) Appunti di Economia Finanziaria, Esculapio

Schafer S, Brealey R, Hodges S (1976) Alternative models of systematic risk. In: Elton EJ, Gruber MJ (eds) International capital markets. North-Holland, Amsterdam

Sharpe W (1971) Mean-absolute-deviation characteristic lines for securities and portfolios. Manage Sci 18(2):1–13

Smith K (1968) Alternative procedures for revising investment portfolios. J Financ Quant Anal III (4):371–403

Statman M (1987) How many stocks make a diversified portfolio? J Financ Quant Anal 22 (3):353–363

Sunder S (1980) Stationary of market risk: random coefficients tests for individual stocks. J Financ 35(4):883–896

Tobin J (1958) Liquidity preference as behaviour towards risk. Rev Econ Stud 25(2):65–86

Wagner W, Lau S (1971) The effect of Diversification on Risk. Financ Anal J 27(5):48–53

Chapter 6
Capital Asset Pricing Model

Abstract The Capital Asset Pricing Model (CAPM) is the most well-known equilibrium model in the capital market. The standard form of CAPM provides a clear description of capital market behaviour if its basic assumptions are respected. There are two main problems. The first one is that some of the basic assumptions are very far from conditions of reality. This is not a problem in itself. The fact that these differences from reality are irrelevant enough, they do not have a material affect on the model's explanatory power. Secondly, the CAPM describes the conditions of equilibrium about returns on the macro level. It does not describe this equilibrium of micro level with regards to individual investor behaviour. Indeed, most investors and institutions have a risky assets portfolio different from the market portfolio. Therefore, while the model can explain the capital markets behaviour as an entity, it is unable to explain the investors behaviour. In fact, the investor's portfolio is usually different from the market portfolio. For this reason, different versions from the CAPM standard are developed, by changing the basic assumptions. The aim is to understand and to explain the standard version of the CPM in greater detail, with the investor's behaviour on the one hand, and the assets price on the other hand. In this context, on the basis of the purpose of this book, the CAPM in its standard version only is considered.

6.1 Baseline Assumptions

The standard form of the Capital Asset Pricing Model (CAPM) probably the most common and the easiest of the equilibrium models. It was developed independently by Sharpe (1964), Lintner (1965a), Mossin (1966). Indeed, it is usually referred to as the Sharpe-Linter-Mossin form of the Capital Asset Pricing Model.

The standard form of CAPM (Back 2017; Bawa and Lindenburg, 1977; Benninga and Protopapadakis 1991; Bernstein 1973; Cochrane 2001; Duffi 2001; Elton et al. 2013; Fama 1968, 1971, 1976, 1998; Kroll and Levy 1992; Lehari and Levy 1977; Levy 1973; Lintner 1965b, 1969, 1970; Modigliani and Pogue 1974a, b; Mossin 1966; Pettit and Westerfield 1972; Ross 1978a, b, c; Rubinstein 1973, 1974;

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_6

Sharpe 1963, 1964, 1966, 1970, 1973, 1991, 1992, 1994; Stapleton 1971; Turnbull 1977) is based on several baseline assumptions. They can be summarized as follows:

1. *no transaction costs:* it implies that there is no cost (no friction) in the buying or selling of any asset. Introduction of the transaction costs adds to the model complexity. Otherwise, the benefits are not relevant because the transaction costs are usually low. Therefore, costs of their introduction in the model are greater than the benefits;
2. *infinite divisibility of the assets:* it implies that the quantity of assets are defined and each asset can be divided without any limitations. Therefore, each asset and for each quantity, is tradable. Consequently, investors could take any position in an investment, regardless of the size of wealth. So, it is possible to invest in each asset regardless of the portfolio dimension;
3. *marketability of all assets*: it implies that all assets, are marketable consequently any asset can be sold and bought on the market;
4. *no personal income taxes:* it implies that there are no differences between the income from obligations, dividends and capital gains;
5. *investor is price-takers:* it implies that investor cannot affect the price of the asset through his buying or selling activities. The asset price is defined by the interactions between demand and supply aggregate of all investors' decisions of simultaneously. It is worth noting that since the investor's choices are unable to influence the asset price in the capital markets, this assumption is analogous to the assumption of the perfect competition;
6. *mean-variance approach in the portfolio choices:* it implies that the portfolio choices are based on the statistical characteristics of the assets with regards to the mean, variance and covariance of their expected returns;
7. *homogeneity of expectations:* it implies three main consequences. First of all, the investors are assumed to be concerned with the mean and variance of returns by using the same mean-variance approach. Secondly, all investors have the same information and expectations about the statistical characteristics of the assets with regards to their expected returns, the variance of returns and the correlation matrix based on the correlation structure between all stock pairs. Thirdly, all investors define the relevant period in the same manner and then the investment time is equal for all investors;
8. *unlimited short sales are allowed*: it implies that there are no restrictions on short selling. Consequently, investors can sell short any number of any assets in a given period of time;
9. *unlimited lending and borrowing at the risk-free rate:* it implies that investors can lend or borrow any amount of funds at a no risk rate.

The CAPM's basic assumptions are rigid and sometimes highly restrictive. In the years several non-standard forms have been tested (Alexander 1977; Arzac and Bawa 1977; Black 1972; Borch 1969; Breeden 1979; Breeden and Litzenberger 1978; Brennan 1971; Brenner and Subrahmanyam 1978; Chamberlain and Rothschild 1983; Chen et al. 1986; Connor 1984; Constantinides 1980; Fama 1970; Fama et al. 1977, 1979; Friend and Westerfield 1980; Grossman and Shiller 1982;

Grossman et al. 1985; Guiso et al. 1996; Hagerman and Kim 1976; Hansen and Singleton 1982; Heckerman 1972; Hilliard 1980; Jarrow 1980; Kraus and Litzenberger 1975; Landskroner 1977a, b; Lintner 1971; Litzenberg and Ronn 1986; Long 1974; Lucas 1978; Mayers 1973, 1976; Merton 1973; Milne and Smith 1980; Rabinovitch and Owel 1978; Ross 1977, 1978a, b, c; Samuelson 1969; Samuelson and Merton 1974; Stapleton and Subrahmanyam 1977).

Despite the assumptions are rigid, they are not the problem in itself. It is not important how realistic they are but, on the contrary, how much the model based on them is able to explain the reality. In this perspective, the right question is how much reality is distorted by the assumption introductions?

The CAPM can be derived in several forms involving different degrees of rigour and mathematical complexity. In this context the two easier approaches are preferred (Elton et al. 2013):

- *Intuitive approach*: it is based on the economic intuition;
- *Rigorous approach*: it is based on a simple mathematical form.

6.2 Intuitive Approach

The intuitive approach to generate the CAPM is based on the economic intuition (Elton et al. 2013).

Assume that short sales are permitted. Assume the opportunity of investor of lending and borrowing at the risk-free rate. Each investor defines the efficient frontier as shown in Fig. 6.1 (adapted from Elton et al. 2013).

All investors require a portfolio of risky assets (M) regardless of their risk preferences. This portfolio lies at the tangent point between the efficient frontier of risky assets and a straight line passing throught the risk-free rate return.

The investor satisfies his risk preferences by combining the portfolio M (portfolio of risky assets) with lending and borrowing at a risk-free rate return.

Investors are characterized by homogeneous expectations. They face the same efficient frontier. Considering that there can only be one risk-free rate in the capital markets (and therefore it is the same for all investors) they also face the same straight line. Therefore, the diagram is the same for all investors. If all investors hold the same risky portfolio (M), then in equilibrium it must be the market portfolio. This portfolio holds all risky assets in the market; each asset is held in the proportion that the market value of that asset represents of the total market value of all risky assets (Elton et al. 2013).

The straight line tangent to the efficient frontier by defining the market portfolio (M) and passing throught the risk-free rate, is the *Capital Market Line* (*CML*). All efficient portfolios are positioned along the CML. Otherwise, all non-efficient portfolios are positioned below the *CML*. Finally, there are no portfolios above the *CML* due to the statistical characteristics of the assets.

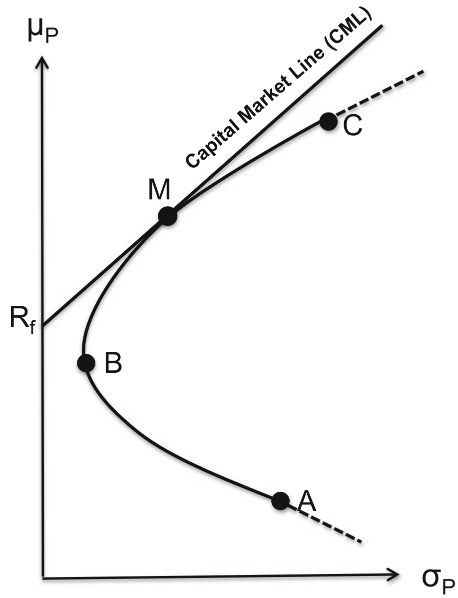


Fig.. 6.1 Efficient frontier with short selling, and lending and borrowing at a risk free rate

Therefore, based on two mutual fund theorem all investors hold a combination of only two portfolios: the portfolio of risky assets that is the market portfolio (*M*) and a risk-free asset. Consequently, all investors can satisfy their risk preferences by considering the market fund and their ability to lend or borrow a risk-free asset.

Note that the risk-free rate is a theoretical construction only. Indeed, a risk-free investment by definition is impossible. But it is necessary to define the market portfolio (M): it is the portfolio on the efficient frontier obtained from the tangent between the efficient frontier and the straight line with interception equal to the risk-free rate. In order to provide a value, the risk-free rate measures the time value of money and it is assumed equal to the government bond return of the country with a default risk low such as to assume it equal to zero.

Denoting with R_{pe} the expected return of the efficient portfolio positioned on the *CML*, σ_{pe} the standard deviation of the efficient portfolio; R_F the risk-free rate, the CML's equation can be defined as follows (Elton et al. 2013):

$$\mathrm{R_{pe}} = \mathrm{R_F} + \left(\frac{\mathrm{R_M} - \mathrm{R_F}}{\sigma_\mathrm{M}}\right)\sigma_\mathrm{pe} \tag{6.1}$$

where:

- $\left(\frac{R_M - R_F}{\sigma_M}\right)$: it is the slope of the *CML* and it can be defined as the *Sharpe ratio* and therefore it can be defined as the "*market price of risk*" for all efficient portfolios. It can be considered as the extra return that can be gained by increasing the risk (standard deviation) of efficient portfolio by unit (Elton et al. 2013);

– $\left(\frac{R_M - R_F}{\sigma_M}\right)\sigma_{pe}$: it can be interpreted as the premium risk and therefore it is the "*market price of risk times the amount of risk in portfolio*". Therefore, it measures the return required for the portfolio's risk equal to the price of risk multiplied by the amount of risk (Elton et al. 2013).

It is relevant to note that Eq. (6.1) describes the equilibrium returns of the efficient portfolio only. It does not describe the equilibrium returns of non-efficient portfolios or individual asset.

In well diversified portfolios, the coefficient beta is the right measure of the asset's risk. In this case, the asset's specific risk (non-systematic risk) tends to go to zero and the only reliance is its systemic risk as measured by the coefficient beta.

Considering the homogeneous expectations of the investors and unlimited risk-free lending and borrowing, all investors hold the market portfolio. Considering that the market portfolio is a well-diversified portfolio, the only relevant characteristics of the assets are its expected returns and beta (Elton et al. 2013).

Therefore, all assets and portfolios have to stand on a straight line in the beta-return space $(\beta; R_p)$. Each point (asset or portfolio) outside of the straight line generates an arbitrage opportunity that moves forward until the price dynamic pushes the point onto a straight line. Indeed, two assets or portfolios that are equivalent cannot sell at different prices.

In Fig. 6.2 (adapted from Elton et al. 2013), point *D* (asset or portfolio) has the same risk level as point *C* but has a greater expected return. Point *D* generates an arbitrage opportunity: point *C* is short sold and point *D* is acquired. In the same way, if point E has the same risk level as point *C* but has a lower expected return, investors sell short E and buy *D* until point *D* goes up onto the line. Generally, the arbitrage opportunities go forward to points *C*, *D* and *E* positioned on the line. Furthermore, note that point *C* represents any combination of points *A* and *B*. Every combination between two points on the line, generates a new point on the line itself.

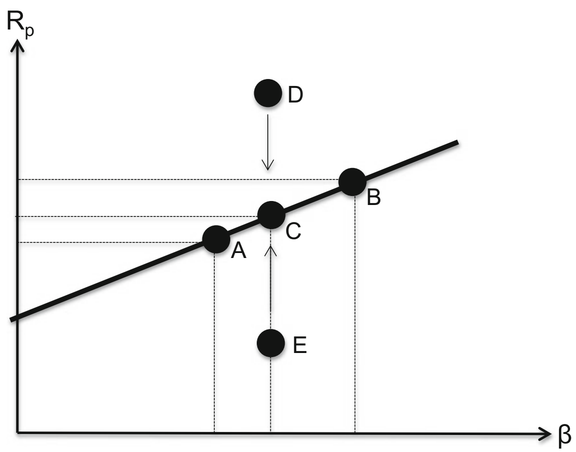


Fig. 6.2 Combination of portfolios in the beta-return space

Based on these considerations, all investments and all portfolios of investments must lie along a straight line in the beta-return space. If any investment is positioned above or below that straight line, there is an opportunity for risk-free arbitrage. This arbitrage continues until all investments converge towards the straight line.

There are many ways of defining this straight line. The easiest one is to identify two points (Elton et al. 2013):

– the first is the market portfolio (M). Under the assumptions of CAPM, each investor holds the market portfolio because all portfolios must lie on the straight line. The market portfolio has a beta equal to 1;
– the second is the risk-free rate. It is the point of interception on the vertical axis when the coefficient beta is equal to zero or, in an equivalent form, when the asset's systemic risk is equal to zero. If the asset is characterised by zero systematic risk, it is a risk-free asset (risk-less asset).

These two points identify the straight line, as shown in Fig. 6.3.

The general equation of the straight line is the following:

$$R = a + b\beta \tag{6.2}$$

where R is the expected return.

The first point can be defined by considering that the beta of the market portfolio is equal to 1. Therefore:

$$\beta_M = 1 \rightarrow R_M = a + b(1) \tag{6.3}$$

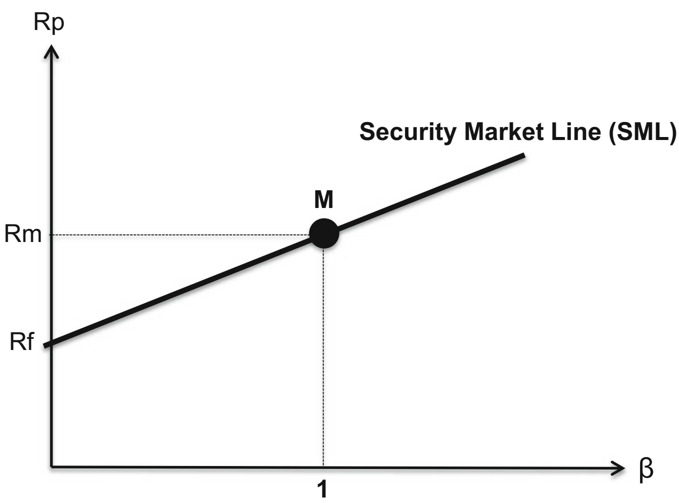


Fig. 6.3 The security market line (SML)

The second point can be defined by considering the interception of the straight line with the vertical axis. This occurs when beta is equal to zero. Therefore:

$$\beta = 0 \rightarrow R_F = a + b(0) \rightarrow R_F = a \tag{6.4}$$

By solving Eq. (6.3) for b, and substituting a with R_F as indicated by Eq. (6.4), we have:

$$b = (R_M - R_F) \tag{6.5}$$

Once the two points are joined, and substituting Eqs. (6.4) and (6.5) in Eq. (6.2), we have:

$$R = R_F + \beta(R_M - R_F) \tag{6.6}$$

Equation (6.6) identifies the *Security Market Line (SML)*. This equation estimates the expected return of each efficient and non-efficient portfolio. This equation is considered one of the most relevant in the field of finance.

It is worth noting that R_M and R_F are not function of the assets considered. Consequently, the difference between assets is only function of the coefficient beta that is the measure of the asset's systematic risk. The relationship between the expected return and beta is linear. Therefore, considering that R_M and R_F are not function of the assets, and considering that the expected return is linear function of the beta, consequently beta and therefore the systematic risk are the only relevant measure and the specific risk (non-systematic risk) is not relevant. In other words, the expected return of the investor is function of the systematic risk bearing. Therefore, the asset' risk is function of the part of the variance in the expected returns that cannot be diversified. Consequently, the investor cannot expect an increase in the expected return function of the asset's specific risk. This is an important economic intuition (Elton et al. 2013):

It is relevant to note, that the definition of the *SML* is function of the definition of beta. If beta is not defined, we have a general straight line and not the *SML*. In this sense, the expected return of the k-th asset or portfolio is equal to:

$$R_K = R_F + \beta_K(R_M - R_F) \tag{6.7}$$

only if:

$$\beta_K = \frac{\sigma_{K,M}}{\sigma_M^2} \tag{6.8}$$

and therefore:

$$R_K = R_F + \left(\frac{\sigma_{K,M}}{\sigma_M^2}\right)(R_M - R_F) \tag{6.9}$$

Equation (6.9) can be expressed in a different form as follows:

$$R_K = R_F + \left(\frac{\sigma_{K,M}}{\sigma_M}\right)\left(\frac{R_M - R_F}{\sigma_M}\right) \tag{6.10}$$

where:

- $\left(\frac{R_M - R_F}{\sigma_M}\right)$ is the Sharpe ratio and can be interpreted as the "*market price for the risk*";
- $\left(\frac{\sigma_{K,M}}{\sigma_M}\right)$ is a risk measurement of each asset or portfolio and therefore measures the risk amount of k-th asset. It is the contribution margin to the standard deviation of the portfolio.

If the "*market price of the risk*" is expressed by variance instead of standard deviation, Eq. (6.9) can be rewritten as follows:

$$R_K = R_F + \sigma_{K,M}\left(\frac{R_M - R_F}{\sigma_M^2}\right) \tag{6.11}$$

where:

- $\left(\frac{R_M - R_F}{\sigma_M^2}\right)$ is the "*market price of the risk*" expressed by variance instead of standard deviation;
- $\left(\sigma_{i,M}\right)$ is the risk amount of the k-th asset.

There are four main aspects emerging from Eq. (6.9) and its different forms as in Eqs. (6.10) and (6.11) (Elton et al. 2013):

- First, the difference between the expected return of two different assets is due to their coefficient beta only. In fact, R_F and R_M are equal for both of them. Then, the higher the coefficient beta, the higher the expected return of the asset. It is relevant to note that CAPM is an equilibrium model. The asset with the higher coefficient beta should achieve greater expected returns. But it does not mean that it is always true in all periods. On the contrary, it means that the asset with the higher coefficient beta can generate a low return; the source of the risk lies in this possibility. It is reasonable to expect higher returns on the long term, but not necessarily on the short term;
- second, the relationship between expected return of the asset and its coefficient beta is linear;

- third, the coefficient beta measures systemic risk only. Then, the expected return of the asset is function of the risk systemic only.
- fourth, such as CML, SML estimates the expected return of the asset equal to the sum between the risk-free rate and the market price of the risk multiplied by the risk amount;
- fifth, the systemic risk is the appropriate measure of the asset's risk. Consequently, two assets with the same systemic risk could not offer different rates of return. The non-systemic risk of large diversified portfolio is essentially zero.

6.3 Rigorous Approach

The CAPM can be derived on the basis of a rigorous approach by using a mathematical form (Elton et al. 2013). In this context the easier one is considered.

Assume that short sales are allowed. Assume that investors can lend and borrow an unlimited amount of money at the risk-free rate.

The existence of a risk-free lending and borrowing rate implies that there is a single portfolio of risky assets that it is preferred over all other portfolios.

Considering the slope of the straight line connecting a free-risk rate and a risky portfolio and considering that the efficient frontier is the entire length of the ray extending between the risk-free rate and the tangent point between the straight line and the frontier of the efficient portfolios, the slope is equal to (Elton et al. 2013):

$$\theta = \frac{R_P - R_F}{\sigma_P} \tag{6.12}$$

In this case, it is necessary to define the portfolio that maximizes the slope of the straight line passing through the risk-free rate (on the vertical axes) and the portfolio itself. Therefore, we have:

$$\begin{cases} \max \theta = \frac{R_P - R_F}{\sigma_P} \\ \sum_{i=1}^{n} \alpha_i = 1 \end{cases} \tag{6.13}$$

Considering the constraints in the function, it is possible to maximise directly instead of using the Lagrangian to solve the system (Elton et al. 2013). In this case, the constraint could be substituted with the objective function; then the objective function maximized as in an unconstrained problem.

The risk-free rate (R_F) can be rewritten as R_F times 1, so that:

$$R_F = 1R_F = \left(\sum_{i=1}^{n} \alpha_i\right) R_F = \sum_{i=1}^{N} (\alpha_i R_F) \tag{6.14}$$

The portfolio's expected return, its variance and standard deviation are the following:

$$R_P = \sum_{i=1}^{n} \alpha_i R_i$$

$$\sigma_P^2 = \sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}$$

$$\sigma_P = \sqrt{\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}} = \left[\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}\right]^{\frac{1}{2}}$$

Explaining, Eq. (6.13) to be maximized, it can be re-written as follows:

$$\theta = \frac{R_P - R_F}{\sigma_P} = \frac{\sum_{i=1}^{n} \alpha_i R_i - \sum_{i=1}^{n} \alpha_i R_F}{\left[\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}\right]^{\frac{1}{2}}}$$

and then:

$$\theta = \left[\sum_{i=1}^{n} \alpha_i \left(\overline{R}_i - R_F\right)\right] \left[\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}\right]^{-\frac{1}{2}} \tag{6.15}$$

Remembering that:

$$\frac{\partial}{\partial \alpha_k} \left[\sum_{i=1}^{n} \alpha_i (R_i - R_F)\right] = R_K - R_F$$

$$\frac{\partial}{\partial \alpha_k} \left[\sum_{i=1}^{N} \alpha_i^2 \sigma_i^2\right] = 2\alpha_k \sigma_k^2$$

$$\frac{\partial}{\partial \alpha_k} \left[\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \alpha_i \alpha_j \sigma_{i,j}\right] = 2 \sum_{\substack{j=1 \\ j \neq k}}^{N} \alpha_j \sigma_{k,j}$$

the derivative of Eq. (6.15) is equal to:

$$\frac{\partial}{\partial\alpha_k}\left(\left[\sum_{i=1}^{n}\alpha_i(R_i-R_F)\right]\left[\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right]^{-\frac{1}{2}}\right)$$
$$=\left\{\left[\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)\right]\left[\left(-\tfrac{1}{2}\right)\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{-\frac{3}{2}}\left(2\alpha_k\sigma_k^2+2\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)\right]\right\}$$
$$+\left\{\left[\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{-\frac{1}{2}}\right]\left[R_K-R_f\right]\right\} \tag{6.16}$$

Pointing the first derivative equal to zero, we have:

$$\left\{\left[\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)\right]\left[\left(-\tfrac{1}{2}\right)\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{-\frac{3}{2}}\left(2\alpha_k\sigma_k^2+2\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)\right]\right\}$$
$$+\left\{\left[\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{-\frac{1}{2}}\right]\left[R_K-R_f\right]\right\}=0$$

multiplying the derivative by $\left[\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{\frac{1}{2}}\right]$, we have:

$$\left\{\left[\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)\right]\left[\left(-\tfrac{1}{2}\right)\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{-\frac{3}{2}}\left(2\alpha_k\sigma_k^2+2\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)\right]\left[\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{\frac{1}{2}}\right]\right\}$$
$$+\left\{\left[\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{-\frac{1}{2}}\right]\left[\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{\frac{1}{2}}\right]\left[R_K-R_f\right]\right\}=0$$

and then:

$$\left[\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)\right]\left[\left(-\frac{1}{2}\right)\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)^{-1}\left(2\alpha_k\sigma_k^2+2\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)\right]$$
$$+\left[R_K-R_f\right]=0$$

and then:

$$\frac{\left[\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)\right]\left[\left(-\frac{1}{2}\right)\left(2\alpha_k\sigma_k^2+2\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)\right]}{\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)}+\left[R_K-R_f\right]=0$$

and then:

$$\frac{\left[\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)\right]\left[(-1)\left(\alpha_k\sigma_k^2+\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)\right]}{\left(\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}\right)}+\left[R_K-R_f\right]=0$$

and then:

$$-\left(\frac{\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)}{\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}}\right)\left(\alpha_k\sigma_k^2+\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)+\left(R_K-R_f\right)=0 \tag{6.17}$$

Defining λ as:

$$\lambda=\frac{\sum_{i=1}^{N}\alpha_i\left(R_i-R_f\right)}{\sum_{i=1}^{N}\alpha_i^2\sigma_i^2+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\alpha_i\alpha_j\sigma_{i,j}}=\frac{R_P-R_F}{\sigma_P^2} \tag{6.18}$$

and substituting constant λ in Eq. (6.17), we have:

$$-\lambda\left(\alpha_k\sigma_k^2+\sum_{\substack{j=1\\j\neq k}}^{N}\alpha_j\sigma_{k,j}\right)+\left(R_K-R_f\right)=0$$

and then:

$$\left(\lambda\alpha_k\sigma_k^2 + \sum_{\substack{j=1 \\ j\neq k}}^{N} \lambda\alpha_j\sigma_{k,j}\right) = \left(R_K - R_f\right) \tag{6.19}$$

Considering this equation with regards to each asset, a system of simultaneous equations is achieved, as follows:

$$\begin{aligned}&\left(\lambda\alpha_1\sigma_{1,k} + \lambda\alpha_2\sigma_{2,k} + \lambda\alpha_3\sigma_{3,k} + \cdots + \lambda\alpha_k\sigma_k^2 + \cdots + \lambda\alpha_{N-1}\sigma_{N-1,k} + \lambda\alpha_N\sigma_{N,k}\right)\\ &\quad = R_K - R_F\end{aligned}$$

and then (Elton et al. 2013):

$$\lambda\left(\alpha_1\sigma_{1,k} + \alpha_2\sigma_{2,k} + \alpha_3\sigma_{3,k} + \cdots + \alpha_k\sigma_k^2 + \cdots + \alpha_{N-1}\sigma_{N-1,k} + \alpha_N\sigma_{N,k}\right) = R_K - R_F \tag{6.20}$$

Therefore, when the derivative of function θ was created with regards to each asset in the portfolio and each equation was set equal to zero, a set of simultaneous equations of the form indicated in Eq. (6.20) was created.

For each asset this equation must be true. Therefore the equation number in the system is equal to the number of assets.

If investors are characterized by homogeneous expectations, each of them define the same optimum portfolio. In conditions of equilibrium, the weight of each asset in this portfolio is equal to its weight in the market. Therefore, the portfolio reproduces the market composition with regards to the assets and their weight (Elton et al. 2013).

Considering that α_i' is the weight of i-asset in the market, the market expected return is equal to:

$$R_M = \sum_{i=1}^{N} \alpha_i' R_i \tag{6.21}$$

And considering that the covariance between the expected return of the market and the expected return of the k-th asset is equal to:

$$\begin{aligned}\sigma_{K,M} &= E\left[\left(R_K - \overline{R_K}\right)\left(R_M - \overline{R_M}\right)\right] = \left[\left(R_K - \overline{R_K}\right)\left(\sum_{i=1}^{N}\alpha_i' R_i - \sum_{i=1}^{N}\alpha_i'\overline{R}_i\right)\right]\\ &= E\left[\left(R_K - \overline{R_K}\right)\left(\sum_{i=1}^{N}\alpha_i'\left(R_i - \overline{R}_i\right)\right)\right]\end{aligned} \tag{6.22}$$

where $\overline{R_K} \equiv \mu_K$ and $\overline{R_M} \equiv \mu_M$, and solving:

$$\begin{aligned}\sigma_{K,M} &= E\left[\alpha_1'\left(R_K-\overline{R_K}\right)\left(R_1-\overline{R_1}\right)+\alpha_2'\left(R_K-\overline{R_K}\right)\left(R_2-\overline{R_2}\right)+\cdots\right.\\&\quad\left.+\alpha_k'\left(R_K-\overline{R_K}\right)\left(R_k-\overline{R_K}\right)+\cdots+\alpha_N'\left(R_K-\overline{R_K}\right)\left(R_N-\overline{R_N}\right)\right]\\&=\alpha_1'E\left[\left(R_K-\overline{R_K}\right)\left(R_1-\overline{R_1}\right)\right]+\alpha_2'E\left[\left(R_K-\overline{R_K}\right)\left(R_2-\overline{R_2}\right)\right]+\cdots\\&\quad+\alpha_k'E\left[\left(R_K-\overline{R_K}\right)^2+\cdots+\alpha_N'E\left[\left(R_K-\overline{R_K}\right)\left(R_N-\overline{R_N}\right)\right]\right]\end{aligned}$$

and then:

$$\sigma_{K,M} = \alpha_1'\sigma_{1,K}+\alpha_2'\sigma_{2,K}+\cdots+\alpha_{N-1}'\sigma_{N-1,K}+\alpha_N'\sigma_{N,K} \tag{6.23}$$

The left side member of the equation is equal to:

$$\lambda\left(\alpha_1\sigma_{1,K}+\alpha_2\sigma_{2,K}+\alpha_3\sigma_{3,K}+\cdots+\alpha_K\sigma_K^2+\cdots+\alpha_{N-1}\sigma_{N-1,K}+\alpha_N\sigma_{N,K}\right)=\lambda\sigma_{K,M}$$

And therefore Eq. (6.23) can be re-written as follows (Elton et al. 2013):

$$\lambda\sigma_{K,M} = \overline{R_K} - R_F \tag{6.24}$$

This equation must be true for each k-th asset. It implies that the equation must be true for each portfolio and therefore also for the market portfolio (Elton et al. 2013).

Particularly for the market portfolio the relationship is the following:

$$\sigma_{m,m} = \sigma_M^2 \rightarrow \lambda\sigma_M^2 = \overline{R_M} - R_F$$

and therefore:

$$\lambda = \frac{\overline{R_M} - R_F}{\sigma_M^2}$$

By substituting λ, the Eq. (6.24) can be re-written as follows:

$$\frac{\overline{R_M} - R_F}{\sigma_M^2}\sigma_{K,M} = \overline{R_K} - R_F$$

and solving by R_k:

$$\overline{R_K} = R_F + \frac{\overline{R_M} - R_F}{\sigma_M^2}\sigma_{K,M} \tag{6.25}$$

or in equal form:

$$\overline{R_K} = R_F + \left(\overline{R_M} - R_F\right)\frac{\sigma_{K,M}}{\sigma_M^2} \tag{6.26}$$

Considering that:

$$\beta_k = \frac{\sigma_{K,M}}{\sigma_M^2} \tag{6.27}$$

Equation (6.26) can be re-written as follows:

$$\overline{R_K} = R_F + \beta_k\left(\overline{R_M} - R_F\right) \tag{6.28}$$

Equation (6.28) is the equation of the *Security Market Line* (*SML*).

Therefore, the intuitive and rigorous construction of the CAPM achieve the same result.

6.4 CAPM in Terms of Prices

The CAPM is the equilibrium model of the capital markets in terms of prices and also in terms of expected returns.

The asset's return is function of its price over time. Therefore, it is possible to define the CAPM in terms of prices. Consequently the CAPM is useful to describe the equilibrium condition in the capital markets in terms of both expected returns and prices (Elton et al. 2013).

It is relevant in many situations as the evaluation or definition of the emission price of a new asset for example. The derivation of the equilibrium condition based on prices starts with the equilibrium condition based on expected returns.

Using P_i to denote the current price of the asset and Y_i the expected price of the asset in the future (next year) it also includes the dividends that will be paid and therefore measures its value in monetary terms. The expected return of the i-th asset (R_i) is equal to (Elton et al. 2013):

$$R_i = \frac{Y_i - P_i}{P_i} = \frac{Y_i}{P_i} - \frac{P_i}{P_i} = \frac{Y_i}{P_i} - 1 \tag{6.29}$$

Similarly, by using P_M to denote the current market price and Y_M the expected price of the market portfolio in the future (next year) also including the dividends that will be paid and therefore measures its value in monetary terms, the expected return of the market (R_M) is equal to (Elton et al. 2013):

$$R_M = \frac{Y_M - P_M}{P_M} = \frac{Y_M}{P_M} - \frac{P_M}{P_M} = \frac{Y_M}{P_M} - 1 \quad (6.30)$$

Based on Eq. (6.26), the expected return of the i-th asset (R_i) is equal to:

$$R_i = R_F + \left(\frac{\sigma_{i,M}}{\sigma_M^2}\right)(R_M - R_F) \quad (6.31)$$

substituting R_i and R_M with their formalization respectively, and explaining it with regards to the expected returns of i-th asset and portfolio market the covariance $(\sigma_{(R_i;R_M)})$ and variance $(\sigma^2_{(R_M)})$, the equation can be rewritten as follows:

$$\frac{Y_i}{P_i} - 1 = R_F + \left(\frac{\sigma_{(R_i;R_M)}}{\sigma^2_{(R_M)}}\right)\left(\frac{Y_M}{P_M} - 1 - R_F\right) \quad (6.32)$$

The covariance between the expected returns of the i-th asset and the expected returns of the market portfolio $(\sigma_{(R_i;R_M)})$, (where $\overline{R_i} \equiv \mu_i$ and $\overline{R_M} \equiv \mu_M$), can be re-written as follows (Elton et al. 2013):

$$\begin{aligned}
\sigma_{(R_i;R_M)} &= E\left[\left(R_i - \overline{R}_i\right)\left(R_M - \overline{R_M}\right)\right] = E\left[\left(\frac{Y_i - P_i}{P_i} - \frac{\overline{Y}_i - P_i}{P_i}\right)\left(\frac{Y_M - P_M}{P_M} - \frac{\overline{Y}_M - P_M}{P_M}\right)\right] \\
&= E\left[\left(\frac{Y_i}{P_i} - 1 - \left(\frac{\overline{Y}_i}{P_i} - 1\right)\right)\left(\frac{Y_M}{P_M} - 1 - \left(\frac{\overline{Y_M}}{P_M} - 1\right)\right)\right] = E\left[\left(\frac{Y_i}{P_i} - \frac{\overline{Y}_i}{P_i}\right)\left(\frac{Y_M}{P_M} - \frac{\overline{Y_M}}{P_M}\right)\right] \\
&= E\left[\left(\frac{Y_i - \overline{Y}_i}{P_i}\right)\left(\frac{Y_M - \overline{Y_M}}{P_M}\right)\right] = E\left[\frac{1}{P_i}\left(Y_i - \overline{Y}_i\right)\frac{1}{P_M}\left(Y_M - \overline{Y_M}\right)\right] \\
&= \frac{1}{P_i}\frac{1}{P_M}E\left[\left(Y_i - \overline{Y}_i\right)\left(Y_M - \overline{Y_M}\right)\right]
\end{aligned}$$

and then:

$$\sigma_{(R_i;R_M)} = \frac{1}{P_i}\frac{1}{P_M}\sigma_{(Y_i;Y_M)} \quad (6.33)$$

where $\sigma_{(Y_i;Y_M)}$ is the covariance between the future price of the i-th asset and the market portfolio.

Similarly, the variance of the market portfolio expected returns $\left(\sigma^2_{(R_M)}\right)$ can be re-written as follows (Elton et al. 2013):

$$\sigma^2_{(R_M)} = E\left[\left(R_M - \overline{R_M}\right)^2\right] = E\left[\left(\frac{Y_M - P_M}{P_M} - \frac{\overline{Y_M} - P_M}{P_M}\right)^2\right] = E\left[\left(\frac{Y_M}{P_M} - 1 - \left(\frac{\overline{Y_M}}{P_M} - 1\right)\right)^2\right]$$
$$= E\left[\left(\frac{Y_M}{P_M} - \frac{\overline{Y_M}}{P_M}\right)^2\right] = E\left[\left(\frac{1}{P_M}\right)^2 \left(Y_M - \overline{Y_M}\right)^2\right] = E\left[\frac{1}{P_M^2}\left(Y_M - \overline{Y_M}\right)^2\right]$$

and then:

$$\sigma^2_{(R_M)} = \frac{1}{P_M^2}\sigma^2_{(Y_M)} \tag{6.34}$$

where $\sigma^2_{Y_M}$ is the variance of the future price of the market portfolio.

On the basis of Eqs. (6.33) and (6.34), Eq. (6.32) can be re-written as follows:

$$\frac{\overline{Y}_i}{P_i} - 1 = R_F + \left(\frac{\overline{Y_M}}{P_M} - 1 - R_F\right)\frac{\frac{1}{P_i}\frac{1}{P_M}\sigma_{(Y_i;Y_M)}}{\frac{1}{P_M^2}\sigma^2_{(Y_M)}} \tag{6.35}$$

Adding 1 to each term, we have:

$$\frac{\overline{Y}_i}{P_i} - 1 + 1 = R_F + \left(\frac{\overline{Y_M}}{P_M} - 1 - R_F\right)\frac{\frac{1}{P_i}\frac{1}{P_M}\sigma_{(Y_i;Y_M)}}{\frac{1}{P_M^2}\sigma^2_{(Y_M)}} + 1$$

defining $r_F = 1 + R_F$ and substituting, we have:

$$\frac{\overline{Y}_i}{P_i} = r_F + \left(\frac{\overline{Y_M}}{P_M} - r_F\right)\frac{\frac{1}{P_i}\frac{1}{P_M}\sigma_{(Y_i;Y_M)}}{\frac{1}{P_M^2}\sigma^2_{(Y_M)}}$$

Multiplying each term for P_i, we have:

$$P_i\frac{\overline{Y}_i}{P_i} = \left[r_F + \left(\frac{\overline{Y_M}}{P_M} - r_F\right)\frac{\frac{1}{P_i}\frac{1}{P_M}\sigma_{(Y_i;Y_M)}}{\frac{1}{P_M^2}\sigma^2_{(Y_M)}}\right]P_i$$

and then:

$$\overline{Y}_i = r_F P_i + \left(\frac{\overline{Y_M} - r_F P_M}{P_M}\right)\frac{\frac{1}{P_i}\frac{1}{P_M}\sigma_{(Y_i;Y_M)}P_i}{\frac{1}{P_M^2}\sigma^2_{(Y_M)}} = r_F P_i + \left(\frac{\overline{Y_M} - r_F P_M}{P_M}\right)\frac{\frac{1}{P_M}\sigma_{(Y_i;Y_M)}}{\frac{1}{P_M^2}\sigma^2_{(Y_M)}}$$
$$= r_F P_i + \left(\frac{\overline{Y_M} - r_F P_M}{P_M}\right)\frac{\sigma_{(Y_i;Y_M)}}{\frac{1}{P_M}\sigma^2_{(Y_M)}} = r_F P_i + \left(\frac{\overline{Y_M} - r_F P_M}{P_M}\right)P_M\frac{\sigma_{(Y_i;Y_M)}}{\sigma^2_{(Y_M)}}$$
$$= r_F P_i + \left(\overline{Y_M} - r_F P_M\right)\frac{\sigma_{(Y_i;Y_M)}}{\sigma^2_{(Y_M)}}$$

and by solving for P_i, we have (Elton et al. 2013):

$$P_i = \frac{1}{r_F}\left[\overline{Y}_i - \left(\overline{Y_M} - r_F P_M\right)\frac{\sigma_{(Y_i;Y_M)}}{\sigma^2_{(Y_M)}}\right] \tag{6.36}$$

Equation (6.36) shows how the current price of the i-th asset (P_i) is function of its expected price and therefore of its expected value in monetary terms $\left(\overline{Y}_i\right)$ less an amount for risk $\left(\left(\overline{Y_M} - r_F P_M\right)\frac{\sigma_{(Y_i;Y_M)}}{\sigma^2_{(Y_M)}}\right)$; this difference is achieved by using r_F and then $\left(\frac{1}{r_F}\right)$. r_F can be used in order to estimate the present value because the term $\left[\overline{Y}_i - \left(\overline{Y_M} - r_F P_M\right)\frac{\sigma_{(Y_i;Y_M)}}{\sigma^2_{(Y_M)}}\right]$ can be considered as the certain equivalent of the future value (Elton et al. 2013).

Equation (6.36) can be written in a different form as follows:

$$P_i = \frac{1}{r_F}\left[\overline{Y}_i - \left(\frac{\overline{Y_M} - r_F P_M}{\sigma_{(Y_M)}}\right)\frac{\sigma_{(Y_i;Y_M)}}{\sigma_{(Y_M)}}\right] \tag{6.37}$$

where:

- $\frac{\overline{Y_M} - r_F P_M}{\sigma_{(Y_M)}}$: is the "market price for the risk";
- $\frac{\sigma_{(Y_i;Y_M)}}{\sigma_{(Y_M)}}$: is the measure of risk for any i-th asset.

References

Alexander G (1977) An algorithmic approach to deriving the minimum-variance zero-beta portfolio. J Financ Econ 4(2):231–236

Arzac E, Bawa V (1977) Portfolio choice and equilibrium in capital markets with safety-first investors. J Financ Econ 4(3):277–288

Back KE (2017) Asset pricing and portfolio choice theory, 2nd edn. Oxford University Press, Oxford

Bawa V, Lindenburg E (1977) Capital market equilibrium in a mean-lower partial moment framework. J Financ Econ 5:189–200

Benninga S, Protopapadakis A (1991) The stock market premium, production and relative risk aversion. Am Econ Rev 81(3):591–599

Bernstein PL (1973) What rate of return can you "reasonably" expect? J Financ XXVIII(2):272–282

Black F (1972) Capital market equilibrium with restricted borrowing. J Bus 45(3):444–455

Borch K (1969) Equilibrium, optimum and prejustices in capital markets. J Financ Quant Anal IV (1):4–14

Breeden D (1979) An intertemporal asset pricing model with stochastic consumption and investment opportunities. J Financ Econ 7(3):265–296

Breeden D, Litzenberger R (1978) Prices of state-contingent claims implicit in option prices. J Bus 51:621–651
Brennan MJ (1971) Capital market equilibrium with divergent borrowing and lending rates. J Financ Quant Anal VI(5):1197–1205
Brenner M, Subrahmanyam M (1978) Portfolio selection in an economy with marketability and short sales restrictions. J Financ XXXIII(2):589–601
Chamberlain G, Rothschild M (1983) Arbitrage, factor structure, and mean-variance analysis on large asset markets, Econometrica 51:1281–1304
Chen N, Roll R, Ross S (1986) Economic forces and the stock market. J Bus 59:386–403
Cochrane J (2001) Asset pricing. Princeton University Press, Princeton, NJ
Connor G (1984) A unified beta pricing theory. J Econ Theor 34:13–31
Constantinides GM (1980) Admissible uncertainty in the intertemporal asset pricing model. J Financ Econ 8(1):71–87
Duffi D (2001) Dynamic asset pricing theory, 3rd edn. Princeton University Press, Princeton, NJ
Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN (2013) Modern portfolio theory and investment analysis, 9th edn. Wiley, New Jersey
Fama E (1968) Risk, return and equilibrium: some clarifying comments. J Financ XXIII(1):29–40
Fama E (1970) Multi-period consumption investment decision. Am Econ Rev 60:163–164
Fama E (1971) Risk, return and equilibrium. J Polit Econ 79(1):30–55
Fama E (1976) Foundations of finance. Basic Books, New York
Fama E, MacBeth J, Schwert G (1977) Asset returns and inflation. J Financ Econ 5:115–146
Fama E, MacBeth J, Schwert G (1979) Inflation, interest and relative prices. J Bus 52:183–209
Fama E (1998) Determining the number of priced state variable in the ICAPM. J Financ Quant Anal 33(2):271–231
Friend I, Westerfield R (1980) Co-skewness and capital assets pricing. J Financ 35(4):897–914
Grossman S, Shiller R (1982) Consumption correlatedness and risk measurement in economies with non-traded assets and heterogeneous information. J Financ Econ 10:195–210
Grossman S, Melino A, Shiller R (1985) Estimating the continuous-time consumption-based asset-pricing model. J Bus Econ Stat 5:315–328
Guiso L, Jappelli T, Terlizzese D (1996) Income risk, borrowing constraints, and portfolio choice. Am Econ Rev 86(1):158–172
Hagerman R, Kim H (1976) Capital asset pricing with price level change. J Financ Quant Anal XI (3):381–391
Hansen L, Singleton K (1982) Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50:1269–1286
Heckerman D (1972) Portfolio selection and the structure of capital asset prices when relative prices of consumption goods may change. J Financ XXVII(1):47–60
Hilliard JE (1980) Asset pricing under a subset of linear risk tolerance functions and log-normal market returns. J Financ Quant Anal XV(5):1041–1062
Jarrow R (1980) Heterogeneous expectations, restrictions and short sales, and equilibrium asset prices. J Financ 35(5):1105–1114
Kraus A, Litzenberger R (1975) Market equilibrium in a multi-period state preference model with logarithmic utility. J Financ XXX(5):1213–1227
Kroll Y, Levy H (1992) Further tests of the separation theorem and the capital asset pricing model. Am Econ Rev 82(3):664–670
Landskroner Y (1977a) Nonmarketable assets and the determinants of the market price of risk. Rev Econ Stat LIX(4):482–514
Landskroner Y (1977b) Intertemporal determination of the market price of risk. J Financ XXXII (5):1671–1681
Lehari D, Levy H (1977) The capital asset pricing model and the investment horizon. Rev Econ Stat LIX(1):92–1904
Levy H (1973) The demand for assets under condition of risk. J Financ XXVIII(1):79–81
Lintner J (1965a) Security prices, risk, and maximal gains from diversification. J Financ 20 (4):587–615

Lintner J (1965b) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev Econ Stat 47(1):13–37
Lintner J (1969) The aggregation of investor's diverse judgments and preferences in purely competitive security markets. J Financ Quant Anal IV(4):347–400
Lintner J (1970) The market price of risk, size of market and investor's risk aversion. Rev Econ Stat LII(1):87–99
Lintner J (1971) The effect of short selling and margin requirements in perfect capital markets. J Financ Quant Anal VI(5):1173–1195
Litzenberg R, Ronn E (1986) A utility based model of common stock return. J Financ 41:67–92
Long J (1974) Stock prices, inflation, and the term structure of interest rates. J Financ Econ 1 (2):131–170
Lucas R (1978) Asset prices in an exchange economy. Econometrica 46:1429–1445
Mayers D (1973) Nonmarketable assets and the determination of capital asset prices in the absence of a riskless asset. J Bus 46(2):258–267
Mayers D (1976) Nonmarketable assets, market segmentation and the level of asset prices. J Financ Quant Anal XI(1):1–37
Merton R (1973) An intertemporal capital asset pricing model. Econometrica 41(5):867–888
Milne F, Smith C (1980) Capital asset pricing with proportional transaction cost. J Financ Quant Anal XV(2):253–266
Modigliani F, Pogue J (1974a) An introduction to risk and return. Financ Anal J 30(2):68–80
Modigliani F, Pogue J (1974b) An introduction to risk and return: part II. Finan Anal J 30(3):69–86
Mossin J (1966) Equilibrium in a capital asset market. Econometrica 34(4):768–783
Pettit RR, Westerfield R (1972) A model of capital asset risk. J Financ Quant Anal VII(2):1649–1668
Rabinovitch R, Owel J (1978) Non-homegeneous expectations and information in capital asset market. J Financ XXXIII(2):575–587
Ross S (1977) The capital asset pricing model (CAPM), short-sale restrictions and related issues. J Financ XXXII(1):177–183
Ross S (1978a) Mutual fund separation in financial theory—the separating distributions. J Econ Theor 17(2):254–286
Ross S (1978b) The current status of the capital asset pricing model (CAPM). J Financ XXXIII (3):885–901
Ross S (1978c) A simple approach to the valuation of risky streams. J Bus 51(3):453–475
Rubinstein M (1973) A mean-variance synthesis of corporate financial theory. J Financ XXXVIII (1):167–181
Rubinstein M (1974) An aggregation theorem for security markets. J Financ Econ 1(3):225–244
Samuelson P (1969) Lifetime portfolio selection by dynamic stochastic programming. Rev Econ Stat LI(3):239–246
Samuelson P, Merton R (1974) Generalized mean-variance tradeoffs for best perturbation corrections to approximate portfolio decisions. J Financ XXIX(1):27–40
Sharpe WF (1963) A simplified model for portfolio analysis. Manage Sci 9(2):277–293
Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. J Financ 19(3):425–442
Sharpe WF (1966) Mutual fund performance. J Bus 39(1), Part. 2: Supplement of Security Prices, 119–138
Sharpe WF (1970) Portfolio theory and capital markets. McGraw-Hill, New York
Sharpe WF (1973) Bonds versus stocks: some lessons from capital market theory. Financ Anal J 29(6):74–80
Sharpe WF (1991) Capital asset prices with and without negative holdings. J Financ 46(2):489–509
Sharpe WF (1992) Asset allocation: management style and performance measurements. J Portf Manag 18(2):7–19
Sharpe WF (1994) The sharpe ratio. J Portf Manag 21(1):49–58

Stapleton CR (1971) Portfolio analysis, stock valuation and capital budgeting decision rules for risky projects. J Financ XXVI(1):95–117

Stapleton R, Subrahmanyam M (1977) Multi-period equilibrium asset pricing model. Econometrica 46:1077–1096

Turnbull S (1977) Market value and systematic risk. J Financ XXXII(4):1125–1142

Part III
Company Valuation

Chapter 7
Capital Structure and the Cost of Capital

Abstract The cost of capital is one of the most relevant variables in the company's valuation models. It is probably one of the most relevant topics for managers and financial economists. For decades several studies have focused on the relationship between capital structure, cost of capital and company value. Despite a broad experience approach in both academic and practices, it should not be surprising that the method for estimation of the cost of capital is still under intensive discussion. In this context, starting with the Modigliani and Miller theories, whose studies are considered the starting point of the modern theory of capital structure, the cost of equity, debt and company capital are estimated.

7.1 Capital Structure Choices

The capital structure of a company refers to the capital sources invested in it. There are two main sources of capital: equity, through stock emission; debt, through the emission of bonds.

Equity and debt are different in their nature, level of risk and rights.

By purchasing equity, the stockholders become the owners of the company. They have the right to vote and share and future profits of the company. More specifically, equity is a source of capital characterised by:

– *heterogeneity*: it can be internal, arising from self-financing due to the retention of earnings, or external, deriving from the emission of new shares;
– *long-period*: it is stable over time in the capital structure of the company and it is intended to fund long-term company activities;
– *full riskiness*: the owners bear the income risk, function of the probability of "non-congruity" of remuneration and capital risk, function of the probability to lose capital invested in the case of default;
– *variable remuneration*: the dividends for stockholders are uncertain in their amount and time of achievement;

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_7

– *indirect refund*: the stakeholders achieve capital gain/loss by selling the stocks in the capital markets based on the difference between the sales prices and the buying price.

On the other hand, by purchasing the debt the bondholders become company creditors. They do not share in the profit of the company but they receive the main plus defined interest. More specifically, the debt is a capital source characterised by:

– *homogeneity*: it is an external source of capital;
– *short and long period*: it can be a long-term debt, and therefore it is a stable source of capital in the capital structure, or a short-term debt;
– *limited riskiness*: the owners bear the default risk only. The obligations on debts must be paid regardless of the company's income performance. Also, in the case of bankruptcy, bondholders must be paid before stockholders;
– *fixed remuneration*: the interest on debt for bondholders is certain in the amount and in time of achievement;
– *direct refund*: the company must reimburse the principal at maturity.

Based on these main characteristics, stocks are riskier than bonds. Consequently, the expected return on equity must be higher than the expected return on debt.

One of the most relevant problems of capital structure, includes the effects of its choices on company value. The main question refers to both positive and negative debt effects on the value of the company.

The relationship between capital structure choices and company value has confused financial economists and managements for decades. Despite the vast theoretical models and decades of empirical tests, the problem still exists and it is feasible to say that the capital structure choices of companies are still "*puzzle*" (Myers 1984). The question continued to involve the financial and managerial economists, with regards to the techniques of company financing and capital budgeting problems and the economic theorists, with regards to the explanation of interest, savings and investment behaviours on both the micro and macro levels.

The starting point of the modern capital structure theory, is usually considered the theory of Modigliani and Miller (MM). It was developed over time on the basis of five main papers:

(1) the first and the most popular paper is "*The Cost of Capital, Corporate Finance and the Theory of Investment*" (1958), in which MM define the well-known Propositions I and II;
(2) the second is the "*Dividend Policy, Growth, and the Valuation of Shares*" (1961), in which MM try to complete and refine the analysis of the Propositions;
(3) the third is a "*Corporate Income Taxes and the Cost of Capital: a Correction*" (1963), that is "*a correction*" of the previous Paper (1958), in which MM defines new conclusions on the Propositions I and II after the introduction of corporate taxes;

(4) the fourth is the "*Debt and Taxes*" (1977) by Miller only, in which he introduces personal taxes in addition to corporate taxes;
(5) the fifth is the "*Debt, Dividend Policy, Taxes, Inflation and Market Valuation*" (1982) by Modigliani only, in which he revises the 1963 version of the Propositions.

The Propositions I and II are the core of their theory. The basic idea of MM is that company value is function of its business operating activities only, and the capital structure defines the way in which this value is distributed between the investors in equity and debt. Thus, the capital structure choices are irrelevant both with regards to the company value and its capital cost (Propositions I and II).

Note that the Propositions of MM comply with the law of preservation of investment value: in the absence of corporate taxes, the company value is function of operating cash-flow only; the way in which these cash-flows are distributed among different investors is irrelevant in the perspective of company value. Therefore, the company's assets cannot be influenced by the composition of capital structure in terms of equity and debt amounts. Based on this argument, the cost of capital is generated. If the value of the assets is independent of the capital structure choices, the discount rate of the operating cash-flows cannot be influenced by the relationship between equity and debt.

The Propositions I and II (MM 1958), are based on strong and restrictive assumptions. To derive these two Propositions, it is useful to follow strictly MM's argumentations sometimes by using the same words and symbols used in their papers. The main basic assumptions can be schematically summarized as follows:

(1) assume an economy in which all physical assets are owned by companies;
(2) assume that these companies finance their assets by issuing common stocks only. Assets will provide stockholders with a flow of profit over time. However, this flow of profit will not be constant and it is uncertain. At a later stage, this assumption is eliminated and is replaced with assumption 3;
(3) assume that the company can be financed by equity and debt. Specifically, assume that: (i) all bonds yield a constant income per unit of time, and this income is regarded as certain by all traders regardless of the issuer. Therefore, all bonds are perfect substitutes up to a scale factor; (ii) all bonds, as well as stocks, are traded in a perfect capital market. Therefore, they must all sell at the same price per dollar's worth of return, or equivalently, they must yield the same rate of return. This rate of return referred to as the rate of interest, or equivalently, as the capitalization rate for sure streams;
(4) assume that this stream of income (and hence the stream of income accruing to any share of common stock) can be extended indefinitely into the future but its mean value over time (the average profit per unit of time) is infinite and represents a random variable subject to a subjective probability distribution. Therefore: (i) the *return of the share* is defined equal to the average value over time of the stream of income accruing to a given share; (ii) the *expected return of the share* is defined equal to the mathematical expectation of this average value. Therefore the uncertainty connects with the mean value over time of the

stream of income and it must not confused with the variability over time of the successive elements of the stream;

(5) assume that the investors agree on the expected return of any share despite the fact that they may have a different probability distribution of the return of any share;

(6) assume that firms can be divided into "equivalent return classes": the return on the shares issued by any company in the same class is proportional to the return on the shares issued by any other company in the same class; the returns on shares as perfectly correlated between them. Therefore, the shares within the same class differ between them by one scale factor at the most. Consequently, if we consider the ratio between the return and the expected return, the probability distribution of the ratio is the same for all shares within the same class. Consequently, there are only two relevant properties of a share: (i) the reference class; (ii) its expected return;

This assumption is probably the most relevant of the model, because it allows for classification of the companies within which the shares of different companies are homogeneous, and therefore perfect substitutes among them. In other words, the assumption allows for the creation of homogeneous classes of stock;

(7) assume that the shares are traded in the perfect capital markets under conditions of atomistic competition;

(8) assume the opportunity for an investor to borrow at the same conditions as the company. Therefore, the investor can reproduce the same company leverage in its portfolio by borrowing on a personal level.

Based on these basic assumptions, in conditions of equilibrium, the price of every share in any given class must be proportional to its expected return. This "*proportionality factor*" for any class is the same for all companies within the class.

Using p_j to denote the price per share of the j-th company in the k-th class and with x_j the expected return per share of the j-th firm in the k-th class, we have:

$$p_j = \frac{1}{p_k} x_j \rightarrow p_k = \frac{x_i}{p_j} \tag{7.1}$$

where $1/p_k$ is the proportionality factor for any k-th class. Specifically, p_k is a constant for all j-firms in the k-class and it is one for each of the k-classes; then, it is the same for all companies in the same k-th class.

Therefore in the same class, the price per share of the j-th company (p_j) is proportional to its expected return (x_j) based on the proportionality factor $(1/p_k)$ that is the same for all companies within the same k-th class.

Based on the equation, p_k it can be interpreted as the price that an investor has to pay for a dollar's worth of expected return in the k-th class. In equivalent terms, p_k can be regarded as the market rate of capitalization for the expected value of uncertain streams of the kind generated by the k-th class of companies. Consequently, p_k can be interpreted as the expected rate of return of any share in the k-th class.

Having defined the concept of homogeneous classes of shares, and therefore having defined the instruments to face the uncertain stream, it is possible to remove assumption n.2 by replacing it with assumption n.3. Therefore, assume that the company can be financed by equity and debt. The introduction of debt modifies market shares. Companies may have a different capital structure and therefore a different debt level. Therefore, shares of different companies can give rise to different probability return distributions, even if they are in the same class. The shares will be subject to different degrees of leverage and therefore financial risk. Consequently, they will no longer be perfect substitutes among them.

Based on these assumptions and considerations, MM (1958) have derived the two Propositions.

Proposition I: *the market value of any company is independent of its capital structure and is created by capitalizing its expected return at the rate* (p_k) *appropriate to its class.*

Equivalently, the average cost of capital to any company is completely independent of its capital structure and is equal to the capitalization rate of a pure equity stream of its class.

In conditions of equilibrium, the company levered value (W_L) is equal to the company unlevered value (W_U), equal to the market value of the company (W_j), as follows:

$$W_L = W_U = W_j \tag{7.2}$$

and by denoting with X_j the expected return of the j-company's assets and therefore the expected return of income; D_j the market value of the debt of the j-company; E_j is the marker value of the equity of the j-company; p_k is the expected rate of return of the stock of the k-th class of the j-company referenced that is constant for all companies within the same k-th class and it is one for each k-th class, we have:

$$W_j \equiv (E_j + D_j) = \frac{X_j}{p_k} \rightarrow p_k = \frac{X_j}{(E_j + D_j)} \equiv \frac{X_j}{W_j} \tag{7.3}$$

where X_j/W_j is the "*average cost of capital*". It is constant for all firms in the same k-th class, and therefore it is independent from the capital structure of the company.

The Proposition I can be proved by considering two companies:

- Company 1: is financed by equity only, and therefore it is defined as an Unlevered Firm (F_U);
- Company 2: is financed by equity and debt, and therefore it is defined as a Levered Firm (F_L).

Assume that these two companies are in the same k-class. Therefore, the expected returns of Company 1 (X_1) and Company 2 (X_2) are equal among them: the two companies in the same k-th class have the same expected return (X) as follows:

$$X_1 \equiv X_2 = X \tag{7.4}$$

Now it is possible to analyse two main cases:

– (*Case 1*) assume that the value of the Company 2 (W_2) is higher than the value of Company 1 (W_1), and therefore the value of the levered company (W_L) is higher than the value of the unlevered company (W_U). This first case can be formalized as follows:

$$W_2 > W_1 \leftrightarrow W_L > W_U$$

– (*Case 1)* assume that the value of Company 1 (W_1) is higher than the value of Company 2 (W_2), and therefore the value of the unlevered company (W_U) is higher than the value of the levered company (W_L). This second case can be formalized as follows:

$$W_1 > W_2 \leftrightarrow W_U > W_L$$

(*Case 1*)
Assume the following basic relationship between the two companies:

$$W_2 \equiv W_L > W_1 \equiv W_U \tag{7.5}$$

Assume that the investor holds an amount (e_2) of shares of Company 2 (Levered Company). This amount represents a fraction (α) of the total equity (E_2) of Company 2 as follows:

$$e_2 = \alpha E_2 \rightarrow \alpha = \frac{e_2}{E_2} \tag{7.6}$$

Therefore, the return on investment and therefore the return of the investor's portfolio (Y_2) is a fraction (α) of the income of Company 2 available for the stockholders that is equal to the total return (X) less the interest charge that is equal to rate return of debt (r) multiplied by the debt (D) in the capital structure, as follows:

$$Y_2 = \alpha(X - rD) \tag{7.7}$$

Assume that the investor decides to sell his share of Company 2 $(e_2 = \alpha E_2)$ and acquire an amount of the equity of Company 1 $\left(\alpha_1 = \frac{e_1}{E_1}\right)$. To acquire a portion of the equity of Company 1, the investor uses the amount derived from the sale of shares of Company 2 (αE_2) plus a personal debt by using new shares in Company 1 as a collateral (αD_2). In this context the assumption refers to the opportunity for the investor to borrow at the same conditions as the company plays a central role.

Based on this, the investor can reproduce the same company leverage in its portfolio by borrowing on a personal level.

Therefore, the investor acquires a fraction of the equity of Company 1 and its income, as follows:

$$\frac{e_1}{E_1} = \frac{\alpha E_2 + \alpha D_2}{E_1} = \frac{\alpha(E_2 + D_2)}{E_1} \tag{7.8}$$

where $\alpha(E_2 + D_2)$ is the amount invested to acquire a fraction of the equity of Company 1.

By considering the interest payments on personal debt (αD_2), the return of the investor's portfolio is equal to:

$$Y_1 = \frac{\alpha(E_2 + D_2)}{E_1} X - r(\alpha D_2) \tag{7.9}$$

Company 1 is financed by equity only. Subsequently, the equity value (E_1) is equal to the company value (W_1) that it is equal to the company unlevered value (W_U) as follows:

$$E_1 = W_1 \equiv W_U \tag{7.10}$$

Company 2 is financed by debt and equity. Subsequently, the sum of equity value (E_2) and debt value (D_2) is equal to the company value (W_2) that it is equal to the company levered value (W_L) as follows:

$$E_2 + D_2 = W_2 \equiv W_L \tag{7.11}$$

On the basis of Eqs. (7.10) and (7.11), Eq. (7.9) can be re-written as follows:

$$Y_1 = \alpha \frac{W_2}{W_1} X - r\alpha D_2 = \alpha \left(\frac{W_2}{W_1} X - rD_2 \right) \tag{7.12}$$

The Case 1 basic assumption is that the value of Company 2 (W_2) is higher than the value of Company 1 (W_1). Consequently, the ratio between the value of Company 2 and the value of Company 1 is positive as follows:

$$W_2 > W_1 \rightarrow \frac{W_2}{W_1} > 0 \tag{7.13}$$

By looking at the returns of Portfolio 1 (Y_1) and Portfolio 2 (Y_2), the relationship is the following:

$$Y_1 = \alpha\left(\frac{W_2}{W_1}X - rD_2\right) > Y_2 = \alpha(X - rD) \tag{7.14}$$

Therefore, as long as the value of Company 2 (Levered Value) is higher than the value of Company 1 (Unlevered Value), the return of Portfolio 1 (Y_1) is higher than the return of Portfolio 2 (Y_2) as follows:

$$W_2 \equiv W_L > W_1 \equiv W_U \to Y_1 > Y_2 \tag{7.15}$$

In this situation, there are conditions for arbitrage: the investor sells the shares of Company 2 (with subsequent depreciation of E_2 and therefore W_2) and acquires shares of Company 1 (with subsequent increase of E_1 and therefore W_1). This movement continues until the value of the two firms are aligned in a new equilibrium condition that will be equal:

$$W_2 \equiv W_L = W_1 \equiv W_U \tag{7.16}$$

It is worth noting, that the levered company (Company 2) cannot command a premium over an unlevered company (Company 1) because investors can replace the company leverage (of Company 2) into their portfolio directly by borrowing on a personal account. Indeed, investors and companies have the opportunity to borrow at the same conditions.

(*Case 2*)
Assume the following baseline relationship between the two companies:

$$W_2 \equiv W_L < W_1 \equiv W_U \tag{7.17}$$

Assume that the investor initially owns an amount (e_1) of shares of Company 1 (Unlevered Company). It is a fraction (α) of the total equity (E_1) of Company 1, as follows:

$$e_1 = \alpha E_1 \to \alpha = \frac{e_1}{E_1} \tag{7.18}$$

The investor's return based on the Portfolio 1 (Y_1) is equal to:

$$Y_1 = \alpha X \to Y_1 = \frac{e_1}{E_1}X \tag{7.19}$$

Assume that the investor decides to exchange his portfolio with another one of the same value equal to e_1 and consisting of e_2 dollars of shares of Company 2 and of d dollars of bonds. In this context, e_2 and d are given by:

$$e_2 = \frac{E_2}{W_2} e_1; \quad d = \frac{D_2}{W_2} e_1 \tag{7.20}$$

Therefore, the investor's wealth (e_1) is invested in a portfolio consisting of shares of Company 2 (e_2) and bonds (d) proportionally respectively to the value of equity on the total levered value (E_2/W_2), and the value of debt on the total levered value (D_2/W_2) of Company 2.

In this case:

- the return on shares (R_S) will be a fraction (α) (equal to the shares owned by the investor (e_2) on the total value of the equity of Company 2 (E_2) so that: $\alpha = e_2/E_2$) of the total return of the stockholders of Company 2 (equal to the total return of assets (X) less the rate return (r) on debt (D_2) of Company 2 so that: $X - rD_2$) as follows:

$$R_S = \frac{e_2}{E_2}(X - rD_2) \tag{7.21}$$

- the return from the bonds (R_B) will be equal to the rate return (r) on bonds (d) as follows:

$$R_B = rd \tag{7.22}$$

Therefore, the total return of the Portfolio 2 (Y_2) is equal to:

$$Y_2 = \frac{e_2}{E_2}(X - rD_2) + rd \tag{7.23}$$

Substituting e_2 and d with their expressions as defined in Eq. (7.20) in Eq. (7.23) we have

$$\begin{aligned} Y_2 &= \frac{\frac{E_2}{W_2} e_1}{E_2}(X - rD_2) + r\frac{D_2}{W_2} e_1 = \frac{E_2}{W_2 E_2} e_1 (X - rD_2) + r\frac{D_2}{W_2} e_1 = \frac{e_1}{W_2}(X - rD_2) + r\frac{D_2}{W_2} e_1 \\ &= \frac{e_1}{W_2} X - \frac{e_1}{W_2} rD_2 + r\frac{D_2}{W_2} e_1 = \frac{e_1}{W_2} X - \frac{1}{W_2} e_1 rD_2 + \frac{1}{W_2} e_1 rD_2 = \frac{e_1}{W_2} X \end{aligned}$$

Remembering that $e_1 = \alpha E_1$, we have:

$$Y_2 = \frac{\alpha E_1}{W_2} X = \alpha \frac{E_1}{W_2} X \tag{7.24}$$

Company 1 is unlevered. Therefore, the equity value (E_1) is equal to the company value (W_1), so that:

$$Y_2 = \alpha \frac{W_1}{W_2} X \tag{7.25}$$

The Case 2 basic assumption is that the value of Company 1 (W_1) is higher than the value of Company 2 (W_2). Then, the ratio between the value of Company 2 and the value of Company 1 is positive as follows:

$$W_1 > W_2 \rightarrow \frac{W_1}{W_2} > 0 \tag{7.26}$$

By looking at the returns of Portfolio 1 (Y_1) and Portfolio 2 (Y_2), the relationship is the following:

$$Y_2 = \alpha \frac{W_1}{W_2} X > Y_1 = \alpha X \tag{7.27}$$

Therefore, by assuming that the value of Company 1 (Company Unlevered Value) is higher than the value of Company 2 (Company Levered Value), the portfolio's return of Company 2 is higher than the portfolio's return of Company 1 as follows:

$$W_2 \equiv W_L < W_1 \equiv W_U \rightarrow Y_2 > Y_1 \tag{7.28}$$

In this situation, there are conditions for arbitrage: the investor sells the shares of Company 1 (with subsequent depreciation of E_1 and therefore W_1) and acquires a mixed portfolio containing an appropriate fraction of the shares of Company 2 (with a consequent increase in E_2 and therefore W_2). This movement continues until the value of the two companies is aligned in new conditions of equilibrium that will be equal to:

$$W_2 \equiv W_L = W_1 \equiv W_U \tag{7.29}$$

In this Case 2, the acquisition of a mixed portfolio of stock of a levered company (Company 2) and bonds in the proportions equal to the value of equity to total levered value (E_2/W_2) and the value of debt on total levered value (D_2/W_2) of the Levered Company (Company 2) respectively, can be regarded as an operation that "undoes" the leverage, providing access to an appropriate fraction of total return of the unlevered company. This possibility of undoing the leverage prevents the value of levered companies from being consistently less than those of unlevered companies; or equivalently, it prevents the average cost of capital (X/W) from being systematically higher for a levered company than for an unlevered company in the same class.

Proposition II: *the expected yield of a stock share is equal to the appropriate capitalization rate* (p_k) *for a pure equity stream in the class, plus a premium related to financial risk equal to the debt-to-equity ratio times the spread between* p_k *and* r.

Denoting with i_j the expected rate of return (or expected yield) on the j-th company's stocks (the after-tax yield on equity capital) belonging to the k-th class; p_k the expected rate of return of pure equity for the k-th class in which the j-company is included (it is constant for any company in the k-th class and it is one for each k-th class); r_D the rate return of debt; D_j the market value of the debt of the j-firm; E_j the market value of the equity of the j-company, we have:

$$i_j = p_k + (p_k - r_D)\frac{D_j}{E_j} \tag{7.30}$$

The market price of any stock is equal to the capitalization of its expected return at the variable rate i_j. Then, the capital structure does not affect the cost of capital because the return on equity is linear function of the relationship between debt and equity.

Equation (7.30) shows that for all companies in the same k-class, the relation between the return on stock and capital structure, as measured by the ratio D_j/E_j, will approximate a straight line with slope equal to $(p_k - r_D)$ and point of interception equal to p_k.

It is relevant to note that by definition the expected rate of return of the stock of the j-th company (i_j) is equal to the ratio by the net profit of the company (equal to the expected return of the assets of the company (X_j) less the rate return on debt (r_D) multiplied for the debt amount (D_j) in the capital structure), and the equity value of the company (E_j). It is the yield on equity capital, and it is equal to:

$$i_j \equiv \frac{X_j - r_D D_j}{E_j} \tag{7.31}$$

From Eq. (7.3) regards the Proposition I, we have:

$$\frac{X_j}{W_j} \equiv \frac{X_j}{(E_j + D_j)} = p_k \rightarrow X_j = p_k(E_j + D_j)$$

and substituting:

$$\begin{aligned} i_j &= \frac{p_k(E_j + D_j) - r_D D_j}{E_j} = \frac{p_k E_j + p_k D_j - r_D D_j}{E_j} = \frac{p_k E_j + D_j(p_k - r_D)}{E_j} \\ &= \frac{p_k E_j}{E_j} + \frac{D_j(p_k - r_D)}{E_j} \end{aligned}$$

and then:

$$i_j = p_k + (p_k - r)\frac{D_j}{E_j}$$

Representing Proposition II. Note that Proposition II is a direct consequence of Proposition I.

In the same paper (1958), MM introduced corporate taxes. The conclusions they reach in this paper will be reviewed and modified in the subsequent paper (1963).

MM (1958) stated that in conditions of equilibrium, the market value of companies in each class must be proportional to their expected return net of corporate taxes. These expected net returns are equal to the sum of the interest paid and expected net income for stockholders.

By introducing the corporate taxes, the expected net income (total income net of taxes generated by the company) changes.

Denoting with $\mathrm{X}_\mathrm{j}^\mathrm{t}$ the expected net income (total income net of taxes generated by the company) of the j-th company; X_j the expected income before taxes; t the average rate of income corporate tax; r_D the rate return of debt; D_j the market value of the debt of the j-company. The expected net income is equal to:

$$X_j^t \equiv \left(X_j - r_D D_j\right)(1 - t) + r_D D_j \tag{7.32}$$

The second part of Eq. (7.32) defines the expected return of bondholders, while the first part defines the expected net income of the stockholders $\left(\pi_j^t\right)$, as follows:

$$\pi_j^t \equiv \left(X_j - r_D D_j\right)(1 - t) \tag{7.33}$$

Therefore, the expected net income of the company is equal to the expected return of both stockholders and bondholders, as follows:

$$X_j^t \equiv \pi_j^t + r_D D_j \tag{7.34}$$

By using X_j^t Propositions I and II can be rewritten.

Proposition I can be modified as follows:

$$\frac{X_j}{\left(E_j + D_j\right)} \equiv \frac{X_j}{W_j} = p_k \quad becomes \Rightarrow \frac{\left(X_j - r_D D_j\right)(1 - t) + r_D D_j}{\left(E_j + D_j\right)} \equiv \frac{X_j^t}{W_j} = p_k^t \tag{7.35}$$

where p_k^t is the expected rate of net return of the k-th class in which is included the j-company. Also in this case it is constant for any companies in the k-class, and one for each k-th class.

Proposition II can be modified as follows:

$$i_j = \frac{X_j^t - r_D D_j}{E_j} = \frac{\pi_j^t + r_D D_j - r_D D_j}{E_j} = \frac{\pi_j^t}{E_j} \tag{7.36}$$

and then:

$$i_j \equiv \frac{X_j - r_D D_j}{E_j} = p_k + (p_k - r_D)\frac{D_j}{E_j} \; becomes \Rightarrow i_j \equiv \frac{\pi_j^t}{E_j} = p_k^t + \left(p_k^t - r_D\right)\frac{D_j}{E_j} \tag{7.37}$$

where i_j is the *yield on equity capital*, and p_k^t is the capitalization rate for net income (income net of corporate taxes) for the j-th company in the k-class. This definition is equivalent to the expected rate of net return of the k-th class in which is included the j-company.

Equations (7.35) and (7.37) show that the form and the structure of Propositions I and II are the same.

In their subsequent paper (1963), MM proposed "*a correction*" by modifying the conclusions which were reached in the original paper (1958). The introduction of corporate taxes modifies Propositions I and II in their structure.

With regards to Proposition I, MM affirmed that the arbitrage makes values within any class a function not only of expected net returns (after tax returns), but also of the tax rate and degree of leverage. It implies that the tax advantages of debt financing are relevant and they generate a quantitative difference between levered and unlevered value of the company.

Therefore, Proposition I can be redefined as follows:

Proposition I with Corporate Taxes: *the value of the levered company is equal to sum of the unlevered company plus the value of the tax shields due to the interest on debt deducibility.*

The distribution of net income (after tax income) is affected by leverage. To see how, denote by the random variable X the long-run average EBIT generated by current assets of the j-company within the k-th risk class.

Denoting with $\bar{X}$ the expected value of X and with Z a random variable that has the same value for all companies in the same k-th class, and therefore it is a constant, and it is a drawing from a distribution, say $f_k(Z)$.

In the same risk-class, X can be defined in the following form:

$$X = \bar{X}Z \rightarrow Z = \frac{X}{\bar{X}} \leftrightarrow f_k(Z) \tag{7.38}$$

Therefore, the random variable of the Net Return (X^t) (after-tax return), can be expressed as function of EBIT (X), the interest bill and therefore the amount of interest on debt (R), the marginal corporate income tax rate (assumed equal to the average) (t) as follows:

$$X^t = (X - R)(1 - t) + R = (1 - t)X + tR \tag{7.39}$$

Substituting X for its random variable (equal for all companies in the same k-th class), as defined in Eq. (7.38), Eq. (7.39) becomes:

$$X^t = (1 - t)\bar{X}Z + tR \tag{7.40}$$

Considering that the expected Net Return is equal to:

$$E(X^t) \equiv \bar{X}^t = (1 - t)\bar{X} + tR$$

and then:

$$(1 - t)\bar{X} = \bar{X}^t - tR$$

$$X^t = (\bar{X}^t - tR)Z + tR$$

and then Eq. (7.40) can be rewritten as follows:

$$X^t = \bar{X}^t\left(1 - \frac{tR}{\bar{X}^t}\right)Z + tR \tag{7.41}$$

Equation (7.41) shows that if the tax rate is different from zero, the shape of the distribution of Net Return (X^t) not only depends on the scale of the expected Net Income $(\bar{X}^t)$ and on distribution (Z), but also on the tax rate (t) and the degree of leverage according to interest on debt as measured by the ratio between the interests paid (R) and the expected Net Income $(\bar{X}^t)$ so that: $R/\bar{X}^t$.

From an investors' perspective, the long-run average stream of Net Returns (after-tax returns) is equal to the sum of components:

- an uncertain stream equal to $(1 - t)\bar{X}Z$: that is all of the expected net returns;
- a sure stream equal to tR: that is the extra net return (extra after tax return) due to the tax advantages of debt as function of the deducibility of interest payments on debt. It is usually defined as the tax savings on interest payments.

The first component, defines the unlevered value of the firm (W_U) as follows:

$$W_U = \frac{(1 - t)\bar{X}}{p_t} \leftrightarrow p_t = \frac{(1 - t)\bar{X}}{W_U} \tag{7.42}$$

where p_t is the rate at which the market capitalizes the expected Net Returns of an unlevered company of size $\bar{X}$ within k-th class.

The second component, defines the value of the tax shields due to the tax advantages of debt function of deducibility of interest payments on debt. Assume that the rate of interest is a constant independent of the debt amount so that:

$$r = \frac{R}{D} \rightarrow D = \frac{R}{r} \rightarrow R = Dr \tag{7.43}$$

In this case, the value the tax savings is equal to:

$$W_{TS} = \frac{tR}{r} \leftrightarrow r = \frac{tR}{W_{TS}} \tag{7.44}$$

where r is the rate at which the market capitalizes the sure streams generated by tax savings on interest payments.

Therefore, the value of a levered company of size $\bar{X}$ with a permanent level of debt (D_L) in the capital structure, is equal to:

$$W_L = \frac{(1-t)\bar{X}}{p_t} + \frac{tR}{r} = W_U + tD_L \tag{7.45}$$

It is worth noting that $r < p_t$ because the extra after tax earnings (tR) is a sure income while the expected after tax earnings $((1-t)\bar{X})$ is uncertain.

Generally, the levered value of the company (W_L) is equal to its unlevered value (W_U) plus the value of tax savings (W_{TS}) as follows:

$$W_L = W_U + W_{TS} \tag{7.46}$$

Based on redefinition of Proposition I, Proposition II can be redefined.

Proposition II with Corporate Taxes: *the cost of equity is positively correlated to the degree of leverage.*

Remember that:

$$(1-t)\bar{X} = \bar{X}^t - tR; \quad r = \frac{R}{D} \rightarrow \begin{matrix} D = \frac{R}{r} \\ R = Dr \end{matrix}$$

And by substituting in Eq. (7.45), we have:

$$W_L = \frac{(1-t)\bar{X}}{p_t} + \frac{tR}{r} = \frac{\bar{X}^t - tR}{p_t} + tD = \frac{\bar{X}^t - tDr}{p_t} + tD = \frac{\bar{X}^t}{p_t} - \frac{tDr}{p_t} + tD = \frac{\bar{X}^t}{p_t} + tD\left(1 - \frac{r}{p_t}\right)$$
$$= \frac{\bar{X}^t}{p_t} + tD\left(\frac{p_t - r}{p_t}\right)$$

And then:

$$W_L = \frac{\bar{X}^t}{p_t} + tD\left(\frac{p_t - r}{p_t}\right) \tag{7.47}$$

Solving Eq. (7.47) for $(\bar{X}^t/W_L)$, we have:

$$\frac{\bar{X}^t}{p_t} = W_L - tD\left(\frac{p_t - r}{p_t}\right) \rightarrow \bar{X}^t = W_L p_t - p_t tD\left(\frac{p_t - r}{p_t}\right) \rightarrow \frac{\bar{X}^t}{W_L} = \frac{W_L}{W_L}p_t - \frac{tD(p_t - r)}{W_L}$$

$$\frac{\bar{X}^t}{W_L} = p_t - t(p_t - r)\frac{D}{W_L} \tag{7.48}$$

Equation (7.47) shows how the after-tax yield is affected by leverage.

Now it is possible to calculate the after-tax yield on equity capital that is equal to the ratio between net profit (profit after taxes) and the value of the shares.

Subtracting D from both side of Eq. (7.47), we have:

$$W_L = \frac{\bar{X}^t}{p_t} + tD\left(\frac{p_t - r}{p_t}\right) \rightarrow W_L - D = \frac{\bar{X}^t}{p_t} + tD\left(\frac{p_t - r}{p_t}\right) - D$$

and by explicating $\bar{X}^t$ into its two components of (i) expected net profit after taxes (π_t) and (ii) the interest payments on debt $(R = rD)$, we have:

$$W_L - D = \frac{\pi_t + rD}{p_t} + tD\left(\frac{p_t - r}{p_t}\right) - D$$

and solving:

$$W_L - D = \frac{\pi_t}{p_t} + \frac{rD}{p_t} + tD - \frac{tDr}{p_t} - D;\ W_L - D = \frac{\pi_t}{p_t} + \frac{rD}{p_t}(1-t) - D(1-t);\ W_L - D = \frac{\pi_t}{p_t} + (1-t)\left(\frac{rD}{p_t} - D\right);\ W_L - D = \frac{\pi_t}{p_t} - (1-t)\left(D - \frac{rD}{p_t}\right);\ W_L - D = \frac{\pi_t}{p_t} - (1-t)\left(D\left(1 - \frac{r}{p_t}\right)\right)$$

and then:

$$W_L - D = \frac{\pi_t}{p_t} - (1-t)D\left(\frac{p_t - r}{p_t}\right) \tag{7.49}$$

Considering that the company levered value (W_L) less debt (D) is equal to the company unlevered value (W_U), we have:

$$W_U = W_L - D \rightarrow W_U = \frac{\pi_t}{p_t} - (1-t)D\left(\frac{p_t - r}{p_t}\right)$$

and by solving for π_t, we have:

$$\pi_t = W_U p_t + p_t(1-t)D\left(\frac{p_t - r}{p_t}\right)$$

and dividing both terms by W_U:

$$\frac{\pi_t}{W_U} = p_t + (1-t)(p_t - r)\frac{D}{W_U}$$

Consider that the after-tax yield on equity capital (i_E) is equal to the ratio between expected net profit after taxes (π_t) and the value of shares and therefore the unlevered value of the company (W_U), the equation con be rewritten as follows:

$$i_E = p_t + (1-t)(p_t - r)\frac{D}{W_U} \tag{7.50}$$

The equation shows an increase in the after-tax yield on equity capital (i_E) as leverage increases which is smaller than the original version of Proposition II by a factor of $(1-t)$. But again, the linear increasing relation of this equation is still fundamentally different from the original in which the cost of equity is completely independent from the leverage. In this case the cost of equity capital is dependent from leverage.

It is possible to summarize two Propositions over time as follows (Table 7.1):

At a later stage, Miller (1977) introduced personal taxes in addition to corporate tax.

By introducing personal taxes, the value of tax savings can be rewritten as follows:

Table 7.1 MM propositions

	Paper 1958		Paper 1963 "A Correction" With corporate tax
	No corporate tax	Corporate tax	
P. I	$\frac{X_j}{(E_j + D_j)} \equiv \frac{X_j}{W_j} = p_k$	$\frac{(X_j - rD_j)(1-t) + rD_j}{(E_j + D_j)} \equiv \frac{X_j^t}{W_j} = p_k^t$	$W_L = \frac{(1-t)\bar{X}}{p_t} + \frac{tR}{r} = W_U + tD_L$
P. II	$i_j \equiv \frac{X_j - rD_j}{E_j} = p_k + (p_k - r)\frac{D_j}{E_j}$	$i_j \equiv \frac{\pi_j^t}{E_j} = p_k^t + (p_k^t - r)\frac{D_j}{E_j}$	$i_E = p_t + (1-t)(p_t - r)\frac{D}{W_U}$

$$W_{TS} = D\left[1 - \frac{(1-t_c)(1-t_e)}{(1-t_d)}\right] \tag{7.51}$$

The company levered value (W_L) is equal to the company unlevered value (W_U) plus the value of tax savings (W_{TS}) as defined by Eq. (7.46). On the basis of specification of Eq. (7.51), Eq. (7.46) can be rewritten as follows:

$$W_L = W_U + W_{TS} \rightarrow W_L = W_U + D\left[1 - \frac{(1-t_c)(1-t_e)}{(1-t_d)}\right] \tag{7.52}$$

In order to demonstrate Eq. (7.52) (Ross et al. 1997), assume that company is financed by equity and debt.

Denoting with D the debt level in the capital structure; K_D the cost of debt; t_c the corporate tax rate (in average); t_e the personal tax rate on equity investor; and by calculating the operating income on the basis of EBIT, the return for stockholders (r_E) and the return for bondholders (r_D) can be defined respectively as follows:

$$r_E = (EBIT - DK_D)(1-t_c)(1-t_e) \tag{7.52}$$

$$r_D = K_D D(1-t_d) \tag{7.53}$$

On the basis of Eqs. (7.52) and (7.53) it is possible to calculate the total return for investors (r_I), as follows:

$$r_I = r_E + r_D = (EBIT - DK_D)(1-t_c)(1-t_e) + K_D D(1-t_d)$$

and then:

$$r_I = EBIT(1-t_c)(1-t_e) + DK_D(1-t_d)\left[1 - \frac{(1-t_c)(1-t_e)}{(1-t_d)}\right] \tag{7.54}$$

The first part of Eq. (7.54) is the unlevered value of the company (W_U). Indeed, it is equal to the stockholders' return for an unlevered company:

$$W_U = EBIT(1-t_c)(1-t_e) \tag{7.55}$$

Considering the second part of Eq. (7.54), if the investor acquires a bond equal to D, his return is equal to: $DK_D(1-t_d)$. Therefore, the value of debt is equal to:

$$D = DK_D(1-t_d) \tag{7.56}$$

On the basis of Eqs. (7.54), (7.55) and (7.56), Eq. (7.54) the levered value of the company (W_L) is equal to:

$$W_L = W_L + D\left[1 - \frac{(1-t_c)(1-t_e)}{(1-t_d)}\right]$$

That is Eq. (7.52).

Note that if the tax rate on equity is the same as the tax rate on debt $(t_d = t_e)$ or if they are both equal zero $(t_d = t_e = 0)$, Eq. (7.52) becomes:

$$W_L = W_L + Dt_c \tag{7.57}$$

Representing Proposition II with corporate tax and without personal taxes.

If personal taxes are different from zero and the tax rates on equity and debt are different among them, two main cases can be generated:

- $t_e < t_d$: personal taxes are higher for a levered company than for an unlevered company. Tax savings are lower than tax savings due to corporate taxes only. Therefore, the lower corporate taxes for the unlevered company are offset by an increase in personal taxes on investors;
- $t_e > t_d$: tax savings are higher than tax savings due to corporate taxes only. Therefore, the levered value of the company is increased by savings due to both corporate and personal taxes;
- $(1-t_d) = (1-t_c)(1-t_e)$: there are no tax savings. In this case tax savings due to corporate taxes are equal to an increase in personal taxes. Therefore, the value of a levered company is equal to the value of an unlevered company in accordance with Proposition I.

Also if:

- $(1-t_d) < (1-t_c)(1-t_e)$: the levered value of the company is lower than its unlevered value: $W_L < W_U$
- $(1-t_d) > (1-t_c)(1-t_e)$: the levered value of the company is higher than its unlevered value: $W_L > W_U$.

It is important that Propositions I and II of MM are based on very restrictive assumptions and are not feasible in a real world. Over the years by removing the restrictions the assumptions used by MM have developed many theories and empirical researches (Fan et al. 2012; Rauh and Sufi 2010; Frank and Goyal 2009; De Jong et al. 2008; Brounen et al. 2006; Claessens and Klapper 2005; Bancel and Mittoo 2004; Hall et al. 2004; Berkowitz et al. 2003; La Porta et al. 2002; Beck et al. 2002; Rajan and Zingales 2003, 1998, 1995; Myers 2001, 2003, 1984; Booth et al. 2001; Graham and Harvey 2001; Wurgler 2000; Demirgüç-Kunt and Maksimivic 1998, 1999; Harris and Raviv 1991).

Specifically, the introduction of many variables in the model leads to postulate the relevance of capital structure choices on company value and consequently the existence of the "optimal" capital structure.

Several theories have been developed over time. Among these, there are three main ones (De Luca 2015, 2017a, b): (i) trade-off theory extended with agency theory; (ii) packing order theory; (iii) market time theory.

The *trade-off theory* (Kraus and Litzenberger 1973) tries to find "*optimal*" capital structure by assuming that the market is not perfect, considering taxes and costs of financial distress. Therefore the capital structure is a result of the trade-off between debt benefits and costs.

In this sense the higher the taxation on dividends, indicates more debt (Miller and Scholes 1978; Modigliani and Miller 1963); the higher the non-debt tax shields, the lower the debt; higher costs of financial distress indicate more equity. Bankruptcy hypothesis can force managers to forgo profitable investment opportunities (Myers 1977).

Usually another relevant element considered in the definition of debt level based on trade-off between benefits and costs of debt are the agency costs (Morellec 2004; Morellec and Schurhoff 2010).

The *agency theory* (Jensen 1986; Jensen and Meckling 1976) focuses attention on the debt effects on the relationship between: (i) stockholders and management, and (ii) stockholders and bondholders. Capital structure choices have effects on agency cost of equity and debt. Considering agency costs of equity and debt, the effects of debt on company performance is non-unique. Often agency concerns are included in the trade-off framework so that it is broadly interpreted.

Debt has a positive effect on *agency cost of equity*. Indeed, it reduces conflicts between stockholders and management due to the difference in their utility functions, with consequent different behaviour, targets, information (managers have more information and better quality than stockholders with the creation of asymmetric information) and operating decisions. While the first want to maximize equity value, the second want to maximize firm value because it increases their control of resources, power and compensation.

Usually conflicts are strong in the presence of relevant free cash flows (Lane et al. 1998; Bergh 1995; Hoskisson et al. 1994; Bethel and Liebeskind 1993; Jensen 1986; Amihud and Lev 1981). Therefore, the managers undertake various forms of action including designing the right controls (Aghion and Bolton 1992), choosing securities to raise financing (Hart and Moore 1995) and therefore determining capital structure (Jensen 1986; Grossman and Hart 1982; Jensen and Meckling 1976).

Debt is a discipline tool for managers (Jensen 1986). It increases the company's default risk according to the cash-out related debt obligations, requiring to maximisation of efficiency, therefore increasing equity value (Lane et al. 1998; Noe and Rebello 1996; Bergh 1995; Hoskisson et al. 1994; Bethel and Liebeskind 1993; Harris and Raviv 1990, 1991; Stulz 1990; Friend and Hasbrouck 1988; Jensen 1986; Amihud and Lev 1981). The positive effects of debt are clearer for manager-owner who have their personal wealth invested in the company because bankruptcy has a greater impact on his interests (Noe and Rebello 1996; Leland and Toft 1996; Firend and Hasbrouck 1988; Amihud and Lev 1981).

The existence and the consistence of agency costs of equity depends mainly on: (i) the culture and characteristics of the managers; (ii) the way in which a manager can exercise his own preferences without considering the value maximization in decision making; (iii) the nature and the costs of monitoring and bonding activities; (iv) the costs of measuring and evaluating the performance of the managers; (v) the costs of devising and applying an index for compensating the manager which correlates with the owner's welfare; (vi) the costs of devising and enforcing specific behavioural rules or policies; (vii) the costs of managers' replacement (Bhagat et al. 2011).

On the contrary, debt has a negative effect on *agency costs of debt*. It increases conflicts between stakeholders and bondholders according to the different company claims (Jensen 1986). Equity offers holders residual claims on company cash flow while debt offers holders a fixed claim over a borrowing company's cash flow. Therefore, stakeholders' moral hazard and asset-substitutions are possible (Harris and Raviv 1991; Diamond 1989; Jensen and Meckling 1976).

If projects are financed by debt, stockholders have an incentive to the sub-optimal investments. Bondholders are damaged in the presence of high-risk investments because they support the risk. If the project fails, bondholders bear most of the consequence mainly due to the stockholders limited liability; otherwise, if the project is successful, the bondholders will be paid a fixed amount and stockholders capture most of the gains (Harris and Raviv 1991; Jensen and Meckling 1976).

Bondholders do not know what type of projects will be chosen by the company (safe projects, risky projects, average safe-risky projects) with no subsequent alignment between risk and debt interest rate in presence of asset substitutions effects (Harris and Raviv 1991). The company can choose a risky project to maximize equity value after raising funds from bondholders shifting from lower-risk to higher-risk investments without alignment of debt interest rate to major risk (Diamond 1989).

Bondholders try to anticipate the stockholders' opportunistic behaviour by lending less debt or increasing its cost.

Based on these considerations, too much equity can lead to conflicts between managers and stockholders (Jensen 1986). Too much debt can lead to conflicts between stockholders and bondholders (Jensen and Meckling 1976).

The *pecking order theory* (Baker and Wurgler 2002; Fama and French 1998, 2002; Shyam-Sunder and Myers 1999; Myers 1984; Myers and Majluf 1984) is based on the adverse selection and argues that it does not define an "*optimal capital structure*" resulting from the balance between benefits and costs of debt and equity. The company structure is based on a source hierarchy. The company initially prefers sources of due to self-finance adapting the dividend policy to the investment opportunities. If external sources are required, the company prefers to resort initially to debt, then hybrid instruments and only finally to equity. Therefore the internal sources of finance (self-finance) are preferred to external ones where debt is preferred to equity.

There are two kinds of equity: internal, regarding self-financing, that is at the top of the packing order; external, regarding the emission of new shares, that is at the bottom of the packing order. Without investment's opportunities the company retains profits and builds up financial slack to avoid having to raise external finance in the future (Zwiebel 1996; Myers and Majluf 1984).

The *market time theory* (Baker and Wurgler 2002) based on inefficiency of the capital market and asymmetric information suggests that the capital structure choices of the company depend on capital market conditions. The capital structure of the company evolves as the cumulative outcome of the pasts attempts to time the equity market. Managers look at current conditions in both debt and equity markets. Therefore if they need financing they use whichever market currently looks more favourable (Frank and Goyal 2009).

The theory highlights that market timing has large and persistent effects on capital structure of the company. Therefore the company issues shares at higher prices and repurchases at a lower price. In this regard, the company tends to exploit temporary fluctuations in the cost of equity related to the cost of other capital forms. Therefore first company tends to issue equity (instead of debt) when market value is high (relative to book value and past market values) and repurchase equity when market value is low. Second, the company issues equity when the cost of equity is relatively low and repurchase equity when the cost is relatively high. Third, the company tends to issue equity at times when investors are too enthusiastic about earnings perspectives. Fourth, managers consider the market timing is very relevant in the capital structure choices (see the survey of Graham and Harvey 2001). Therefore low leverage companies are the ones that raised funds when their market valuations were high, as measured by the market-to-book ratio, while high leverage companies are the ones that raised funds when their market valuations were low.

The theory states that the existing capital structure of the company is a cumulative result of the company's past experience, which it has attempted time to time under equity market. The theory states that there is no optimal capital structure where capital structure is an outcome of various decisions, which have been made over time. Therefore the market time theory, as well as packing order theory, is structured more on the manager reactions to the environment rather than to a trade-off of specific determinants.

Several researches find proof of market timing in different forms and degrees (see Ritter 2003 for a detailed list of papers that provide evidence of market timing. Among others: Hovakimian 2006; Pagano et al. 1998; Taggart 1977; Jung et al. 1996; Loughran and Ritter 1995; Ikenberry et al. 1995; Korajczyk et al. 1991; Marsh 1982).

There are two versions of equity market timing that lead to similar capital structure dynamics (Baker and Wurgler 2002).

The first, based on the rational managers and investors and adverse selection costs. The company tends to announce equity issue following the release of information which may reduce information asymmetry. Equity issues cluster around periods of somewhat smaller announcement effects (Korajczyk et al. 1991, 1992, 2003; Baker and Wurgler 2002; Bayless and Chaplinsky 1996; Choe et al. 1993; Lucas and McDonald 1990; Myers and Majluf 1984).

The second, based on irrational investors and managers and time-varying mispricing or perceptions of mispricing. Therefore managers issue equity when they believe its cost is irrationally low and repurchase equity when they believe its cost is irrationally high. Market-to-book is well known to be inversely related to future equity returns, and extreme values of market-to-book have been connected to extreme investor expectations (Frankel and Lee 1998; La Porta et al. 1997; Shleifer and Vishny 1997).

Empirical researches have highlighted many determinants, in addition to the models, that could affect the capital structure choices (Frank and Goyal 2003, 2009; Rajan and Zingales 1995; Harris and Raviv 1991; Titman and Wessels 1988; De Luca 2015; La Porta 1996; La Porta et al. 1998, 1999; Noe 1988; Graham 1999, 2000). Generally, the empirical studies show, with different degrees and little difference, that leverage increases with fixed assets, tax shields, growth opportunities and company size; otherwise, leverage decreases with volatility, advertising expenditures, research and development expenditures, bankruptcy probability, and uniqueness of the product. There is no clear relationship between company performance and leverage.

Empirical studies show that the difficulties associated with determinants are not only due to a correct identification but also to their positive or negative effects on the choices of capital structure. It is also due to the specific characteristics of the company (Myers 2003). It is not unusual that a single determinant has a positive impact on the capital structure choices in some studies while negative in other.

It seems to say that theories and empirical researches are able to explain some aspects under certain conditions of company behaviour. However, there is still no theory can fully explain company behaviour on capital structure or, even more, able to define the optimal capital structure.

In this context the main problem of the capital structure can be defined in terms of capital cost. Specifically, the main problem is to estimate the cost of capital.

Considering that capital structure is defined by equity and debt, it is necessary to estimate the cost of equity and the cost of debt. In both cases, they measure the expected return of investors in equity and debt related to risk bearing. Specifically, the expected return for investors in equity is the cost of equity for the company, and the expected return for investors in debt is the cost of debt for the company (Damodaran 2012).

In general terms, risk-return models in finance are built around the rate that investors can require for risk-free and risky investments. Indeed, most risk-return models start off with an asset defined as risk-free and use its expected return to define the risk-free rate. Then, a risk-premium function of the investments risk is added to the risk-free rate.

Denoting with R_f the risk-free rate and RP the risk premium, the R_I the investment expected return is equal to:

$$R_I = R_f + RP \tag{7.58}$$

There are three main considerations about Eq. (7.58):

- The first: in a model of equilibrium such as the CAPM only the market risk is considered and it is captured in the market portfolio. The premium risk is the premium that investors can require when they invest in this portfolio. Also, in the multifactor models only market risk is considered. However, in this case there are multiple risk premiums, each one measuring the premium required by investors to bear a specific market risk factor;
- The second: there is no investment that can be defined as risk-free by definition. The asset free-risk is characterized by a risk level so low as to be approximated to zero. In this context the expected returns can be assumed as certain returns. The securities that have a chance of being risk-free are government bonds. Note that it is not because they cannot fail, but because the government controls the printing of currency, unlike the largest and safest private entity;
- The third: the higher the risk, the higher the premium to the risk. Therefore, riskier investments have a higher expected return than risk-free and the equity risk premium is higher than the risk-free rate $\left(RP > R_f\right)$ by definition.

Equation (7.58) can be declined for investors in debt and investor in equity. Specifically, the expected return by bondholders (R_D) is equal to the risk-free rate $\left(R_f\right)$ plus a spread able to measure the default risk of the firm (D_F) as follows:

$$R_D = R_f + D_F \tag{7.59}$$

Differently, the expected return by stockholders (R_E) is equal to the risk-free rate $\left(R_f\right)$ plus an equity risk premium for the investment in equity (ERP) as follows:

$$R_E = R_f + ERP \tag{7.60}$$

The expected return for investors in equity (R_E) is the cost of equity (K_E) for the company, as well as, the expected return for investors in debt (R_D) is the cost of debt (K_D) for the company (Damodaran 2012).

Indeed, the expected return for investors in both equity and debt reflects risk bearing. Consequently, it is the cost of capital in equity and debt for the company.

In terms of cost of capital, Eqs. (7.59) and (7.60) can be rewritten as follows:

$$R_D \equiv K_D\text{: } R_D = R_f + D_F \leftrightarrow K_D = R_f + D_F \tag{7.61}$$

$$R_E \equiv K_E\text{: } R_E = R_f + ERP \leftrightarrow K_E = R_f + ERP \tag{7.62}$$

7.2 Cost of Equity

There are several competitive models to estimate the equity risk premium. Among these, the most well-known is the equilibrium model of Capital Asset Pricing Model (CAPM), Arbitrage Pricing Model (APM), Multifactor Model, Proxy Model. Also if the models are different among them, they share some common views on risk (Damodaran 2012):

- risk is defined in terms of variance in actual returns around the expected return. Consequently, the higher the variance of the expected return, the higher the volatility of the asset and the higher the risk. In this sense, the investment is riskless when actual returns are always equal to the expected return and when the variance of expected return is null;
- risk has to be measured from the perspective of the marginal investor in an asset, and then in the perspective of a diversified investor. Therefore, only the risk that an investment adds to a diversified portfolio should be measured and compensated. In this sense, only the systematic-risk is considered because it is not diversifiable, while the company specific risk is not considered because it is diversifiable.

In this context, the cost of equity is derived by CAPM on the basis of SML. The expected return of the investment in equity in the j-th company $\left(R_{E_j}\right)$, and therefore the cost of equity for the j-th company $\left(K_{E_j}\right)$, is equal to the risk-free rate R_f plus the equity risk premium equal to coefficient beta of the j-th company $\left(\beta_j\right)$ multiplied by the difference between expected return on equity market as measured by the market portfolio (R_m) and the return (certain) on risk-free investments as measured by the risk-free rate $\left(R_f\right)$, as follows:

$$R_{E_j} \equiv K_{E_j}: \quad R_{E_j} = R_f + \beta_j\left(R_m - R_f\right) \leftrightarrow K_{E_j} = R_f + \beta_j\left(R_m - R_f\right) \tag{7.63}$$

The coefficient beta is equal to the covariance between the expected return of the j-th company $\left(R_j\right)$ and the expected return of market portfolio (R_m), and the variance of the expected return of the market portfolio (R_m), as follows:

$$\beta_j = \frac{Cov\left(R_j; R_m\right)}{Var(R_m)} = \frac{\sigma_{\left(R_j;R_m\right)}}{\sigma^2_{(R_m)}} \tag{7.64}$$

On the basis of Eq. (7.64), Eq. (7.63) in terms of cost of equity $\left(K_E\right)$ can be rewritten as follows:

$$K_{E_j} = R_f + \frac{\sigma_{\left(R_j;R_m\right)}}{\sigma^2_{(R_m)}}\left(R_m - R_f\right) \tag{7.65}$$

It is relevant to note that Eq. (7.63) is not the CAPM but its derivation. Indeed, in CAPM the risk premium measures what investors, on average, demand as extra return for investing in the market portfolio only with respect to the risk-free asset.

In Eq. (7.63) the key role is played by coefficient beta. It has a central role in the risk-return models because it measures the asset's market risk based on the market portfolio in the CAPM, as well as, based on specific factors in the Arbitrage Pricing Model and Multifactor Models (among others: Alexander and Benston 1982; Alexander and Chervany 1980; Beaver et al. 1970; Blume 1975; Fama and French 1993; Elton et al. 1978; Gonedes 1973; Handa et al. 1989; Hill and Stone 1980; Klemkosky and Martin 1975; Levy 1971, 1974; Logue and Merville 1972; Roenfeldt et al. 1978; Rosenberg and Guy 1976a, b; Rosenberg and McKibben 1973; Scholes and Williams 1977; Scott and Brown 1980; Theobald 1981; Young et al. 1991; Elton and Gruber 1974; La Porta 1996).

There are two main baseline approaches to estimate the coefficient beta. The first is based on an estimate of future betas on the basis of an estimate of correlations between the asset's expected returns and the index's expected returns. In this case, the beta estimate tends to be subjected to the previsions of the analyst about expected returns. The second is based on an estimate of future betas on the basis of historical betas. This approach is more useful and there is proof that historical betas provide useful information on future betas (Blume 1970, 1975; Levy 1971). This approach is usually preferred by analysts to the first one.

In order to estimate the coefficient beta in a corporate assessment, there are two main approaches that are normally used: (i) *Regression Beta;* (ii) *Bottom-Up Beta.*

Regression Beta

The *Regression Beta* estimates the coefficient beta based on the regression of returns on the investment against returns on a market. This approach is one of the most commonly used by analysts (Damodaran 2012). Obviously, it can be used only if the assets have been traded and have market prices.

In theory, stock returns on assets should be related to returns on a market portfolio that includes all traded assets. In practice, a stock index is used as a proxy for the market portfolio and the beta is estimated for stocks against the stock index.

Specifically, by considering time (t), the expected return of j-th company can be estimated on the basis of regression equation as follows:

$$R_{jt} = \alpha_j + \beta_j R_{mt} + e_j \tag{7.66}$$

And by applying the Ordinary Least Square (OLS), we have:

$$\beta_j = \frac{\sum_{t=1}^{n}\left[\left(R_{jt} - \mu_{jt}\right)\left(R_{mt} - \mu_{mt}\right)\right]}{\sum_{t=1}^{n}\left(R_{jt} - \mu_{jt}\right)^2} = \frac{Cov\left(R_j; R_m\right)}{Var\left(R_m\right)} = \frac{\sigma_{\left(R_j; R_m\right)}}{\sigma_{R_m}^2} \tag{7.67}$$

and

$$\alpha_j = R_{jt} - \beta_j R_{mt} \tag{7.68}$$

The value of β_i and α_i obtained by using the regression analysis, is used to estimate the true β_i and α_i that exist for an asset. Obviously, the estimates are subject to error and therefore, the estimated β_i and α_i could not be equal to their real values that existed in the period. Furthermore, β_i and α_i are not perfectly stationary over time.

Equation (7.67) shows how the coefficient beta is equal to the covariance between the expected return of the j-th company $\left(R_j\right)$ and the expected return of market portfolio (R_m), and the variance of the expected return of the market portfolio (R_m).

The intercept $\left(\alpha_j\right)$ as defined in Eq. (7.68) provides simple measures of performance of the investment in the j-th company during the period of regression when returns on investment in the j-firm are measured against the expected returns on the basis of the CAPM (Damodaran 2012).

Note that Eq. (7.65) can be rewritten as follows:

$$K_{E_j} = R_f\left(1 - \beta_j\right) + \beta_j R_m \leftrightarrow K_{E_j} = R_f\left(1 - \frac{\sigma_{\left(R_j;R_m\right)}}{\sigma^2_{(R_m)}}\right) + \frac{\sigma_{\left(R_j;R_m\right)}}{\sigma^2_{(R_m)}} R_m \tag{7.69}$$

The difference between the intercept a_j from the regression and $R_f(1 - \beta)$ is known as Jensen's Alpha (A) (Damodaran 2012):

$$A = a_j - R_f\left(1 - \beta_j\right) \tag{7.70}$$

It measures the stock performance of the j-th company compared with companies with similar beta during the regression period. So that, if (Damodaran 2012):

- $a_j > R_f\left(1 - \beta_j\right)$: stock did better than expected during regression period. The j-company has earned more than companies with a similar beta during the regression period;
- $a_j = R_f\left(1 - \beta_j\right)$: stock did as well as expected during regression period. The j-company has earned equal the same as companies with a similar beta during the regression period;
- $a_j < R_f\left(1 - \beta_j\right)$: stock did worse than expected during regression period. The j-company has earned lower than companies with a similar beta during the regression period.

Note that one of the most important problems to estimate beta by regression model concerns the length of the estimation period. Indeed, there is a normal trade-off: a longer estimation period provides more data, but the company may have changed in its risk characteristics over time.

Consequently, there could be positive or negative sampling errors. Therefore, the beta will be partially a function of the true underlying beta and, partially, a function of the sampling error. Generally, on the long term, the beta tends to converge, on average, to 1. Therefore, in the forecast period the beta tends to be closer to 1 than the estimate obtained from historical data (Blume 1975; Levy 1974). On the basis of this observation, it is possible to modify the past betas to capture the tendency. There are two main techniques used: (i) Blume's technique; (ii) Vasicek's Technique.

Blume's technique (Blume 1975) corrects the past betas by measuring this adjustment directly towards 1. This technique arising from the empirical evidence suggests that the betas for most companies tend to move towards the average beta over time, which is 1.

It assumes that the adjustment in one period is a good estimate of the adjustment in the next. Specifically, in the second period $(t = 2)$ the beta of i-th asset $\left(\beta_{i,2}\right)$ can be estimated on the basis of the beta in the first period $\left(\beta_{i,1}\right)$ adjusted as follows:

$$\beta_{i,2} = 0.343 + 0.677\beta_{i,1} \tag{7.71}$$

Note that Eq (7.71) lowers the high value of beta and raises the low value of beta. The process pushes all estimated betas towards 1.

Therefore, if the average beta increased over these two periods, it assumes that average betas will increase over the next period. Obviously, if there are no expectations about this trend, and then if there are not expectations to assume the next average beta of the period will be more than this period's, this change in beta definitely represents an undesirable property.

This technique is one of the most commonly used by analysts. Specifically, the Adjusted Beta (β_A) (that measures the expected beta) is equal to the Raw Beta (β_{RB}) (that is the historical beta) considered for $2/3(0.67)$ plus the Beta of Market Portfolio (β_M) (that is equal 1 by definition) considered for $1/3(0.33)$, as follows (Damodaran 2012):

$$\beta_A = \beta_{RB}(0.67) + \beta_M(0.33) = \beta_{RB}(0.67) + 1.00(0.33) \tag{7.72}$$

Vesicek's technique (Vesicek 1973) assumes that the adjustment depends on the size of the uncertainty about beta. Consequently, the larger the sampling error, the greater the change in large differences from the average being due to sampling error the greater the adjustment.

Specifically, it assumes that the adjusted beta can be modified based on two main considerations: (i) many sectors are characterized by structurally lower or higher betas than the market. Therefore, the average beta of the sector should be considered rather than the market; (ii) the correction of historical beta in order to obtain the expected beta should be proportional to the relevance of the standard error of the beta.

Denoting with β_1 the average beta across the sample in the historical period; $\sigma^2_{\beta_1}$ the variance of the distribution of the historical beta over the sample (it measures the variation of beta across the sample of assets under consideration); $\sigma^2_{\beta_{j1}}$ is the square of the standard error of the estimate of beta for the j-th asset measured in the period 1 (it measures the uncertainty associated with the measurement of the individual assets beta). In the second period $(t = 2)$ the j-th beta $\left(\beta_{j,2}\right)$ can be estimated as follows:

$$\beta_{j,2} = \frac{\sigma^2_{\beta_{j,1}}}{\sigma^2_{\beta_1} + \sigma^2_{\beta_{j,1}}}\beta_1 + \frac{\sigma^2_{\beta_1}}{\sigma^2_{\beta_1} + \sigma^2_{\beta_{j,1}}}\beta_{j,1} \tag{7.73}$$

Equation (7.71) can be simplified by analysts in its application. Denoting with β_S the beta of the sector, $\sigma^2_{\beta_S}$ the variance of the beta of the sector, β_{MS} the average beta of the sector, $\sigma^2_{\beta_{MS}}$ the variance of the average beta of the sector, SE^2 the square of the standard error of the historical beta, β_{RB} the Raw Beta that is the historical beta, the Adjusted Beta (β_A) that measures the expected beta is equal to:

$$\beta_A = \beta_{RB}\frac{\sigma^2_{\beta_S}}{\frac{\left(\sigma^2_{\beta_S} + SE^2\right) + \left(\sigma^2_{\beta_{MS}} + SE^2\right)}{\left(\sigma^2_{\beta_S}\right) + SE^2}} \tag{7.74}$$

Equation (7.74) shows that the higher the standard error of the historical beta, the higher the adjustment of the beta of the company in the estimate process.

Note that the weighted procedure adjusts observations with large standard errors further towards the mean rather than observations with small standard errors. Indeed, this technique suffers from the potential source of bias of Blume's techniques although it does not forecast a trend in betas.

Based on the considerations of these two techniques, it is relevant to note that by using different periods, market index, and procedures to adjust the regression beta, often the analysis provides different beta for the same company at the same point in time. All beta estimates have a standard error. The main relevant point is that all of the betas reported for a company fall within the range of standard errors from the regressions.

Bottom-Up Beta

The *Bottom-Up Beta* estimates the beta by looking at the fundamentals of the business and the characteristics of the company with regards to its financial and operating leverage. Specifically, three are the main determinants of the beta (Damodaran 2012):

(a) *business of the company*: beta is affected by the business characteristics of the company. It measures the level of risk of the business of the company regarding the market. In this regard, the higher the company's business sensitivity to market conditions, the higher the beta. Consequently, the company involves in sectors of economy that are very sensitive to economic conditions, have a high beta. Other variables being equal, cyclical companies can be expected to have higher betas than non-cyclical companies;

(b) *degree of operating leverage of the firm*: beta is affected by the degree of operating leverage of the company. The operating leverage is function of the cost structure of the company as it defines on the basis of relationship between fixed costs and variables costs: the higher the fixed costs related to total costs, the higher the operating leverage of the company. The relationship between beta and operating leverage is due to its effects on operating and net income: the higher the operating leverage, the greater the fixed costs, the higher the volatility of operating income due to the volatility of revenues and, other things remaining equal, the higher the variability of net income.

(c) *degree of financial leverage of the company*: beta is affected by the degree of financial leverage of the company. Financial leverage is function of the capital structure of the company as it defines, in terms of relationship between debt and equity: the higher the debt in capital structure, the higher the financial leverage of the company.

The Bottom-Up Beta approach is based on the relationship between beta of levered companies and beta of unlevered companies (Hamada 1972). The equity beta of levered companies (β_L) is equal to the weighted average of the equity beta of unlevered companies (β_U) and the beta of debt (β_D); the weights are function of the debt and equity in the capital structure (Brealey et al. 2016). Assuming no corporate taxes, the relationship can be defined as follows:

$$\beta_L = \beta_U\left(\frac{E}{E+D}\right) + \beta_D\left(\frac{D}{E+D}\right) \tag{7.75}$$

and solving for β_U, we have:

$$\begin{aligned}\beta_U &= \left[\beta_L - \beta_D\left(\frac{D}{E+D}\right)\right]\left(\frac{E+D}{E}\right) = \beta_L\left(\frac{E+D}{E}\right) - \beta_D\left(\frac{D}{E+D}\right)\left(\frac{E+D}{E}\right) \\ &= \beta_{EL}\left(\frac{E}{E}\right) + \beta_{EL}\left(\frac{D}{E}\right) - \beta_D\left(\frac{D}{E}\right)\end{aligned}$$

and then:

$$\beta_{EU} = \beta_{EL} + (\beta_{EL} - \beta_D)\frac{D}{E} \tag{7.76}$$

Note that Eq. (7.76) is the Proposition II of MM on betas. The relationship between levered and unlevered beta can be defined by combining the Proposition of MM and MM's Proposition and the CAPM (Hamada 1972). Specifically, assume that (Hamada 1972): (i) the MM formulation the tax shield value for constant debt over time; (ii) the debt risk level is very low and then the beta of debt is assumed equal to zero; (iii) the interest on debt are tax deductions by generating tax savings; (iv) the discount rate used to estimate the value of tax shield is assumed equal to the cost of debt. It implies that the value of the tax shield is proportionate to the market value of debt: $W_{TS} = Dt_c$.

Denote with E_U the equity of unlevered company; E_L the equity of the levered company; $EBIT$ the Earnings before tax and interest; t_c the corporate tax rate; I the interest payments on debt; ΔIC the changes in capital invested in the business.

The expected return on equity for an unlevered company (R_{E_U}) is equal to:

$$R_{E_U} = \frac{EBIT(1 - t_c) - \Delta IC}{E_U} \tag{7.77}$$

and the expected return on equity for a levered company (R_{E_L}) is equal to:

$$R_{E_L} = \frac{EBIT(1 - t_c) - I - \Delta IC}{E_L} \tag{7.78}$$

Substituting Eqs. (7.77) and (7.78) in the equation of j-th company beta $\left(\beta_j\right)$, and assuming that the covariance between return on market (R_M), the components of capital invested in the company and interests on debt are zero $(\beta_{\Delta IC} = \beta_I = 0)$, we have:

$$\beta_j = \frac{Cov\left(R_j; R_M\right)}{Var(R_M)} \rightarrow \begin{matrix} \beta_U = \dfrac{Cov\left(\frac{EBIT(1-t_c)}{E_U}; R_M\right)}{Var(R_M)} \\ \beta_L = \dfrac{Cov\left(\frac{EBIT(1-t_c)}{E_L}; R_M\right)}{Var(R_M)} \end{matrix} \tag{7.79}$$

The equity, both levered and unlevered, does not have a covariance with market portfolio. Therefore:

$$\beta_U = \frac{\frac{1}{E_U} Cov(EBIT(1 - t_c); R_M)}{Var(R_M)} = \frac{Cov(EBIT(1 - t_c); R_M)}{E_U Var(R_M)} \rightarrow \beta_U E_U$$
$$= \frac{Cov(EBIT(1 - t_c); R_M)}{Var(R_M)}$$

as well as:

$$\beta_L = \frac{\frac{1}{E_L} Cov(EBIT(1-t_c);R_M)}{Var(R_M)} = \frac{Cov(EBIT(1-t_c);R_M)}{E_L Var(R_M)} \rightarrow \beta_L E_L = \frac{Cov(EBIT(1-t_c);R_M)}{Var(R_M)}$$

Therefore, we have:

$$\beta_U E_U = \beta_L E_L = \frac{Cov(EBIT(1-t_c);R_M)}{Var(R_M)}$$

Consequently:

$$\beta_L = \beta_U \frac{E_U}{E_L} \tag{7.80}$$

Assuming no corporate taxes. In this case if the company is financed by equity only its unlevered value (W_U) is equal to the value of unlevered equity (E_U) as follows:

$$W_U = E_U \tag{7.81}$$

Otherwise, if the company is financed by equity and debt its levered value (W_L) is equal to the value of equity levered (E_L) plus the value of debt (D), as follows:

$$W_L = E_L + D \tag{7.82}$$

On the basis of no corporate taxes, the Proposition I of MM shows as the value of unlevered company (W_U) is equal to the value of levered firm (W_L). On the basis of Eq. (7.29) we have:

$$W_U = W_L$$

Consequently on the basis of Eqs. (7.81) and (7.82), we have:

$$E_U = E_L + D \tag{7.83}$$

On the basis of Eq. (7.83), Eq. (7.80) can be rewritten as follows:

$$\beta_L = \beta_U \frac{E_U}{E_L} \rightarrow \beta_L = \beta_U \left(\frac{E_L + D}{E_L} \right) = \beta_U \left(\frac{E_L}{E_L} + \frac{D}{E_L} \right) = \beta_U \left(1 + \frac{D}{E_L} \right)$$

and then:

$$\beta_L = \beta_U \left(1 + \frac{D}{E}\right) \tag{7.84}$$

and inverse:

$$\beta_U = \frac{\beta_L}{\left(1 + \frac{D}{E}\right)} \tag{7.85}$$

Assume corporate taxes. The relation between levered and unlevered beta must be changed because the Proposition I of MM is not applicable. In this case, the value of the levered company (W_L) is greater than the value of an unlevered company (W_U). The difference is related to the value of tax shield (W_{TS}). On the basis of Eq. (7.46) we have:

$$W_L = W_U + W_{TS}$$

Assume that the expected tax savings are in perpetuity and assume they are discounted to the cost of debt (K_D). The value of tax shield (W_{TS}) is equal to:

$$W_{TS} = \sum_{t=1}^{n} \frac{Dt_c K_D}{(1 + K_D)^t} = \frac{Dt_c K_D}{K_D} = Dt_c \tag{7.86}$$

By considering Eqs. (7.46), (7.81), (7.82) and (7.86), Eq. (7.46) can be rewritten as follows:

$$\begin{array}{l} W_L = W_U + W_{TS} \\ W_U = E_U \\ W_{TS} = Dt_c \\ W_L = E_L + D \end{array} \quad \rightarrow E_L + D = E_U + Dt_c$$

and by solving by E_U, we have:

$$E_U = E_L + D(1 - t_c)$$

By substituting the unlevered equity (E_U) in Eq. (7.80), we have:

$$\begin{aligned} \beta_L = \beta_U \frac{E_U}{E_L} \rightarrow \beta_U \left[\frac{E_L + D(1 - t_c)}{E_L}\right] = \beta_U \left[\frac{E_L}{E_L} + \frac{D(1 - t_c)}{E_L}\right] \\ = \beta_U \left[1 + (1 - t_c)\frac{D}{E_L}\right] \end{aligned}$$

and then:

$$\beta_L = \beta_U\left[1 + (1 - t_c)\left(\frac{D}{E}\right)\right] \tag{7.87}$$

and inverse:

$$\beta_U = \frac{\beta_L}{\left[1 + (1 - t_c)\left(\frac{D}{E}\right)\right]} \tag{7.88}$$

that is the Hamada's equation.

Note that by considering debt risk, it is necessary to introduce the beta of debt (β_D). In this case Eq. (7.87) can be rewritten as follows (Damodaran 2012):

$$\beta_L = \beta_U\left[1 + (1 - t_c)\frac{D}{E}\right] - \beta_D(1 - t_c)\left(\frac{D}{E}\right) \tag{7.89}$$

Equations (7.84) and (7.87) show how the unlevered beta measures the risk of unlevered company and therefore the risk function of the company of its business only. Otherwise, the levered beta is function of the capital structure choices and therefore the financial risk of leverage. Consequently, other variables being equal, an increase in financial leverage increases the levered beta while the unlevered beta does not change. This effect of debt on beta tends to explain why companies operating in the same business, with the same business risk, can have different levered beta as a result of the different capital structure (Berk and DeMarzo 2008). The bottom-up beta, is different from the regression betas and it is based on the company and business fundamentals (Damodaran 2012).

7.3 Cost of Debt

The expected return for bondholders (R_D) is equal to the risk-free rate (R_f) plus a premium for default risk of the company (D_F) that is function of the probability of the company's incapacity to face debt obligation. On the basis of Eq. (7.59) we have:

$$R_D = R_f + D_F$$

As defined, the expected return of investors in debt (R_D) is the cost of debt (K_D). Therefore, on the basis of Eq. (7.61) we have:

$$R_D \equiv K_D\colon R_D = R_f + D_F \leftrightarrow K_D = R_f + D_F$$

This chapter analyses the default risk of the company while Chap. 10 analyses the risk-free rate.

With regards to the default risk, it is important to note that unlike investment in equity where the risk relates to the likelihood that cash flows on investments can be different from the expected cash flow, in the investment in debt, the cash flows are promised when the investment is made. Therefore, in this case the risk is not function of the variance of expected return but it is due to the company's default probability only.

While the risk-return models for equity are based on the effects of the market risk only by assuming a diversified investor (marginal investor), differently models of default risk are based on the effects of company-specific default risk on promised return for investors in debt.

Therefore, while thanks to the diversification process the specific risk of the company is not priced into expected return for equity, the same rationale cannot be applied to bonds because they have limited upside potential and a greater downside potential arising from company-specific conditions. By investing in debt, the interests are fixed at the time of investment, and they represent the promised cash flow on debt. The worst-scenario for bondholders, is that he does not receive, with a different degree, the payment of interest and the reimbursement of capital.

There is a positive proportional relationship between cost of debt and default risk of the company: higher the company's default risk, the higher the return on debt required by investors, and then the higher the cost of debt for the company.

Generally, the company's default risk is function of three main variables:

- the company's ability to generate cash flows from operations;
- the volatility of cash flows from operations;
- the financial obligations with regards to interest payments and principal reimbursement.

Therefore, the higher the company's ability to generate cash flows from operations able to face debt obligations, the lower the company's default risk. Then, companies that generate high operating cash flows related to their financial obligations should have lower default risk than companies that generate low operating cash flows related to obligations.

Also, with equal conditions of operating cash flows they are able to cover the obligations on debt, the lower the operating cash flows volatility, higher their stability, and therefore the lower the default risk.

An estimate of the company's default risk is the main problem to solve. There are two main approaches to be followed:

- *Agency rating approach*: it is the most widely used. It is based on the bond rating assigned by a specialized Agency Rating to company bonds.
  The market price of the bonds, together with their coupons and maturities, allow for capturing of their yields (or more usually, their yields to maturity) that are used as the cost of debt of the company.
  The rating assigned to company bonds by a specialized Agency Rating is mainly based on the financial ratios of the company with the objective of assessing the company's ability to face debt obligations.

The most important problem is that this approach requires that outstanding bonds are widely liquid, trade frequent and have a rating as issued by a specialized rating agency.

– *Synthetic rating approach*: it measures the default risk of the company based on the analysis of its fundamentals. Many companies are not rated, by choice or because they are smaller. Generally, when there is no rating available to estimate the cost of debt, synthetic rating can be estimated. In this case the analyst plays a role of the rating agency by assigning a rating to a company based on the analysis of its fundamentals.

There are two main types of study in literature about the default risk of the company also if they are strictly related: (i) studies on structural-form model and reduced-form model and (ii) studies on accounting models (De Luca 2017a, b).

The structural-form models and the reduced-form models modelling the credit risk. These two classes of studies are characterized by their own advantages and disadvantages. Generally, while the reduced-form models have a lot of room for calibrating of historical data but they lack the financial elements for the mode parameters, the structural models are intuitive and they have a nice explanation in financial terms but they lack measuring in particular the short-term credit risk and they are much harder to apply whenever there is more than one debt instrument (Gupta 2013). Specifically, while the structural models have the advantage to explicitly model a company's assets by assuming that the company's assets follow a geometric Brownian motion, reduced models do not formally link a company's assets to the corporate credit issued by it. They are based on exogenously specified parameters, such as recovery rates to value corporate credit (Rajaratnam et al. 2017).

Structural-form models originate from the study of Merton (1974), Black and Scholes (1973). In general terms, with Merton's work that postulates company value follows a diffusion process with constant volatility as described by the Geometric Brownian Motion, the volatility of share prices and debt values exhibit plays a key role in the modern financial literature.

The approach of structural-form models assume that the bankruptcy process is explicated. It defines that both of the events can trigger the default and the pay-offs to the bondholders at a default state in terms of company assets and liabilities. Therefore, this approach is only able to produce price under the extremely simplistic assumption on capital structure. However, these models require a certain level of abstraction about bankruptcy according to its traceability The main problem is that the asset volatility employed in structural models is not directly observable and it has to be estimated indirectly by using different indicators such as equity volatility. Several recent studies show the difficulty in structural models to explain yield spreads (Huang and Huang 2012; Bao 2009; Cremers et al. 2008; Eom et al. 2004; Schaefer and Strebulaev 2008; Chen et al. 2007).

The reduced form models (or statistical approach) originated with the study of Jarrow and Turnbull (1995), Duffie and Singleton (1999) among others, are based on a statistical approach and they are abstracted away completely from the economic definition of bankruptcy and default treats that are considered as exogenously specified process. This approach is traceable and it is usually used in pricing default risk. In this perspective, any default free term structure model can be used to price bonds with default risk.

The studies on accounting models focus on company fundamentals. They were born with the pioneering works Beaver (1966, 1968) that applied a unique statistical analysis for the prediction of corporate failure and Altman (1968) developed the Z-Score model in order to identify accounting variables and financial ratios to measure the company's default risk. Following several models are developed. The same Z-score model has been revised by starting the second generation of the model with several enhancements (Altman et al. 1977). After, the Z-score model has been modified for its application in the context of corporations in emerging markets (Altman et al. 1995).

By following a similar approach Ohlson (1980) its O-Score Mode has been developed with the identification of nine ratios capable of predicting bankruptcy, Taffler (1984) developed the UK-based Z-Score model that is one of the most commonly used Z-Score model, and Zmijewski (1984) developed the Probit Analysis developed a hybrid logistic model. Also, Bathia (1988) and Shaoo et al. (1996) examined the predictive power of accounting ratios on a sample of sick and non-sick companies by using the multiple discriminating analysis and Lennox (1999) in his study showed that profitability, leverage and cash flows have relevant effects on the probability of company's default as Kim et al. (1993).

Over the time several empirical tests have illustrated the relevance of the models developed (Agarwal and Taffler 2007; Baninoe 2010; Kumar and Kumar 2012; Shumway 2011).

Finally, several studies show the effect of debt on the competitive strategy of the company and its profitability over time. There are two main contrasting theories: the first states that a high debt level gives the company the opportunity to invest many resources in the business by increasingly acquiring a share of the market (Brander and Lewis 1986); the second states that a high debt level leads the company to increase prices on the short-term in order to secure short-term profit. It is due to the higher discounting rate associated with the higher default probability of the levered company (Chevalier and Scharfstein 1996; Dasgupta and Titman 1998). Several empirical studies have found support for each of these theories (see Parson and Titman 2008).

In this context, the cost of debt (K_D) is structured on three parts (Elton et al. 2013):

- the first, is the financial cost of time and it can be calculated approximately according to the risk-free rate. It can be approximated to the return of risk-free government bond and therefore it measures the return on default-free bonds;
- the second, is the market *risk-premium* due to the higher volatility of the corporate bonds than government bonds;
- the third, is the default risk of corporate bonds due to the risk of the bond-holder's loss in the case of company insolvency.

Among these, the third part is the most relevant in this context. Indeed, debt cost is based on a fundamental analysis of the company rather than the volatility of the asset's value on the market. In this regard, the cost of debt is based on the relationship between debt level and the economic and financial dynamics with regards to Operating and Net Income, Capital Invested and Capital Structure and Free Cash-Flow from Operations and Equity.

On the basis of these three main parts, the cost of debt $(\mathrm{K_D})$ can be defined equal to (De Luca 2017):

$$\mathrm{K_D} = \mathrm{R_F} + (\mathrm{YC_M} - \mathrm{R_F}) + \rho \mathrm{e}^{\delta} \tag{7.90}$$

where:

- R_F: is the risk-free rate;
- YC_M: is the expected medium market yield of corporate bonds;
- $(\mathrm{YC_M} - \mathrm{R_F})$: is the difference between the expected medium market yield and the risk-free rate and it measures the market risk premium of corporate bonds due to their higher risk than government bonds;
- δ: is a composite index that measures the company's ability to face debt obligations based on its fundamentals;
- e^{δ}: is the exponential function that measures the default risk of the company based on its fundamentals. The exponential function is used for two main reasons: first, it highlights the negative effects of debt when debt increases; second, it can be approximated with a quadratic form that is compatible with mean-variance approach. It is expressed on based 100 $\left(e^{\delta} = e^{\delta}/100\right)$;
- ρ: is a discretionary variable. It goes between zero and one $(0 \leq \rho \leq 1)$ and it defines the part of default risk that investor wants to assume.

Equation (7.90) can be divided in three main parts:

- R_F: is the *risk-free rate* and it is the time value of money. It can be approximated to the return of riskless government bonds and therefore it measures the return on default-free bonds;
- $(\mathrm{YC_M} - \mathrm{R_F})$: is the market *risk-premium* of corporate bonds due to their higher risk than government bonds;
- $\rho \mathrm{e}^{\delta}$: is the *default premium* and it measures the specific risk of default of the company due to its incapability to face debt obligations.

Therefore, the second and third parts of Eq. (7.90) measure the *spread on debt* (S_D) and therefore its default risk (D_F):

$$S_D \equiv D_F = (\mathrm{YC_M} - \mathrm{R_F}) + \rho \mathrm{e}^{\delta} \tag{7.91}$$

The *composite index* (δ) measures the company's abilities to face debt obligations on the basis of its fundamentals. It can defined as follows:

$$\delta = \gamma_1 + \gamma_2 + \gamma_3 \tag{7.92}$$

where γ_1, γ_2 and γ_3 are the *composite index coefficients* defined on the company's fundamentals.

The *composite index coefficient* γ_1 is equal to the ratio between Cash-flows out for Debt Commitment (CDC) and the Current Operating Cash-flows $\left(CFO_{(A)}\right)$ realized in the period considered, as follows:

$$\gamma_1 = \frac{CDC}{CFO_{(A)}} \quad 0 \le \gamma_1 \le 1 \tag{7.93}$$

The $CFO_{(A)}$ can be defined as equal to the EBITDA plus the changes in the Net Operating Capital Invested (NOCI) that it is equal to the sum of investments in Capex and Net Working Capital.

The CDC can be defined equal to the Interest on Debt in the period (DK_D) plus the Share of Debt (principal) to be reimbursed in the period based on the debt level at the start of the period (αD), as follows:

$$CDC = \alpha D + DK_D = D(\alpha + K_D) \tag{7.94}$$

On the basis of Eq. (7.94), Eq. (7.93) can be rewritten as follows:

$$\gamma_1 = \frac{D(\alpha + K_D)}{CFO_{(A)}} \quad 0 \le \gamma_1 \le 1 \tag{7.95}$$

Equation (7.95) shows how the lower the distance between CDC and $CFO_{(A)}$, the higher the default risk due to the company's inability to face debt obligations.

Note that if the company is financed by equity only, the coefficient is equal to zero. Otherwise, if the company is financed by equity and debt, the limit of coefficient is equal to one and consequently the CDC is equal to $CFO_{(A)}$. When the value of the coefficient is greater than one, the company is unable to face debt obligations and therefore it can be considered in default.

The *composite index coefficient* γ_2 is equal to the ratio between the Financial Debt (D) (book value) and Liquid Assets (LA), as follows:

$$\gamma_2 = \frac{D}{LA} \quad 0 \le \gamma_2 < \varepsilon \tag{7.96}$$

The LA refers to tangible, intangible and financial assets, credit, inventors. In order for an asset to be considered as Liquid, it must be characterized by two elements: (i) a market value; (ii) marketable in the short-time.

If the amount of implemented liquid assets is greater than the amount of debt, they can be used to satisfy the bondholders' expectations; otherwise, if the debt amount is greater than liquid assets implemented, even if they sold them on the market the bondholder's expectations cannot be completely satisfied.

Therefore, this coefficient measures the coverage of debt obligations in the case of insolvency case and therefore it is a measure of the bondholders' guaranty. The more the liquid assets are in place, the higher the bondholder's guaranty in the case of insolvency. Otherwise, the greater the debt than the liquid assets, the greater the bondholders' risk.

Note that if the company is all-equity financed, the coefficient is equal to zero. Otherwise, if the company is financed by equity and debt, the coefficient can change between zero and value ε: it is defined by the bondholders on the basis of their risk aversion. In any case, $\varepsilon = 1$ can be considered the alert point.

The *composite index coefficient* γ_3, is equal to the ratio between Cash-flows to Equity Net Current Debt $\left(CFNE_{(A)}\right)$ and Cash-flows to Equity Net Debt Expected $\left(CFNE_{(E)}\right)$ with regards to the same period, as follows:

$$\gamma_3 = \left(1 - \frac{\mathrm{CFNE_{(A)}}}{\mathrm{CFNE_{(E)}}}\right) \quad 0 \leq \gamma_3 < \varepsilon \tag{7.97}$$

Note the dividends may be paid in debt. In this case the increase in debt is not used to increase the investments but to pay dividends. The company reduces its capabilities to increase the future CFO and CFE. Therefore, Cash-flows to Equity Net of Debt Increases (CFNE) are considered. In this sense, it is possible to define CFNE equal to the CFO minus the CDC, as follows:

$$\mathrm{CFNE} = \mathrm{CFO} - \mathrm{CDC} \tag{7.98}$$

And by considering Eq. (7.94), we have:

$$\mathrm{CFNE} = \mathrm{CFO} - \mathrm{D}(\alpha + \mathrm{K_D}) \tag{7.99}$$

The positive amount of the CFNE indicates the dividend amount for the shareholders; otherwise its negative amount indicates the increase in equity to cover the company's needs.

The smaller the difference between CFNE Current and Expected, the greater is the shareholders' satisfaction and higher is the company's capability to raise capital on favourable conditions in financial markets. Therefore, the coefficient γ_3 can be considered a proxy of the discipline effects of debt on management (Jensen and Meckling 1976; Jensen 1986). High debt level allows for management to invest capital in positive net present value projects that increase the CFO and the CFE with maximization of the equity value.

In this context, it is reasonable to assume:

(a) CFNE Expected (CFNE $_{(E)}$) is the maximum value for the CFNE;
(b) CFNE Current (CFNE $_{(A)}$) cannot be higher than CFNE$_{(E)}$.

Therefore, $\gamma_3 = 0$ is the maximum value of the coefficient and represents the best conditions. In this case the company maximizes the equity value because the shareholders' expectations are achieved.

Otherwise, the lower the CFNE $_{(A)}$ compared with the CFNE $_{(E)}$, the higher the coefficient value. In absence of the CFNE $_{(A)}$ the coefficient is equal $\gamma_3 = 1$. Note that the coefficient value is higher than 1 $(\gamma_3 > 1)$ if the CFNE $_{(A)}$ is negative.

By considering Eq. (7.99), Eq. (7.97) can be rewritten as follows:

$$\gamma_3 = \left[1 - \frac{CFO_{(A)} - D(\alpha + K_D)}{CFO_{(E)} - D(\alpha + K_D)}\right]$$

and then:

$$\gamma_3 = \left[\frac{CFO_{(E)} - CFO_{(A)}}{CFO_{(E)} - D(\alpha + K_D)}\right] \tag{7.100}$$

On the basis of the three *composite index coefficients* $\gamma_1, \gamma_2, \gamma_3$ as defined by Eqs. (7.95), (7.96) and (7.100) respectively, the composite index δ as defined in Eq. (7.92) can be explicated as follows:

$$\delta = \frac{D(\alpha + K_D)}{CFO_{(A)}} + \frac{D}{LA} + \frac{CFO_{(E)} - CFO_{(A)}}{CFO_{(E)} - D(\alpha + K_D)} \tag{7.101}$$

Denoting the difference between $CFO_{(E)}$ and $CFO_{(A)}$ as follows:

$$\Delta^{CFO}_{(E-A)} = CFO_{(E)} - CFO_{(A)} \tag{7.102}$$

And substituting, Eq. (7.101) can be rewritten as follows:

$$\delta = D\left[\frac{(\alpha + K_D)}{CFO_{(A)}} + \frac{1}{LA}\right] + \left[\frac{\Delta^{CFO}_{(E-A)}}{CFO_{(E)} - D(\alpha + K_D)}\right] \tag{7.103}$$

Equation (7.103) is the *composite index* (δ) able to measures the default risk of the firm on the basis of its fundamentals.

The cost of debt as defined by Eq. (7.90) can be explicated on the basis of the composite index (δ) as defined in Eq. (7.103), as follows:

$$K_D = R_F + (YC_M - R_F) + \rho e^{\left\{D\left[\frac{(\alpha + K_D)}{CFO_{(A)}} + \frac{1}{LA}\right] + \left[\frac{\Delta^{CFO}_{(E-A)}}{CFO_{(E)} - D(\alpha + K_D)}\right]\right\}} \tag{7.104}$$

Equation (7.104) shows how the cost of debt is function of three parts:

- the *risk-free rate* (R_F) that is the time value of money and is measured on the basis of the expected return of government bonds. Its analysis is developed in Chap. 10;
- the market *risk-premium* $(YC_M - R_F)$ of corporate bonds compared with government bonds on the basis of higher risk of corporate bonds than government bonds;

– the *default premium* $\left(\rho e^{\left\{D\left[\frac{(\alpha+K_D)}{CFO_{(A)}}+\frac{1}{LA}\right]+\left[\frac{\Delta^{CFO}_{(E-A)}}{CFO_{(E)}-D(\alpha+K_D)}\right]\right\}}\right)$ that measures the specific risk of default of the firm due to its inability to face debt obligations.

7.4 Cost of Capital of the Company

The estimate of capital cost of the company is one of the most difficult topics for managers and financial economists.

The most widely used model to estimate the cost of capital is the *Weighted Average Cost of Capital (WACC)*. Its equation is simple.

Denote with K_{EL} the cost of equity of the levered company (the Levered Cost of Equity) that measures the expected return of investors in equity; K_D the cost of debt of the levered company that measures the expected return of investors in debt; E the value of equity in the capital structure and $E/(E+D)$ the weight of equity in the capital structure; D the value of debt in the capital structure and $D/(E+D)$ the weight of debt in the capital structure; t_c the corporate tax of the company. The WACC estimates the cost of capital for the company firm (K_A) as follows:

$$WACC \equiv K_A = K_{EL}\left(\frac{E}{E+D}\right)+K_D\left(\frac{D}{E+D}\right) \tag{7.105}$$

and by introducing the corporate taxes (t_c), we have:

$$WACC \equiv K_A = K_{EL}\left(\frac{E}{E+D}\right)+K_D\left(\frac{D}{E+D}\right)(1-t_c) \tag{7.106}$$

The WACC is usually used in assessments to discount the expected operating cash flows of the company. However, its use is not always correct as its logical and methodological derivation is not considered. The WACC, as the most widely used models to estimate the cost of capital, derives from the Modigliani–Miller studies. Specifically, the WACC is a strict derivation of the Propositions of MM (1958).

On the basis of the Propositions, the value of the company is function of the value of its assets only and it is independent of the capital structure. The company's value is estimated on the basis of its expected operating cash flows discounted at a risk rate of the k-th class within which the company is placed. Therefore, this risk rate is completely independent from the capital structure. Consequently, the value of the company is constant compared with the capital structure. Then, the WACC is a constant compared with the capital structure and it only measures the expected return by investors based on the risk undertaken in the capital markets as defined by risk rate of the k-class of risk of the company. Formally:

$$\text{WACC} \equiv K_A \; costant$$

The K_A is the risk rate at which the operating cash flow of the company is discounted. It is defined on the basis of the rick class of the company in the capital market independently by its capital structure.

Note that if the company is financed by equity only, the WACC is equal to the Cost of Equity of Unlevered Firm and therefore the Unlevered Cost of Equity (K_{EU}):

$$\text{WACC} \equiv K_A \equiv K_{EU}$$

On the basis of these considerations, it is possible to explicit Eq. (7.105) in terms of levered cost of equity $(\mathrm{K_{EL}})$ as follows:

$$\mathrm{K_{EL}}\left(\frac{\mathrm{E}}{\mathrm{E+D}}\right) = K_A - \mathrm{K_D}\left(\frac{\mathrm{D}}{\mathrm{E+D}}\right); \quad \mathrm{K_{EL}} = \frac{K_A - \mathrm{K_D}\left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)}{\left(\frac{\mathrm{E}}{\mathrm{E+D}}\right)};$$
$$\mathrm{K_{EL}} = \left[K_A - \mathrm{K_D}\left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\right]\left(\frac{\mathrm{E+D}}{\mathrm{E}}\right); \quad \mathrm{K_{EL}} = K_A\left(\frac{\mathrm{E+D}}{\mathrm{E}}\right) - \mathrm{K_D}\left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\left(\frac{\mathrm{E+D}}{\mathrm{E}}\right);$$
$$\mathrm{K_{EL}} = K_A\left(\frac{\mathrm{E}}{\mathrm{E}}\right) + K_A\left(\frac{\mathrm{D}}{\mathrm{E}}\right) - \mathrm{K_D}\left(\frac{\mathrm{D}}{\mathrm{E}}\right)$$

and therefore:

$$\mathrm{K_{EL}} = K_A + (K_A - \mathrm{K_D})\left(\frac{\mathrm{D}}{\mathrm{E}}\right) \tag{7.107}$$

Similar by explicating Eq. (7.106) on the basis of levered cost of equity $(\mathrm{K_{EL}})$ we have:

$$K_{EL} = K_A + [K_A - K_D(1 - \mathrm{t_c})]\frac{\mathrm{D}}{\mathrm{E}} \tag{7.108}$$

Equations (7.107) and (7.108) define the levered cost of equity of the company as a direct derivation of Propositions I and II of MM.

Considering that K_A is constant and it measures the expected return of investors for operating cash flow based on the risk k-class of the company, it can be interpreted as the expected return of investors in equity of all equity financed companies. Therefore, it is possible to replace K_A with the Unlevered Cost of Equity (K_{EU}) (Massari and Zanetti 2008). In this sense, Eq. (7.107) can be rewritten as follows:

$$K_{EL} = K_{EU} + (K_{EU} - K_D)\frac{D}{E} \tag{7.109}$$

and Eq. (7.108) can be re-written as follows:

$$K_{EL} = K_{EU} + [K_{EU} - K_D(1 - t_c)]\frac{D}{E} \tag{7.110}$$

Obviously in both cases, K_{EU} is a constant in the equation and it is independent of the company's capital structure.

A relevant problem concerns the rate used to discount expected tax savings. Usually, tax savings are discounted to the cost of debt as in the Propositions of MM. The main reason is that tax savings arising from debt and subsequently the risk of expected tax savings is strictly related to the risk of debt. Therefore, the company does not achieve tax savings if it is unable to face debt obligations. Consequently, it is possible to discount the expected tax savings by the cost of debt (K_D). Several studies are placed in a critical way and highlight the need to use the unlevered cost of equity (K_{EU}) to discount the expected tax savings (Miles and Ezzell 1980; Taggart 1991; Kaplan and Ruback 1995; Ruback 2002). The main argument is that the debt level is known in the first year only. Consequently, only in this case is it possible to use the cost of debt to discount the expected tax savings. In each year other than the first, the debt level is unknown because the company levered value is unknown; it is function of the tax savings that, in turn, are function of the debt level that is unknown. Consequently, the risk of expected tax savings is similar to the risk on the value of company assets and therefore they must be discounted to the unlevered cost of equity (K_{EU}).

On the basis of these considerations and Eqs. (7.107) and (7.108) in the WACC's approach, the cost of equity cannot be estimated until the leverage ratio has been defined that, in turn, requires an estimate of the company levered value, representing the aim of the analysis in which the WACC is used as a discount rate. Consequently, the WACC can only be used if the debt level is an exogenous variable and, therefore, it is defined and well-known (Harris and Pringle 1985). In other words, the WACC can be used only if the debt level is defined. Consequently they are not useful to evaluate the change in leverage.

In this context, the *Levered Cost of Capital* (*LCC*) is proposed (De Luca 2017).

The LCC proposed is a theoretical model with normative function that tries to overcome the WACC's limits by defining the relationship, based on an exponential function between the debt level and the cost of debt for the company. This relationship is defined on the basis of the trade-off approach by considering the effects, both positive (mainly with regards to the tax shields) and negative (mainly with regards to the default risk) of the debt level on the cost of debt and therefore on the levered cost of capital. Consequently, it can be used to estimate the effects of the debt level changes on the levered cost of capital and therefore on the company value.

By using the cost of equity and the cost of debt, the LCC is function of the company's capital structure and it measures the expected return of company investors in equity and in debt.

Specifically, LCC estimates the cost of capital of the company on the basis of an exponential function between risk and debt level: on the one hand, the increase in leverage increases the benefits due to the tax shields by reducing the LCC but, on the other hand, it increases the default risk probability. Consequently, on the basis of these two effects, LCC draws a curve with a minimum point identifying the debt level that minimized the cost of capital.

The LCC can be used to define the debt level of the company by measuring its effects on the cost of capital. In fact, a change in the debt level changes the cost of debt and then the cost of capital along the curve.

The basic assumptions of LCC are the following:

(1) the capital structure is based on two capital sources: equity and debt. Hybrid forms are not considered. Also, it is assumed that the company uses a single class of equity and debt;
(2) the cost of capital measures the cost of sources invested in the company. Consequently, it can be interpreted as the expected return of investors: the expected return by investors in equity is the cost of equity for the company as the expected return by investors in debt is the cost of debt for the company;
(3) the investor is diversified;
(4) it is considered a one-period time with no events between the start and the end of the period. In each time period, the company defines its debt level;
(5) there is a single class of zero coupon risky debt of maturity;
(6) the cost of debt is defined based on the debt level at the start of the period;
(7) the investors personal taxes on debt and equity are not considered;
(8) the share of debt paid at the end of the year is based on the debt level from the start of the year.

Based on the portfolio theory (Markowitz 1952) the cost of equity can be derived from the CAPM (Sharpe 1964; Lintner 1965; Mossin 1966) as defined in Eq. (7.65) as follows:

$$\mathrm{K_E} = \mathrm{R_F} + \beta(\mathrm{R_M} - \mathrm{R_F}) \leftrightarrow \mathrm{K_E} = \mathrm{R_F} + \frac{\sigma(R_M; R_F)}{\sigma^2(R_M)}(\mathrm{R_M} - \mathrm{R_F})$$

The cost of debt is defined as in Eq. (7.104) as follows:

$$\mathrm{K_D} = \mathrm{R_F} + (\mathrm{YC_M} - \mathrm{R_F}) + \rho e^{\left\{\mathrm{D}\left[\frac{(\alpha+\mathrm{K_D})}{\mathrm{CFO_{(A)}}} + \frac{1}{\mathrm{LA}}\right] + \left[\frac{\Delta^{\mathrm{CFO}}_{(\mathrm{E-A})}}{\mathrm{CFO_{(E)}} - \mathrm{D}(\alpha+\mathrm{K_D})}\right]\right\}}$$

On the basis of the cost of equity as defined by Eq. (7.65) and the cost of debt as defined by Eq. (7.104), the Levered Cost of Capital (LCC) of the company $(\mathrm{K_L})$ can be estimated as follows:

$$K_L = K_E - L[K_E - K_D(1 - t_c)] \tag{7.111}$$

where:

$$L = \frac{D}{E + D} \rightarrow 1 - L = \frac{E}{E + D} \tag{7.112}$$

Note that on the basis of Eq. (7.112), Eq. (7.111) can be expressed in a form similar to the WACC as follows:

$$K_L = K_E\left(\frac{E}{E + D}\right) + K_D\left(\frac{D}{E + D}\right)(1 - t_c) \tag{7.113}$$

The structure of the LCC (K_L), as defined by Eq. (7.113) is only formally similar to the WACC. Indeed, they are very different in construction. There are three main differences:

- first, the WACC is defined by market as function of the risk class of the company and it is used to discount the unlevered cash flows assuming an all-equity financed company. On the other hand, the LCC is based on an estimate of the cost of equity and the cost of debt on the basis of the expected return of the investors in equity and debt, respectively. Then, the LCC is a direct function of the company capital structure. It is used to discount unlevered cash flows of the company because they are the source to satisfy expectations of investors in equity and debt;
- second, the WACC is based on the assumption of a linear relationship between the cost of capital and leverage. Differently, the LCC is estimated on the basis of a non-linear relationship between leverage and the cost of capital. The LCC is estimated by assuming an exponential function between risk and debt level. The increase of leverage, on the one hand, increases the benefits due to tax shields by reducing the LCC but, on the other hand, it increases the default risk probability. Based on these two effects, the LCC draws a curve with a minimum point that defines the debt level that minimizes the cost of capital;
- third, in WACC the debt level is assumed constant and well-known. Differently, in the LCC the definition of debt level is function of its effects on the cost of capital. By a change in the debt level, there is a change in the cost of debt and therefore the cost of capital along the curve. Also, this is the main difference between LCC and the ME's cost of capital model.

By substituting Eqs. (7.65) and (7.104), in Eq. (7.111) the LCC (K_L) is equal to:

$$\begin{aligned}K_L &= R_F + \beta(R_M - R_F)\\ &\quad - L\Bigg[(R_F + \beta(R_M - R_F))\\ &\quad - \left(R_F + (YC_M - R_F) + \rho e^{D\left(\frac{(\alpha+K_D)}{CFO_{(A)}}+\frac{1}{LA}\right)+\left(\frac{\Delta^{CFO}_{(E-A)}}{CFO_{(E)}-D(\alpha+K_D)}\right)}\right)(1-t_c)\Bigg]\end{aligned} \tag{7.114}$$

Placing:

- $A = K_E = R_F + \beta(R_M - R_F)$;
- $F = R_F + (YC_M - R_F)$;
- $H = \frac{\alpha + K_{D(-1)}}{CFO_{(A)}} + \frac{1}{LA}$;
- $I = \Delta^{CFO}_{(E-A)} = CFO_{(E)} - CFO_{(A)}$
- $M = CFO_{(E)}$;
- $N = \alpha + K_D$.

Equation (7.114) can be simplified as follows:

$$K_L = A - L\Big[A - \Big(\Big(F + \rho e^{\left(DH + \frac{I}{M-DN}\right)}\Big)(1 - t_c)\Big)\Big] \tag{7.115}$$

or in an equivalent form:

$$K_L = A - \left(\frac{D}{E+D}\right)\left\{A - \left[\left(F + \rho e^{\left(DH + \frac{I}{M-DN}\right)}\right)(1 - t_c)\right]\right\} \tag{7.116}$$

Equations (7.115) or (7.116) draws a function continuously and differentiable at least twice by construction. The main problem is to prove that the function has a minimum point. Indeed, in this case it defines the minimum level of the LCC and then the minimum level of the cost of capital of the company. Consequently, on the basis of effects of the debt level changes on the LCC, the company can choose the debt level that minimizes the LCC.

The first derivative compared with the debt of Eq. (7.116) is the following:

$$\begin{aligned}
\mathrm{K_L} &= \mathrm{A} - \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\left\{\mathrm{A} - \left[\left(\mathrm{F} + \rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right)(1 - \mathrm{t_c})\right]\right\} \\
&= \mathrm{A} - \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\left\{\mathrm{A} - \left[\mathrm{F} - \mathrm{Ft_c} + \rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)} - \mathrm{t_c}\rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right]\right\} \\
&= \mathrm{A} - \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\left[\mathrm{A} - \mathrm{F} + \mathrm{Ft_c} - \rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)} + \mathrm{t_c}\rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right] \\
&= \mathrm{A} - \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)A + \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)F - \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\mathrm{Ft_c} + \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)} \\
&\quad - \left(\frac{\mathrm{D}}{\mathrm{E+D}}\right)\mathrm{t_c}\rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}
\end{aligned}$$

By considering that:

- $\frac{\partial}{\partial D}[A] = 0;$
- $\frac{\partial}{\partial D}\left[\frac{D}{E+D}\right] = \frac{E+D-D}{(E+D)^2} = \frac{E}{(E+D)^2};$
- $\frac{\partial}{\partial D}[DH] = H;$
- $\frac{\partial}{\partial D}\left[\frac{I}{M-DN}\right] = \frac{-I(-N)}{(M-DN)^2} = \frac{IN}{(M-DN)^2};$
- $\frac{\partial}{\partial D}\left[\mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right] = \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H} + \frac{IN}{(M-DN)^2}\right);$
- $\frac{\partial}{\partial D}\left[\rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right] = \rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H} + \frac{IN}{(M-DN)^2}\right);$
- $\frac{\partial}{\partial D}\left[\rho \mathrm{t_c} \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right] = \rho \mathrm{t_c} \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H} + \frac{IN}{(M-DN)^2}\right).$

We have:

$$\begin{aligned}
\frac{\partial}{\partial D}[\mathrm{K_L}] = &-\left[\frac{E}{(E+D)^2}\right]A + \left[\frac{E}{(E+D)^2}\right]F - \left[\frac{E}{(E+D)^2}\right]F\mathrm{t_c} \\
&+ \left\{\left[\frac{E}{(E+D)^2}\right]\rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)} + \left(\frac{D}{E+D}\right)\rho \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H} + \frac{IN}{(M-DN)^2}\right)\right\} \\
&- \left\{\left[\frac{E}{(E+D)^2}\right]\rho \mathrm{t_c} \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)} - \left(\frac{D}{E+D}\right)\rho \mathrm{t_c} \mathrm{e}^{\left(\mathrm{DH} + \frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H} + \frac{IN}{(M-DN)^2}\right)\right\}
\end{aligned}$$

And then:

$$\frac{\partial}{\partial D}[\mathrm{K_L}] = -\left[\frac{E}{(E+D)^2}\right]A + \left[\frac{E}{(E+D)^2}\right]F - \left[\frac{E}{(E+D)^2}\right]F\mathrm{t_c} + \left[\frac{E}{(E+D)^2}\right]\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}$$
$$+\left(\frac{D}{E+D}\right)\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H}+\frac{IN}{(M-DN)^2}\right) - \left[\frac{E}{(E+D)^2}\right]\rho \mathrm{t_c} e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}$$
$$-\left(\frac{D}{E+D}\right)\rho \mathrm{t_c} e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H}+\frac{IN}{(M-DN)^2}\right)$$

And then:

$$\frac{\partial}{\partial D}[\mathrm{K_L}] = \left[\frac{E}{(E+D)^2}\right]\left\{-A+F-F\mathrm{t_c}+\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)} - \rho \mathrm{t_c} e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right\}$$
$$+\left(\frac{D}{E+D}\right)\left\{\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H}+\frac{IN}{(M-DN)^2}\right)\right.$$
$$\left.-\rho \mathrm{t_c} e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H}+\frac{IN}{(M-DN)^2}\right)\right\}$$

And then:

$$\frac{\partial}{\partial D}[\mathrm{K_L}] = \left[\frac{E}{(E+D)^2}\right]\left\{F(1-\mathrm{t_c})+\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}(1-\mathrm{t_c})-A\right\}$$
$$+\left(\frac{D}{E+D}\right)\left\{\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H}+\frac{IN}{(M-DN)^2}\right)(1-\mathrm{t_c})\right\}$$

And then:

$$\frac{\partial}{\partial D}[\mathrm{K_L}] = \left[\frac{E}{(E+D)^2}\right]\left\{\left[\left(F+\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\right)(1-\mathrm{t_c})\right]-A\right\}$$
$$+\left(\frac{D}{E+D}\right)\left\{\rho e^{\left(\mathrm{DH}+\frac{\mathrm{I}}{\mathrm{M-DN}}\right)}\left(\mathrm{H}+\frac{IN}{(M-DN)^2}\right)(1-\mathrm{t_c})\right\} \tag{7.117}$$

Assume that the first derivative is a function defined and continued for each debt. By assuming that the debt level can be defined between 0 (all-equity financed) and 1 (all-debt financed), so that the sum of equity and debt is equal to 1, as follows:

$$\forall \mathrm{D} \in [0;1] \leftrightarrow \mathrm{E}+\mathrm{D}=1 \tag{7.118}$$

By using the intermediate value theorem, if the first derivative is negative in $D=0$ and it is positive in $D=1$, there is a debt level D^* between 0 and 1 in which the first derivative is equal to zero. Formally:

$$\left\{\frac{\partial \mathrm{K_L}}{\partial \mathrm{D}}(\mathrm{D}=0)<0 \quad and \quad \frac{\partial \mathrm{K_L}}{\partial \mathrm{D}}(\mathrm{D}=1)>0\right\} \Rightarrow \left\{\exists \mathrm{D}^* \in (0,1):\ \frac{\partial \mathrm{K_L}}{\partial \mathrm{D}}(\mathrm{D}^*)=0\right\} \tag{7.119}$$

Furthermore, if the first derivative is negative on the left of D^* and it is positive on the right, there is a minimum point necessary (D^*).

Based on Eq. (7.118) for $D=0(E=1)$ the company is all-equity financed; otherwise, for $D=1(E=0)$ the company is all debt financed. The value of the first derivative can be studied for $D=0$ and $D=1$. These are the two points that define the function fields.

The value of the first derivative for $D=0$ is always *negative*, as follows:

$$D=0 \rightarrow \frac{\partial K_L}{\partial D}=\left(\frac{1}{E}\right)[F(1-t_c)-A] \tag{7.120}$$

Indeed, the equity risk is greater than the debt risk by definition, and therefore:

$$\mathrm{A} \equiv \mathrm{K_E}=\mathrm{R_F}+\beta_\mathrm{E}(\mathrm{R_M}-\mathrm{R_F})>F=\mathrm{R_F}+(\mathrm{YC_M}-\mathrm{R_F}) \tag{7.121}$$

Differently, the value of the first derivative for $D=1$ is always *positive*, as follows:

$$\mathrm{D}=1 \rightarrow \frac{\partial \mathrm{K_L}}{\partial \mathrm{D}}=\rho e^{\left(\mathrm{H}+\frac{\mathrm{I}}{\mathrm{M-N}}\right)}\left(\mathrm{H}+\frac{\mathrm{IN}}{(\mathrm{M}-\mathrm{N})^2}\right)(1-\mathrm{t_c}) \tag{7.122}$$

Indeed, all terms of Eq. (7.120) are positive:

- $0<\rho\leq 1 \rightarrow \rho>0 \forall D$. Theoretically $\rho \in (0;1)$. But if the company is all-debt financed, assuming $\rho=0$ is equivalent to assume a risk-free debt and it is difficult to assume. In this case it is right to assume $\rho=1$;
- $H=\frac{\alpha+K_D}{CFO_{(A)}}+\frac{1}{LA}>0\ \forall D$, because all element are positive by definition;
- $IN=\Delta^{CFO}_{(E-A)}(\alpha+K_D)>0\ \forall D$, because the first term is positive based on the assumption about CFO Current and Expected and the second term is always positive by definition;
- $e^{\left(H+\frac{I}{M-N}\right)}>0\ \forall D$, by definition;
- $(M-N)^2=\left(CFO_{(E)}-(\alpha+K_D)\right)^2>0\ \forall D$, by definition.

On the basis of Eqs. (7.120) and (7.122), it can be proved that there is a minimum point in the function LCC.

Explicating all variables, Eq. (7.117) can be rewritten as follows:

$$\frac{\partial K_L}{\partial D} = \left(\frac{E}{(E+D)^2}\right)\Bigg\{[(R_F + (YC_M - R_F))(1-t_c) - (R_F + \beta_E(R_M - R_F))].$$
$$+ \rho e^{\left[D\left(\frac{\alpha+K_D}{CFO_{(A)}}+\frac{1}{LA}\right)+\frac{\Delta^{CFO}_{(E-A)}}{CFO_{(E)}-D(\alpha+K_D)}\right]}(1-t_c)\Bigg\}$$
$$+ \left(\frac{D}{E+D}\right)\rho e^{\left[D\left(\frac{\alpha+K_D}{CFO_{(A)}}+\frac{1}{LA}\right)+\frac{\Delta^{CFO}_{(E-A)}}{CFO_{(E)}-D(\alpha+K_D)}\right]}\left[\left(\frac{\alpha+K_D}{CFO_{(A)}}+\frac{1}{LA}\right)\right.$$
$$\left.+\frac{\Delta^{CFO}_{(E-A)}(\alpha+K_D)}{\left(CFO_{(E)}-D(\alpha+K_D)\right)^2}\right](1-t_c) \tag{7.123}$$

The minimum point can be found by searching the root of the derivate in Eq. (7.123) by using numerical methods.

Therefore, the function of the LCC draws a curve with a minimum point as in Fig. 7.1. where the ordinate is the cost of capital and the abscissa is the debt level (D).

The curve reflects the combined effects of the stock market rates and obligations (exogenous variable) and the company's specific default risk (endogenous variable). There are two main movements of the curve:

– the first refers to movement of the D^* point between 0 (all-equity financed) and 1 (all-debt financed) along the abscissa. This movement is mainly due to the company's fundamentals. If the company's default risk increases, point D^* moves to the left; otherwise, if it reduces, point D^* moves to the right. The left

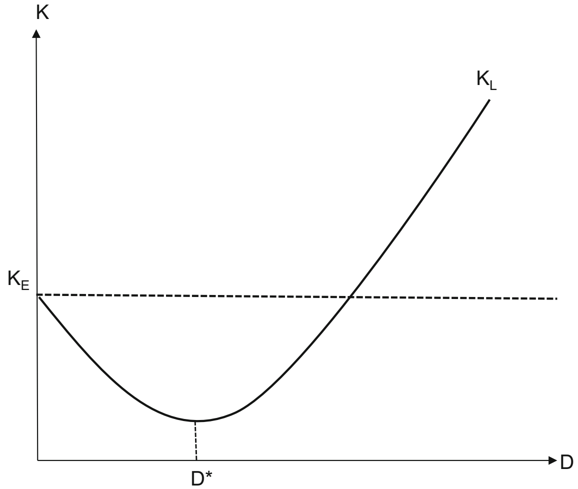


Fig. 7.1 The debt level that minimizes the Levered Cost of Capital (K_L)

movement reduces the company's ability to face debt while the right movements increased it. It is mainly function of the operating cash flows and their capabilities to face debt obligations;

– the second refers to movement of the curve upwards and downwards along the ordinate. In this case it is mainly due to the market rates of stocks and obligations. Generally, by isolating the rate movements if: R_M or YC_M increase, the curve shift upwards; R_M or YC_M decrease, the curve shift downwards. The effect of the change R_F is due to the coefficient beta on equity and debt. It is relevant to note that if the distance between R_M and YC_M increases, ceteris paribus, the curve shifts upwards and point D^* moves to the right.

References

Agarwal V, Taffler RJ (2007) Twenty-five years of the Taffler Z-score model: does it really have predictive ability? Account Bus Res 37(4):285–300

Aghion P, Bolton P (1992) An incomplete contracts approach to financial contracting. Rev Econ Stud 59(3):473–494

Alexander GJ, Benston PG (1982) More on beta as a random coefficient. J Financ Quant Anal 17(1):27–36

Alexander GJ, Chervany NL (1980) On the estimation and stability of Beta. J Financ Quant Anal 15(1):123–138

Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Finance 23(4):589–609

Altman EI, Haldeman RG, Narayanan P (1977) ZETAM analysis: a new model to identify bankruptcy risk of corporations. J Bank Finance 1(1):29–54

Altman EI, Hartzell J, Peck M (1995) Emerging markets corporate bonds: a scoring system. Salomon Brothers, New York, NY

Amihud Y, Lev B (1981) Risk reduction as a managerial motive for conglomerate mergers. Bell J Econ 12:605–6017

Baker M, Wurgler J (2002) Market timing and capital structure. J Finance 57:1–32

Bancel F, Mittoo UR (2004) Cross-country determinants of capital structure choice: a survey of European firms. Financ Manage 33(4):103–132

Baninoe R (2010) Corporate bankruptcy prediction and equity returns in the UK. Cranfield School of Management, Cranfield University

Bao J (2009) Structural models of default and the cross section of corporate bond yield spreads. Working Paper, MIT

Beaver WH (1966) Financial ratios as predictors of failure. J Account Res 4(3):179–192

Beaver WH (1968) Alternative accounting measures as predictors of failure. J Account Res 6(1):113

Beaver W, Kettler P, Scholes M (1970) The association between market determined and accounting determined risk measures. Account Rev 45:654–682

Beck T, Demirgüç-Kunt A, Maksimovic V (2002) Financing patterns around the world: the role of institutions. World Bank Policy Research Working Paper 2905

Bergh DD (1995) Size and relatedness of units sold. Strateg Manage J 16:221–240

Berk J, DeMarzo P (2008) Corporate finance. Pearson Education, Inc., London

Berkowitz D, Pistor D, Richard JF (2003) Economic development, legality, and the transplaneteffect. Eur Econ Rev 47:165–195

Bethel JE, Liebeskind J (1993) The effects of ownership structure on corporate restructuring. Strateg Manage J 14:15–31

Bhagat S, Bolton B, Subramanian A (2011) Manager characteristics and capital structure: theory and evidence. J Financ Quant Anal 46:1581–1627

Bhatia U (1988) Predicting corporate sickness in India. Stud Bank Finance 7:57–71

Black F, Scholes M (1973) The pricing of options and corporate liabilities. J Polit Econ 81:637–654

Blume ME (1970) Portfolio theory: a step toward its practical application. J Bus 43(2):152–1573

Blume MW (1975) Betas and their regression tendencies. J Finance 30(3):785–795

Booth L, Aivazian V, Demirgüç-Kunt A, Maksimovic V (2001) Capital structure in developing countries. J Finance 55:87–130

Brander J, Lewis T (1986) Capital structure and product market behaviour: the limited liability effect. Am Econ Rev 76:956–970

Brealey R.A., Myers S.C., Allen F., (2016), *Principles of Corporate Finance*, 12 ed., McGraw-Hill

Brounen D, De Jong A, Koedijk K (2006) Capital structure policies in Europe: survey evidence. J Bank Finance 30:1409–1442

Chen L, Lesmond DA, Wei J (2007) Corporate yield spreads and bond liquidity. J Finance 62:119–149

Chevalier J, Scharfstein D (1996) Capital markets imperfections and countercyclical markups: theory and evidence. Am Econ Rev 86:703–725

Choe H, Masulis R, Nanda V (1993) Common stock offerings across the business cycle: theory and evidence. J Empir Finance 1(1):3–31

Claessens S, Klapper LF (2005) Bankruptcy around the World: explanations of its relative use. American Law and Economics Review 7(1):253–283

Cremers M, Driessen J, Maenhout P (2008) Explaining the level of credit spreads: option implied jump risk premia in a firm value model. Rev Financ Stud 21:2209–2242

Damodaran A (2012) Investment valuation: tools and techniques for a determining the value of any assets, 3rd edn. Wiley, London

Dasgupta S, Titman S (1998) Pricing strategy and financial policy. Rev Financ Stud 4:705–737

De Jong A, Kabir R, Nguyen TT (2008) Capital structure around the World: the roles of firm and country specific determinants. J Bank Finance 32:1954–1969

De Luca P (2015) The debt choices of the firms in developed countries: evidence from G-7. Int J Econ Finance 7(4):122–134

De Luca P (2017a) Debt level and the firm levered cost of capital. Int J Econ Financ Issue 7(5):475–484

De Luca P (2017b) The company fundamental analysis and the default risk ratio. Int J Bus Manage 12(10):79–90

Demirgüç-Kant A, Maksimovic V (1998) Law, finance, and firm growth. J Finance 53:2107–2137

Demirgüç-Kant A, Maksimovic V (1999) Institutions, financial markets and firm debt maturity. J Financ Econ 54:295–336

Diamond DW (1989) Reputation acquisition in debt markets. J Polit Econ 97:828–862

Duffie D, Singleton K (1999) Modeling term structures of defaultable bonds. Rev Financ Stud 12(4):197–226

Elton EJ, Gruber MJ (1974) Portfolio theory when investment relatives are lognormally distributed. J Finance 29(4):1265–1273

Elton EJ, Gruber MJ, Urich T (1978) Are betas best? J Finance 13(5):1375–1384

Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN (2013) Modern portfolio theory and investment analysis, 9th edn. Wiley, New York

Eom YH, Helwege J, Huang JZ (2004) Structural models of corporate bond pricing: an empirical analysis. Rev Financ Stud 17:499–544

Fama E, French K (1993) Common risk factors in the return on stocks and bonds. J Financ Econ 33:3–56

Fama E, French K (1998) Taxes, financing decisions, and firm value. J Finance 53:819–843

Fama E, French K (2002) Testing trade-off and pecking order predictions about dividends and debt. Rev Financ Stud 15:1–33
Fan J, Titman S, Twite G (2012) An international comparison of capital structure and debt maturity choices. J Financ Quant Anal 47:23–56
Frank MZ, Goyal VK (2003) Testing the pecking order theory of capital structure. J Financ Econ 67:217–248
Frank MZ, Goyal VK (2009) Capital structure decisions: which factors are reliably important? Financ Manage 38:1–37
Frankel R, Lee CMC (1998) Accounting valuation, market expectation, and cross-sectional stock returns. J Account Econ 25:283–319
Friend I, Hasbrouck J (1988) Determinants of capital structure. Res Finance 7:1–19
Gonedes NJ (1973) Evidence on the information content of accounting numbers: accounting-based and market-based estimates of systemic risk. J Financ Quant Anal 8:407–443
Graham JR (1999) Do personal taxes affect corporate financing decisions? J Public Econ 73:41–73
Graham JR (2000) How big are the tax benefits of debt? J Finance 55:1901–1941
Graham JR, Harvey CR (2001) The theory and practice of corporate finance: evidence from the field. J Financ Econ 60:187–243
Grossman S, Hart O (1982) Corporate finance structure and managerial incentives. In: The economics of information and uncertainty, National Bureau of Economic Research, Inc.
Gupta V (2013) A Unified credit risk model for predicting default. Prajnan, 42(3):2015–2045
Hall GC, Hutchinson PJ, Michaelas N (2004) Determinants of the capital structures of European SMEs. J Bus Finance Account 31:711–728
Hamada RS (1972) The effect of the firm's capital structure on the systemic risk of common stocks. J Finance 27(2), 435–452
Handa P, Kothari SP, Wasley C (1989) The relationship between the return interval and betas: implications for the size effect. J Financ Econ 23(1):79–101
Harris M, Raviv A (1991) The theory of capital structure. J Finance 1:297–355
Harris RS, Pringle JJ (1985) Risk-adjusted discount rates-extensions from the average-risk case. J Financ Res 8(3):237–244
Hart O, Moore J (1995) Debt and seniority: an analysis of the role and hard claims in constraining management. Am Econ Rev 85(3):567–585
Hill NC, Stone BK (1980) Accounting Betas, systematic operating risk, and financial leverage: a risk-composition approach to the determinants of systemic risk. J Financ Quant Anal 15(3):595–638
Hoskisson RE, Johnson JRA, Moesel DD (1994) Corporate divestiture intensity in restructuring firms: effects of governance, strategy and performance. Acad Manage J 37:1207–1238
Hovakimian A (2006) Are observed capital structures determined by equity market timing? J Financ Quant Anal 41:221–243
Huang JZ, Huang M (2012) How much of the corporate-treasury yield spread is due to credit risk? Rev Asset Pricing Stud 2(2):153–202
Ikenberry D, Lakonishok J, Vermaelen T (1995) Market under reaction to open market share repurchases. J Financ Econ 39:181–208
Jarrow R, Turnbull MT (1995) Pricing derivatives on financial securities subject to credit risk. J Finance 50(1):53–85
Jensen MC (1986) Agency cost of free cash flow, corporate finance and takeovers. Am Econ Rev 76:323–329
Jensen MC, Meckling W (1976) Theory of the firm: managerial behavior, agency costs, and ownership structure. J Financ Econ 3:305–360
Jung K, Kim YC, Stulz RM (1996) Timing, investment opportunities, managerial discretion, and security issue decision. J Financ Econ 42:159–185
Kaplan S, Ruback R (1995) The valuation of cash flow forecast: an empirical analysis. J Finance 4:1059–1093
Kim J, Ramaswamy K, Sundaresan S (1993) Does default risk in coupons affect the valuation of corporate bonds? A contingent claims model. Financ Manage 177–131

Klemkosky R, Martin J (1975) The effect of market risk on portfolio diversification. J Finance 10(1):147–153
Korajczyk RA, Levey A (2003) Capital structure choice: macroeconomic conditions and financial constraints. J Financ Econ 68:75–109
Korajczyk RA, Lucas D, McDonald R (1991) The effects of information releases on the pricing and timing of equity issue. Rev Financ Stud 4:685–708
Korajczyk RA, Lucas D, McDonald R (1992) Equity issue with time-varying asymmetric information. J Financ Quant Anal 27:397–417
Kraus A, Litzenberger RH (1973) A state-preference model of optimal financial leverage. J Finance 28:911–922
Kumar RG, Kumar K (2012) A comparison of bankruptcy models. Int J Market Financ Serv Manage Res 1(4)
La Porta R (1996) Expectations and the cross section of stock returns. J Finance 51:1715–1742
La Porta R, Lopez-De-Silanes F, Shleifer A, Vishny RW (1997) Legal determinants of external finance. J Finance 52:1131–1150
La Porta R, Lopez-De-Silanes F, Shleifer A, Vishny RW (1998) Law and finance. J Polit Econ 106:1113–1155
La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1999) The quality of government. J Law Econ Organ 15:222–279
La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (2002) Investor protection and corporate valuation. J Finance 57:1147–1170
Lane PJ, Cannella AA Jr, Lubatkin MH (1998) Agency problems as antecedents to unrelated mergers and diversification: Amihud and lev reconsidered. Strateg Manage J 19:555–578
Leland HE, Toft KB (1996) Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. J Finance 51:987–1019
Lennox C (1999) Identifying failing companies: a re-evaluation of the logit, probit, and DA approaches. J Econ Bus 15:347–364
Levy RA (1971) On the short-term, stationary of beta coefficients. Financ Anal J 27(5):55–62
Levy RA (1974) Beta coefficients as predictors of return. Financ Anal J 30(1):61–69
Lintner J (1965) Security prices, risk, and maximal gains from diversification. J Finance 4:587–615
Logue D, Merville L (1972) Financial policy and market expectations. Financ Manage 1:37–44
Loughran T, Ritter J (1995) The new issues puzzle. J Finance 50:23–51
Lucas D, MacDonald R (1990) Equity issues and stock price dynamics. J Finance 45:1019–1043
Markowitz H (1952) Portfolio selection. J Finance 1:77–91
Marsh P (1982) The choice between equity and debt: an empirical study. J Finance 37:121–144
Massari M, Zanetti L (2008) Valutazione. Fondamenti teorici e best practice nel settore industriale e finanziario. McGraw-Hill, Milano
Merton RC (1974) On the pricing of corporate debt: the risk structure of interest rates. J Finance 29(2):449–470
Miller MH (1977) Debt and taxes. J Finance 32(2):261–275
Miles JA, Ezzell JR (1980) The weighted average cost of capital, perfect capital markets, and project life: a clarification. J Financ Quant Anal 15(3):719–730
Miller MH, Scholes M (1978) Dividends and taxes. J Financ Econ 6(4):333–364
Modigliani F (1982) Debt, dividend policy, taxes, inflation and market valuation. J Finance 37(2): 255–273
Modigliani F, Miller MH (1958) The cost of capital, corporate finance and the theory of investment. Am Econ Rev 48:261–297
Modigliani F, Miller MH (1961) Dividend policy, growth, and the valuation of shares. J Bus 34(4):411–433
Modigliani F, Miller MH (1963) Corporate income taxes and the cost of capital: a correction. Am Econ Rev 53:433–443
Morellec E (2004) Can managerial discretion explain observed leverage ratios? Rev Financ Stud 17:257–294

Morellec E, Schurhoff N (2010) Dynamic investment and financing under personal taxation. Rev Financ Stud 23:101–146
Mossin J (1966) Equilibrium in a capital asset market. Econometrica 34:768–783
Myers SC (1977) Determinants of corporate borrowing. J Financ Econ 5:147–175
Myers SC (1984) The capital structure puzzle. J Finance 39:575–592
Myers SC (2001) Capital structure. J Econ Perspect 15:81–102
Myers SC (2003) Financing of corporations. In: Constantinides G, Harris M, Stulz R (eds) Handbook of the economics of finance, pp 215–253
Myers SC, Majluf NS (1984) Corporate financing and investment decision when firms have information that investors do not have. J Financ Econ 13:187–221
Noe T (1988) Capital structure and signalling game equilibria. Rev Financ Stud 1:331–356
Noe T, Rebello MJ (1996) Asymmetric information, managerial opportunism, financing, and payout policies. J Finance 51:637–660
Ohlson JA (1980) Financial ratios and the probabilistic prediction of bankruptcy. J Account Res 18(1):109–131
Pagano M, Panetta F, Zingales L (1998) Why do companies go public? An empirical analysis. J Finance 53:27–64
Pearson CA, Titman S (2008) Empirical capital structure: a review. Found Trends Finance 3:1–93
Rajan RG, Zingales L (1995) What do we know about capital structure: some evidence from international data. J Finance 50:1421–1460
Rajan RG, Zingales L (1998) Financial dependence and growth. Am Econ Rev 88:559–587
Rajan RG, Zingales L (2003) The great reversals: the politics of financial development in the twentieh century. J Financ Econ 69:559–586
Rajaratnam B, Rajaratnam K, Rajaratnam M (2017) A theoretical model for the term structure of corporate credit based on competitive advantage. Eur Financ Manage 23(2):183–210
Rauh J, Sufi A (2010) Capital structure and debt structure. Rev Financ Stud 23:4242–4280
Ritter J (2003) Investment banking and security issuance. In: Constantinides G, Harris M, Stulz M (eds) Handbook of the economics of finance. Elsevier, Amsterdam
Roenfeldt R, Griepentrog G, Pflaum C (1978) Further evidence on the stationarity of beta coefficients. J Financ Quant Anal 13(1):117–121
Rosenberg B, Guy J (1976a) Predictions of beta from investment fundamentals. Financ Anal J 32(3):60–72
Rosenberg B, Guy J (1976b) Predictions of beta from investment fundamentals: part ii. Financ Anal J 32(3):62–70
Rosenberg B, McKibben W (1973) The prediction of systematic and specific risk in common stocks. J Financ Quant Anal 8(2):317–333
Ross SA, Westerfield R, Jaffe J (1997) Corporate finance, Irwin, Homewood
Ruback R (2002) Capital cash flows: a simple approach to valuing risky cash flow. Financ Manage, Summer, pp 85–103
Sahoo PK, Mishra KC, Mayadnar S (1996) Financial ratios as the forewarning indicators of corporate health. Finance India 10(4):955–965
Schaefer SM, Strebulaev IA (2008) Structural models of credit risk are useful: evidence from hedge ratios on corporare bonds. J Financ Econ 20:1–19
Scholes M, Williams J (1977) Estimating betas from non-synchronous data. J Financ Econ 5(3):309–328
Scott E, Brown S (1980) Biased estimators and unstable betas. J Finance 35(1):49–56
Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. J Finance 3:425–442
Shleifer A, Vishny RW (1997) A survey of corporate governance. J Finance 52:737–783
Shumway T (2011) Forecasting bankruptcy more accurately: a simple hazard model. J Bus 74(1):101–124
Shyam-Sunder L, Myers SC (1999) Testing static trade-off against pecking order models of capital structure. J Financ Econ 51:219–244
Stulz R (1990) Managerial discretion and optimal financing policies. J Financ Econ 26:3–27

Taffler RJ (1984) Empirical models for the monitoring of UK corporations. J Bank Finance 8(2):199–227

Taggart RA (1977) A model of corporate financing decisions. J Finance 32:1467–1484

Taggart RA (1991) Consistent valuation and cost of capital expressions with corporate and personal taxes. Financ Manage 3:8–20

Theobald M (1981) Beta stationary and estimation period: some analytical results. J Financ Quant Anal 16(5):747–758

Titman S, Wessels R (1988) The determinants of capital structure choice. J Finance 43:1–19

Vesicek OA (1973) A note on using cross-sectional information in Bayesian estimation of security betas. J Finance 28(5):1233–1239

Wurgler J (2000) Financial markets and the allocation of capital. J Financ Econ 58:187–214

Young SD, Berry MA, Harvey DW, Page JR (1991) Macroeconomic forces, systemic risk, and financial variables: an empirical investigation. J Financ Quant Anal 26(4):559–565

Zmijewski ME (1984) Methodological issues related to estimation of financial distress predicting models. J Account Res 22:59–82

Zwiebel J (1996) Dynamic capital structure under managerial entrenchment. Am Econ Rev 86:1197–1215

Chapter 8
Equity Valuation

Abstract Company value is a function of its ability to create positive performance in the future. The value of the company is equal to the current value of expected future cash flows and the cost of capital is used as a discount rate. There are three main variables: (i) *Time*: the value of the company is strictly related to future performance rather than to past performance; (ii) *Cash-flows*: the expected future cash-flows from operations and equity; (iii) *Cost of capital*: it defines the discount rate for expected future cash-flows. In the evaluation process, two perspectives can be used: (i) Equity side, in which the equity value is estimated; (ii) Asset side, in which the enterprise value is estimated. This Chapter focuses on the Equity Valuation, while the next Chapter focuses on the Enterprise Valuation. The Equity Value is estimated on the basis of free cash-flows to equity discounted at the cost of equity.

8.1 The General Equation of Value

The company's value is function of its ability to achieve positive performance in the future. Specifically, the value of the company is related to expected future cash flows: the company's value is equal to the present value of expected future cash flows and the cost of capital is used as discount rate (Williams 1938; Modigliani and Miller 1958; Benninga 2014; Berk and DeMarzo 2008; Brealey et al. 2016; Copeland et al. 2004; Damodaran 2012, 2015; Fuller and Farrell 1987; Hillier et al. 2016; Koller et al. 2015; Vernimmen et al. 2014; Altaman 1969; Bower and Bower 1970).

The price of common stock is a function of the company's dividends, cash-flows, risk, cost of capital and growth rate.

The valuation model combines these variables to estimate the value of the company. Specifically, the valuation model is a formalization of the relationships expected to exist between these economic variables.

The formalization of a model requires a lot of inputs and the knowledge of relationships among them in a forecasting perspective. Variables considered are systematically collected among them in the model.

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_8

Therefore, based on this definition, there are three main variables:

(1) *Time*: the referenced time is the future. The value of the company is strictly related to future performance rather than past performance;
(2) *Cash-flows*: company performance is measured in cash-flows terms. Specifically, the expected future cash-flows from operations and to equity;
(3) *Cost of capital*: it is the cost of debt and the cost of equity and it defines the discount rate for expected future cash-flows.

By following a financial approach, the *General Equation of value* can be defined, based on these three main variables as follows:

$$W_F = \sum_{t=1}^{\infty} \frac{CF_t}{(1+k)^t} \tag{8.1}$$

where:

- W_F: is the company's value;
- t: is the period-time of valuation;
- CF_t: is the expected future cash-flows for each year (t). Note that they refer to the expected value of cash-flows but in order to simplify the formalization the operator $\mathrm{E}[CF_t]$ is not used, by the meaning is the same;
- k: is the cost of capital used as a discounted rate.

The Eq. (8.1) has a great theoretical relevance. It estimates the value of the company based on expected cash flows, arising from the fundamental analysis of the company and the cost of capital. Also the equation defines the relationship between company value, the expected cash flow and the cost of capital in the time of valuation: the company's value increases together with an increase in the expected cash flow and decreases together with an the increase in the cost of capital.

Unfortunately, the general equation has a relevant theoretical importance but it is not applicable directly. There are two main problems to be solved before:

- the valuation time-period;
- the valuation perspective.

Valuation Time-Period

The first problem is the definition of the valuation time-period. In Eq. (8.1) time goes from 1 $(t = 1)$ to infinite $(t = n)$. Therefore, the Eq. (8.1) is not directly applicable. The problem can be solved by dividing the valuation time-period in two conceptual parts:

- *definite time-period*: it is the time period of *analytic valuation*. Generally, this time period is equal to 3 or 5 years on the basis of company characteristics and its market, and it defines the time period of the business plan;
- *indefinite time-period*: it is the time period of *synthetic valuation*. It goes from the end of time-period of analytic valuation to infinity by using the Terminal

Value (TV). Generally, the Terminal Value measures the company's value after the analytic valuation. Estimation of the Terminal Value is one of the most relevant problems in company valuation. There are two main approaches to estimate the Terminal Value:

- *going concern approach*: it is assumed that the company continues to deliver cash flows in perpetuity. In this case, the Terminal Value reflects the value of the company after the end of the analytic valuation in perpetuity;
- *liquidation approach*: an end-time of company life is assumed together with the sale of the assets. In this case the Terminal Value reflects the liquidation value of the company. There are two main ways of estimating the liquidation value: the first, is to estimate the book value of the assets at the end of the analytic valuation and adjust it for inflation during the period; the second, is to estimate the value of the earning power of the assets through the present value of their expected cash flows.

Among these two approaches, the first (going concern approach) is more commonly used than the second (liquidation approach). The indefinite life of the company is usually assumed. The company may decide the end of its life or the company may fail, but these two events are not planned. Therefore, in this context the focus will be on an estimation of the Terminal Value based on the assumption of the ongoing concern.

By distinguishing between the analytical valuation and the synthetic valuation, the Eq. (8.1) can be rewritten as follows:

$$W_F = \sum_{t=1}^{n} \frac{CF_t}{(1+K)^t} + \frac{TV_n}{(1+K)^n} \tag{8.2}$$

where TV_n indicates the Terminal Value at the end $(t = n)$ of the period of the analytical valuation.

Therefore, while the first part of the equation estimates the company value in a given time period (analytic value of the company), the second part of the equation estimates the company value in an indefinite time period by using the Terminal Value (synthetic value of the company).

It is possible to summarize these two different periods on value as follows:

$$\begin{aligned} \text{Company Value}(\mathrm{W_F}) = {} & \text{Present Value of Cash-Flows } \textit{during} \text{ the Explicit Forecast Period} \\ & + \text{Present Value of Cash-Flows } \textit{after} \text{ the Explicit Forecast Period} \end{aligned}$$

Consequently, the Terminal Value plays a key role in the equation. Usually, its weight in the estimation of company value is relevant. If it very much relevant, the entire valuation process can be considered unreliable: the greater the Terminal Value, greater the synthetic valuation and lower the analytic valuation; therefore, much of the estimated value is out of the business plan and, then out of the company's fundamental analysis. In our perspective the Terminal Value should be less than 40% of company value.

The Terminal Value estimation, on the basis of the ongoing concern approach, can be estimated on the basis of two main techniques (Damodaran 2012):

- *relative value*: it is based on multiples of earnings, revenues or book value and therefore it is based on relative value. This technique is easier but dangerous for two main reasons: firstly, the multiple is estimated by looking at how comparable companies in business today are priced by the market, and therefore a relative valuation of the Terminal Value is used; secondly, by using a relative valuation of Terminal Value there is a dangerous mix between relative valuation and direct valuation used for the analytical valuation. In this case there are problems of incompatibility between analytic and synthetic valuation with consequent unreliability of the entire valuation of the company;
- *intrinsic value*: it is based on discounted cash flow and therefore valuation of the Terminal Value is based on the intrinsic value of the company. Specifically, it is based on the assumption that future expected cash-flows will grow at a constant rate in perpetuity. Assuming a stable growth rate, the Terminal Value can be estimated using a perpetual growth model. Even if this approach is more complex, nevertheless it is more coherent with the analytical valuation and it provides more homogeneity between the first and the second parts of Eq. (8.2). By following this approach, the Terminal Value can be estimated as follows:

$$TV_t = \frac{CF_{t+1}}{K - g_n} \tag{8.3}$$

where:

- $t+1$: is the first year after the end-time of analytic valuation;
- CF: is the future normalized expected cash-flow from the first year after the end-time of analytic valuation;
- K: is the cost of capital;
- g_n: is the stable growth rate.

In the Terminal Value equation, there are two main problems: (i) estimation of the stable growth rate (g_n) and (ii) the normalized value of the expected cash-flows in perpetuity. Small changes in these value can change the Terminal Value significantly and therefore the company value.

The assumption that the stable growth rate is constant in perpetuity requires strong restrictions about its estimation. Specifically, three main caveats should be kept in mind (Damodaran 2012):

(a) *the company cannot grow in perpetuity at a rate higher than the growth rate of the economy referenced*. The stable growth rate must be lower or equal to the expected growth rate of the market or, in general terms, the economy. If the company is of a domestic dimension, due to internal or external constraints, the expected growth rate of domestic economy represents the limiting value of the expected stable growth rate of the company. On the other hand, if the company has a multinational dimension, or it expects to be so in the future, the expected growth rate in the global economy, or at least in the parts in which the company operates or wants to operate in the future, represents the limiting value of the expected stable growth rate of the company. Also, if the valuation is based on nominal value (nominal valuation), the stable growth rate should be nominal and, therefore it should include the expected inflation rate; however, if the valuation is based on real value (real valuation), the stable growth rate should be real and therefore should be lower than nominal one.

It is worth noting that by estimating the stable growth rate lower or equal to the growth rate of the economy referenced, it ensures that the stable growth rate will be less than the discount rate.

(b) *the period in which the company is able to sustain a high growth rate before laying down on stable growth rate has not been previously defined.* It is function of the market dynamics and its competition level. Generally, a high growth rate of the company comes from its capability to create value that it is function of the return on investments greater than the cost of capital, higher than its competitors. But over time the competitors will react by engaging in competitive actions to increase their value to the detriment of the company. Therefore, the ability of the company to sustain a high growth rate is function of its operations in the market as well as of the market operations of its competitors and the general dynamics of the market. In strategic terms, the ability of the company to sustain a high growth rate for a long period before it will lay down on a stable growth rate is function of two main variables: first, existence of the company's competitive advantage allowing it to obtain and defend a profitable market position; second, the characteristics and competitive dynamics in the market;

(c) *the transition from the high growth scenario to the stable growth scenario can be achieved in different ways.* There are two main ways: (i) firstly, the company maintains a high growth condition for a period and then it suddenly falls into a stable growth condition abruptly. This scenario is captured by a two-stage model; (ii) secondly, the company maintains a high growth condition for a period and then it undergoes a transition period in which it reduces growth gradually towards a stable growth condition. This scenario is captured by a three-stage model. The two-stage and three-stage models represent the basis to evaluate the company and equity.

Finally, it is worth noting that the company in a condition of stable growth characterized by a basically stable capital structure, low or null excess returns on investments and low reinvestments.

By substituting the Eq. (8.3) in the Eq. (8.2), we have:

$$W_F = \sum_{t=1}^{n} \frac{CF_t}{(1+K)^t} + \frac{\left[\frac{CF_{t+1}}{K-g_n}\right]}{(1+K)^n}$$

and therefore:

$$W_F = \sum_{t=1}^{n} \frac{CF_t}{(1+K)^t} + \frac{CF_{t+1}}{(K-g_n)(1+K)^n} \tag{8.4}$$

The Eq. (8.4) has great relevance. It allows for application of the General Equation as formalized by the Eq. (8.1) on the basis of estimation of cash-flows, cost of capital and growth rate.

Obviously, by changing the assumption about the estimation of these variables, it changes the company value.

Note that in the first part of Eq. (8.4), the growth rate (g) is included in the estimation of cash-flow in each time period of valuation. Therefore, it is equal to:

$$g = \frac{CF_t - CF_{t-1}}{CF_{t-1}} = \frac{CF_t}{CF_{t-1}} - 1$$

It is worth noting that while the growth rate g is included in the estimation of cash-flows in each year of analytical valuation, g_n is the growth rate in a steady condition and then it is the growth at which the cash flows will grow in perpetuity.

Valuation Perspective

The second problem is the valuation perspective. Application of the General Equation requires the definition of its variables: identification of the expected cash flow to be discounted and identification of capital cost used to discount the expected cash flow. The solution of the problem requires the definition of the valuation perspective. They could be two: Equity Side and Asset Side.

In the *Equity Side perspective* the Equity Value of the Company is estimated on the basis of Free Cash Flow to Equity (FCFE) discounted at the Cost of Equity.

Use of the Cost of Equity instead of Cost of Capital is due to the nature of the free cash-flow to be discounted: they are the residual cash-flow after the coverage of the company's needs and the debt obligations and destined to equity remuneration.

In the Equity Side perspective, the Eq. (8.2) can be rewritten as follows:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{(1+K_E)^t} + \frac{TV_n}{(1+K_E)^n} \tag{8.5}$$

where:

- W_E: is the Equity Value;
- $FCFE$: is the future expected Free Cash-Flow to Equity;
- K_E: is the cost of equity;
- g_n: is the stable growth rate;
- TV_n: is the terminal value.

Similarly, the Eq. (8.3) can be rewritten as follows:

$$TV_t = \frac{FCFE_{t+1}}{K_{E_{t+1}} - g_n} \tag{8.6}$$

By substituting the Eq. (8.6) in the Eq. (8.5), we have:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{(1+K_E)^t} + \frac{\left[\frac{FCFE_{t+1}}{K_{E_{t+1}} - g_n}\right]}{(1+K_E)^n}$$

and therefore:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{(1+K_E)^t} + \frac{FCFE_{t+1}}{\left(K_{E_{t+1}} - g_n\right)(1+K_E)^n} \tag{8.7}$$

The Eq. (8.7) is the Eq. (8.4) in Equity Side perspective.

In the *Asset Side perspective* the asset value called Enterprise Value is estimated. The Enterprise Value is estimated based on the Free Cash Flow From Operations (FCFO) discounted to the Cost of Capital. Use of the Cost of Capital, including both the cost of equity and the cost of debt, is due to the nature of the free cash flows to be discounted. In fact, these cash flows derive from the operating activities of the company and they are used in remuneration of both equity holders and debt holders.

Therefore, the Enterprise Value is the value generated by the company's operating activities and they must be distributed among the investors in equity and debt.

In the Asset Side perspective, the Eq. (8.2) can be rewritten as follows:

$$W_A = \sum_{t=1}^{n} \frac{FCFF_t}{(1+K_A)^t} + \frac{TV_n}{(1+K_A)^n} \tag{8.8}$$

where:

- W_A: is the Enterprise Value;
- $FCFO$: is the future expected Free Cash-Flow from Operations;
- K_A: is the cost of capital of the company;
- g_n: is the stable growth rate;
- TV_n: is the terminal value.

Similarly, the Eq. (8.3) can be rewritten as follows:

$$TV_t = \frac{FCFF_{t+1}}{K_{A_{t+1}} - g_n} \tag{8.9}$$

By substituting the Eq. (8.9) in the Eq. (8.8), we have:

$$W_A = \sum_{t=1}^{n} \frac{FCFF_t}{(1+K_A)^t} + \frac{\left[\frac{FCFF_{t+1}}{K_{A_{t+1}} - g_n}\right]}{(1+K_A)^n}$$

and therefore:

$$W_A = \sum_{t=1}^{n} \frac{FCFF_t}{(1+K_A)^t} + \frac{FCFF_{t+1}}{\left(K_{A_{t+1}} - g_n\right)(1+K_A)^n} \tag{8.10}$$

The Eq. (8.10) is the Eq. (8.4) in the Asset Side perspective.

It is worth noting that the Equity Value can be estimated from the Enterprise Value, by subtracting the Surplus Asset (SA) and the Net Financial Position (NFP) from the Enterprise Value, as follows:

$$W_E = W_A - NFP \tag{8.11}$$

Note that if there are Surplus Assets (SA), they must be considered in the Enterprise Value estimation by adding the cash flow from surplus assets and the Equity Value from the free cash flow obtained by subtracting the value of surplus assets (SA) and the Net Financial Positions (NFP). Therefore, the Eq. (8.11) can be rewritten as follows:

$$W_E = W_A^* - NFP - SA \tag{8.12}$$

Finally, it is worth noting that if there is no debt there is no valuation perspective problem. It is exactly the same evaluation of the assets or the equity of the company.

The two valuation perspectives, Asset Side and Equity Side, can be schematically represented as in Fig. 8.1.

There are several models to estimate the Equity Value and the Enterprise Value. In this context, following a financial approach, the *Equity Value* is estimated by using:

- Dividend Discounted Model (DDM);
- Free Cash Flow to Equity Discounted Model (E-DM);
- Multiples on Equity Value.

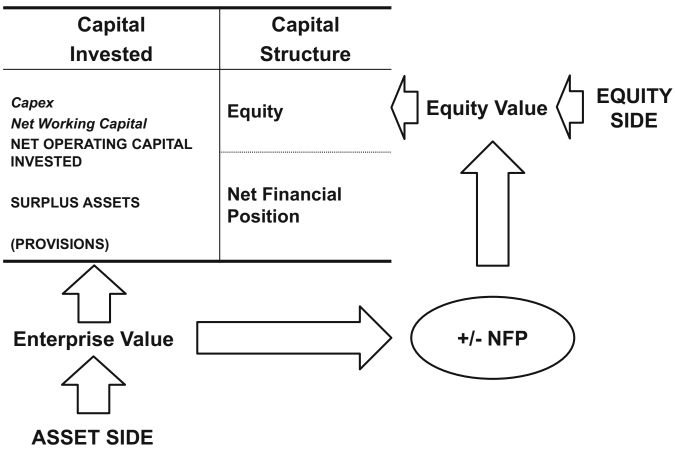


Fig. 8.1 Valuation perspectives: Asset side and equity side

Similarly, the *Enterprise Value* is estimated by using:

- Free Cash Flow from Operations Discounted Model (EV-DM) based on Cost of Capital Approach and on Adjusted Present Value Approach;
- Economic Value Added (EVA);
- Multiples on Enterprise Value.

In both cases, Equity Value and Enterprise Value, the models are defined on the basis of three scenarios:

(a) *Constant growth or one-period model*: it assumes a constant growth over time indefinitely;
(b) *Two-Stage Growth or two-period model*: it assumes an initial period characterised by extraordinary growth (good or bad) that it continuous for a certain number of years; a second period characterised by a steady-state growth rate expected to continue indefinitely;
(c) *Three-Stage Growth or three-period model*: it assumes a first period characterized by growth constant at the same level; a second period characterized by a changing growth from its level in the first period to a long-run steady-state level; a third period characterized by a constant growth indefinitely.

Obviously, it is possible to use others steps by passing from the three-period model to an n-period model.

Generally, by moving from the one-period model to three-period model (or n-period model), more information is required in terms of quantity and complexity. Also, more variables imply more complexity in the forecast process. Otherwise, the use of few variables implies a high level of simplicity but a low level of confidence about the value estimated.

The trade-off between complexity and manageability must be solved on the basis of the information available on the company and the analyst's forecasting skills.

8.2 Dividend Discount Model

Generally, the shareholder's value is function of two main variables:

- *expected dividends*, paid by the company during the period the shares are held;
- *expected capital gain*, at the end of the holding period if the selling price in the financial markets is higher than the buying price.

While dividends are paid by the company, capital gain is paid by the financial market. The expected price of the stock in the financial markets, in order to realize capital gain, is function of two main variables: company's fundamentals with regards to the expected dividends that could be paid; supply and demand in the financial market based on trading analysis.

It is important to note that expected Dividends are equal the expected Free Cash-Flow to Equity $(FCFE)$. The difference between the models refers to how Dividends and FCFE are estimated. Specifically, in the Dividend Discount Model (DDM) they are estimated synthetically by considering the dividend directly and their growth over time. Differently, in the Free Cash-Flow to Equity Discounted Model, they are estimated analytically on the basis of fundamental analysis and then the growth rate is included in their estimation.

The *Dividend Discounted Model (DDM)* estimates the Equity Value (W_E) (or if the Equity Value per Share is preferred (W_{E_S})) equal to the present value of the expected Dividends in the future (Div_t) [or Dividend per Share (DPS_t)] discounted to the Cost of Equity (K_E), as follows:

$$W_E = \sum_{t=1}^{\infty} \frac{Div_t}{(1+K_E)^t} \leftrightarrow W_{E_S} = \sum_{t=1}^{\infty} \frac{DPS_t}{(1+K_E)^t} \tag{8.13}$$

Note that the first of the two Eq. (8.13) considers Dividends in their total amount each year, while the second considers the Dividends per Share (DPS). Consequently, in the first case the Equity Value (W_E) is estimate while in the second the Equity Value per Share (W_{E_S}) is estimated.

Therefore, the DDM is based on two variables: (i) the *expected dividend*; (ii) the *cost of equity*.

The conventional measure of dividend policy is the dividend payout ratio (PR): it is equal to the sum of the dividend (Div) and stock buyback (SB) divided by Net Income (NI). Usually, the stock payback is not considered because it is difficult to estimate for the future. Formally, we have:

$$PR = \frac{Div}{NI} \rightarrow PR = \frac{Div + SB}{NI} \tag{8.14}$$

and per Share we have:

$$PR_S = \frac{DPS}{EPS} \rightarrow PR_S = \frac{DPS + SBS}{EPS} \tag{8.15}$$

where *DPS* is the Dividend per Share, *SBS* is the Stock Buyback per Share and *EPS* is Earnings per Share.

If stock buyback is considered, it could be financed by debt. To avoid this effect, the Eq. (8.14) can be modified by calculating the Payout Ratio Adjusted (PA_A) by subtracting the long-term debt (D_{LT}) as follows:

$$PA_A = \frac{Div + SB - D_{LT}}{NI} \tag{8.16}$$

The *cost of equity* is the expected return of the investor in equity. It is estimated based on: (i) the market beta in CAPM; (ii) the factor betas in the Arbitrage and Multifactor Models. The model is flexible enough to allow for time-varying discount rates due to the expected changes in interest rates or risk (Campbell and Shiller 1989).

The DDM is the simplest model to estimate Equity Value. There are three main advantages:

- it is based on an easy economic intuition: dividends are the only cash-flows relevant and tangible for the shareholders;
- the estimation of dividends is based on few assumptions: dividends are estimated by applying a growth rate to the last dividend paid or by considering the historical trend. Therefore, it is not necessary to estimate the company's fundamentals and the market future trend about the rates;
- dividends are usually assumed stable over time: equity value estimation based on dividends is more stable than the estimation based on cash-flows.

Despite the fact that the DDM model is simple, it is only used in specific cases. Indeed, the model is based on strong restrictions. There are two main restrictions:

- cash generated and held in the company for self-funding, instead of being distributed as dividends, is not considered. Then, the equity value could be underestimated;
- dividends may be paid by issuing new debt or new shares. Then, the share value could be overestimated.

By considering its advantages and disadvantages, the model can be useful in three main cases:

– first case: cash-flows generated by the company are higher than dividends paid. In this case, the definition of dividends reduces the managers' opportunity to invest in low value investments;
– second case: dividends paid by the company are stable over time. In this case the estimation of the share value is realistic;
– third case: estimation of the cash-flows is difficult or even impossible, due to the characteristics of the company's business. In this case, the dividends are the only stable parameter for evaluation.

In this context specific versions of the DDM are analysed, arising from different assumptions on the future growth of the company. Indeed, the Eq. (8.13) has a relevant theoretical value but it is not applicable directly. Therefore, its application requires assumptions about the main variables. There are four main versions of DDM (Damodaran 2012):

(d) constant growth DDM (C-DDM): it assumes that dividends will grow at the same growth rate over time indefinitely;
(e) two-Stage Growth DDM (2S-DDM): it assumes that dividends will grow on the basis of two different growth periods: a first period characterised by extraordinary growth (good or bad) continuous for a certain number of years; a second period characterised by a steady-state growth rate expected to continue indefinitely;
(f) three-Stage Growth DDM (3S-DDM): the model assumes that in the first period growth is assumed to be constant at some level. During the second period the growth changes from its level in the first period to a long-run steady-state level. In the third period constant growth is assumed.

For greater understanding of the model in its three main versions, the Eq. (8.13) can be defined with regards to the Stock Price (P), considering that the stock price (P_0) reflects its equity value per share (W_{E_S}), in $t = 0$ we have:

$$W_{E_{S(0)}} = P_0$$

Constant Growth DDM

The *Constant growth DDM* (*C-DDM*) (Williams 1938; Gordon 1962), also called *Single-Period Model* or *One-Period Model*, is based on the assumption of constant growth of the company over time and the payout ratio aligned to company's capabilities. The C-DDM is the simplest DDM's version because it assumes that dividends will grow at the same growth rate (g_n) into an indefinite future.

Therefore, the C-DDM estimates the Current Stock Price (P_0) on the basis of the expected dividends in the next period (Div_1), the cost of equity (K_E) and the expected growth rate of dividends (g_n), as follows:

$$P_0 = \frac{Div_1}{(1+K_E)} + \frac{Div_1(1+g_n)}{(1+K_E)^2} + \frac{Div_1(1+g_n)^2}{(1+K_E)^3} + \frac{Div_1(1+g_n)^3}{(1+K_E)^4} + \cdots + \frac{Div_1(1+g_n)^{n-1}}{(1+K_E)^n}$$

The growth rate g_n is assumed constant over time.

By using the formula for the sum of a geometric progression, we have:

$$P_0 = \frac{\frac{Div_1}{(1+K_E)}\left[1 - \left(\frac{1+g_n}{1+K_E}\right)^n\right]}{1 - \frac{1+g_n}{1+K_E}}$$

In this case:

$$\lim_{n\to\infty}\left(\frac{1+g_n}{1+K_E}\right)^n = \lim_{n\to\infty}\frac{(1+g_n)^n}{(1+K_E)^n} = 0$$

and then:

$$P_0 = \frac{\frac{Div_1}{1+K_E}}{1 - \frac{1+g_n}{1+K_E}} = \frac{\frac{Div_1}{1+K_E}}{\frac{1+K_E-1-g_n}{1+K_E}} = \frac{Div_1}{1+K_E}\cdot\frac{1+K_E}{K_E-g_n} = \frac{Div_1}{K_E-g_n}$$

and then:

$$P_0 = \frac{Div_1}{K_E - g_n} \tag{8.17}$$

The Eq. (8.17) states that the current stock price (P_0) is equal to the expected dividend of the following year (Div_1) divided by the difference between the Cost of Equity (K_E), used as appropriate discount rate, and the expected constant growth rate (g_n) over time.

Note that the C-DDM can be stated in terms of rate of return (Elton et al. 2013). Considering that the cost of equity is the rate of return expected by investors in equity (r_E), so that $r_E \equiv K_E$, substituting and solving the Eq. (8.17) for r_E, we have:

$$r_E = \frac{Div_1}{P_0} + g_n \tag{8.18}$$

The Eq. (8.18) estimates the rate of return for Equity-holders included in the stock price.

The C-DDM is based on a steady-state growth rate (g_n) of dividends over time. It requires three main basic assumptions:

– the company's characteristics must be in line with the stable growth rate;
– the company has to maintain a stable dividend policy;
– the company has to earn a stable return on new equity investment over time.

By considering these basic assumptions, it is easy to show how the constant growth rate (g_n) can be defined in terms of fraction of earnings retained within the company as self-financing (b), and the rate of return the company will earn on all new investments (r) and therefore the future profitability of investment opportunities (Elton et al. 2013). To derive the growth in dividends by the growth in earnings (E) arising from the return on new investments (I) as follows:

$$E_t = E_{t-1} + rI_{t-1}$$

The retention rate (r) on investments (I) can be assumed constant over time. Therefore, investments are equal to the fraction earnings retained within the company (b), so that:

$$E_t = E_{t-1} + rbE_{t-1} = E_{t-1}(1 + rb)$$

Growth in earnings (g_E) in percentage can be defined as follows:

$$g_E = \frac{E_t - E_{t-1}}{E_{t-1}}$$

and by substituting, we have:

$$g_E = \frac{E_{t-1}(1 + rb) - E_{t-1}}{E_{t-1}} = \frac{E_{t-1} + E_{t-1}rb - E_{t-1}}{E_{t-1}} = rb$$

By defining the part of earnings to be retained (r) and by assuming a constant the rate of return the company will earn on all new investments (b), the growth of earnings (g_E) is equal to the growth of dividends (g_D). Indeed, on the basis of r and b, a constant portion of earnings is assumed to be paid out in each time t. Therefore, we have:

$$g_E = rb = g_D \tag{8.19}$$

On the basis of Eq. (8.19), Eq. (8.17) can be rewritten as follows:

$$P_0 = \frac{Div_1}{K_E - rb} \tag{8.20}$$

And in terms of expected returns of investors' in equity r_E, we have:

$$r_E = \frac{Div_1}{P_0} + rb \tag{8.21}$$

A key role is played by future profitability of investment opportunities. In order to show this, consider that the rate of return on new investments (r) can be expressed as a fraction (α) of the rate of return of investors' in equity (r_E), as follows (Elton et al. 2013):

$$r = \alpha r_E$$

By considering that:

$$Div_1 = (1 - b)E_1$$

And substituting, the expected return of investors in equity (r_E), it is equal to:

$$r_E = \frac{(1 - b)E_1}{P_0} + \alpha r_E b$$

and then:

$$r_E - \alpha r_E b = \frac{(1 - b)E_1}{P_0};\ r_E(1 - \alpha b) = \frac{(1 - b)E_1}{P_0}$$

and:

$$r_E = \frac{(1 - b)E_1}{(1 - \alpha b)P_0} \tag{8.22}$$

The Eq. (8.22) shows that (Elton and Gruber 1976; Elton et al. 2013):

- if there are no extraordinary investment opportunities, $\alpha = 1$ and consequently the Eq. (8.22) becomes $r_E = E_1/P_0$. Therefore, the Equity-holder requires the inverse of Price-Earnings ratio.
- if there are extraordinary investment opportunities, $\alpha > 1$ and consequently the Eq. (8.22) implies that investment opportunities are expected to offer a return above the one required by the Equity holders.

The C-DDM is simple but its utility is limited. The main caveats to keep in mind are the following (Elton et al. 2013; Damodaran 2012):

- first, the model requires an estimation of the dividends of the following year, the company's growth rate in perpetuity and the rate of return required by the Equity holder for holding the stock;
- second, the model can be used with some utility only by the companies characterised by a stable growth rate. Specifically, the model can be useful for the companies with a growth rate in line or lower than the nominal growth rate in

the economy and with a well-established dividend payout policy over time to be continuous in the future;

- third, since the growth rate in dividends is expected to last forever, the other measures of the company's performances (revenues, costs, and earnings) are expected to grow at the same rate. It can generate a trade-off because the growth is not free: when the growth rate is increased, the payout ratio should be decreased in order to increase self-financing. Therefore, there is a trade-off on growth with the net effect on increasing growth being positive, neutral or even negative. Therefore, the model underestimates the company with the self-financing process. In this case the earnings are used to self-finance the company with a reduction in dividends;
- fourth, the growth rate of the stable growth has to be less than or equal to the growth rate of the markets referenced. It is unreasonable to assume that the company can grow at a rate greater than the growth rate of the referenced market in the long-term.

Two-Stage Growth DDM

The *Two-Stage Growth DDM* (*2S-DDM*) (Malkiel 1963; Fuller and Hsia 1984), also called *Two-Period Growth Model*, is based on two stages of growth:

- *extraordinary growth period*: it is the first period of the growth rate of dividend that cannot be considered stable over time. In this period it lasts in the first n years $(t = 1 \rightarrow t = n)$ the growth rate of dividend can be higher or lower than the stable growth rate. The term "extraordinary" is used because the growth rate of dividend in this first period can be greater or lower than the second period;
- *steady-state growth period*: it is the second period growth rate of dividend and is assumed stable over time. In this second period $(t = n + 1 \rightarrow t = \infty)$ the growth rate of dividend is assumed stable over time.

It is reasonable to assume two periods of growth. Indeed, after some years (3, 5, 8 years) it is difficult to make assumptions on future growth. Therefore, after some years, it is reasonable to assume that company will grow with a constant growth rate.

Assume that the length of the first period is equal n years. In this first period, the extraordinary (good or bad) growth rate is equal to g_e while in the second period the stable growth rate is equal to g_n. Denoting with P_N the price at the end of n-periods and then the price of stock in the second period. In both periods, the discount rate is the Cost of Equity (K_E). Denoting with $K_{E,eg}$ the cost of equity in the first period (extraordinary growth) and with $K_{E,st}$ the cost of equity in the second period (constant growth). It is possible to assume that they are equal among them so that $K_{E,st} = K_{E,eg}$.

The current stock Price (P_0) can be estimated as follows (Elton et al. 2013):

$$P_0 = \left[\frac{Div_1}{\left(1+K_{E,eg}\right)} + \frac{Div_1(1+g_e)}{\left(1+K_{E,eg}\right)^2} + \frac{Div_1(1+g_e)^2}{\left(1+K_{E,eg}\right)^3} + \cdots + \frac{Div_1(1+g_e)^{n-1}}{\left(1+K_{E,eg}\right)^n}\right] + \frac{P_n}{\left(1+K_{E,st}\right)^n}$$

By using the formula for the sum of a geometric progression, we have (Elton et al. 2013):

$$P_0 = \left[\frac{\frac{Div_1}{\left(1+K_{E,eg}\right)}\left[1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n\right]}{1-\frac{1+g_e}{1+K_{E,eg}}}\right] + \frac{P_n}{\left(1+K_{E,st}\right)^n}$$

And by considering that:

$$\frac{\frac{Div_1}{\left(1+K_{E,eg}\right)}\left[1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n\right]}{1-\frac{1+g_e}{1+K_{E,eg}}} = \frac{\frac{Div_1\left[1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n\right]}{\left(1+K_{E,eg}\right)}}{\frac{1+K_{E,eg}-1-g_e}{1+K_{E,eg}}}$$

$$= \frac{Div_1\left[1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n\right]}{\left(1+K_{E,eg}\right)} \cdot \frac{\left(1+K_{E,eg}\right)}{K_{E,eg}-g_e}$$

$$= \frac{Div_1\left[1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n\right]}{K_{E,eg}-g_e}$$

We have:

$$P_0 = Div_1\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{P_n}{\left(1+K_{E,eg}\right)^n}$$

The Price in the second period (P_n) can be estimated on the basis of the one-period model as follows:

$$P_n = \frac{Div_{n+1}}{K_{E,st}-g_n}$$

And by substituting, we have:

$$P_0 = Div_1\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{\frac{Div_{n+1}}{K_{E,st}-g_n}}{\left(1+K_{E,eg}\right)^n}$$

and then:

$$P_0 = Div_1\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{Div_{n+1}}{\left(K_{E,st}-g_n\right)\left(1+K_{E,eg}\right)^n} \tag{8.23}$$

The Eq. (8.23) can be rewritten by considering that Dividend in $t = n+1$ (Div_{n+1}) can be expressed in terms of dividend in the first period as follows:

$$Div_{n+1} = Div_1(1+g_e)^{n-1}(1+g_n)$$

In this case, by substituting it, we have:

$$P_n = \frac{Div_1(1+g_e)^{n-1}(1+g_n)}{K_{E,st}-g_n}$$

and then:

$$P_0 = Div_1\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{Div_1(1+g_e)^{n-1}(1+g_n)}{\left(K_{E,st}-g_n\right)\left(1+K_{E,eg}\right)^n} \tag{8.24}$$

The Eq. (8.24) can be rewritten by explicating the expected dividend in the next period (Div_1). Indeed, they can be estimated on the basis of current dividend (Div_0) and the growth rate (g_e) as follows:

$$Div_1 = Div_0(1+g_e)$$

Substituting in the Eq. (8.24), we have:

$$P_0 = Div_0(1+g_e)\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{Div_0(1+g_e)(1+g_e)^{n-1}(1+g_n)}{\left(K_{E,st}-g_n\right)\left(1+K_{E,eg}\right)^n}$$

$$P_0 = Div_0(1+g_e)\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{Div_0(1+g_e)^{n-1+1}(1+g_n)}{\left(K_{E,st}-g_n\right)\left(1+K_{E,eg}\right)^n}$$

and then:

$$P_0 = Div_0(1+g_e)\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{Div_0(1+g_e)^n(1+g_n)}{\left(K_{E,st}-g_n\right)\left(1+K_{E,eg}\right)^n} \tag{8.25}$$

Note that by using an approximate procedure, dividends can be estimated on the basis of Net Income (NI) and Payout Ratio (PR). Similarly, Dividends per Share (DPS) can be estimated on the basis of Earnings per Share (DPS) and Payout Ratio per Share (PRS). The relative equations are the following (Damodaran 2012):

$$Div_t = NI_t \cdot PR_t \leftrightarrow DPS_t = EPS_t \cdot PRS_t$$

On the basis of these equations, the Eq. (8.25) can be rewritten as follows:

$$P_0 = NI_0 \cdot PR_o \cdot (1+g_e)\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{NI_0 \cdot PR_o \cdot (1+g_e)^n(1+g_n)}{\left(K_{E,st}-g_n\right)\left(1+K_{E,eg}\right)^n}$$
$$P_0 = EPS_0 \cdot PRS_o \cdot (1+g_e)\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right] + \frac{EPS_0 \cdot PRS_o \cdot (1+g_e)^n(1+g_n)}{\left(K_{E,st}-g_n\right)\left(1+K_{E,eg}\right)^n} \tag{8.26}$$

Note that there is a strict relationship between the Payout Ratio (PR) and the growth rate (g). If the growth rate is expected to drop drastically in the second period, the payout should be higher in the second period than in the first period. A company in steady-state period (second period) can pay out more of its earnings in dividends than a company in the high growth period (first period).

The relationship between Payout Ratio (PR) and growth rate (g) can be expressed on the basis of the ROE (Damodaran 2012).

The Retention Rate (RR) on ROE is a self-financing measurement. It can be expressed as 1 less Payout Ratio (PR). The growth rate (g) can be expressed as the Retention Rate multiplied by ROE. On the basis of these relationships, we have:

$$RR \equiv 1 - PR$$
$$g = RR \cdot ROE \rightarrow g = (1-PR) \cdot ROE$$

and then:

$$PR = 1 - \frac{g}{ROE} \tag{8.27}$$

The Eq. (8.27) defines the relationship between Payout Ratio and growth rate on the basis of ROE.

The 2S-DDM can be applied in a different form. Usually analysts prefer to estimate dividends for each years in the first period (Div_t), and a constant dividend in the second period (Div_{n+1}). In this case, the 2S-DDM can be applied as follows (Damodaran 2012):

$$P_0 = \sum_{t=1}^{n} \frac{Div_t}{\left(1 + K_{E,eg}\right)^t} + \frac{P_n}{\left(1 + K_{E,sg}\right)^n} \tag{8.28}$$

and by considering that the stock price in the second period (P_n) can be estimated as follows:

$$P_n = \frac{Div_{n+1}}{K_{E,sg} - g_n} \tag{8.29}$$

and substituting, we have:

$$P_0 = \sum_{t=1}^{n} \frac{Div_t}{\left(1 + K_{E,eg}\right)^t} + \frac{\frac{Div_{n+1}}{K_{E,sg} - g_n}}{\left(1 + K_{E,sg}\right)^n}$$

and then:

$$P_0 = \sum_{t=1}^{n} \frac{Div_t}{\left(1 + K_{E,eg}\right)^t} + \frac{Div_{n+1}}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n} \tag{8.30}$$

Note that in the first period dividends include the growth rate. It can be estimated indirectly as follows:

$$g_e = \frac{Div_{t+1}}{Div_t} - 1$$

The Eq. (8.30) can be divided in two parts:

- the first part of equation is the present value of the expected dividends in the first period $(t = 1 \rightarrow n)$. It can be defined as the "*analytical value*" because: (i) the dividends are estimated for each year (Div_t) of the first period; (ii) the growth rate (g_e) is incorporated in the dividend estimation in each year and it can be different over the years $\left(g_{e_{(t)}}\right)$;
- the second part of equation is the present value of the price in the second period and then at the end of the n-years $(t = n + 1 \rightarrow \infty)$. It can be defined as the "*synthetic value*" because: (i) the dividends (Div_{n+1}) are estimated constant over time, (ii) the growth rate (g_n) is assumed constant in perpetuity.

It is worth noting that in both periods, the cost of equity is used to discount the expected dividends. It can be assumed equal in the two periods $\left(K_{E,eg} = K_{E,sg}\right)$. However, it should be assumed different because the two periods define a different structural scenario. In this case, we have $K_{E,eg} \neq K_{E,sg}$.

Generally, the 2S-DDM is useful for the company characterized by different growth levels between the first and the second period. Specifically, the model can be used if the company is characterized by a growth rate in the first period higher or lower than the stable growth rate of the second period.

There are three main problems in the model (Damodaran 2012):

- first, it is difficult to estimate the length of the extraordinary growth period;
- second, the growth rate is assumed high in the first period and declining at the end of the first period until a stable level characterises the second period. Also, assuming that this happens, it is more reasonable to assume that the move from the high growth to stable growth happens gradually over time;
- third, it can lead to skewed estimates of equity value that are not paying out what it can afford in dividends. Specifically, the model underestimates the value of the company, preferring prefer self-financing and the pay out of few dividends.

A specific version of 2S-DDM is the *Two-Stage Growth DDM with H-Model Specification* (*2SH-DDM*) (Fuller and Hsia 1984) that can be considered a specific version of the 2S-DDM. This version is based on three basic assumptions:

- first, the growth rate decreases in a linear manner during the first period (extraordinary growth) characterised by 2 stages (2H) to reach a stable growth rate during the second period (steady-state). The difference between 2S-DDM and 2SH-DDM is the behaviour of the growth rate in the first period: constant in the first model and decreasing in the second model;
- second, cost of equity is constant over time and it is independent of the growth rate;
- third, payout-ratio is constant over time and it is independent of the growth rate.

In this version, the growth rate starts from a high initial level (g_a) and it decreases in a linear manner during the first period (extraordinary growth period) that is assumed to last a stable growth rate (g_n) in the second period (steady-state growth period). The cost of equity (K_E) is assumed constant over time. Formally:

$$P_0 = \frac{Div_0(1+g_n)}{K_E - g_n} + \frac{Div_0(g_a - g_n)H}{K_E - g_n} \tag{8.31}$$

The relationship between the two growth rates $(g_a; g_n)$ can be summarized as in Fig. 8.2.

The model allows to exceed the limit of the 2S-DDM, with regards to the growth rate that drops drastically from high growth to stable growth. Otherwise, the model has two main limitations (Damodaran 2012):

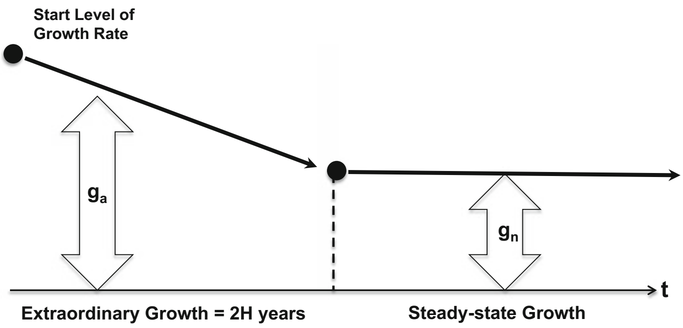


Fig. 8.2 Relationship between two growth rates

– first, the model assumes a linear decrease of the growth rate from a level in perpetuity. Even assuming a fall in the growth rate, it may not be linear. Generally, small differences from the assumption are not relevant on the company value estimation; otherwise, relevant differences have a high impact on this. Based on this assumption, the model can be used for the company characterized by a high growth rate, but with an estimation of its gradual fall mainly due to the growth of the company size and reduction of the competitive advantage;
– second, the model assumes constant dividends (payout ratio) in both periods. In reality, usually when the growth rate decreases, the payout ratio increases. Therefore, the model is not appropriate for companies with low dividends or with no dividends.

Three-Stage Growth DDM

The *Three-Stage Growth DDM (3S-DDM)* (Molodovsky et al. 1965), also called *Three-Period Growth Model*, is based on three different periods:

– *extraordinary growth period*: it is the first period and it is characterized by a high or low growth rate;
– *transitional growth period*: it is the second period and it is characterized by a growth rate declining or increasing to reach a stable level;
– *steady-state growth period*: it is the third period and it is characterized by a steady-state growth rate.

The current stock Price (P_0) is equal to the sum of present values of expected Dividends (Div) in the high or low growth period (first period), transaction period (second period) and steady-state period (third period). Denoting with: g_e and $K_{E,eg}$ the extraordinary growth and the cost of equity respectively in the first period; g_{tr} and $K_{E,tr}$ the transactional growth rate and the cost of equity respectively in the second period; g_n and $K_{E,st}$ the steady-state growth rate and the cost of equity respectively in the third period.

The 3S-DDM can be derived by 2S-DDM as follows:

$$P_0 = \left[\frac{Div_1}{(1+K_{E,eg})} + \frac{Div_1(1+g_e)}{(1+K_{E,eg})^2} + \frac{Div_1(1+g_e)^2}{(1+K_{E,eg})^3} + \cdots + \frac{Div_1(1+g_e)^{n-1}}{(1+K_{E,eg})^n}\right]$$
$$+ \left[\frac{Div_{n+1}}{(1+K_{E,tr})^{n+1}} + \frac{Div_{n+1}(1+g_{tr})^{n+1}}{(1+K_{E,tr})^{n+2}} + \frac{Div_1(1+g_{tr})^{n+2}}{(1+K_{E,tr})^{n+3}} + \cdots + \frac{Div_1(1+g_{tr})^{m-1}}{(1+K_{E,tr})^m}\right]$$
$$+ \frac{P_m}{(1+K_{E,st})^m}$$

and then:

By using the formula for the sum of a geometric progression, we have:

$$P_0 = \left[\frac{\frac{Div_1}{(1+K_{E,eg})}\left[1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n\right]}{1-\frac{1+g_e}{1+K_{E,eg}}}\right] + \left[\frac{\frac{Div_{n+1}}{(1+K_{E,tr})}\left[1-\left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m\right]}{1-\frac{1+g_{tr}}{1+K_{E,tr}}}\right] + \frac{P_m}{(1+K_{E,st})^m}$$

The second part of the equation can be rewritten as follows:

$$\frac{\frac{Div_{n+1}}{(1+K_{E,tr})}\left[1-\left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m\right]}{1-\frac{1+g_{tr}}{1+K_{E,tr}}} = \frac{\frac{Div_{n+1}\left[1-\left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m\right]}{(1+K_{E,tr})}}{\frac{1+K_{E,tr}-1-g_{tr}}{1+K_{E,tr}}}$$
$$= \frac{Div_{n+1}\left[1-\left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m\right]}{(1+K_{E,tr})} \cdot \frac{(1+K_{E,tr})}{K_{E,tr}-g_{tr}}$$
$$= \frac{Div_{n+1}\left[1-\left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m\right]}{K_{E,tr}-g_{tr}}$$

Therefore, by substituting we have:

$$P_0 = \left[Div_1\left[\frac{1-\left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg}-g_e}\right]\right] + \left[Div_{n+1}\left[\frac{1-\left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m}{K_{E,tr}-g_{tr}}\right]\right] + \frac{P_m}{(1+K_{E,eg})^m}$$

The P_m in the third term of equation can be estimated on the basis of the one-period model as follows:

$$P_m = \frac{Div_{m+1}}{K_{E,st}-g_n}$$

And by substituting it, we have:

$$P_0 = \left[Div_1 \left[\frac{1 - \left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg} - g_e}\right]\right] + \left[Div_{n+1} \left[\frac{1 - \left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m}{K_{E,tr} - g_{tr}}\right]\right] + \frac{\frac{Div_{m+1}}{K_{E,st}-g_n}}{\left(1+K_{E,eg}\right)^m}$$

and then:

$$P_0 = \left[Div_1 \left[\frac{1 - \left(\frac{1+g_e}{1+K_{E,eg}}\right)^n}{K_{E,eg} - g_e}\right]\right] + \left[Div_{n+1} \left[\frac{1 - \left(\frac{1+g_{tr}}{1+K_{E,tr}}\right)^m}{K_{E,tr} - g_{tr}}\right]\right] + \frac{Div_{m+1}}{\left(K_{E,st} - g_n\right)\left(1+K_{E,eg}\right)^m} \tag{8.32}$$

The Eq. (8.32) shows how the current stock Price (P_0) is equal to the sum of present values of expected Dividends (Div) in the high or low growth period (first period), transaction period (second period) and steady-state period (third period).

If the analyst prefers to estimate dividends for each years in the first and second periods $\left(Div_{t=1;n}; Div_{t=n+1;m}\right)$, and a constant dividend in the third period (Div_{m+1}), the 3S-DDM can be applied as follows (Damodaran 2012):

$$P_0 = \sum_{t=1}^{n} \frac{Div_t}{\left(1+K_{E,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{Div_t}{\left(1+K_{E,tr}\right)^t} + \frac{P_m}{\left(1+K_{E,st}\right)^m} \tag{8.33}$$

and by considering that the stock price in the third period (P_m) can be estimated as follows:

$$P_m = \frac{Div_{m+1}}{K_{E,sg} - g_n} \tag{8.34}$$

and by substituting it, we have:

$$P_0 = \sum_{t=1}^{n} \frac{Div_t}{\left(1+K_{E,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{Div_t}{\left(1+K_{E,tr}\right)^t} + \frac{\frac{Div_{m+1}}{K_{E,sg}-g_n}}{\left(1+K_{E,st}\right)^m}$$

and then:

$$P_0 = \sum_{t=1}^{n} \frac{Div_t}{\left(1+K_{E,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{Div_t}{\left(1+K_{E,tr}\right)^t} + \frac{Div_{m+1}}{\left(K_{E,sg} - g_n\right)\left(1+K_{E,st}\right)^m} \tag{8.35}$$

Note that in the first and second periods, dividends include the growth rate. It can be estimated indirectly as follows:

$$g_e = \frac{Div_{t+1}}{Div_t} - 1 \leftrightarrow g_t = \frac{Div_{t+1}}{Div_t} - 1$$

The Eq. (8.35) can be divided in three parts:

- the first part is the present value of the expected dividends in the first period $(t = 1 \rightarrow n)$. It can be defined as the "*analytical value*" because: (i) the dividends are estimated for each year (Div_t) of the first period; (ii) the growth rate (g_e) is incorporated in the dividend estimation in each year and the it can be different across the years $(g_{e_{(t)}})$;
- the second part is the present value of the expected dividends in the second period $(t = n + 1 \rightarrow m)$. It can be defined as the "*analytical value*" because: (i) the dividends are estimated for each year (Div_t) of the second period; (ii) the growth rate (g_{tr}) is incorporated in the dividend estimation in each year and the it can be different across the years $(g_{tr_{(t)}})$;
- the third part is the present value of the price in the third period and then at the end of the m-years $(t = m + 1 \rightarrow \infty)$. It can be defined as the "*synthetic value*" because: (i) the dividends (Div_{m+1}) are estimated constant over time, (ii) the growth rate (g_n) is assumed constant in perpetuity.

It is worth noting that in both periods, the cost of equity is used to discount the expected dividends. It can be assumed equal in the three periods $\left(K_{E,eg} = K_{E,tr} = K_{E,sg}\right)$. However, it should be assumed different because the three periods define structural different scenarios. In this case, we have $K_{E,eg} \neq K_{E,tr} \neq K_{E,sg}$.

The relationship among three growth rate can be summarized as in Fig. 8.3.

The 3S-DDM is more flexible than the 2S-DDM. It is useful in a scenario characterized by changes in growth rate, risk profile and dividend policy.

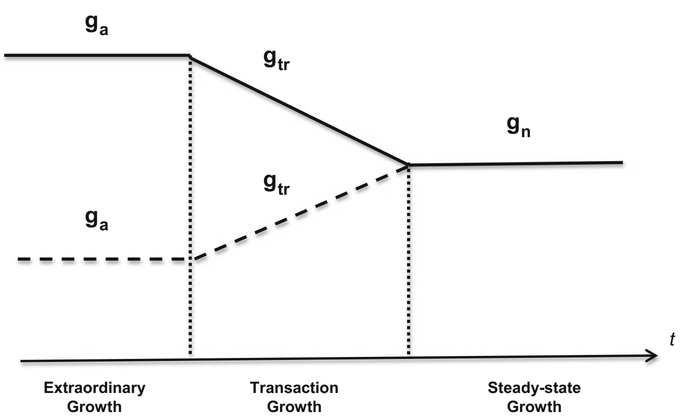


Fig. 8.3 Relationship between three growth rates

Note that the pay-out ratio is normally assumed as inverse to the growth rate; in the first period, high growth rate requires a low pay-out ratio, while in the third period a low-stable growth rate is compatible with a high pay-out ratio. In the second period their dynamics are inverted.

It is useful for companies characterised by a high growth rate expected to maintain for an initial period after which it is expected to gradually decrease until reaching a stable growth rate.

8.3 Free Cash-Flow to Equity Discounted Model

The *Free Cash-flow to Equity Discounted Models (EDM)* is one of the most useful models to estimate Equity Value.

The model is based on a more extensive definition of the dividends than DDM: while in DDM expected dividends paid synthetically and estimated on the basis of the past are considered without considering the company's fundamentals, in EDM expected Free Cash-Flow to Equity (FCFE) is estimated analytically on the basis of economic and financial dynamics of the company and then on its fundamental analysis.

The EDM can be considered as a DDM generalization. Indeed, dividend payments and FCFE may not be aligned among them. Specifically, dividend payments can be higher or lower than FCFE. In these cases, EDM is more effective than the DDM. Indeed, if (Damodaran 2012):

– the company pays dividends higher than FCFE, the Equity Value estimated by EDM is lower than the one estimated by DDM;
– the company pays dividends lower than the FCFE, the Equity Value estimated by EDM is higher than the one estimated by DDM.

Finally, EDM is more effective than the DDM whenever there is a change in the dividend policy as in a takeover where the bidder acquires control of the company.

There are three main problems of EDM:

– first, the free cash-flows generated, after the coverage of needs (actual and expected) and the payments of debt and taxes are used to pay dividends. Therefore, there is no cash accumulation (self-financing) in the company. Then, the model assumes that the FCFE generated each year are distributed in their entirety to the shareholders resulting in no cash accumulation. Also the FCFE may be negative. In this case, they represent the company's capital needs;
– second, it requires an estimate of the growth rate of each variable. Therefore, the FCFE's growth rate is incorporated into the analytical estimation of the expected FCFE;
– third, for external analysts it is not always easy to estimate FCFE. It implies an analytical analysis of the company's fundamentals in order to estimate economic and financial dynamics over time. This procedure requires a lot of quantitative

and qualitative information on the company and its expectations about the future. On the other hand, the dividends paid and its estimation by models used in DDM are easier to obtain and use.

Usually, the Equity Value estimated by EDM is different from the Equity Value estimated by DDM. Otherwise, there are two conditions under which two models generate the same Equity Value:

- first, the expected dividends and expected FCFE are equal. In this case, the company gives the shareholders the free cash flow generated by the company each year without cash accumulation;
- second, the FCFE are higher than the dividends paid but the cash accumulation is invested in the project with a net present value equal to zero $(VAN = 0)$. In this case, cash accumulation is neutralized by the Equity Value.

The general formulation of EDM estimates the Equity Value equal to the present value of expected FCFE. Specifically, the Equity Value (W_E) is estimated equal to the present value of expected FCFE in each time (t) by using the Cost of Equity (K_E) as discounted rate, as follows:

$$W_E = \sum_{t=1}^{\infty} \frac{FCFE_t}{(1 + K_E)^t} \tag{8.36}$$

The Eq. (8.36) has a conceptual value but it is not directly applicable. The problem is the indefinite time of valuation.

The correct application of EDM as formally defined by Eq. (8.36) can be achieved on the basis of three main versions (Damodaran 2012):

(a) constant Growth EDM (C-EDM);
(b) two-Stage Growth EDM (2S-EDM);
(c) three-Stage Growth EDM (3S-EDM).

Constant Growth EDM

The *Constant Growth EDM (C-EDM)* is based on a stable growth rate of the company assumed in a steady-state growth.

In this condition, the Equity Value (W_E) is estimated on the basis of the expected FCFE in the next period $(FCFE_1)$, a stable growth rate (g_n), and the cost of equity (K_E) used as discounted rate. Formally (Damodaran 2012):

$$W_E = \frac{FCFE_1}{K_E - g_n} \tag{8.37}$$

The Eq. (8.37) is similar to the Gordon's model. Consequently it is characterized by the same limitations. Specifically, there are two main caveats (Damodaran 2012):

- the expected growth rate in perpetuity (g_n) must be equal or lower than the expected growth rate of the economy or the business references of the company.

Therefore, the model can be used for the company with a growth rate in line with that of the business or of the economy;

– the company's characteristics must be in line with the assumption of stable growth. Specifically, the stable condition requires that the company's investments must be in line with amortizations and depreciations. Indeed, in this case the company's investments activities are not relevant because the growth margin is low.

Two-Stage EDM

The *Two-Stage EDM (2S-EDM)* can be used to estimate the Equity Value of company characterized by an extraordinary growth rate in the first period characterized by a high or low growth rate until to reach a steady-state growth rate in a second period (Damodaran 2012).

Therefore, there are two periods:

– *extraordinary growth period*: it is the first period characterized by higher or lower growth than steady-state growth. This period lasts n years.
– *steady-state growth period*: it is the second period characterized by a steady-state growth in which growth is stable over time.

On the basis of these two periods, the 2S-EDM estimates the Equity Value (W_E) equal to the present value of FCFE during the extraordinary growth period plus the present value of Terminal Value (TV) estimated at the end of the extraordinary period for the steady-state growth period (second period). The discounted rate is the Cost of Equity ($K_{E,eg}$ in the first period "extraordinary growth" and $K_{E,st}$ in the second period "Steady-state growth"). Formally:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{\left(1+K_{E,eg}\right)^t} + \frac{TV_n}{\left(1+K_{E,st}\right)^n} \tag{8.38}$$

Note that the grow rate in the first period, is equal to:

$$g = \frac{FCFE_{t+1}}{FCFE_t} - 1$$

The *extraordinary growth rate* (g) in the first period is different from *steady-state growth rate* (g_n) of the second period: g can be higher or lower than g_n.

The Terminal Value is calculated by using the infinite growth rate model as follows:

$$TV_n = \frac{FCFE_{n+1}}{K_{E,sg} - g_n} \tag{8.39}$$

Therefore, by substituting Eq. (8.39) in Eq. (8.38), we have:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{\left(1+K_{E,eg}\right)^t} + \frac{\frac{FCFE_{n+1}}{K_{E,st}-g_n}}{\left(1+K_{E,sg}\right)^n}$$

and then:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{\left(1+K_{E,eg}\right)^t} + \frac{FCFE_{n+1}}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n} \tag{8.40}$$

The Eq. (8.40) estimates the Equity Value on the basis of two parts:

- the first part, is the present value of the expected FCFE estimated in the first period for each year. Therefore, it is an analytical valuation;
- the second part, is the present value of expected Terminal Value estimated in the second period from the end of the first period in perpetuity. Therefore, it is a synthetic valuation.

The Cost of Equity (K_E) used to discount both the expected FCFE and Terminal Value, may be the same in the two periods.

Three-Stage EDM

The *Three-Stage EDM (3S-EDM)* can be used to evaluate companies characterized by three different stages of growth:

- *extraordinary growth period*: it is the first period and it is characterized by a high or low growth rate;
- *transitional growth period*: it is the second period and it is characterized by a declining or increasing growth rate to reach a stable level;
- *steady-state growth period*: it is the third period and it is characterized by a steady-state growth rate.

The Equity Value is equal to the sum of present values of expected FCFE in a high or low growth period (first period), transaction period (second period) and steady-state period (third period) (Damodaran 2012). By considering the extraordinary growth rate (g_a), transitional growth rate (g_{tr}) and the steady-state growth rate (g_n) and the cost of equity (K_E) in the first period $\left(K_{E,eg}\right)$, second period $\left(K_{E,tr}\right)$ and third period $\left(K_{E,st}\right)$, the Equity Value (W_E) is equal to:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{\left(1+K_{E,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFE_t}{\left(1+K_{E,tr}\right)^t} + \frac{TV_m}{\left(1+K_{E,st}\right)^m} \tag{8.41}$$

where the growth rates referring to the first and second periods are the following:

$$g_a = \frac{Div_{t+1}}{Div_t} - 1 \leftrightarrow g_a = \frac{DPS_{t+1}}{DPS_t} - 1$$
$$g_{tr} = \frac{Div_{t+1}}{Div_t} - 1 \leftrightarrow g_{tr} = \frac{DPS_{t+1}}{DPS_t} - 1$$

The Terminal value (TV) can be calculated by using the infinite growth rate model:

$$TV_m = \frac{FCFE_{m+1}}{K_{E,st} - g_m} \tag{8.42}$$

By substituting Eq. (8.42), the Eq. (8.41) can be rewritten as follows:

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{\left(1 + K_{E,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFE_t}{\left(1 + K_{E,tr}\right)^t} + \frac{\frac{FCFE_{m+1}}{K_{E,st} - g_m}}{\left(1 + K_{E,st}\right)^m}$$

and then

$$W_E = \sum_{t=1}^{n} \frac{FCFE_t}{\left(1 + K_{E,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFE_t}{\left(1 + K_{E,tr}\right)^t} + \frac{FCFE_{m+1}}{\left(K_{E,st} - g_m\right)\left(1 + K_{E,st}\right)^m} \tag{8.43}$$

The Eq. (8.43) is based on three different parts:

- The first part, is the present value of expected FCFE in the first period (from $t = 1$ to $t = n$) and it is characterized by a high or low growth rate;
- The second part, is the present value of the expected FCFE in the second period (from $t = n + 1$ to $t = m$) and it is characterized by a transitory growth rate;
- The third part, is the present value of expected Terminal Value in the third period (from $t = m + 1$ in perpetuity) and it is characterized by a steady-state growth rate.

While the first and second periods can be defined "analytically" because the expected FCFE's are estimated in each period of valuation, the third part can be defined "synthetically" because the Terminal Value is estimated in perpetuity.

8.4 Multiples on Equity Value

The relative valuation estimates the Equity Value and Enterprise Value on the basis of multiples (Copeland et al. 2004; Damodaran 2012, 2015; Koller et al. 2015; Vernimmen et al. 2014; Graham et al. 1962; Beidelman 1971; Foster 1970;

Whitbeck and Kisor 1963; Gordon 1962; Malkiel and Cragg 1970; Joy and Jones 1970; Hawkins 1977; Chen 1998; Blume 1977; Corelli 2016; Beaver 1978).

While in the Discounted Cash-Flow Models (DDM, EDM, and EPDM) both Equity Value and Enterprise Value are estimated on the basis of the company's fundamentals, in the multiples approach they are estimated on the basis of current market price of comparable companies.

Although the Discounted Cash-Flow Models (DCF) are far more reliable in the asset valuation, they require many forecasting variables. On the other hand, the multiple approach requires few variables and it is easier than the DCF (Bing 1971).

Specifically, the multiples approach:

- requires less time, efforts and information than a direct valuation of Discounted Cash-Flow Models;
- basic assumptions are not required;
- is easy to understand for analyst, advisors and investors;
- is easy to defend the reliability of evaluation in the market;
- reflects the market dynamics more than the direct valuation because the objective is the relative value of the company based on market price and not its intrinsic value as direct valuation.

Unfortunately, these strong points define weak points at same time. Indeed, this approach is unsophisticated and has several problems. Among these, the main ones can be summarized as follows (Damodaran 2012):

- the basic reasoning according to which if the assets are equal, their price in the market must be equal, and therefore the price of the asset can be defined based according to the market price of an equal asset, is very limited. The main problem is due to the "comparable asset": it is a great simplification because there are no two assets, and therefore two companies, that can be defined equal in the market;
- the simply use of multiples, can generate an inconsistent valuation compared to the company's fundamentals mainly with regards to its expectations about the expected free cash flow and risk;
- the multiples reflect the market trend. Then, the company's value reflects the market trends: the company's valuation can be overestimated (or underestimated) if the market is overestimated (or underestimated);
- the multiples can be easily manipulated because the basic assumptions cannot be transparent. It is not strange that even easy multiples are calculated differently by different analysts;
- if the companies use a different accounting principle, there are distortions in the multiples;
- not always can the value of the multiple be calculated in a useful manner. It is not useful to calculate the multiple if one of the two variables considered is negative;

- the valuation based on the multiples is unstable in the long time. The value of the multiples in the business and in the market, change over time. These changes are due to the changes in the market structure, in the company's fundamentals or in the financial markets with regards to the cost of capital. Therefore, it is not always easy to compare the value of a multiple at different times. Also, one company may be cheaper than another company today, but not tomorrow;
- the use of subjective judgments is high in all phases of the multiple definition and its application;
- in the use of the multiples, both equity and enterprise, it is necessary to consider that the multiples change in time for the same company and each multiple is different according to the market and the State of the company.

Even if the multiples approach is easy, its correct application requires a rigorous procedure (Damodaran 2012). This procedure can be scheduled in 6 steps:

- *(step 1) to find a comparable company*. The basic reasoning is simple: if the assets are equal, their price in the market must be equal. Therefore, the company's value can be estimated based on the market value of other comparable companies. Generally, the "comparable company" is defined as a company with similar expected cash-flows, growth rate and risk profile in the same business;
- *(step 2) to relate the market price with common variables by generating the well-known multiples* (such as net income, EBITDA, EBIT, Capital invested, etc.). It is necessary to standardize the price for comparison;
- *(step 3) to define the multiple coherently according to the relationship between numerator and denominator*. If the numerator refers to equity (or enterprise) the denominator has to refer to equity (or enterprise) too;
- *(step 4) to define the multiple evenly for all comparable companies. It must be defined in the same way for all comparable companies*. Also, it is useful to understand the companion variable of the multiple in order to measure the sensitivity of the multiple to the company's fundamental changes. Finally, it is necessary to understand the range of the "normal value" of the multiple with regards to the business references;
- *(step 5) to adjust the company's value estimated by multiples according to the differences between the company evaluated and the company comparable*;
- *(step 6) to understand the statistic characteristics of the distribution of the multiple with regards to the specific business and the market in general*. The first is useful to understand the position of the company in the business; the second is useful to understand the position of the business in the market.

In this context the multiples derive from the discounted cash flow approach (Discount Dividend Model (DDM), Free Cash Flow to Equity Discount Models (EDM), and Free Cash Flow to Company Discount Models (EFDM)). In these models the value of equity and the value of enterprise are function of the expected cash flows (FCFO and FCFE), the expected growth rate of these cash flow and their uncertainty and the cost of capital. Therefore, each multiple is a function of the same variables of the models such as expected cash-flow, growth and risk.

In this context, as in the case of direct valuation, also for the relative valuation the multiples approach is used by distinguishing the two perspectives: Asset Side, to estimate the Enterprise Value; Equity Side, to estimate Equity Value. Therefore, the most useful multiples are distinguished in these two perspectives as follows:

– Equity Side, to estimate Equity Value the four main multiples are considered:
  - Price-to-Earnings (P/E or PE) Ratio;
  - PE-to-Growth (PE/G or PEG) Ratio;
  - Price-to-Book Value (P/BV or PBV) Ratio;
  - Price-to-Sales (P/S or PS) Ratio.
– Asset Side, to estimate Enterprise Value, three main multiples are considered:
  - Enterprise Value-to-EBITDA (EV/EBITDA) Ratio;
  - Enterprise Value-to-Book Value (EV/BV) Ratio;
  - Value of Operating Assets-to-Sales (EV/Sales) Ratio.

Figure 8.4 summarizes the multiples approach by distinguishing between the Asset Side and the Equity Side in order to estimate Enterprise Value and Equity Value respectively.

Note that the PS Ratio and the EV/Sales Ratio are two multiples based on revenues. Specifically, these multiples estimate the Equity and Enterprise Values on the basis of the company's ability to generate revenues. Their simplicity generates relevant advantages:

– first, they can be easily used to evaluate a young company with negative earnings or for new-economy companies in which their value can be estimated based on the specific sector measures (such as the number of customers, subscribers, web site visitors);
– second, while the earnings and book value ratios can be negative and therefore not meaningful, revenues multiples are available even for the most troubled company and for young company;

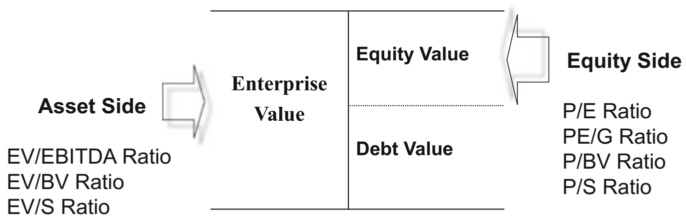


Fig. 8.4 Multiples to estimate equity value and enterprise value

– third, while the earnings and book values are heavily affected by accounting standards and therefore they can be manipulated, the revenue is relatively difficult to manipulate;
– fourth, revenue multiples tend to have a low volatility and therefore they are more stable over time.

The main problem of these multiples is that they may assign high value to a company that is generating high revenue but losing a relevant amount of money.

In this paragraph multiples on Equity Value are analysed while in the next chapter multiples on Enterprise Value will be analysed.

Price-to-Earnings Ratio (P/E)

The *Price-earnings (PE) Ratio* is the most commonly used multiple to estimate Equity Value. Three main approaches are used. The simplest way to use the multiple is to define a measure of earnings: present earnings, normalized earnings, predicted earnings. A more complex way is to investigate and to discuss large numbers of factors that should affect the multiple in order to find which of them really affects the multiple and their weights (Graham et al. 1962; Beidelman 1971; Foster 1970). In a larger view of this approach, common stock price, earnings, dividends, risk, growth and time value of money are considered. All of these variables are put together by considering their weights in order to estimate the multiple. The relationship between the multiple and variables identified is usually defined by using a regression analysis and multiple regression analysis (Whitbeck and Kisor 1963; Gordon 1962; Malkiel and Cragg 1970). This is probably one of the most popular in the standard texts of security analysis. One of the earlier equations shows that the P/E ratio is related to earnings, dividends, growth, risk (Whitbeck and Kisor 1963). The relationship can be formalized as follows: the higher the growth, the higher the dividends (by assuming constant growth), and the lower the risk (measured as standard deviation of growth), the higher the P/E ratio.

It is worth noting that all models developed over time on this approach are able to explain stock price at a given point in time. However, they are not able to select the appropriate stocks to buy or sell short at the same time. In other words, models are able to find which variables and their weights are important in the stock price at a given point in time but they cannot identify which stock will be successful. There are three main reasons (Elton et al. 2013): (i) market changes: the importance of certain variables and their weight change over time, sometimes rapid and drastic; (ii) the values of each variable change over time: even if the market preference on variables was to remain constant over time, the theoretical value of stock changes because the estimation of the value of variables changes; (iii) the model is not able to capture all company fundamentals and their effects on stock price: the actual price of stocks can be above or below their theoretical prices. Indeed, the theory behind their use in order to find stocks under-valued and over-valued is that the market price will converge to a theoretical price before this theoretical price itself changes. The main problem is that the parameters that determine theoretical price might change.

Note that even if the multiple use seems simple, several errors are usually made in its application. These errors are mainly due to an unclear relationship with company's fundamentals.

Formally, it is equal to the ratio of the Market Price per Share (MPS) to the Earnings per Share (EPS):

$$PE = \frac{MPS}{EPS} \tag{8.44}$$

There are several approaches to estimate both numerator and denominator. Indeed (Damodaran 2012):

- the market price can be defined in terms of (i) the *current* market price or (ii) the *mean* market price based on the last quarter or the last year;
- the earnings per share can be defined in terms of (i) the *current* earnings per share, with regards to the most recent book value earnings; (ii) the *trailing* earnings per share, with regards to the book value earnings of the last quarter; (iii) the *forward* earnings per share, with regards to the expected book value earnings for the next year.

By combining these determinations of numerator and denominator, several configurations of the multiple are achieved:

- price current or mean/earnings *current* (*PE current*): is the ratio between the market price and the current earnings per share of the company;
- price current or mean/earnings *trailing* (*PE trailing*): is the ratio between the market price and the earnings per share of the last year;
- price current or mean/earnings *forward* (*PE forward*): is the ratio between the market price and the expected earnings per share for the next year.

Note that there is a relationship among these configurations as follows:

$$PE\,forward < PE\,trailing < PE\,current$$

A relationship between the PE ratio and the company's fundamentals can be found. Specifically, this relationship can be defined by using the Discounted Dividend Model (DDM) in two scenarios (Damodaran 2012):

- steady-state growth over time scenario;
- two-stage growth scenario: extraordinary growth in the first period and steady-state growth in the second period.

In the *steady-state growth over time scenario*, the Equity Value can be estimated on the basis of C-DDM as follows:

$$P_0 = \frac{DPS}{K_E - g_n}$$

where P_0 is the current stock price, DPS are the Dividends per Share that the company can pay in perpetuity in steady-state scenario and g_n is the steady-state growth rate.

Dividing both sides by Earnings per Share (EPS), we have:

$$\frac{P_0}{EPS} = \frac{\frac{DPS}{K_E - g_n}}{EPS} \rightarrow \frac{P_0}{EPS} = \frac{DPS}{(K_E - g_n)EPS}$$

The first member of equation is the PE Ratio, and then we have:

$$PE = \frac{DPS}{(K_E - g_n)EPS}$$

The Dividends per Share (DPS) can be defined on the basis of Earnings per Share (EPS) by considering the Payout Ratio (PR) as follows:

$$DPS = EPS \cdot PR$$

The expected Dividends per Share (DPS) can be defined on the basis of Earnings per Share (EPS), its expected growth rate in a steady-state scenario (g_n), and the applied payout ratio (PR), as follows:

$$DPS = [EPS(1 + g_n)] \cdot PR$$

Substituting in the equation, we have:

$$PE = \frac{[EPS(1 + g_n)]PR}{(K_E - g_n)EPS}$$

and therefore:

$$PE = \frac{PR(1 + g_n)}{K_E - g_n} \tag{8.45}$$

The Payout Ratio (PR) can be explicated on the basis of the steady-state growth rate (g_n) and ROE (Damodaran 2012). In order to define this relationship, the following should be considered in a steady-state scenario: (i) the growth rate (g_n) can be estimated equal to ROE multiply Retention Ratio (RR); (ii) the Retention Ratio (RR) can be defined as 1 less Payout Ratio (PR). Formally, the Payout Ratio (PR) can be estimated equal to:

$$\begin{cases} g_n = RR \cdot ROE \\ RR = 1 - PR \end{cases} \rightarrow g_n = (1 - PR) \cdot ROE \rightarrow PR = 1 - \frac{g_n}{ROE} \tag{8.46}$$

On the basis of Eq. (8.46), the Eq. (8.45) can be rewritten as follows:

$$PE = \frac{\left(1 - \frac{g_n}{ROE}\right)(1 + g_n)}{K_E - g_n} \tag{8.47}$$

The Eq. (8.47) shows the relationship between PE Ratio and the company's fundamental in steady-state growth condition.

In the *two stage growth scenario* (extraordinary growth in the first period and steady-state grow in the second period) the Equity Value can be estimated on the basis of 2S-DDM as follows:

$$P_0 = \frac{EPS \cdot (1 + g_e) \cdot PR\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{K_{E,eg} - g_e} + \frac{EPS \cdot PR \cdot (1 + g_e)^n (1 + g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}$$

where g_e is the growth rate in the extraordinary growth period (first period) and g_n is the growth rate in the steady-state period (second period).

By dividing the first and second member by Earning per Share (EPS), we have:

$$\frac{P_0}{EPS} = \frac{EPS \cdot (1 + g_e) \cdot PR\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{EPS\left(K_{E,eg} - g_e\right)} + \frac{EPS \cdot PR \cdot (1 + g_e)^n (1 + g_n)}{EPS\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}$$

The first member is the PE multiple. Therefore, we have:

$$PE = \frac{PR \cdot (1 + g_e)\left[1 - \frac{(1+g)^n}{(1+K_{E,eg})^n}\right]}{\left(K_{E,eg} - g_e\right)} + \frac{PR \cdot (1 + g_e)^n (1 + g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n} \tag{8.48}$$

By estimating the Payout Ratio (PR) on the basis of Eq. (8.46), the Eq. (8.48) can be rewritten as follows:

$$PE = \frac{\left(1 - \frac{g_n}{ROE}\right)(1 + g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{\left(K_{E,eg} - g_e\right)} + \frac{\left(1 - \frac{g_n}{ROE}\right)(1 + g_e)^n (1 + g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}$$

and then:

$$PE = \left(1 - \frac{g_n}{ROE}\right)\left[\frac{(1 + g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{\left(K_{E,eg} - g_e\right)} + \frac{(1 + g_e)^n (1 + g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}\right] \tag{8.49}$$

The Eq. (8.49) shows the relationship between PE multiple and company's fundamental analysis in two stage growth rate.

PE-to-Growth Ratio (PEG)

The *PE to Growth Ratio (PEG)* is equal to PE (Price to Earnings Ratio) divided by the expected Growth Rate of Earnings per Share (g) as follows:

$$PEG = \frac{\mathrm{PE}}{\mathrm{g}} \tag{8.50}$$

Analysts and investors usually compare the PE Ratio with the expected growth rate for the earnings per share to identify underestimated or overestimated stocks (Damodaran 2012):

- companies with PE less than their expected growth rate are considered as undervalued;
- companies with PE more than their expected growth rate are considered as overvalued.

Therefore, the greater the PEG, the higher the overvaluation of stock; on the other hand, the lower the PEG, the higher the undervaluation of stock.

It is worth noting that the growth rate of earnings per share is used rather than operating income because the consistency requires coherence between numerator and denominator and the first is PE. Therefore, the PEG is an equity multiple.

The PE Ratio has different versions as shown previously. The correct configuration to use is function of the growth rate is calculated. Therefore, the following are used:

- current earnings, the *current PE Ratio* should be used;
- trailing earnings, the *trailing PE Ratio* should be used.

The *forward PE Ratio* should not be used because the growth may be considered twice.

As in the PE Ratio, also in this case it is possible to define a relationship between the PEG Ratio and company's fundamentals. Specifically, this relationship can be defined by using the Discounted Dividend Model (DDM) in two scenarios (Damodaran 2012):

- steady-state growth over time scenario;
- two-stage growth scenario: extraordinary growth in the first period and steady-state growth in the second period.

In the *steady-state growth over time scenario*, the Equity Value can be estimated on the basis of C-DDM as follows:

$$P_0 = \frac{DPS}{K_E - g_n}$$

where *DPS* are the Dividends per Share that company can paid in perpetuity in steady-state scenario and g_n is the steady-state growth rate.

On the basis of Eq. (8.45) and by dividing both terms by the steady-state growth rate (g_n), we have:

$$PE = \frac{PR(1+g_n)}{K_E - g_n} \rightarrow \frac{PE}{g_n} = \frac{PR(1+g_n)}{g_n(K_E - g_n)}$$

The first member is the PEG multiple:

$$PEG = \frac{PR(1+g_n)}{g_n(K_E - g_n)} \tag{8.51}$$

By estimating the Payout Ratio (PR) on the basis of Eq. (8.46), the Eq. (8.51) can be rewritten as follows:

$$PEG = \frac{\left(1 - \frac{g_n}{ROE}\right)(1+g_n)}{g_n(K_E - g_n)} \tag{8.52}$$

Note that the same result is obtained directly by dividing the first and second member of the Eq. (8.47) by the steady-state growth rate (g_n) as follows:

$$PE = \frac{\left(1 - \frac{g_n}{ROE}\right)(1+g_n)}{K_E - g_n} \rightarrow \frac{PE}{g_n} = PEG = \frac{\left(1 - \frac{g_n}{ROE}\right)(1+g_n)}{g_n(K_E - g_n)}$$

The Eq. (8.52) shows the relationship between PEG Ratio and company's fundamentals in a steady-state growth scenario over time.

In *two stage growth scenario* (extraordinary growth in the first period and steady-state growth in the second period) the Equity Value can be estimated on the basis of 2S-DDM as follows:

$$P_0 = \frac{EPS \cdot (1+g_e) \cdot PR\left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg} - g_e} + \frac{EPS \cdot PR \cdot (1+g_e)^n(1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1+K_{E,sg}\right)^n}$$

where g_e is the growth rate in the extraordinary growth period (first period) and g_n is the growth rate in the steady-state period (second period).

By dividing both terms of Eq. (8.48) by the steady-state growth rate (g_n), we have:

$$PE = \frac{PR \cdot (1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{(K_{E,eg} - g_e)} + \frac{PR \cdot (1+g_e)^n (1+g_n)}{(K_{E,sg} - g_n)(1+K_{E,sg})^n}$$

$$\frac{PE}{g_n} = \frac{PR \cdot (1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{g_n(K_{E,eg} - g_e)} + \frac{PR \cdot (1+g_e)^n (1+g_n)}{g_n(K_{E,sg} - g_n)(1+K_{E,sg})^n}$$

The first member is the PEG multiple:

$$PEG = \frac{PR \cdot (1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{g_n(K_{E,eg} - g_e)} + \frac{PR \cdot (1+g_e)^n (1+g_n)}{g_n(K_{E,sg} - g_n)(1+K_{E,sg})^n} \tag{8.53}$$

By estimating the Payout Ratio (PR) on the basis of Eq. (8.46), the Eq. (8.53) can be rewritten as follows:

$$PEG = \frac{\left(1 - \frac{g_n}{ROE}\right)(1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{g_n(K_{E,eg} - g_e)} + \frac{\left(1 - \frac{g_n}{ROE}\right)(1+g_e)^n (1+g_n)}{g_n(K_{E,sg} - g_n)(1+K_{E,sg})^n} \tag{8.54}$$

Note that the same result is obtained directly by dividing the first and second member of the Eq. (8.49) by the steady-state growth rate (g_n) as follows:

$$\frac{PE}{g_n} = PEG = \frac{\left(1 - \frac{g_n}{ROE}\right)(1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{g_n(K_{E,eg} - g_e)} + \frac{\left(1 - \frac{g_n}{ROE}\right)(1+g_e)^n (1+g_n)}{g_n(K_{E,sg} - g_n)(1+K_{E,sg})^n}$$

The Eq. (8.54) shows the relationship between PEG Ratio and company's fundamentals in a condition of two stage growth: extraordinary growth (first period) and steady-state growth (second period).

Price-to-Book Value Ratio (PBV)

The *Price-Book Value Ratio (PBV)* is equal to the ratio between the Equity Market Value defined on the basis of the Share Market Price and the Equity Book Value.

Formally, the Price-Book Value (PBV) is equal to the Market Price per Share (MPS) divided by the Equity Book Value per Share (EBVS) as follows:

$$PBV = \frac{\text{MPS}}{\text{EBVS}} \tag{8.55}$$

While the Equity Market Value, as measured by the Market Price per Share (MPS), is function of the investors' expectation about the company's ability to generate cash-flows on earnings per share in the future, the Equity Book Value is equal to the difference between the book value of assets and liabilities defined on the basis of accounting principles.

There are three main advantages of the multiple (Damodaran 2012):

- first, the book value is simple to calculate and it is generally stable variable; therefore, it is easy to compare with the market price of the company;
- second, by assuming consistent accounting standards across companies, the multiple allows for easier comparison of companies and estimation of whether or not the company is undervalued or overvalued. Specifically, companies with a market price lower than equity book value are considered undervalued, while those with market price higher than equity book value are considered overvalued;
- third, the probability that companies have a negative equity book value is lower than if they have negative earnings. Therefore, it is more likely that it is impossible to calculate the PE ratio (due to the negativity of the earnings) than the PBV (due to the negativity of the equity book value).

Nevertheless, it is necessary to keep in mind that the book value is function of the accounting principles. Therefore, all variables used are affected by accounting standards principles and policies. Obviously, if the equity book value is negative due to the string of negative earnings, the multiple is negative and therefore meaningless.

Note there may be a problem regarding the portion of the equity that is attributable to preferred stock, or in case there are multiple classes of shares outstanding. To resolve the majority of these problems directly, the multiple is calculated by considering Total Equity Market Value (EMV) and, consequently, the Total Equity Book Value (EBV) as follows:

$$PBV = \frac{\text{EMV}}{\text{EBV}} \tag{8.56}$$

As in PE and PEG Ratios, also in this case it is possible to define a relationship between PBV Ratio and company's fundamentals. Specifically, this relationship can be defined by using the Discounted Dividend Model (DDM) in two scenarios (Damodaran 2012):

- steady-state growth over time scenario;
- two-stage growth scenario: extraordinary growth in the first period and steady-state growth in the second period.

In the *steady-state growth over time scenario*, the Equity Value can be estimated on the basis of C-DDM as follows:

$$P_0 = \frac{DPS}{K_E - g_n}$$

where DPS are the Dividends per Share that the company can pay in perpetuity in steady-state scenario and g_n is the steady-state growth rate.

Remember that the Dividends per Share (DPS) can be expressed on the basis of Earning per Share (EPS) and Payout Ratio (PR) as follows:

$$DPS = EPS \cdot PR$$

by substituting we have:

$$P_0 = \frac{EPS \cdot PR}{K_E - g_n}$$

Remembering that ROE in t can be expressed as the ratio between the Earnings per Share (EPS) and Equity Book Value (EBV), it is possible to define the Earning per Share (EPS) in term of ROE as follows:

$$ROE = \frac{EPS}{EBV} \rightarrow EPS = ROE \cdot EBV$$

substituting we have:

$$P_0 = \frac{ROE \cdot EBV \cdot PR}{K_E - g_n}$$

By dividing both terms by Equity Book Value (EBV), we have:

$$\frac{P_0}{EBV} = \frac{ROE \cdot EBV \cdot PR}{EBV(K_E - g_n)} = \frac{ROE \cdot PR}{(K_E - g_n)}$$

The first member is the PBV multiple:

$$PBV = \frac{ROE \cdot PR}{(K_E - g_n)} \tag{8.57}$$

By estimating the Payout Ratio (PR) on the basis of Eq. (8.46), the Eq. (8.57) can be rewritten as follows:

$$PBV = \frac{ROE \cdot \left(1 - \frac{g_n}{ROE}\right)}{(K_E - g_n)}$$

and then:

$$PBV = \frac{ROE - g_n}{K_E - g_n} \tag{8.58}$$

The Eq. (8.58) shows the relationship between PEG Ratio and company's fundamentals in steady-state growth scenario over time.

In the *two stage growth scenario* (extraordinary growth in the first period and steady-state growth in the second period) the Equity Value can be estimated on the basis of 2S-DDM as follows:

$$P_0 = \frac{EPS \cdot (1+g_e) \cdot PR\left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg} - g_e} + \frac{EPS \cdot PR \cdot (1+g_e)^n (1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1+K_{E,sg}\right)^n}$$

where g_e is the growth rate in the extraordinary growth period (first period) and g_n is the growth rate in the steady-state period (second period).

Remembering that:

$$EPS = ROE \cdot EBV$$

and substituting, we have:

$$P_0 = \frac{ROE \cdot EBV \cdot PR \cdot (1+g_e) \cdot \left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg} - g_e} + \frac{ROE \cdot EBV \cdot PR \cdot (1+g_e)^n (1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1+K_{E,sg}\right)^n}$$

and dividing both terms by current Equity Book Value (EBV), the equation can be rewritten as follows:

$$\frac{P_0}{EBV} = \frac{ROE \cdot EBV \cdot PR \cdot (1+g_e) \cdot \left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{EBV\left(K_{E,eg} - g_e\right)} + \frac{ROE \cdot EBV \cdot PR \cdot (1+g_e)^n (1+g_n)}{EBV\left(K_{E,sg} - g_n\right)\left(1+K_{E,sg}\right)^n}$$

$$\frac{P_0}{EBV} = \frac{ROE \cdot PR \cdot (1+g_e) \cdot \left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{\left(K_{E,eg} - g_e\right)} + \frac{ROE \cdot PR \cdot (1+g_e)^n (1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1+K_{E,sg}\right)^n}$$

$$\frac{P_0}{EBV} = ROE\left\{\frac{PR \cdot (1+g_e) \cdot \left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{\left(K_{E,eg} - g_e\right)} + \frac{PR \cdot (1+g_e)^n (1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1+K_{E,sg}\right)^n}\right\}$$

The first member is the PBV multiple, and therefore:

$$PBV = ROE\left\{\frac{PR \cdot (1+g_e) \cdot \left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{\left(K_{E,eg} - g_e\right)} + \frac{PR \cdot (1+g_e)^n (1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1+K_{E,sg}\right)^n}\right\} \tag{8.59}$$

The Eq. (8.59) shows the relationship between the PBV Ratio and company's fundamentals in condition of two stage growth: extraordinary growth (first period) and steady-state growth (second period).

Price-to-Sales Ratio (PS)

The *Price-to-Sales Ratio (PS)* is equal to the ratio between the Equity Market Value (EMV) and the Revenues (Rev) of the company as follows:

$$PS = \frac{EMV}{Rev} \tag{8.60}$$

The relationship between the multiple and the company's fundamental by deriving the multiple from the DDM can be defined.

As in PE, PEG and PBV Ratios, also in this case it is possible to define a relationship between PS Ratio and company's fundamentals. Specifically, this relationship can be defined by using the Discounted Dividend Model (DDM) in two scenarios (Damodaran 2012):

- steady-state growth over time scenario;
- two-stage growth scenario: extraordinary growth in the first period and steady-state grow in the second period.

In the *steady-state growth over time scenario*, the Equity Value can be estimated on the basis of C-DDM as follows:

$$P_0 = \frac{DPS}{K_E - g_n}$$

The Dividend per Share (DPS) can be explicated according to the Earnings per Share (EPS), steady-state growth rate (g_n) and Payout Ratio (PR) as follows:

$$DPS = EPS \cdot PR$$

Substituting, we have:

$$P_0 = \frac{EPS(1 + g_n)PR}{K_E - g_n}$$

Note that the Net Profit Margin of the company can be defined as the ratio between the Earnings and Sales. Therefore, the Net Profit Margin per Share $(NPMS)$ is equal to Earnings per Share (EPS) and Sales (S). On the basis of this relationship, it is possible to define Earnings per Share as follows:

$$NPMS = \frac{EPS}{S} \rightarrow EPS = NPMS \cdot S$$

Substituting *EPS*, we have:

$$P_0 = \frac{NPMS \cdot S \cdot (1 + g_n) \cdot PR}{K_E - g_n}$$

Dividing both terms by Sales (S), we have:

$$\frac{P_0}{S} = \frac{NPMS \cdot S \cdot (1 + g_n) \cdot PR}{S(K_E - g_n)} = \frac{NPMS \cdot (1 + g_n) \cdot PR}{(K_E - g_n)}$$

The first member is the PS Ratio, as follows:

$$PS = \frac{NPMS \cdot (1 + g_n) \cdot PR}{(K_E - g_n)} \tag{8.61}$$

The Eq. (8.61) shows the relationship between PEG Ratio and company's fundamentals in steady-state growth scenario over time.

In the *two stage growth scenario* (extraordinary growth in the first period and steady-state growth in the second period) the Equity Value can be estimated on the basis of 2S-DDM as follows:

$$P_0 = \frac{EPS \cdot (1 + g_e) \cdot PR\left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg} - g_e} + \frac{EPS \cdot PR \cdot (1 + g_e)^n (1 + g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}$$

where g_e is the growth rate in the extraordinary growth period (first period) and g_n is the growth rate in the steady-state period (second period).

By considering that:

$$EPS = NPMS \cdot S$$

and substituting, we have:

$$P_0 = \frac{NPMS \cdot S \cdot (1 + g_e) \cdot PR\left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg} - g_e} + \frac{NPMS \cdot S \cdot PR \cdot (1 + g_e)^n (1 + g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}$$

Dividing both terms by Sales (S), we have:

$$\frac{P_0}{S}=\frac{NPMS\cdot S\cdot(1+g_e)\cdot PR\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{S_{(0)}\left(K_{E,eg}-g_e\right)}+\frac{NPMS\cdot S\cdot PR\cdot(1+g_e)^n(1+g_n)}{S_{(0)}\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}$$

$$\frac{P_0}{S}=\frac{NPMS\cdot(1+g_e)\cdot PR\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{\left(K_{E,eg}-g_e\right)}+\frac{NPMS\cdot PR\cdot(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}$$

$$\frac{P_0}{S}=NPMS\left\{\frac{(1+g_e)\cdot PR\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{\left(K_{E,eg}-g_e\right)}+\frac{PR\cdot(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\}$$

The first member is the PS ratio, and then:

$$PS=NPMS\left\{\frac{(1+g_e)\cdot PR\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{\left(K_{E,eg}-g_e\right)}+\frac{PR\cdot(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\}\quad(8.62)$$

The Eq. (8.62) shows the relationship between PS Ratio and company's fundamentals in a condition of two stage growth: extraordinary growth (first period) and steady-state growth (second period).

References

Altaman E (1969) Bankrupt company's equity securities as an investment alternative. Financ Anal J 25(4):129–133

Beaver W, Morse D (1978) What determines price-earnings ratios? Financ Anal J 34(4):65–76

Benninga S (2014) Financial modeling, 4th edn. MIT Press

Beidelman C (1971) Pitfalls of the price-earnings ratio. Financ Anal J 27(4):86–91

Berk J, DeMarzo P (2008) Corporate finance. Pearson Education, Inc

Bing R (1971) Survey of practitioners' stock evaluation methods. Financ Anal J 27(3):55–69

Blume MKJ, Kraft A (1977) Determinants of common stock prices: a time series analysis. J Finance XXXII(2):417–425

Bower D, Bower SR (1970) Test of a stock valuation model. J Financ XXV(2):483–492

Brealey RA, Myers SC, Allen F (2016) Principles of corporate finance, 12th edn. McGraw-Hill

Chen N (1998) Risk and return of value stock. J Bus 71(4):501–535

Copeland T, Weston F, Shastri K (2004) Financial theory and corporate policy, 4th edn. Addison-Wesley, Reading, MA

Campbell J, Shiller R (1989) The dividend price ratio and expectation of future dividends and discount factors. Rev Financ Stud 1:195–228

Corelli A (2016) Analytical corporate finance. Springer

Damodaran A (2012) Investment valuation: tools and techniques for a determining the value of any assets, 3rd edn. Wiley

Damodaran A (2015) Applied corporate finance, 4th edn. Wiley

Elton EJ, Gruber MJ (1976) Valuation and the asset selection under alternative investment opportunities. J Finance XXXI(2):525–539

Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN (2013) Modern portfolio theory and investment analysis, 9th edn. Wiley

Foster E (1970) Price-earnings ratio and corporate growth: a revision. Financ Anal J 26 (4):115–118

Fuller RJ, Farrell JL (1987) Modern investments and security analysis. McGraw-Hill, Inc

Fuller RJ, Hsia CC (1984) A simplified common stock valuation model. Financ Anal J:49–56 (September–October)

Gordon M (1962) The investment, financing, and valuation of the corporation. Richard D. Irwin, Homewood, IL

Graham B, Dodd D, Cottle S (1962) Security analysis principles and techniques, 4th edn. McGraw-Hill, New York

Hawkins D (1977) Toward an old theory of equity valuation. Financ Anal J 33(6):48–53

Hillier D, Ross S, Westerfield R, Jaffe J, Jordan B (2016) Corporate finance, 3rd edn. McGraw-Hill

Joy M, Jones C (1970) Another look at the value of P/E ratios. Financ Anal J 26(4):61–64

Koller T, Goedhart M, Wessels D (2015) Valuation, 6th edn. Wiley

Malkiel B (1963) Equity yields, growth, and the structure of share prices. Am Econ Rev 53:1004–1031

Malkiel B, Cragg J (1970) Expectations and the structure of share price. Am Econ Rev LX (4):601–617

Modigliani F, Miller MH (1958) The cost of capital, corporate finance and the theory of investment. Am Econ Rev 48:261–297

Molodovsky N, May C, Chottinger S (1965) Common stock valuation. Financ Anal J 21:104–123

Vernimmen P, Quiry P, Dallocchio M, Le Fur Y, Salvi A (2014) Corporate finance: theory and practice. Wiley

Whitbeck V, Kisor M (1963) A new tool in investment decision making. Financ Anal J:55–62 (May–June)

Williams JB (1938) The theory of investment value. Harvard University Press, Cambridge, MA

Chapter 9
Enterprise Valuation

Abstract The company's value is function of its ability to achieve positive performance in the future. The value of the company is equal to the present value of future expected cash flows and the cost of capital is used as a discount rate. In the previous Chapter an Equity Side perspective is adopted and the Equity Value is estimated. In this chapter the Asset Side perspective is adopted and the Enterprise Value is estimated. It is estimated on the basis of free cash-flows from operations discounted to the cost of capital based on the cost of equity and the cost of debt.

9.1 Free Cash Flow from Operations Discounted Model

The Enterprise Value is the value of operating assets of the company. It can be estimated by several models. Similar to Equity Value estimation the Enterprise Value is estimated by following a financial approach on the basis of Cash-Flow Discounted model (Altaman 1969; Arditti and Pinkerton 1978; Baron 1975; Bonnes and Jatusipitak 1972; Beaver and Morse 1978; Benninga 2014; Beidelman 1971; Berk and DeMarzo 2008; Bing 1971; Blume and Kraft 1977; Bower and Bower 1970; Brealey et al. 2016; Chen 1998; Copeland et al. 2004; Campbell and Shiller 1989; Corelli 2016; Damodaran 2012, 2015; Elton and Gruber 1971, 1976; Elton et al. 2013; Foster 1970; Fuller and Farrell 1987; Fuller and Hsia 1984; Gordon 1962; Graham et al. 1962; Haugen and Kumar 1974; Haugen and Pappas 1971; Hawkins 1977; Hillier et al. 2016; Joy and Jones 1970; Koller et al. 2015; Litzenberger and Budd 1970; Malkiel 1963; Malkiel and Cragg 1970; Modigliani and Miller 1958; Molodovsky et al. 1965; Myers 1974; Sydsaeter et al. 2012; Titman 1984; Vernimmen et al. 2014; Whitbeck and Kisor 1963; Williams 1938). Specifically, in this context the Free Cash-Flow from Operations (FCFO) are considered instead of Free Cash-Flow to Equity (FCFE) used in Equity Value. Indeed, the FCFO's are the cash flows generated by company operations and they are used for the remuneration of investors' both in equity and debt and to pay taxes.

Therefore, while in the Equity valuation models the equity-holders' cash flows is directly evaluated, in the Enterprise valuation models the cash flows available to all

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_9

investors are evaluated: equity-holders, debt-holders, and any other non-equity investors.

The Enterprise valuation models are especially useful for multi-business companies. In this case, the Enterprise value is equal to the sum of individual operating units value less the present value of the corporate costs plus the value of non-operating assets.

Based on the FCFO, the Enterprise Value is estimated by ***Free Cash-Flow from Operations Discounted Model (EV-DM)***.

The EV-DM can be applied by following two main approaches (Damodaran 2012):

1. *Cost of Capital (CC) approach*: the Enterprise Value is estimated on the basis of Free Cash Flow from Operations (FCFO) discounted at a Cost of Capital by considering either the cost of equity and the cost of debt. In this approach the costs and benefits of debt are considered directly in the Cost of Capital used as a discounted rate of the future expected FCFO;
2. *Adjusted Present Value (APV) approach*: the Enterprise Value is estimated by distinguishing and summing three parts: (i) the *unlevered value of enterprise*, equal to the present value of the unlevered Free Cash Flow from Operations (FCFO); (ii) the *value of the benefits on debt*, equal to the present value of the tax savings due to the interests on debt; (iii) the *value of the default risk* due to the bankruptcy costs as function of leverage.

These two approaches do not necessarily come up with the same Enterprise Value. There are three main reasons (Damodaran 2012; Koller et al. 2015):

– first, bankruptcy costs are considered in different ways in the two approaches. Specifically, indirect costs are more clearly defined in the APV approach than in the CC approach;
– second, the tax benefits from debt value are considered by the APV approach usually on the basis of the existing debt level. Differently, the CC approach estimates tax benefits after the definition of the debt level that is defined by finding the "optimal" leverage. Therefore, the cost of capital is used to define the debt level;
– third, in the APV approach the value of tax shield is estimated equal to the present value of the tax savings. The expected cash-flows of tax savings may be discounted at a rate different from Cost of Debt (such as the Cost of Unlevered Equity or Cost of Capital) by changing the Enterprise Value.

Generally, the CC approach is more useful in the case of ongoing company evaluation characterized by a sustainable capital structure. Differently, the APV approach is more useful in transactions based on a disproportionate debt level and where the debt repayment is negotiated or well known.

There is a relationship between the Enterprise Value and Equity Value. At any time, it is possible to estimate the Equity Value on the basis of the Enterprise Value and vice versa. Specifically, to the Equity Value (Enterprise Value) the following is sufficient:

- add (subtract) the value of non-operating activities, that are not included in the FCFO;
- subtract (add) the value of the Net Financial Position.

It is worth noting that the Equity Value estimated on the basis of the Enterprise Value and the Equity Value estimated directly are the same values only if consistent assumptions are made about financial leverage. Generally, the E-DM is preferred to the EV-DM if the company is characterized by a very high or low leverage. In both cases, it increases the difficulty to use the EV-DM while the effects of debts are captured directly in the FCFE.

There are three main problems of the EV-DM (Damodaran 2012):

- first, the FCFO's are less intuitive than the FCFE. While the FCFE represents the cash-flows used as dividends, the FCFO is an intermediate dimension of cash-flow used to analyse the company's ability to face debt obligations and to pay dividends;
- second, the FCFO can hide the real problem of company survival. If the company is characterized by a positive FCFO but a negative FCFE, the focus on FCFO could hide the problem. While the positivity of the first measures the company's ability to face debt obligations, the negativity of the second not only measures the company's inability to pay dividends, but the company's need for recapitalization;
- third, the EV-DM requires the inclusion in the cost of capital of all costs and benefits of debt. It requires the use of several assumptions with relevant effect on Enterprise Value.

The *Free Cash-Flow from Operations Discounted Model* (*EV-DM*) in the two versions, are analysed here below:

- Free Cash-Flow from Operations Discounted Model based on Cost of Capital approach ($EV\text{-}DM_{CC}$);
- Free Cash-Flow from Operations Discounted Model based on Adjusted Present Value approach ($EV\text{-}DM_{APV}$);
- Discounted Economic Profit (DEP).

9.2 Free Cash Flow from Operations Discounted Model: Cost of Capital Approach

The *Free Cash-flow from Operations Discounted Model based on Cost of Capital approach* (*$EV\text{-}DM_{CC}$*) estimates the Enterprise Value equal to the present value of the Free Cash Flow from Operation (FCFO) discounted to the Cost of Capital (Damodaran 2012).

The cost of capital is estimated on the basis of the cost of equity and the cost of debt. Then it represents the expected returns of investors in equity and debt.

Therefore, this approach considers the capital structure problem and it translates the choices about debt level in the cost of capital. Consequently, the cost of capital estimation is the key variable in the model.

The cost of capital is usually estimated on the basis of the Weighted Average Cost of Capital (WACC). The Leverage Cost of Capital (LCC) as developed in Chap. 7, can be used instead of WACC. Indeed, while the WACC is based on the stable debt level, the LCC defines the cost of debt on the basis of debt level changes. Consequently, the LCC can be more useful than the WACC to define the "optimal" capital structure and to estimate the cost of capital. In any case, the equations use general definition of the cost of capital, denoted by K_A, by leaving the reader the choice to use WACC or LCC.

The general version of the EV-DM_{CC} estimates the Enterprise Value (W_A) equal to the present value of expected FCFO discounted at the Cost of Capital (K_A) as follows:

$$W_A = \sum_{t=1}^{\infty} \frac{FCFF_t}{(1+K_A)^t} \tag{9.1}$$

Equation (9.1) has a theoretical value only. Its applications require assumptions about its variables: FCFO, time (t) and the Cost of Capital (K_A).

Specifically, as shown in previous chapter, Eq. (9.1) can be applied as follows:

$$W_A = \sum_{t=1}^{n} \frac{FCFF_t}{(1+K_A)^t} + \frac{TV_n}{(1+K_A)^n} \tag{9.2}$$

There are three main versions of the EV-DM_{CC} (Damodaran 2012):

(a) Steady-state Growth EV-DM_{CC};
(b) Two-Stage Growth EV-DM_{CC};
(c) Three-Stage Growth EV-DM_{CC}.

Therefore, conceptually these three versions are the same as in E-DM. Therefore, considerations developed in E-DM can be applied also in EV-DM_{CC}.

(A) Steady-State Growth EV-DM_{CC}

The *Steady-State Growth EV-DM_{CC}* estimates the Enterprise Value by assuming a constant growth rate in perpetuity in a steady-state scenario.

Therefore, the Enterprise Value (W_A) is equal to the present value of di Free Cash-Flow from Operations in the next year and assumed in perpetuity $(FCFO_1)$ discounted at the cost of capital (K_A) less the growth rate in perpetuity (g_n) as follows:

$$W_A = \frac{FCFO_1}{K_A - g_n} \tag{9.3}$$

Note that the $FCFO_1$ must be estimated by assuming that they can be achieved each year over time in perpetuity. In this model, the estimation is synthetic and then it is not based on the analytical estimation of each FCFO item.

Therefore, the Enterprise Value is evaluated synthetically on the basis of the assumptions with regards to the growth rate and the cost of capital without any analytical estimation of the FCFO.

Equation (9.3) correct application requires two main conditions:

- the expected growth rate in perpetuity (g_n) must be equal or lower than the expected growth rate of the economy or the business referencing of the company;
- the company characteristics are assumed unchangeable over time and it must be in line with the assumption of the stable growth. In this case the company's investments are not relevant because the growth margin is low. The best way to evaluate this second condition is to derivate the re-investment rate by the stable growth rate and the return on investment in perpetuity (Damodaran 2012). Specifically, the Re-Investment Rate in stable growth condition (IR_{SG}) can be estimated equal to the growth rate in a steady-state condition (g_n) divided by Return on Investment assumed stable over time (ROI_{SG}), as follows:

$$IR_{GS} = \frac{g_n}{ROI_{GS}}$$

(B) **Two-Stage Growth EV-DM$_{CC}$**

The *Two-Stage Growth EV-DM$_{CC}$* (*2S-EV-DM$_{CC}$*) estimates the Enterprise Value by assuming two different growth periods:

- *extraordinary growth period*: the first period is the growth rate of dividend that cannot be considered stable over time. In this period lasting in the first n years $(t = 1 \rightarrow t = n)$ the growth rate of dividend can be higher or lower than the stable growth rate. The term "extraordinary" is used because the growth rate of dividends in this first period can be greater or lower than the second period;
- *steady-state growth period*: the second period is the growth rate of dividend and is assumed stable over time. In this second period $(t = n + 1 \rightarrow t = \infty)$ growth rate of dividend is assumed stable over time.

On the basis of these two periods, the 2S-EV-DM$_{CC}$ estimates the Enterprise Value (W_A) equal to the present value of Free Cash Flow from Operations $(FCFO)$ during the extraordinary growth period (g_e) plus the present value of Terminal Value (TV) estimated at the end of the extraordinary period for the steady-state growth period (second period) by assuming a constant growth rate in perpetuity

(g_n). The discounted rate is the Cost of Capital $\left(K_{A,eg}\right.$ in the first period "extraordinary growth" and $K_{A,sg}$ in the second period "Steady-state growth"). Formally:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{A,eg}\right)^t} + \frac{TV_n}{\left(1 + K_{A,st}\right)^n} \tag{9.4}$$

In the first period, the FCFO's are analytically estimated for each year (from $t = 1$ to $t = n$). Consequently, the extraordinary growth rate (g_e) is incorporated in the analytical estimation of FCFO and it is equal in each year as follows:

$$g_{e_t} = \frac{FCFO_t - FCFO_{t-1}}{FCFO_{t-1}} = \frac{FCFO_t}{FCFO_{t-1}} - 1 \tag{9.5}$$

The terminal value (TV) can be calculated by using steady-state growth rate (g_n) in perpetuity as follows:

$$TV_n = \frac{FCFO_{n+1}}{K_{A,st} - g_n} \tag{9.6}$$

Note that the $FCFO_{n+1}$ must be estimated by assuming their achievement over time in perpetuity. Therefore, it can be estimated independently to the FCFO of the first period.

Substituting Eq. (9.6), Eq. (9.4) can be rewritten as follows:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{A,eg}\right)^t} + \frac{\frac{FCFO_{n+1}}{K_{A,st} - g_n}}{\left(1 + K_{A,st}\right)^n}$$

and therefore:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{A,eg}\right)^t} + \frac{FCFO_{n+1}}{\left(K_{A,st} - g_n\right)\left(1 + K_{A,st}\right)^n} \tag{9.7}$$

Equation (9.7) can be divided in two parts:

- the first part is the present value of the expected FCFO estimated analytically for each year in the first period $(t = 1 \rightarrow t = n)$. In this first period, the growth rate (g_e) can be defined extraordinary because it is different (higher or lower) from the steady-state growth rate (g_n) used in the second period in perpetuity. Also g_e can be changed each year with regards to the previous year. Therefore, the first part of equation defines the "*analytical value*";
- the second part is the present value of the Terminal Value estimated on the basis of a synthetic estimation of the FCFO capable of being achieved each year in perpetuity. Consequently, in this second period the steady-state growth rate (g_n)

is used, assumed constant in perpetuity. Therefore, the second part of equation defines the "syntactical value".

Note that the Cost of Capital should be different in the two periods $\left(K_{A,eg} \neq K_{A,sg}\right)$. In fact, it is based on the cost of equity and the cost of debt. Even if it is assumed that the cost of equity is the same, the cost of debt reflects the choices about capital structure. Therefore, it is reasonable to assume that the capital structure in the second period characterized by a steady-state condition may be different from the first period characterized by an extraordinary growth.

The 2S-EVDM$_{CC}$ is useful for the company characterized by different growth rates. Specifically, the model can be used if the company is characterized by a growth rate in the first period higher or lower than the stable growth rate of the second period.

There are two main problems of the model (Damodaran 2012):

- first, it is difficult to estimate the length of the extraordinary growth period;
- second, the growth rate is assumed higher or lower in the first period and in declining or increasing at the end of the first period until a stable level characterises the second period. Even if it happens, it is reasonable to assume that the shift happens gradually over time.

By considering this second limit, it can be more useful the three-stage model.

(C) Three-Stage Growth EV-DM$_{CC}$

The *Three-Stage Growth EV-DM$_{CC}$* (*3S-EV-DM$_{CC}$*) estimates the Enterprise Value by assuming two different growth periods:

- *Extraordinary growth period*: it is the first period and it is characterized by high or low growth rate;
- *Transitional growth period*: it is the second period and it is characterized a growth rate declining or increasing to reach a stable level;
- *Steady-state growth period*: it is the third period and it is characterized by a steady-state growth rate.

Therefore, the 3S-EV-DM$_{CC}$ exceeds the limit of the 2S-EV-DM$_{CC}$ by introducing a new step of growth. Indeed, the transitional growth period is used to gradually stabilize the extraordinary growth in the first period until the steady-state condition of the third period.

The 3S-EV-DM$_{CC}$ estimates the Enterprise Value equal to the sum of present values of expected Free Cash Flow from Operations (FCFO) in the extraordinary growth period (first period), in the transaction period (second period) and in the steady-state period (third period).

By considering the extraordinary growth rate (g_e), transitional growth rate (g_{tr}) and the steady-state growth rate (g_n) and the cost of equity (K_A) in the first period $\left(K_{A,eg}\right)$, second period $\left(K_{A,tr}\right)$ and third period $\left(K_{A,st}\right)$, the Enterprise Value (W_A) is equal to:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1+K_{A,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFO_t}{\left(1+K_{A,tr}\right)^t} + \frac{TV_m}{\left(1+K_{A,st}\right)^m} \tag{9.8}$$

where:

$$g_{e_t} = \frac{FCFO_t}{FCFO_{t-1}} - 1 \; for\, t = 1 \rightarrow t = n$$

$$g_{tr_t} = \frac{FCFO_t}{FCFO_{t-1}} - 1 \; for\, t = n+1 \rightarrow t = m$$

The terminal value (TV_m) can be computed by using the infinite growth rate model:

$$TV_m = \frac{FCFO_{m+1}}{K_{A,st} - g_n} \tag{9.9}$$

Substituting Eq. (9.9), Eq. (9.8) can be rewritten as follows:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1+K_{A,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFO_t}{\left(1+K_{A,tr}\right)^t} + \frac{\frac{FCFO_{m+1}}{K_{A,st}-g_n}}{\left(1+K_{A,st}\right)^m}$$

and therefore:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1+K_{A,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFO_t}{\left(1+K_{A,tr}\right)^t} + \frac{FCFO_{m+1}}{\left(K_{A,st} - g_n\right)\left(1+K_{A,st}\right)^m} \tag{9.10}$$

Equation (9.10) can be divided in two parts:

- The first part and second, they are the present values of the expected FCFO estimated analytically for each year in the first period $(t = 1 \rightarrow t = n)$ and in the second period $(t = n+1 \rightarrow t = m)$. In this first period, the extraordinary growth rate (g_e) can be mitigated in the second period by passing to the transitional growth rate (g_{tr}) in the second period until it becomes stable and constant over time (g_n) in the steady-state condition of the third period. Both g_e and g_{tr} can be changed each year compared with the previous year. Therefore, the first part and second part of the equation define the "*analytical value*";
- The third part is the present value of the Terminal Value estimated on the basis of a synthetic estimation of the FCFO capable of being achieved each year in perpetuity. Consequently, in this third period the steady-state growth rate (g_n) is used, assumed constant in perpetuity. Therefore, the second part of equation defines the "*syntactical value*".

As discussed in the 2S-EV-DM_{CC} the cost of capital may be different across the three periods $\left(K_{A,eg} \neq K_{A,tr} \neq K_{A,st}\right)$. It is reasonable to assume that the company's capital structure changes over time until it stabilises during the last period. Therefore, the debt level changes as well as the cost of capital.

9.3 Free Cash Flow from Operations Discounted Model: Adjusted Present Value Approach

The *Free Cash Flow from Operations Discounted Model based on the Adjusted Present Value approach* (*EV-DM$_{APV}$*) derived from the Arbitrage Price Theory (APT) (Myers 1974). Generally, the cost of capital approach, as discussed previously, is preferred when the debt level is stable over time. This is even more true when the WACC is used as the cost of capital. Differently, when the debt level changes over time, the Adjusted Present Value (APV) approach is more useful to estimate enterprise value (Koller at al. 2015).

The EV-DM$_{\text{APV}}$ estimates the Enterprise Value by distinguishing between:

- the *unlevered value*: it is the value of the company without debt
- the *leverage net value*: it is the value of the company related to the benefits and costs of debt.

Consequently, the EV-DM$_{\text{APV}}$ estimates the Enterprise Value on three main steps (Damodaran 2012):

- the unlevered value and then the enterprise value without debt;
- the value of tax shields due to the interests on debt benefits. It is the value of advantage of debt and it is added to the unlevered value;
- the value of default risk related to the debt. It is the value of costs of debt and it is subtracted from the unlevered value.

Therefore, the Enterprise Value (W_A) is equal to the Unlevered Value (W_{AU}) plus the Value of the Tax Shields (W_{TS}) minus the Value of the Default Risk and Bankruptcy Costs (W_{DB}) related to the debt. Formally:

$$W_A = W_{AU} + W_{TS} - W_{DB} \tag{9.11}$$

Equation (9.11) valuates the debt effects, both positive and negative, separately from the unlevered value. On the other hand, in the Cost of Capital Approach the positive and negative effects of debt are considered directly in the cost of capital used to discount FCFO.

Unlevered Value

The first term in the model is the *Unlevered Enterprise Value*. It is the Enterprise Value without debt in the capital structure. The expected Free Cash Flow from Operations (FCFO) are discounted at the Unlevered Cost of Capital (K_A) that is equal to the Unlevered Cost of Equity (K_{EU}) because there is no debt:

$$K_A = K_{EU}$$

Therefore, the Unlevered Enterprise Value (W_{AU}) is equal to:

$$W_{AU} = \sum_{t=1}^{\infty} \frac{FCFO_t}{(1 + K_{EU})^t} \tag{9.12}$$

It is possible to use the versions (steady-state growth, two-stage growth and three-stage growth) of EV-DM_{CC}, to apply Eq. (9.12).

In the Steady-State growth scenario, the Unlevered Enterprise Value is equal to:

$$W_{AU} = \frac{FCFO_1}{K_{EU} - g_n} \tag{9.13}$$

In the Two-Stage growth scenario, the Unlevered Enterprise Value is equal to:

$$W_{AU} = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{EU,eg}\right)^t} + \frac{TV_n}{\left(1 + K_{EU,st}\right)^n}$$

where

$$TV_n = \frac{FCFO_{n+1}}{K_{EU,st} - g_n}$$

and then:

$$W_{AU} = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{EU,eg}\right)^t} + \frac{FCFO_{n+1}}{\left(K_{EU,st} - g_n\right)\left(1 + K_{EU,st}\right)^n} \tag{9.14}$$

In the Three-Stage growth scenario, the Unlevered Enterprise Value is equal to:

$$W_{AU} = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{EU,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFO_t}{\left(1 + K_{EU,tr}\right)^t} + \frac{TV_m}{\left(1 + K_{EU,st}\right)^m}$$

where:

$$TV_m = \frac{FCFO_{m+1}}{K_{EU,st} - g_n}$$

And then:

$$W_{AU} = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{EU,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFO_t}{\left(1 + K_{EU,tr}\right)^t} + \frac{FCFO_{m+1}}{\left(K_{EU,st} - g_n\right)\left(1 + K_{EU,st}\right)^m} \tag{9.15}$$

The most relevant difference between the Enterprise Value and the Unlevered Enterprise Value is due to the Unlevered Cost of Equity.

Specifically, the main problem is related to the determination of the beta coefficient. In an Unlevered Enterprise Value the Unlevered Beta may be considered (Hamada 1972). Specifically, the Unlevered Beta (β_U) can be defined on the basis of Levered Beta (β_L) that is the current beta if the capital structure is based on equity and debt, the corporate taxes (t_c) and the Leverage equal to the ratio between debt level and equity (D/E), as follows:

$$\beta_U = \frac{\beta_L}{1 + (1 - t_c)\frac{D}{E}} \tag{9.16}$$

Therefore, the Unlevered Cost of Equity (K_{EU}) can be estimated as follows:

$$K_{EU} = R_f + \beta_U \left(R_m - R_f\right)$$

and then:

$$K_{EU} = R_f + \frac{\beta_L}{1 + (1 - t_c)\frac{D}{E}} \left(R_m - R_f\right) \tag{9.17}$$

Value of Tax Shields

The second term in the model, is the *Value of the Tax Shields*. Generally, interests on debt are tax deductible. They generate a positive cash-flow due to tax savings. Therefore, they are function of corporate tax rate and the amount of the interests on debt.

It is worth noting that the tax shields can be estimated for a given debt level. Indeed, they can be estimated only if the debt level is known.

The value of the tax shields can be estimated equal to the present value of the expected cash-flow of the tax savings discounted to the appropriate rate on the basis of their risk level.

The cash-flow from tax savings (CF_{TS}) can be estimated equal to the corporate taxed (t_c) multiplies interest on debt (I_D). Specifically, interest on debt can be defined as the cost of debt (K_D) multiplied by the debt level (D). Therefore, we have:

$$CF_{TS} = t_c I_D \rightarrow I_D = K_D D \rightarrow CF_{TS} = t_c K_D D$$

Denoted with K_{TS} the discount rate of the cash-flow from tax savings (CF_{TS}), the value of the tax shields (W_{TS}) can be estimated as follows:

$$W_{TS} = \frac{t_c K_D D}{K_{TS}} \tag{9.18}$$

The K_{TS} definition is the main problem of Eq. (9.18). It measures the function of the risk level of the cash-flow from tax savings. Usually, the value of this discount rate is assumed equal to the cost of debt (K_D). The basic reason for this relationship is that tax shields exist because there is a debt. If the company does not pay interest on debt, there are no cash-flow from tax savings. Therefore, the risk level of cash-flow from tax shields can be assumed equal to the risk level of debt. By assuming that $K_{TS} = K_D$, Eq. (9.18) can be rewritten as follows:

$$W_{TS} = \frac{t_c K_D D}{K_D} = t_c D \tag{9.19}$$

Value of Bankruptcy Costs

The third term of the model is the *Value of Bankruptcy Costs* in case of the company's default related to the leverage. In the case of default, the company has to bear in several bankruptcy costs, both direct and indirect. Therefore, the value of default risk is assumed equal to the bankruptcy direct and indirect costs in case of default.

It is worth noting that these costs can be measured for a given debt level. Therefore, also in this case (as for cash-flows from tax savings) the value of default risk can be estimated only if the debt level is known.

Generally, the value of bankruptcy costs (W_{BC}) is estimated equal to the probability (π_D) of default related to the increase in debt from a given level, multiplied by the expected cash-flow related to the bankruptcy direct and indirect costs (CF_{BC}), as follows:

$$W_{BC} = \pi_D \cdot CF_{BC} \tag{9.20}$$

The Eq. (9.20) is the most complex to apply because neither the probability of the company's default (π_D) nor the bankruptcy costs (CF_{BC}) can be estimated directly. There are two main approaches:

- the first approach estimates a bond rating and it uses the empirical estimates of default probabilities for the rating;
- the second approach uses the statistical techniques based on the company's fundamental characteristics that can be observed.

With regards to bankruptcy costs, several studies show that the direct costs are lower than indirect costs in company value. Generally, indirect costs are estimated equal to 25%-30% of company value (Titman 1984).

The defined Unlevered Enterprise Value (W_{AU}), Value of Tax Shields (W_{TS}) and Value of Bankruptcy Costs (W_{BC}), Eq. (9.11) can be applied.

While the first part (W_{BC}) can be declined in three growth scenarios, the second part (W_{TS}) and the third part (W_{BC}) of the equation are the same. Therefore, Eq. (9.11):

$$W_A = W_{AU} + W_{TS} - W_{DB}$$

in steady-state growth scenario, we have:

$$W_A = \frac{FCFO_1}{K_{EU} - g_n} + t_c D - \pi_D \cdot CF_{BC} \tag{9.21}$$

in two-stage growth scenario, we have:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{EU,eg}\right)^t} + \frac{FCFO_{n+1}}{\left(K_{EU,st} - g_n\right)\left(1 + K_{EU,st}\right)^n} + t_c D - \pi_D \cdot CF_{BC} \tag{9.22}$$

in three-stage growth scenario, we have:

$$W_A = \sum_{t=1}^{n} \frac{FCFO_t}{\left(1 + K_{EU,eg}\right)^t} + \sum_{t=n+1}^{m} \frac{FCFO_t}{\left(1 + K_{EU,tr}\right)^t} + \frac{FCFO_{m+1}}{\left(K_{EU,st} - g_n\right)\left(1 + K_{EU,st}\right)^m} + t_c D - \pi_D \cdot CF_{BC} \tag{9.23}$$

The relationship between the capital structure choices and the company's value can be analysed more efficiently in an APV approach rather than the cost of capital approach.

The main APV approach advantage with regards to the identification of debt effects on company value through the second and third parts. However, it is not able to define clearly how and in which measure changes in the debt level affect the default risks and the bankruptcy costs. This is the main reason why the APV approach is commonly used with a given debt level.

9.4 Discounted Economic Profit

The Economic Profit (EP) model (also called Economic Value Added (EVA)) is a direct measurement of the company's ability to create value. Specifically, the EPM highlights how and when the company creates value (Koller et al. 2015).

The EP is based on a simple reasoning: the company generates value if the return on investments is greater than the cost of capital invested. The positive difference between the investment returns and the cost of capital, defines the excess returns. Therefore, the excess returns measure the value creation.

Denoting with ER_A the excess returns of the company, RI_A the Return on Investments of the company, and with K_A the Cost of Capital (both equity and debt) of the company, we have:

$$ER_A = RI_A - K_A \tag{9.24}$$

As shown by Eq. (9.24), the EPM measures the value created by the company in a single period.

Denoting with *CIN* the Net Capital Invested by the company, *NOPAT* the Net Operating Profit After Taxes (also called NOPLAT—Net Operating Profit Less Adjusted Taxes), *ROIC* the Return on Invested Capital, and K_A the Cost of Capital of the company, Eq. (9.24) can be applied to estimate Economic Profit (EP) as follows (Koller et al. 2015):

$$EP = CIN \cdot (ROIC - K_A)$$

Considering that:

$$ROIC = \frac{NOPAT}{CIN}$$

and substituting, we have:

$$EP = CIN \cdot \left(\frac{NOPAT}{CIN} - K_A\right)$$

and then:

$$EP = NOPAT - K_A \cdot CIN \tag{9.25}$$

Equation (9.25) measures the value created by the firm in a single period. It is possible to estimate the Enterprise Value (W_A) by using the Discounted Economic Profit (DEP) as follows:

$$W_{A_{(t=0)}} = CIN_{(t=0)} + \sum_{t=1}^{\infty} \frac{EP_t}{(1+K_A)^t} \tag{9.26}$$

In order to apply Eq. (9.26), the analytical and synthetic valuation can be distinguished, as follows:

$$W_{A_{(t=0)}} = CIN_{(t=0)} + \sum_{t=1}^{n} \frac{EP_t}{(1+K_A)^t} + \frac{PV(EP_n)}{(1+K_A)^n}$$

where $PV(EP_n)$ is the Present Value of the Economic Profit at the end of the analytical period $(t = n)$. Considering the steady-state growth rate (g_n), it can be estimated as follows:

$$PV(EP_n) = \frac{EP_{n+1}}{K_A - g_n}$$

and substituting, we have:

$$W_{A_{(t=0)}} = CIN_{(t=0)} + \sum_{t=1}^{n} \frac{EP_t}{(1+K_A)^t} + \frac{\frac{EP_{n+1}}{K_A - g_n}}{(1+K_A)^n}$$

and then:

$$W_{A_{(t=0)}} = CIN_{(t=0)} + \sum_{t=1}^{n} \frac{EP_t}{(1+K_A)^t} + \frac{EP_{n+1}}{(K_A - g_n)(1+K_A)^n} \tag{9.27}$$

Equation (9.27) shows how the Discounted Economic Profit is based on three terms:

- the net capital invested in the assets in place;$(t = 0)$;
- the present value of the expected Economic Profit in analytical period $(t = 1 \rightarrow n)$;
- the present value of the expected Economic Profit in steady-state condition $(t = n \rightarrow \infty)$.

The use of DEP to evaluate the Enterprise Value is also due to compatibility between two models. Indeed, the DEP can be considered equivalent, under defined assumptions, to the Enterprise Discounted Cash-Flow (EV-DM). It can be demonstrated as follows (Koller et al. 2015).

Considering the following baseline relationships:

(1) Net Investment (NI) can be defined as the increase in Net Capital Invested (NCI) in the core-business of the company as follows:

$$NI = \Delta NCI = NCI_{t+1} - NCI_t$$

(2) Return on Invested Capital (ROIC) can be defined as the ratio between the NOPAT and the Net Capital Invested as follows:

$$ROIC = \frac{NOPAT}{NCI} \rightarrow \begin{cases} NOPAT = ROIC \cdot NCI \\ NCI = \frac{NOPAT}{ROIC} \end{cases}$$

(3) Investment Rate (IR) can be defined as a portion of the NOPAT invested back into the core-business. Consequently, it can be estimated equal to the ratio between Net Investment (NI) and NOPAT as follows:

$$IR = \frac{NI}{NOPAT} \rightarrow \begin{cases} NI = IR \cdot NOPAT \\ NOPAT = \frac{NI}{IR} \end{cases}$$

(4) The growth rate (g) can be estimated on the basis of ROIC and IR as follows:

$$g = ROIC \cdot IR \rightarrow \begin{cases} IR = \frac{g}{ROIC} \\ ROIC = \frac{g}{IR} \end{cases}$$

(5) Net Investment (NI) in core-business is the capital invested in Capex plus Net Working Capital (NWC) less Provisions (Pro). If the Operating Taxes (OT) are considered, then the NOPAT can be considered rather than EBITDA (NOPAT = EBITDA-OT). In this case, the Free Cash Flow From Operations (FCFO) can be defined as follows:

$$FCFO = EBITDA - OT + \Delta Capex + \Delta NWC - Pro$$
$$EBITDA - OT = NOPAT; \Delta Capex + \Delta NWC - Pro = \Delta NCI = NCI_{t+1} - NCI_t = NI$$
$$FCFO = NOPAT - NI \leftrightarrow FCFO = NOPAT - (NCI_{t+1} - NCI_t)$$

This relationship can be rewritten on the basis of the relationship n.3, as follows:

$$FCFO = NOPAT - (NOPAT \cdot IR) = NOPAT \cdot (1 - IR)$$

and on the basis of relationship n.4, as follows:

$$FCFO = NOPAT \cdot \left(1 - \frac{g}{ROIC}\right)$$

Now consider the Enterprise Value Discounted Model in the version of constant growth rate (one-period growth):

$$W_A = \frac{FCFO_{(1)}}{K_A - g_n}$$

The FCFO's are constant over time if all variables are constant. Therefore, the FCFO in the next period $FCFO_{(1)}$ can be defined on the basis of a constant growth rate (g_n), cost of capital (K_A), and ROIC as follows:

$$FCFO_{(1)} = NOPAT_{(1)} \cdot \left(1 - \frac{g_n}{ROIC}\right)$$

and by substituting we have:

$$W_A = \frac{NOPAT_{(1)} \cdot \left(1 - \frac{g_n}{ROIC}\right)}{K_A - g_n}$$

On the basis of the relationship n.2, we have:

$$NOPAT_{(1)} = NCI_{(0)} \cdot ROIC$$

and substituting, we have:

$$W_A = \frac{NCI_{(0)} \cdot ROIC \cdot \left(1 - \frac{g_n}{ROIC}\right)}{K_A - g_n}$$

and by considering that the ROIC is constant over time, we have:

$$W_A = NCI_{(0)} \cdot \left(\frac{ROIC - g_n}{K_A - g_n}\right) \tag{9.28}$$

Equation (9.28) is useful to show the value drivers. Indeed, the Enterprise Value is function of growth rate, ROIC and the cost of capital. This equation is not used in practice because it assumes a constant value of growth rate (g_n), ROIC and the cost of capital in perpetuity.

Note that Eq. (9.28) can be transformed into a multiple Value-to-Invested Capital by dividing both terms by Net Capital Invested (NCI), as follows:

$$\frac{W_A}{NCI_{(0)}} = \frac{ROIC - g_n}{K_A - g_n} \tag{9.29}$$

In Eq. (9.29) all variables $(ROIC, K_A, g_n)$ are constant over time.

Now it is possible to rewrite Eq. (9.28) in a formula on Economic Profit (Koller et al. 2015). To do this it is necessary to demonstrate that the Enterprise Value Discounted Cash-Flow model is equivalent to the current book value of capital invested plus the present value of future economic profits. Proof can be achieved by considering the version of one-period growth as follows.

$$W_A = \frac{FCFO_1}{K_A - g_n} \rightarrow W_A = \frac{NOPAT_{(1)} \cdot \left(1 - \frac{g_n}{ROIC}\right)}{K_A - g_n} \rightarrow W_A = NCI_{(0)} \cdot \left(\frac{ROIC - g_n}{K_A - g_n}\right)$$

By subtracting and adding the Cost of Capital (K_A) in the numerator, we have:

$$W_A = NCI_{(0)} \cdot \left(\frac{ROIC - K_A + K_A - g_n}{K_A - g_n}\right)$$

and separating the fraction in two terms, we have:

$$W_A = NCI_{(0)} \cdot \left(\frac{ROIC - K_A}{K_A - g_n}\right) + NCI_{(0)} \cdot \left(\frac{K_A - g_n}{K_A - g_n}\right)$$
$$= NCI_{(0)} + NCI_{(0)} \cdot \left(\frac{ROIC - K_A}{K_A - g_n}\right)$$

and by considering that:

$$NCI_{(0)} \cdot (ROIC - K_A) = Economic\,Profit\,(EP)$$

substituting, we have:

$$W_A = NCI_{(0)} + \frac{EP_{(1)}}{K_A - g_n} \tag{9.30}$$

Equation (9.30) shows how the Enterprise Value is equal to the book value of capital invested $\left(NCI_{(0)}\right)$ plus the present value of the future economic profit $\left(EP_{(1)}\right)$. Note that the second term of the equation, is the one-period growth of the Discounted Cash-Flow models. Finally, note that:

- if the future economic profits are expected to be zero $\left(EP_{(1)} = 0\right)$, the Enterprise Value is equal to the book value of capital invested. Consequently, the intrinsic value of the Enterprise is equal to the book value;
- if the future economic profits are expected less than zero $\left(EP_{(1)} < 0\right)$, the investments destroy value and then the Enterprise Value is less than its book value.

Equation (9.30) can be generalized (Koller et al. 2015). In general terms, we have:

$$W_A = \sum_{t=1}^{\infty} \frac{FCFO_t}{(1 + K_A)^t}$$

Add and subtract the cumulative sum of Net Capital Invested $\left(\sum_{t=0}^{\infty} \frac{NCI_t}{(1+K_A)^t}\right)$ as follows:

$$W_A = \sum_{t=0}^{\infty} \frac{NCI_t}{(1 + K_A)^t} - \sum_{t=0}^{\infty} \frac{NCI_t}{(1 + K_A)^t} + \sum_{t=1}^{\infty} \frac{FCFO_t}{(1 + K_A)^t}$$

By considering the starting point $t = 0$, the first cumulative sum becomes $\left(NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NCI_t}{(1+K_A)^t}\right)$ while the second starts from $t = 1$ to infinity and then it changes each t inside in the second cumulative sum to $t - 1\left(\sum_{t=1}^{\infty} \frac{NCI_{t-1}}{(1+K_A)^{t-1}}\right)$, as follows:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NCI_t}{(1+K_A)^t} - \sum_{t=1}^{\infty} \frac{NCI_{t-1}}{(1+K_A)^{t-1}} + \sum_{t=1}^{\infty} \frac{FCFO_t}{(1+K_A)^t}$$

Multiply and divide the third term by $\frac{1+K_A}{1+K_A}$ as follows:

$$\sum_{t=1}^{\infty} \frac{NCI_{t-1}(1+K_A)}{(1+K_A)^{t-1}(1+K_A)} = \sum_{t=1}^{\infty} \frac{NCI_{t-1}(1+K_A)}{(1+K_A)^{t-1+1}} = \sum_{t=1}^{\infty} \frac{NCI_{t-1}(1+K_A)}{(1+K_A)^t}$$

and then, we have:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NCI_t}{(1+K_A)^t} - \sum_{t=1}^{\infty} \frac{NCI_{t-1}(1+K_A)}{(1+K_A)^t} + \sum_{t=1}^{\infty} \frac{FCFO_t}{(1+K_A)^t}$$

and:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NCI_t - NCI_{t-1}(1+K_A) + FCFO_t}{(1+K_A)^t}$$

and:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NCI_t - NCI_{t-1} - NCI_{t-1} \cdot K_A + FCFO_t}{(1+K_A)^t}$$

On the basis of relationship n.5, we have:

$$FCFO = NOPAT - (NCI_t - NCI_{t-1})$$

and substituting, we have:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NCI_t - NCI_{t-1} - NCI_{t-1} \cdot K_A + NOPAT_t - (NCI_t - NCI_{t-1})}{(1+K_A)^t}$$

and:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NCI_t - NCI_{t-1} - NCI_{t-1} \cdot K_A + NOPAT_t - NCI_t + NCI_{t-1}}{(1+K_A)^t}$$

and:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{NOPAT_t - NCI_{t-1} \cdot K_A}{(1+K_A)^t}$$

Considering that:

$$Economic\,Profit\,(EP) = NOPAT_t - NCI_{t-1} \cdot K_A$$

substituting, we have:

$$W_A = NCI_{(0)} + \sum_{t=1}^{\infty} \frac{EP_t}{(1 + K_A)^t} \tag{9.31}$$

Equation (9.31) shows that the Enterprise Value is equal to the book value of Capital Invested $\left(NCI_{(0)}\right)$ plus the present value of all future economic profits $\left(\sum_{t=1}^{\infty} \frac{EP_t}{(1+K_A)^t}\right)$. Equation (9.31) is the generalization of Eq. (9.30) and then:

- if the future economic profits are expected to be zero $\left(\sum_{t=1}^{\infty} \frac{EP_t}{(1+K_A)^t} = 0\right)$, the Enterprise Value is equal to the book value of capital invested. Consequently, the intrinsic value of the Enterprise is equal to the book value;
- if the future economic profits are expected less than zero $\left(\sum_{t=1}^{\infty} \frac{EP_t}{(1+K_A)^t} < 0\right)$, the investments destroy value and then the Enterprise Value is less than its book value.

The equivalence between Discounted Economic Profit and the Enterprise Value Discounted Cash-Flows, as shown by Eqs. (9.30), (9.31), requires defined assumptions (Koller et al. 2015):

- use beginning-of-year invested capital instead of average or current-year invested capital;
- define invested capital for both economic profit and ROIC using the same value;
- use a constant cost of capital as discount rate.

Finally, note that also in this case by using the EPM it is possible to move from the Enterprise Value (W_A) to Equity Value (W_E). In this case, it is necessary to subtract the Enterprise Value from the Net Financial Position (NFP) as follows:

$$W_E = W_A - PFN \tag{9.32}$$

and by considering Eq. (9.30), we have:

$$W_{E_{(t=0)}} = \left[NCI_{(0)} + \frac{EP_{(1)}}{K_A - g_n}\right] - PFN_{(t=0)} \tag{9.33}$$

and by considering Eq. (9.31), we have:

$$W_{E_{(t=0)}} = \left[NCI_{(0)} + \sum_{t=1}^{\infty} \frac{EP_t}{(1+K_A)^t}\right] - PFN_{(t=0)} \tag{9.34}$$

and by considering Eq. (9.27), we have:

$$W_{E_{(t=0)}} = \left[CIN_{(t=0)} + \sum_{t=1}^{n} \frac{EP_t}{(1+K_A)^t} + \frac{EP_{n+1}}{(K_A - g_n)(1+K_A)^n}\right] - PFN_{(t=0)} \tag{9.35}$$

9.5 Multiplies on Enterprise Value

The relative valuation to estimate the Enterprise Value can be achieved on the basis of several multiples. Among these there are three most relevant ones in the financial approach:

(a) Enterprise Value-to-EBITDA (EV/EBITDA) ratio;
(b) Enterprise Value-to-Book Value (EV/BV) ratio;
(c) Enterprise Value-to-Sales (EV/S) ratio;

Enterprise Value-to-EBITDA

The multiple *Enterprise Value-to-EBITDA* (*EV/EBITDA*) ratio is one of the most commonly used to estimate the Enterprise Value. The main advantages of this multiple refer to the use of EBITDA. Indeed:

– the EBITDA is always positive. Its negative value identifies an anomalous condition of the company;
– the EBITDA has a direct impact on Free Cash Flow and then on the financial dynamic of the company;
– the EBITDA does not consider the depreciation and amortization process. Therefore, it allows for the comparison of companies with different level of investments;
– the EBITDA does not consider the effects of the leverage. Therefore, it allows for the comparison of companies with different capital structures.

The Enterprise Value-to-EBITDA is equal to the ratio between Enterprise Value (EV) and the EBITDA as follows:

$$EV/EBITDA = \frac{EV}{EBITDA} \tag{9.36}$$

A relationship can be defined between the company's fundamentals and the multiple. This relationship can be analysed on the basis of the Free Cash-Flow from Operations Discounted Model in the version of Cost of Capital (EV-DM_{CC}). Specifically, this relationship can be defined in two scenarios (Damodaran 2012):

– steady-state growth over time scenario;
– two-stage growth scenario: extraordinary growth in the first period and steady-state growth in the second period.

In the steady-state growth over time scenario, the Enterprise Value (W_A) can be estimated on the basis of EV-DM_{CC} as follows:

$$W_A = \frac{FCFO}{K_A - g_n}$$

where *FCFO* are the Free Cash-Flow from Operations that the company can achieve in perpetuity in steady-state scenario and g_n is the steady-state growth rate.

By estimating the Enterprise Value with multiple $EV/EBITDA$, we have:

$$W_A \equiv EV$$

Remembering that Free Cash Flow from Operations (FCFO) is equal to EBITDA, plus/(minus) changes in Net Capex $(\Delta Capex)$, plus/(minus) changes in Net Working Capital (ΔNWC), minus cash-out due to negative changes in Provisions (ΔP), we have:

$$FCFO = EBITDA \pm \Delta Capex \pm \Delta NWC - \Delta P$$

and substituting, we have:

$$W_A \equiv EV = \frac{EBITDA \pm \Delta Capex \pm \Delta NWC - \Delta P}{K_A - g_n}$$

By dividing both terms by *EBITDA*, we have:

$$\frac{EV}{EBITDA} = \frac{EBITDA \pm \Delta Capex \pm \Delta NWC - \Delta P}{EBITDA(K_A - g_n)}$$

and then:

$$\frac{EV}{EBITDA} = \frac{1}{(K_A - g_n)}\left[1 + \frac{1}{EBITDA}(\pm \Delta Capex \pm \Delta NWC - \Delta P)\right] \tag{9.37}$$

Equation (9.37) defines the relationship between the multiple $EV/EBITDA$ and the company's fundamentals. Specifically, the equation shows as the multiple is affected by (Damodaran 2012):

– *Depreciation and Amortization process*: if the company derives a greater portion of its EBITDA from depreciation and amortization it should trade at lower multiples of EBITDA than similar companies;

– *Net Capex and Net Working Capital*: the greater the portion of the EBITDA that needs to be reinvested in the company to generate expected future growth, the lower multiple of EBITDA;
– *Cost of Capital*: a company with lower cost of capital should trade at much higher multiples of EBITDA;
– *Steady-state growth rate*: the higher the steady-state growth rate, higher the multiplis of EBITDA.

The EBIT can be used instead of the EBITDA. In this case the multiple becomes EV-to-EBITDA and Eq. (9.36) can be rewritten as follows:

$$EV/EBIT = \frac{EV}{EBIT} \tag{9.38}$$

In this case, the relationship between the multiple $EV/EBIT$ and the company's fundamentals defined in Eq. (9.37) does not change. Indeed, in cash-flow perspective the EBIT is equal to the EBITDA.

In the two-stage growth scenario (extraordinary growth in the first period and steady-state growth in the second period) the Enterprise Value (W_A) can be estimated on the basis of 2S-EVDM$_{CC}$ as follows:

$$W_A = \frac{FCFO(1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{K_{E,eg} - g_e} + \frac{FCFO(1+g_e)^n(1+g_n)}{(K_{E,sg} - g_n)(1+K_{E,sg})^n}$$

where g_e is the growth rate in the extraordinary growth period (first period) and g_n is the growth rate in the steady-state period (second period).

By considering that:

$$W_A \equiv EV$$

and:

$$FCFO = EBITDA \pm \Delta Capex \pm \Delta NWC - \Delta P$$

Substituting, we have:

$$EV = \frac{[EBITDA \pm \Delta Capex \pm \Delta NWC - \Delta P](1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{K_{E,eg} - g_e} + \frac{[EBITDA \pm \Delta Capex \pm \Delta NWC - \Delta P](1+g_e)^n(1+g_n)}{(K_{E,sg} - g_n)(1+K_{E,sg})^n}$$

and then:

$$EV = [EBITDA \pm \Delta Capex \pm \Delta NWC - \Delta P]\left[\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e} + \frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right]$$

And dividing both terms of equation by EBITDA, we have:

$$\frac{EV}{EBITDA} = \left[1+\frac{1}{EBITDA}(\pm\Delta Capex \pm \Delta NWC - \Delta P)\right]\left[\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e} + \frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right] \quad (9.39)$$

Equation (9.39) shows the relationship between the multiple $EV/EBITDA$ and the company's fundamentals in the two-stage growth scenario.

Value-to-Book Value

The multiple *Enterprise Value-to-Book* (*EV/BV*) ratio estimates the Enterprise Value on the basis of the market value and book value. Specifically, the multiple is equal to the ratio between the Enterprise Value (EV) and the book value of equity and debt (BV), as follows:

$$EVBV = \frac{EV}{BV} \quad (9.40)$$

Therefore, it is similar to the multiple Price-to-Book (PBV).

A relationship between the company's fundamentals and the multiple can be defined. This relationship can be analysed on the basis of the Free Cash-Flow from Operations Discounted Model in the version of Cost of Capital (EV-DM_{CC}). Specifically, this relationship can be defined in two scenarios (Damodaran 2012):

- steady-state growth over time scenario;
- two-stage growth scenario: extraordinary growth in the first period and steady-state growth in the second period.

In steady-state growth over time scenario, the Enterprise Value (W_A) can be estimated on the basis the basis of EV-DM_{CC} as follows:

$$W_A = \frac{FCFO}{K_A - g_n}$$

where $FCFO$ are the Free Cash-Flow from Operations that the company can achieve in perpetuity in steady-state scenario and g_n is the steady-state growth rate.

Also in this case, we have:

$$W_A \equiv EV$$

Assuming that a cash-flow needs to be invested in Capex, Net Working Capital and Provisions is a part of EBITDA, the FCFO can be expressed in terms of EBITDA and Reinvestment Rate (RR) as follows:

$$FCFO = EBITDA(1 - RR)$$

Substituting, we have:

$$EV = \frac{EBITDA(1 - RR)}{K_A - g_n}$$

Dividing both terms by the Book Value of the Capital Invested, we have:

$$\frac{EV}{BV} = \frac{EBITDA(1 - RR)}{BV(K_A - g_n)}$$

$$\frac{EV}{BV} = \frac{EBITDA(1 - RR)}{BV(K_A - g_n)} \tag{9.41}$$

Equation (9.41) shows the relationship between the multiple and company fundamentals. This equation can be further developed.

The Book Value of Equity and Debt defines the Capital Structure (CS) of the company. By definition it must be equal to the Capital Invested (CI). By assuming that the entire capital is invested in core-business of the firm, we have:

$$ROI = \frac{EBITDA}{BV}$$

And substituting, we have:

$$\frac{EV}{BV} = \frac{ROI(1 - RR)}{K_A - g_n}$$

In a steady-state growth scenario, the Reinvestment Rate (RR) is equal to the ratio between the steady-state growth rate (g_n) and ROI. Indeed, it is the requirement of the reinvestment to increase the ROI by following the growth rate:

$$g_n = RR \cdot ROI \rightarrow RR = \frac{g_n}{ROI}$$

and substituting, we have:

$$\frac{EV}{BV} = \frac{ROI\left(1 - \frac{g_n}{ROI}\right)}{K_A - g_n}$$

And then:

$$\frac{EV}{BV} = \frac{ROI - g_n}{K_A - g_n} \tag{9.42}$$

Equation (9.42) defines further the relationship between the multiple EV/BV and the company's fundamentals.

In the two-stage growth scenario (extraordinary growth in the first period and steady-state growth in the second period) the Enterprise Value (W_A) can be estimated on the basis of 2S-EVDM$_{CC}$ as follows:

$$W_A = \frac{FCFO(1 + g_e)\left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg} - g_e} + \frac{FCFO(1+g_e)^n(1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}$$

By considering that:

$$W_A \equiv EV$$

and:

$$FCFO = EBITDA(1 - RR)$$

Substituting, we have:

$$EV = \frac{EBITDA(1 - RR)(1 + g_e)\left[1 - \frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg} - g_e} + \frac{EBITDA(1 - RR)(1+g_e)^n(1+g_n)}{\left(K_{E,sg} - g_n\right)\left(1 + K_{E,sg}\right)^n}$$

and then:

$$EV = EBITDA(1-RR)\left\{\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e}+\frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\}$$

By dividing both terms by the Book Value of the Capital Invested, we have:

$$\frac{EV}{BV}=\frac{EBITDA(1-RR)}{BV}\left\{\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e}+\frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\} \tag{9.43}$$

Equation (9.43) shows the relationship between the multiple EV/BV and company's fundamentals. This equation can be further developed (Damodaran 2012).

The Book Value of Equity and Debt defines the Capital Structure (CS) of the company. By definition it must be equal to the Capital Invested (CI). By assuming that the entire capital is invested in the core business of the company, we have:

$$ROI=\frac{EBITDA}{BV}$$

And substituting, we have:

$$\frac{EV}{BV}=ROI(1-RR)\left\{\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e}+\frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\}$$

By considering that the reinvestment rate (RR) must be equal to the growth rate (g_e) to achieve the ROI, we have:

$$g_e=RR\cdot ROI\rightarrow RR=\frac{g_e}{ROI}$$

Therefore, by substituting we have:

$$\frac{EV}{BV}=ROI\left(1-\frac{g_e}{ROI}\right)\left\{\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e}+\frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\}$$

and then:

$$\frac{EV}{BV} = (ROI - g_e)\left\{\frac{(1+g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{K_{E,eg} - g_e} + \frac{(1+g_e)^n(1+g_n)}{(K_{E,sg} - g_n)(1+K_{E,sg})^n}\right\} \tag{9.44}$$

Equation (9.44) is a further relationship between the multiple EV/BV and company's fundamentals.

Enterprise Value-to-Sales

The multiple Value-to-Sales (EV/S or EVS) ratio is calculated based on the value of the operating assets to revenues. Specifically, it is equal to the ratio between Enterprise Value (EV) and Sales Revenues (S) as follows:

$$EVS = \frac{EnterpriseValue\,(EV)}{Sales} \tag{9.45}$$

Also in this case, it is possible to define a relationship between the company's fundamentals and the multiple. This relationship can be analysed on the basis of the Free Cash-Flow from Operations Discounted Model in the version of Cost of Capital (EV-DM_{CC}). Specifically, this relationship can be defined in two scenarios (Damodaran 2012):

- steady-state growth over time scenario;
- two-stage growth scenario: extraordinary growth in the first period and steady-state growth in the second period.

In the steady-state growth over time scenario, the Enterprise Value (W_A) can be estimated on the basis the basis of EV-DM_{CC} as follows:

$$W_A = \frac{FCFO}{K_A - g_n}$$

where $FCFO$'s are the Free Cash-Flow from Operations that the company can achieve in perpetuity in steady-state scenario and g_n is the steady-state growth rate.

By considering that:

$$W_A \equiv EV$$

And by considering the to define the FCFO based on EBITDA and Retention Rate (RR), we have:

$$FCFO = EBITDA(1 - RR)$$

Substituting, we have:

$$W_A \equiv EV = \frac{EBITDA(1 - RR)}{K_A - g_n}$$

Dividing both terms by the Sales Revenues (S), we have:

$$\frac{EV}{S} = \frac{EBITDA(1 - RR)}{S(K_A - g_n)}$$

The ratio between EBITDA and Revenues Sales defines the Operating Margin (OM):

$$OM = \frac{EBITDA}{S}$$

and substituting, we have:

$$\frac{EV}{S} = OM \frac{(1 - RR)}{K_A - g_n} \tag{9.46}$$

Equation (9.46) defines the relationship between the EV/S multiple and the company's fundamentals in the steady-state growth scenario.

In the two-stage growth scenario (extraordinary growth in the first period and steady-state growth in the second period) the Enterprise Value (W_A) can be estimated on the basis of 2S-EVDM$_{CC}$ as follows:

$$W_A = \frac{FCFO(1 + g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{K_{E,eg} - g_e} + \frac{FCFO(1 + g_e)^n(1 + g_n)}{(K_{E,sg} - g_n)(1 + K_{E,sg})^n}$$

By considering that:

$$W_A \equiv EV$$

and:

$$FCFO = EBITDA(1 - RR)$$

Substituting, we have:

$$EV = \frac{EBITDA(1 - RR)(1 + g_e)\left[1 - \frac{(1+g_e)^n}{(1+K_{E,eg})^n}\right]}{K_{E,eg} - g_e} + \frac{EBITDA(1 - RR)(1 + g_e)^n(1 + g_n)}{(K_{E,sg} - g_n)(1 + K_{E,sg})^n}$$

Dividing both terms by Revenues Sales (S), we have:

$$\frac{EV}{S} = \frac{EBITDA(1-RR)(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{S\left(K_{E,eg}-g_e\right)} + \frac{EBITDA(1-RR)(1+g_e)^n(1+g_n)}{S\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}$$

and then:

$$\frac{EV}{S} = \frac{EBITDA}{S}(1-RR)\left\{\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e} + \frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\}$$

Remembering that Operating Margin is equal to:

$$OM = \frac{EBITDA}{S}$$

and substituting, we have:

$$\frac{EV}{S} = OM(1-RR)\left\{\frac{(1+g_e)\left[1-\frac{(1+g_e)^n}{\left(1+K_{E,eg}\right)^n}\right]}{K_{E,eg}-g_e} + \frac{(1+g_e)^n(1+g_n)}{\left(K_{E,sg}-g_n\right)\left(1+K_{E,sg}\right)^n}\right\} \tag{9.47}$$

Equation (9.47) shows the relationship between the EV/S multiple and the company's fundamentals in the two-stage growth scenario.

References

Altaman E (1969) Bankrupt company's equity securities as an investment alternative. Financ Anal J 25(4):129–133
Arditti F, Pinkerton J (1978) The valuation and cost of capital of the levered firm with growth opportunities. J Financ XXXIII(1):54–73
Baron D (1975) Firm valuation, corporate taxes, and default risk. J Financ XXX(5):1251–1264
Beaver W, Morse D (1978) What determines price-earnings ratios? Financ Anal J 34(4):65–76
Beidelman C (1971) Pitfalls of the price-earnings ratio. Financ Anal J 27(4):86–91
Benninga S (2014) Financial modeling, 4th edn. MIT Press, Cambridge
Berk J, DeMarzo P (2008) Corporate finance. Pearson Education, Inc
Bing R (1971) Survey of practitioners' stock evaluation methods. Financ Anal J 27(3):55–69
Blume MKJ, Kraft A (1977) Determinants of common stock prices: a time series analysis. J Financ XXXII(2):417–425
Bonnes JCA, Jatusipitak S (1972) On relations among stock price behaviour a d changes in the capital structure of the firm. J Financ Quant Anal VII(4):1967–1982
Bower D, Bower SR (1970) Test of a stock valuation model. J Financ XXV(2):483–492

Brealey RA, Myers SC, Allen F (2016) Principles of corporate finance, 12th edn. McGraw-Hill
Campbell J, Shiller R (1989) The dividend price ratio and expectation of future dividends and discount factors. Rev Financ Stud 1:195–228
Chen N (1998) Risk and return of value stock. J Bus 71(4):501–535
Copeland T, Weston F, Shastri K (2004) Financial theory and corporate policy, 4th edn. Addison-Wesley, Reading, MA
Corelli A (2016) Anal Corp Financ. Springer, Berlin
Damodaran A (2012) Investment valuation: tools and techniques for a determining the value of any assets, 3rd edn. Wiley, New Jersey
Damodaran A (2015) Applied corporate finance, 4th edn. Wiley, New Jersey
Elton EJ, Gruber MJ (1971) Valuation and the cost of capital for regulated industries. J Financ XXVI(3):661–670
Elton EJ, Gruber MJ (1976) Valuation and the asset selection under alternative investment opportunities. J Financ XXXI(2):525–539
Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN (2013) Modern portfolio theory and investment analysis, 9th edn. Wiley, New Jersey
Foster E (1970) Price-earnings ratio and corporate growth: a revision. Financ Anal J 26(4):115–118
Fuller RJ, Farrell JL (1987) Modern investments and security analysis. McGraw-Hill, Inc, New York
Fuller RJ, Hsia CC (1984) A simplified common stock valuation model. Financ Anal J, 49–56 (September–October)
Gordon M (1962) The investment, financing, and valuation of the corporation. Richard D. Irwin, Homewood, IL
Graham B, Dodd D, Cottle S (1962) Security analysis principles and techniques, 4th edn. McGraw-Hill, New York
Hamada RS (1972) The effect of the firm's capital structure on the systematic risk of common stocks. J Financ 27(2):435–452
Haugen R, Kumar P (1974) The traditional approach to valuing levered-growth stocks: a clarification. J Financ Quant Anal IX(6):1031–1044
Haugen R, Pappas JL (1971) Equilibrium in the pricing of capital assets, risk-bearing debt instruments, and the question of optimal capital structure. J Financ Quant Anal, VI(3):943–953
Hawkins D (1977) Toward an old theory of equity valuation. Financ Anal J 33(6):48–53
Hillier D, Ross S, Westerfield R, Jaffe J, Jordan B (2016) Corp Financ, 3rd ed. McGraw-Hill, New York
Joy M, Jones C (1970) Another look at the value of p/e ratios. Financ Anal J 26(4):61–64
Koller T, Goedhart M, Wessels D (2015) Valuation, 6th edn. Wiley, Hoboken, New Jersey
Litzenberger R, Budd A (1970) Corporate investment criteria and the valuation of risk assets. J Financ Quant Anal V(4):385–419
Malkiel B (1963) Equity yields, growth, and the structure of share prices. Am Econ Rev 53:1004–1031
Malkiel B, Cragg J (1970) Expectations and the structure of share price. Am Econ Rev LX(4):601–617
Modigliani F, Miller MH (1958) The cost of capital. Corp Financ Theor Invest Am Econ Rev 48:261–297
Molodovsky N, May C, Chottinger S (1965) Common stock valuation. Financ Anal J 21:104–123
Myers SC (1974) Interactions of corporate financing and investment decisions—implications for capital budgeting. J Financ 29(1):1–25
Sydsaeter K, Hammond P, Strom A (2012) Essential mathematics for economic analysis, 4th edn. Pearson
Titman S (1984) The effect of capital structure on a firm's liquidation decision. J Financ Econ 13(1):137–151
Vernimmen P, Quiry P, Dallocchio M, Le Fur Y, Salvi A (2014) Corporate finance: theory and practice. Wiley, New Jersey
Whitbeck V, Kisor M (1963) A new tool in investment decision making. Financ Anal J, 55–62 (May–June)
Williams JB (1938) The theory of investment value. Harvard University Press, Cambridge, MA

Chapter 10
Debt Valuation

Abstract Governments and companies can raise the capital needed to finance their activities by issuing bonds to a public market. A bond is nothing more than a loan. There are several types of bonds. However, there are some elements common to all of them. In reality, any bond can be defined on the basis of five main elements: face value, price, coupon, maturity date, issuer. There is a strict relationship between price and risk of the bond. Specifically, the lower the grade of the bond, the higher the risk and therefore higher the return offered by the issuer to the investors in the bond. Therefore, not all bonds are inherently safer than stocks. Certain types of bonds can be just as risky, if not riskier in certain conditions, than stocks. Usually it is normal to measure the free-risk rate on the basis of the government bonds. Indeed, the default risk of the government tend to be small (mostly for the developed countries). It is because the government will always be able (or should be able) to bring in future revenues through taxation. On the other hand, companies must be able to generate profit in order to survive and face their debt obligations. The difference in risk between government and corporate bonds implies that the corporate bonds must offer a higher yield than government bonds. Therefore, it is necessary to evaluate the government bonds in order to estimate the free-risk rate, as well the company bonds on the basis of its default risk.

10.1 Interest Rate Rules

Any financial operations can be obtained by combining three main mechanisms for the transaction of goods and assets (Hicks 1939):

– *spot transaction*: when the two parts of the transaction perform their obligations at the date of the agreement or signing of the contract;
– *forward transaction*: when the two parts of the transaction perform their obligations at a future date;
– *loan transaction*: when one of the parts perform its obligations immediately and the other part in a future time.

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_10

Generally, any financial operation can be defined on the basis of two variables: *time* and *money*.

On the basis of these two variables, two main basic capital financial operations can be defined:

- *loan*: at an initial time t the investor obtains a capital C (cash-in) and at the end of the period T he has to reimburse the capital M (cash-out);
- *investment*: at an initial time t the investor lends a capital C (cash-out) and at the end of the time T he receives a capital M (cash-in).

The time of the financial operation plays a key role in the investment valuation. The main problem refers to the equivalence between the C-capital in t-time (K_t) and the M-capital in T-time (M_T):

$$C_t \sim M_T$$

There are two mechanisms to move money over time:

- *investment*: it defines the value of M-capital in T-time related to the C-capital invested in t-time. In this case, M-capital in T-time is equal to the principal (C-capital in t-time) plus accrued interests (I);
- *discounting*: it defines the present value of the C-capital in t-time related to the M-capital that will be received in T-time. In this case, C-capital in t-time is the present value of the M-capital in T-time.

The investor always prefers (in a strict sense $(\succ)$) a C-capital in t-time rather than M-capital in T-time (the capital of today is preferred to capital tomorrow):

$$C_t \succ M_T$$

For further understanding of these two operations, consider a time equal to one year.

In the *investment operation*, the capital (C) invested today (t_0) will be equal from one year (t_1) to the capital invested (C) plus the interest (I) as follows:

$$M = C + I \tag{10.1}$$

On the basis of Eq. (10.1), we have the *Interest* (I) as difference between M and C as follows:

$$I = M - C \tag{10.2}$$

The *Interest Rate* (i) can be defined as the ratio between the Interest (I) and the Capital invested (C), as follows:

$$i = \frac{I}{C} \tag{10.3}$$

Therefore, on the basis of Interest Rate (i) as defined in Eq. (10.3), it is possible to obtain the Interest (I) as follows:

$$I = Ci \tag{10.4}$$

On the basis of Eqs. (10.2) and (10.4), we have:

$$Ci = M - C$$

and by solving for M, we have:

$$M = C(1 + i) \tag{10.5}$$

The *factor* $(1 + i)$ is called the *capitalization factor* and it is usually denoting with r, so that:

$$r = (1 + i) \tag{10.6}$$

Therefore, Eq. (10.5) can also be rewritten as follows:

$$M = Cr \tag{10.7}$$

In the case of *discounting operation*, the present value (V) today (t_0) of the capital (C) that will be received from one year (t_1) is equal to the capital (C) minus a discount factor (D) as follows:

$$V = C - D \tag{10.8}$$

On the basis of Eq. (10.8) it gets the *Discount* (D) as difference between C and V as follows:

$$D = C - V \tag{10.9}$$

The *Discount Rate* (d) can be defined as the ratio between the Discount (D) and the Capital (C), as follows:

$$d = \frac{D}{C} \tag{10.10}$$

Therefore, on the basis of Discount Rate (d) as defined in Eq. (10.10), it is possible to obtain the Discount (D) as follows:

$$D = Cd \tag{10.11}$$

On the basis of Eqs. (10.9) and (10.11), we have:

$$Cd = C - V$$

and solving for V, we have:

$$V = C(1 - d) \tag{10.12}$$

The *factor* $(1 - d)$ is called *discount factor* and it is usually denoting by v as follows:

$$v = (1 - d) \tag{10.13}$$

On the basis of Eq. (10.13), Eq. (10.12) can also rewritten as follows:

$$V = Cv \tag{10.14}$$

These two operations can be summarized as in Fig. 10.1.

Between the Interest Rate (i) and Discount Rate (d) a relationship can be defined. Assume receiving a capital (C) from one year (t_1). Its value today (t_0) is equal to $C(1 - d)$. Assume investing this capital from one year (t_1). Its value is equal to $C(1 - d)(1 + i)$. These two financial operations can be summarized as in Fig. 10.2.

The amount of the capital at the start must be equal to the amount of the capital at the end, so that:

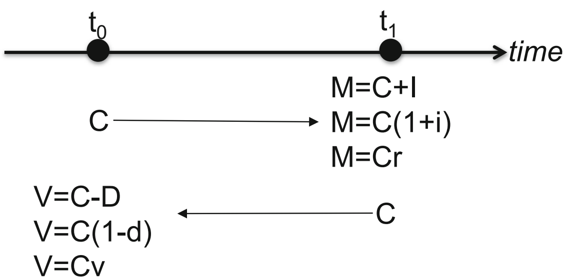


Fig. 10.1 Financial operations

Fig. 10.2 The relationship between interest rate and discount rate

$$C(1 - d)(1 + i) = C$$

and then:

$$(1 - d)(1 + i) = 1 \rightarrow \begin{matrix} i = \frac{d}{1-d} \\ d = \frac{i}{1+i} \end{matrix} \tag{10.15}$$

Therefore, for a time period longer than one $(t > 1)$, the interest and discount rates as well as interest and discount factors are function of the time: $i(t), d(t), r(t), v(t)$. They are defined on the basis of the following rules:

- *Simple Interest*;
- *Compound Interest*;
- *Continuously Compound Interest*;
- *Commercial Interest*.

It is relevant to note that the time t can be expressed on the basis of the year, month and day. For the second and the third a ratio must be used in which the denominator is equal to 12 for the months and 360 for the days.

Obviously, time and interest rates must be expressed in the same unit-time period: time in years, and interest rate in year; time in months and interest rates in months.

Simple Interest

Interest is paid at the end of the accrued period without any capital reinvestment. Consequently, there is no new-interest earned on the matured interest.

The basic equation of capitalization factor $(r(t))$ is the following:

$$r(t) = 1 + i \cdot t \tag{10.16}$$

and then:

$$M = Cr(t) \rightarrow M = C(1 + i \cdot t) \tag{10.17}$$

The interest rate $(i(t))$ is equal to:

$$i(t) = \frac{I}{C} = \frac{M - C}{C} = \frac{C(1 + i \cdot t) - C}{C} = \frac{C[(1 + i \cdot t) - 1]}{C} = (1 + i \cdot t) - 1 = i \cdot t$$

and then:

$$i(t) = i \cdot t \tag{10.18}$$

The relationship between the capitalization factor $(r(t))$ and discount factor $(v(t))$ can be defined as follows:

$$r(t) \cdot v(t) = 1 \tag{10.19}$$

and consequently:

$$v(t) = \frac{1}{r(t)} \leftrightarrow r(t) = \frac{1}{v(t)} \tag{10.20}$$

On the basis of Eq. (10.16), the discount factor $(v(t))$ as defined in Eq. (10.20) on the basis of Eq. (10.19) can be defined as follows:

$$v(t) = \frac{1}{r(t)} \rightarrow v(t) = \frac{1}{1 + i \cdot t} \tag{10.21}$$

Consequently, we have:

$$V = Cv(t) \rightarrow V = C\frac{1}{r(t)} \rightarrow V = C \cdot \frac{1}{1 + i \cdot t} \tag{10.22}$$

The discount factor $(v(t))$ can be defined as follows:

$$v(t) = 1 - d \cdot t \tag{10.23}$$

On the basis of Eqs. (10.16), (10.20) and (10.23), it is possible to define $d \cdot t$ as follows:

$$\begin{matrix} v(t) = \frac{1}{r(t)} \\ r(t) = 1 + i \cdot t \\ v(t) = 1 - d \cdot t \end{matrix} \rightarrow \frac{1}{1 + i \cdot t} = 1 - d \cdot t$$

and then:

$$d \cdot t = \frac{i \cdot t}{1 + i \cdot t} \tag{10.24}$$

On the basis of $r(t)$ and the interest rate (i), it is possible to define the time by solving Eq. (10.16) for time t as follows:

$$t = \frac{r(t) - 1}{i} \tag{10.25}$$

It is possible to define the rules of Simple Interest in terms of discount rate (d) rather than the interest rate (i). On the basis of Eq. (10.15), we have:

$$i = \frac{d}{1-d}; \quad d = \frac{i}{1+i}$$

and on the basis of Eq. (10.18), and by considering Eq. (10.15), it is possible to define $i(t)$ in terms of $d(t)$ as follows:

$$i(t) = i \cdot t \rightarrow i = \frac{d}{1-d} \rightarrow i(t) = \left(\frac{d}{1-d}\right) \cdot t$$

$$i(t) = \frac{d}{1-d} \cdot t \tag{10.26}$$

On the basis of Eq. (10.26) it is possible to redefine $r(t)$, $v(t)$, $d(t)$, t as follows:

$$r(t) = 1 + i \cdot t \rightarrow r(t) = 1 + \frac{d}{1-d} \cdot t = \frac{1-d+d \cdot t}{1-d}$$

$$v(t) = \frac{1}{r(t)} = \frac{1}{1+i \cdot t} \rightarrow v(t) = \frac{1}{\frac{1-d+d \cdot t}{1-d}} = \frac{1-d}{1-d+d \cdot t}$$

$$d(t) = \frac{i \cdot t}{1+i \cdot t} \rightarrow d(t) = \frac{\frac{d \cdot t}{1-d}}{\frac{1-d+d \cdot t}{1-d}} = \frac{d \cdot t}{1-d} \cdot \frac{1-d}{1-d+d \cdot t} = \frac{d \cdot t}{1-d+d \cdot t}$$

$$t = \frac{r(t)-1}{i} \rightarrow t = \frac{r(t)-1}{\frac{d}{1-d}} = \frac{[r(t)-1] \cdot (1-d)}{d}$$

In the simple interest, the basic relationship can be summarized as follows:

$$\begin{array}{rr} r(t) = 1 + i \cdot t; & r(t) = \frac{1-d+d \cdot t}{1-d} \\ v(t) = \frac{1}{1+i \cdot t}; & v(t) = \frac{1-d}{1-d+d \cdot t} \\ i(t) = i \cdot t; & i(t) = \frac{d}{1-d} \cdot t \\ d(t) = \frac{i \cdot t}{1+i \cdot t}; & d(t) = \frac{d \cdot t}{1-d+d \cdot t} \\ t = \frac{r(t)-1}{i}; & t = \frac{[r(t)-1] \cdot (1-d)}{d} \end{array} \tag{10.27}$$

Therefore, by considering the initial capital C and the final capital C, in the simple interest rule we have:

$$M = C(1 + i \cdot t) \rightarrow \begin{array}{c} C = M \cdot \frac{1}{(1+i \cdot t)} \\ i = \frac{M-C}{C.t} \\ t = \frac{M-C}{C.i} \end{array} \tag{10.28}$$

It is relevant to know the dynamics of $v(t)$ and $d(t)$ as function of time t:

- $\lim_{t\to\infty} r(t) = \infty$; $\lim_{t\to 0} r(t) = 1$
- $\lim_{t\to\infty} v(t) = \frac{1}{\infty} = 0$; $\lim_{t\to 0} v(t) = \frac{1}{1} = 1$
- $\lim_{t\to\infty} d(t) = \frac{\infty}{\infty} = \frac{i}{i} = 1$; $\lim_{t\to 0} d(t) = \frac{0}{1} = 0$.

Otherwise, $r(t)$ draws a straight-line with slope equal to i and intercept equal to 1. Therefore, i measures the increase of $r(t)$ to the increase of the time (t). The relationship shows how the interest is always the same and the capital increases in order to the sum of interests. It is possible to summarize these relationships as follows as in Fig. 10.3.

Compound Interest

Interest generates interest. At the end of the compounding period accrued interest is reinvested with capital. Consequently, in each period interests are calculated based on capital and interest matured in the previous period with increasing of the amount earned at the end of the periods.

The basic equation of capitalization factor $(r(t))$ is as follows:

$$r(t) = (1+i)^t \tag{10.29}$$

and then:

$$M = Cr(t) \to M = C(1+i)^t \tag{10.30}$$

The interest rate $(i(t))$ is equal to:

$$i(t) = \frac{M-C}{C} = \frac{C(1+i)^t - C}{C} = \frac{C\left[(1+i)^t - 1\right]}{C} = (1+i)^t - 1$$

and then:

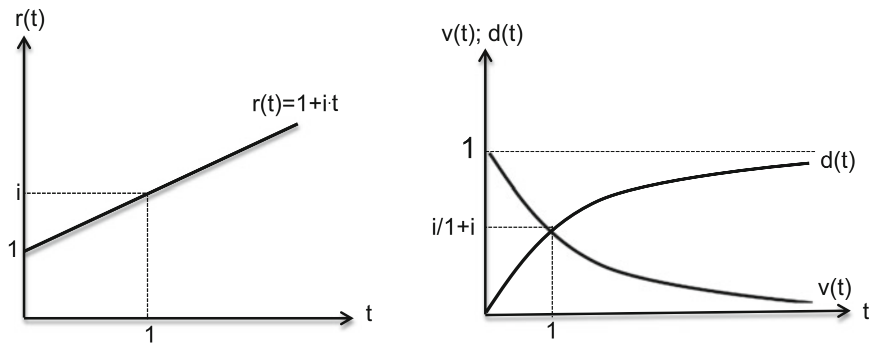


Fig. 10.3 Relationship between time t and $r(t)$, $v(t)$; $d(t)$ in simple interest rule

$$i(t) = (1 + i)^t - 1 \tag{10.31}$$

The relationship between the capitalization factor $(r(t))$ and discount factor $(v(t))$ are defined by Eqs. (10.19) and (10.20) as follows:

$$r(t) \cdot v(t) = 1 \rightarrow \begin{matrix} v(t) = \frac{1}{r(t)} \\ r(t) = \frac{1}{v(t)} \end{matrix}$$

and in this context we have:

$$v(t) = \frac{1}{(1 + i)^t} \tag{10.32}$$

Consequently, we have:

$$V = Cv(t) \rightarrow V = C\frac{1}{r(t)} \rightarrow V = C \cdot \frac{1}{(1 + i)^t} \tag{10.33}$$

On the basis of Eq. (10.23), the discount factor $(v(t))$ is equal to:

$$v(t) = 1 - d \cdot t$$

On the basis of Eqs. (10.20), (10.23) and (10.29), it is possible to define $d \cdot t$ as follows:

$$\begin{matrix} v(t) = \frac{1}{r(t)} \\ r(t) = (1 + i)^t \\ v(t) = 1 - d \cdot t \end{matrix} \rightarrow \frac{1}{(1 + i)^t} = 1 - d \cdot t$$

and then:

$$d \cdot t = \frac{(1 + i)^t - 1}{(1 + i)^t} \tag{10.34}$$

On the basis of $r(t)$ and the interest rate (i), it is possible to define the time t by solving Eq. (10.29) as follows:

$$r(t) = (1 + i)^t \rightarrow \ln r(t) = \ln(1 + i)^t; \quad \ln r(t) = t \ln(1 + i)$$

and then:

$$t = \frac{\ln r(t)}{\ln(1 + i)} \tag{10.35}$$

It is possible to define the rules of Compound Interest in terms of discount rate (d) rather than the interest rate (i). On the basis of Eq. (10.15) we have:

$$i = \frac{d}{1-d}; \quad d = \frac{i}{1+i}$$

and on the basis of Eq. (10.31) and on the basis of Eq. (10.15), it is possible to define $i(t)$ in terms of $d(t)$ as follows:

$$i(t) = (1+i)^t - 1 \rightarrow i(t) = \left(1 + \frac{d}{1-d}\right)^t - 1 = \frac{1-(1-d)^t}{(1-d)^t}$$

$$i(t) = \frac{1-(1-d)^t}{(1-d)^t} \tag{10.36}$$

On the basis of Eq. (10.36) it is possible to redefine $r(t)$, $v(t)$, $d(t)$, t as follows:

$$r(t) = (1+i)^t \rightarrow r(t) = \left(1 + \frac{d}{1-d}\right)^t = \frac{1}{(1-d)^t}$$

$$v(t) = \frac{1}{r(t)} = \frac{1}{(1+i)^t} \rightarrow v(t) = \frac{1}{\left(1+\frac{d}{1-d}\right)^t} = \frac{1}{\frac{1}{(1-d)^t}} = (1-d)^t$$

$$d(t) = \frac{(1+i)^t - 1}{(1+i)^t} \rightarrow d(t) = \frac{\left(1+\frac{d}{1-d}\right)^t - 1}{\left(1+\frac{d}{1-d}\right)^t} = \frac{\frac{1}{(1-d)^t} - 1}{\frac{1}{(1-d)^t}} = (1-d)^t$$

$$t = \frac{\ln r(t)}{\ln(1+i)} \rightarrow t = \frac{\ln r(t)}{\ln\left(1+\frac{d}{1-d}\right)} = \frac{\ln r(t)}{\ln\left(\frac{1}{1-d}\right)}$$

In the compound interest, the baseline relationship can be summarized as follows:

$$\begin{array}{c} r(t) = (1+i)^t; \quad r(t) = \frac{1}{(1-d)^t} \\ v(t) = \frac{1}{(1+i)^t}; \quad v(t) = (1-d)^t \\ i(t) = (1+i)^t - 1; \quad i(t) = \frac{1-(1-d)^t}{(1-d)^t} \\ d(t) = \frac{(1+i)^t - 1}{(1+i)^t}; \quad d(t) = 1-(1-d)^t \\ t = \frac{\ln r(t)}{\ln(1+i)}; \quad t = \frac{\ln r(t)}{\ln\left(\frac{1}{1-d}\right)} \end{array} \tag{10.37}$$

Therefore, by considering the initial capital C and the final capital C, in the compound interest rule we have:

$$M = C(1+i)^t \rightarrow \begin{array}{l} C = M \cdot \frac{1}{(1+i)^t} \\ i = \left(\frac{M}{C}\right)^{\frac{1}{t}} - 1 \\ t = \frac{\ln\left(\frac{M}{C}\right)}{\ln(1+i)} \end{array} \tag{10.38}$$

With regards to the interest rate and time, it is worth noting that:

$$M = C(1+i)^t \rightarrow \begin{array}{l} \left(\frac{M}{C}\right)^{\frac{1}{t}} = \left[(1+i)^t\right]^{\frac{1}{t}} \rightarrow \left(\frac{M}{C}\right)^{\frac{1}{t}} = (1+i)^{t \cdot \frac{1}{t}} \rightarrow i = \left(\frac{M}{C}\right)^{\frac{1}{t}} = i \\ \ln\left(\frac{M}{C}\right) = \ln(1+i)^t \rightarrow \ln\left(\frac{M}{C}\right) = t\ln(1+i) \rightarrow t = \frac{\ln\left(\frac{M}{C}\right)}{\ln(1+i)} \end{array} \tag{10.39}$$

The $r(t)$ equation draws an exponential curve with intercept equal to 1.

It is relevant to know the dynamics of $v(t)$ and $d(t)$ as function of time t:

- $\lim_{t \to \infty} r(t) = \infty$; $\lim_{t \to 0} r(t) = 1$
- $\lim_{t \to \infty} v(t) = \frac{1}{\infty} = 0$; $\lim_{t \to 0} v(t) = \frac{1}{1} = 1$
- $\lim_{t \to \infty} d(t) = \frac{\infty}{\infty} = \frac{i}{i} = 1$; $\lim_{t \to 0} d(t) = \frac{0}{1} = 0.$

It is possible to summarize these relationships as follows as in Fig. 10.4.

Continuously Compound Interest

Interest is generated continuously on the basis of a succession of instantaneous intervals. The interest matured in the previous instant is added to the capital by generating interest in the next period. It is based on the compound interest by reducing the period at the infinitesimal level (ds).

By considering the instant s and the instant ds immediately following, the value of capital M can be defined as follows:

$$M(s+ds) = M(s) + M(s) \cdot ds \cdot i \tag{10.40}$$

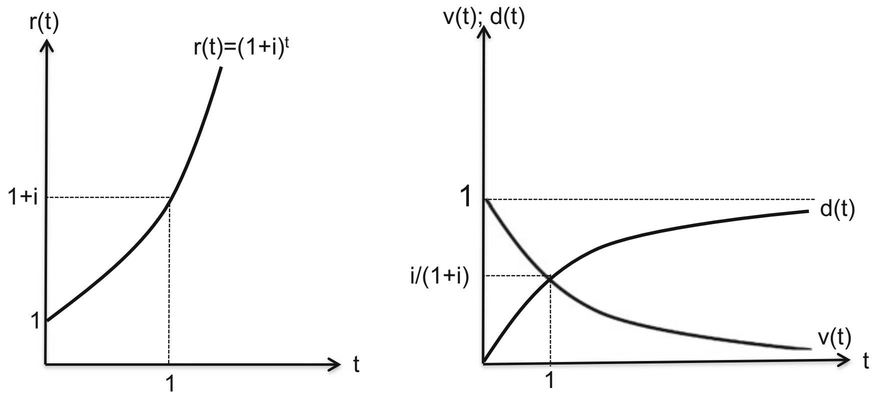


Fig. 10.4 Relationship between time t and $r(t)$, $v(t)$; $d(t)$ in compound interest rule

It is worth noting that:

$$\frac{M(s+ds)-M(s)}{ds}=M(s)\cdot i \tag{10.41}$$

is the derivative of the M in the instant s compared with time s, so that

$$\frac{dM(s)}{ds}=M(s)\cdot i \tag{10.42}$$

The solution of this differential equation is the following:

$$M=C\cdot e^{t\cdot i} \tag{10.43}$$

Solving Eq. (10.43) by initial capital C we have:

$$C=\frac{M}{e^{t\cdot i}}\leftrightarrow C=M\cdot e^{-t\cdot i} \tag{10.44}$$

In interest rate (i) is equal to:

$$M=C\cdot e^{t\cdot i}\rightarrow\frac{M}{C}=e^{t\cdot i}\rightarrow\ln\left(\frac{M}{C}\right)=\ln e^{t\cdot i}\rightarrow\ln\left(\frac{M}{C}\right)=t\cdot i$$

and then:

$$i=\frac{\ln\left(\frac{M}{C}\right)}{t} \tag{10.45}$$

On the basis of Eq. (10.45) it is possible to define the time (t) as follows:

$$t=\frac{\ln\left(\frac{M}{C}\right)}{i} \tag{10.46}$$

Therefore, the continuously compound interest rule can be summarized as follows:

$$M=C\cdot e^{t\cdot i}\rightarrow\begin{array}{c}C=M\cdot e^{-(t\cdot i)}\\ i=\frac{\ln\left(\frac{M}{C}\right)}{t}\\ t=\frac{\ln\left(\frac{M}{C}\right)}{i}\end{array} \tag{10.47}$$

Bank Discount Rate

The *bank discount rate rule* is normally used by banks to discount short-time marketable assets.

In this case the discount is function of the discount rate and the time: $d(t)=d\cdot t$.

The baseline equation is the following:

$$v(t) = 1 - d \cdot t \tag{10.48}$$

and by considering the relationship between $v(t)$ and $r(t)$, we have:

$$r(t) = \frac{1}{v(t)}$$

Therefore:

$$r(t) = \frac{1}{1 - d \cdot t} \tag{10.49}$$

The interest rate can be obtained as follows:

$$i(t) = \frac{C\left(\frac{1}{1-d\cdot t}\right) - C}{C} = \frac{C\left[\left(\frac{1}{1-d\cdot t}\right) - 1\right]}{C} = \left(\frac{1}{1 - d \cdot t}\right) - 1 = \frac{1 - 1 + d \cdot t}{1 - d \cdot t}$$
$$= \frac{d \cdot t}{1 - d \cdot t}$$

and then:

$$i(t) = \frac{d \cdot t}{1 - d \cdot t} \tag{10.50}$$

On the basis of Eq. (10.48), it is possible to define the time (t) as follows:

$$t = \frac{1 - v(t)}{d} \tag{10.51}$$

On the basis of the relationship between i and d as defined in Eq. (10.15):

$$i = \frac{d}{1 - d}; \quad d = \frac{i}{1 + i}$$

it is possible to define $r(t), v(t), i(t), d(t)$ in terms of the interest rate i rather than in terms of discount rate d as follows:

$$d(t) = d \cdot t \rightarrow d(t) = \frac{i}{1 + i} \cdot t = \frac{i \cdot t}{1 + i}$$

$$r(t) = \frac{1}{1 - d \cdot t} \rightarrow r(t) = \frac{1}{1 - \frac{i \cdot t}{1+i}} = \frac{1}{\frac{1+i-i\cdot t}{1+i}} = \frac{1 + i}{1 + i - i \cdot t}$$

$$v(t) = 1 - d \cdot t \rightarrow v(t) = 1 - \frac{i \cdot t}{1 + i} = \frac{1 + i - i \cdot t}{1 + i}$$

$$i(t) = \frac{d \cdot t}{1 - d \cdot t} \to i(t) = \frac{\frac{i \cdot t}{1+i}}{1 - \frac{i \cdot t}{1+i}} = \frac{\frac{i \cdot t}{1+i}}{\frac{1+i-i \cdot t}{1+i}} = \frac{i \cdot t}{1 + i - i \cdot t}$$

$$t = \frac{1 - v(t)}{d} \to t = \frac{1 - v(t)}{\frac{i}{1+i}} = \frac{(1 - v(t))(1+i)}{i}$$

Therefore, it is possible summarize the bank discount rate rule as follows:

$$\begin{array}{c} v(t) = 1 - d \cdot t; \quad v(t) = \frac{1+i-i \cdot t}{1+i} \\ r(t) = \frac{1}{1-d \cdot t}; \quad r(t) = \frac{1+i}{1+i-i \cdot t} \\ d(t) = d \cdot t; \quad d(t) = \frac{i \cdot t}{1+i} \\ i(t) = \frac{d \cdot t}{1-d \cdot t}; \quad i(t) = \frac{i \cdot t}{1+i-i \cdot t} \end{array} \tag{10.52}$$

Therefore, by considering an initial capital C and the final capital M, it is possible to define the bank discount rate rule as follows:

$$C = M \cdot (1 - d \cdot t) \to \begin{array}{c} M = \frac{C}{1-d \cdot t} \\ d = \frac{M-C}{M \cdot t} \\ t = \frac{M-C}{M \cdot d} \end{array} \tag{10.53}$$

It is worth noting that:

- $\lim_{t \to \infty} r(t) = 0$; $\lim_{t \to 0} r(t) = 1$
- $\lim_{t \to \infty} v(t) = \infty$; $\lim_{t \to 0} v(t) = 1$
- $\lim_{t \to \infty} d(t) = \frac{\infty}{\infty} = \frac{d}{d} = 1$; $\lim_{t \to 0} d(t) = \frac{0}{1} = 0$.

Therefore, it is relevant to note that there is a vertical asymptote. Indeed, the existence of the $r(t)$ equation requires that:

$$1 - d \cdot t \neq 0 \to t \neq \frac{1}{d} \to t = \frac{1}{d}$$

Also:

- $\lim_{t \to 0} \frac{1}{1-d \cdot t} = \frac{1}{1} = 1$
- $\lim_{t \to \infty} \frac{1}{1-d \cdot t} = \frac{1}{\infty} = 0$
- $\lim_{t \to \frac{1}{d}^{\pm}} \frac{1}{1-d \cdot t} = \pm\infty$.

Also, with regards to the $r(t)$ equation, it is equal to zero for $t = \frac{1}{d}$ and positive before and negative after this value, as follows:

$$v(t) = 0 \rightarrow 1 - d \cdot t = 0 \rightarrow t = \frac{1}{d}$$
$$v(t) > 0 \rightarrow 1 - d \cdot t > 0 \rightarrow t < \frac{1}{d}$$
$$v(t) < 0 \rightarrow 1 - d \cdot t \langle 0 \rightarrow t \rangle \frac{1}{d}$$

Therefore, equations $r(t)$ and $v(t)$ in the bank discount rate can be represented as in Fig. 10.5.

It is possible to summarize the simple interest, compound interest, continuously compound interest and bank discount rate rules as in Table 10.1:

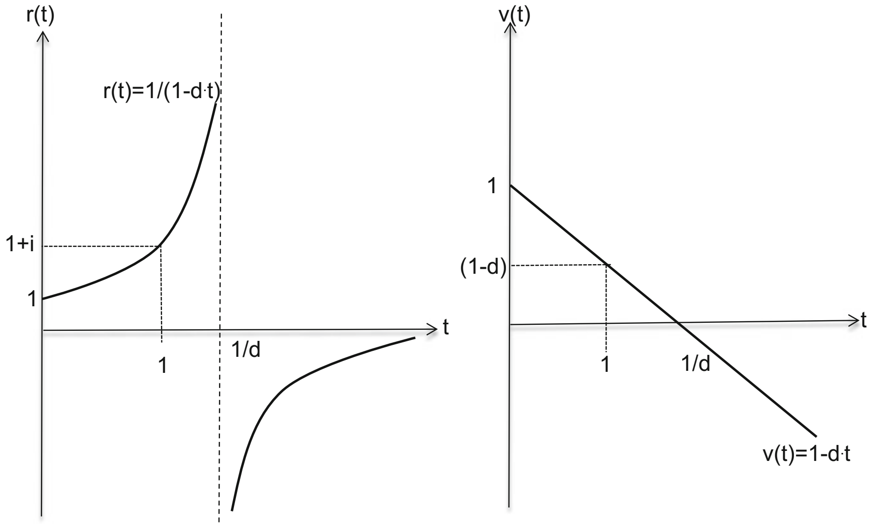


Fig. 10.5 Relationship between time t and $r(t)$, $v(t)$; $d(t)$ in bank discount rule

Table 10.1 Interest rules

Simple Interest	Compound interest	Continuously compound interest	Bank discount rate
$M = C(1 + i \cdot t)$ $\downarrow$ $C = M \cdot \frac{1}{(1+i \cdot t)}$ $i = \frac{M-C}{C \cdot t}$ $t = \frac{M-C}{C \cdot i}$	$M = C(1+i)^t$ $\downarrow$ $C = M \cdot \frac{1}{(1+i)^t}$ $i = \left(\frac{M}{C}\right)^{\frac{1}{t}} - 1$ $t = \frac{\ln\left(\frac{M}{C}\right)}{\ln(1+i)}$	$M = C \cdot e^{t \cdot i}$ $\downarrow$ $C = M \cdot e^{-t \cdot i}$ $i = \frac{\ln\left(\frac{M}{C}\right)}{t}$ $t = \frac{\ln\left(\frac{M}{C}\right)}{i}$	$C = M \cdot (1 - d \cdot t)$ $\downarrow$ $M = \frac{C}{1-d \cdot t}$ $d = \frac{M-C}{M \cdot t}$ $t = \frac{M-C}{M \cdot d}$

Relationships Between Interest Rates

The relationships with regards to time t can be defined among the simple interest, compounded interest and bank discount rate.

For $t = 1$ these three rules provide the same result, but the results are different for $t < 1$ and $t > 1$ as shown in Fig. 10.6.

Denoting with the pedicle "S" the simple interest, "C" the compound interest, "D" the bank discount rate, relationships among them as shown in Fig. 10.6. can be summarized as follows:

$$\begin{aligned}
&0 < t < 1 \rightarrow r(t)_S > r(t)_C > r(t)_D \Leftrightarrow v(t)_D > r(t)_C > r(t)_S \\
&t = 1 \rightarrow r(t)_S = r(t)_C = r(t)_D \Leftrightarrow v(t)_S = v(t)_C = v(t)_D \\
&t > 1 \ \rightarrow r(t)_D > r(t)_C > r(t)_S \Leftrightarrow v(t)_S > v(t)_C > v(t)_D
\end{aligned} \tag{10.54}$$

These results can be confirmed on the basis of the following *Table* 10.2. It is possible to define the relationships among the different rules in order to have *equivalent interest rates*.

Denote with: i_S the interest rate in the simple interest rule; i_C the interest rate in the compound interest rule; i_{CC} the interest rate in the continuously compound interest rule; d the discount rate in the bank discount rate rule.

Assume equal initial capital C and consider the same time period t. Assume use of the same interest rates $i_S = i_C = i_{CC}$. At the end of the period the capital obtained by applying i_{CC} (M_{CC}) is higher than the capital obtained by applying i_C (M_C) that in turn is higher than the capital obtained by applying i_S (M_S). The relationship can be defined as follows:

$$M_{CC} > M_C > M_S \tag{10.55}$$

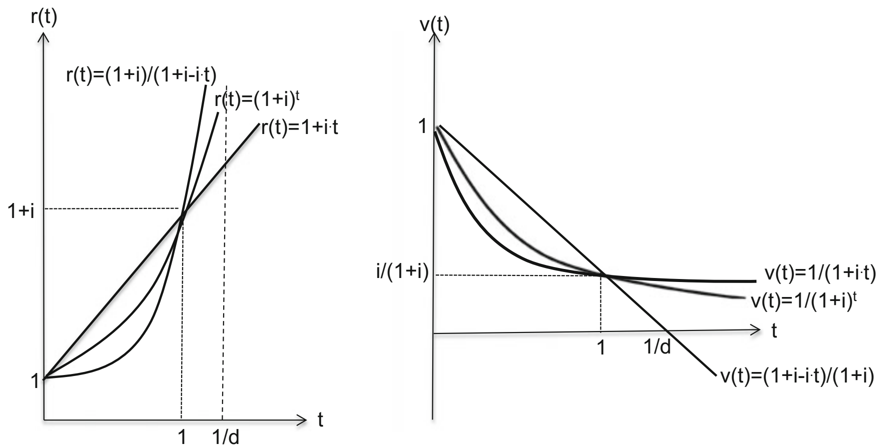


Fig. 10.6 Relationships among interest rates

Table 10.2 Relationships among interest rates

Interest rate (i) (%)	Time (t)		r(t) Simple interest	r(t) Compound interest	r(t) Bank discount rate	v(t) Simple interest	v(t) Compound interest	v(t) Bank discount rate
10	1 month	0.083	1.0083	1.0080	1.0076	0.9917	0.9921	0.9924
10	2 months	0.167	1.0167	1.0160	1.0154	0.9836	0.9842	0.9848
10	3 months	0.250	1.0250	1.0241	1.0233	0.9756	0.9765	0.9773
10	4 months	0.333	1.0333	1.0323	1.0313	0.9677	0.9687	0.9697
10	5 months	0.417	1.0417	1.0405	1.0394	0.9600	0.9611	0.9621
10	6 months	0.500	1.0500	1.0488	1.0476	0.9524	0.9535	0.9545
10	7 months	0.583	1.0583	1.0572	1.0560	0.9449	0.9459	0.9470
10	8 months	0.667	1.0667	1.0656	1.0645	0.9375	0.9384	0.9394
10	9 months	0.750	1.0750	1.0741	1.0732	0.9302	0.9310	0.9318
10	10 months	0.833	1.0833	1.0827	1.0820	0.9231	0.9236	0.9242
10	11 months	0.917	1.0917	1.0913	1.0909	0.9160	0.9163	0.9167
10	**1 year**	**1**	**1.1000**	**1.1000**	**1.1000**	**0.9091**	**0.9091**	**0.9091**
10	1 year and 1 month	1.083	1.1083	1.1088	1.1092	0.9023	0.9019	0.9015
10	1 year and 2 months	1.167	1.1167	1.1176	1.1186	0.8955	0.8948	0.8939
10	1 year and 3 months	1.250	1.1250	1.1265	1.1282	0.8889	0.8877	0.8864
10	1 year and 4 months	1.333	1.1333	1.1355	1.1379	0.8824	0.8807	0.8788
10	1 year and 5 months	1.417	1.1417	1.1446	1.1478	0.8759	0.8737	0.8712
10	1 year and 6 months	1.500	1.1500	1.1537	1.1579	0.8696	0.8668	0.8636
10	1 year and 7 months	1.583	1.1583	1.1629	1.1681	0.8633	0.8599	0.8561
10	1 year and 8 months	1.667	1.1667	1.1722	1.1786	0.8571	0.8531	0.8485

(continued)

Table 10.2 (continued)

Interest rate (i) (%)	Time (t)		r(t) Simple interest	r(t) Compound interest	r(t) Bank discount rate	v(t) Simple interest	v(t) Compound interest	v(t) Bank discount rate
10	1 year and 9 months	1.750	1.1750	1.1815	1.1892	0.8511	0.8464	0.8409
10	1 year and 10 months	1.833	1.1833	1.1909	1.2000	0.8451	0.8397	0.8333
10	1 year and 11 months	1.917	1.1917	1.2004	1.2110	0.8392	0.8330	0.8258
10	2 years	2	1.2000	1.2100	1.2222	0.8333	0.8264	0.8182

By starting with the same capital C and considering the same time period t and by assuming the same capital M at the end of the period, it is possible to define the relationships between the different interest rates and therefore the *equivalent interest rates*.

Consider the simple interest (i_S) and the compound interest (i_C). Consider the same time period t and the initial capital C. Assume that at the end of period the capital obtained with simple interest (M_S) must be equal to the capital obtained with compound interest (M_C) as follows:

$$M_S = M_C \rightarrow C(1 + i_S \cdot t) = C(1 + i_C)^t$$

and then:

$$(1 + i_S \cdot t) = (1 + i_C)^t \tag{10.56}$$

On the basis of Eq. (10.56), by knowing i_S it is possible to define the *equivalent compound interest rate* (i_C) as follows:

$$(1 + i_S \cdot t)^{\frac{1}{t}} = \left[(1 + i_C)^t\right]^{\frac{1}{t}} \rightarrow (1 + i_S \cdot t)^{\frac{1}{t}} = 1 + i_C \rightarrow i_C = (1 + i_S \cdot t)^{\frac{1}{t}} - 1$$

and then:

$$i_C = (1 + i_S \cdot t)^{\frac{1}{t}} - 1 \tag{10.57}$$

Similarly, by knowing i_C it is possible to define the *equivalent simple interest rate* (i_S) as follows:

$$(1 + i_S \cdot t) = (1 + i_C)^t \rightarrow i_S = \frac{(1 + i_C)^t - 1}{t}$$

and then:

$$i_S = \frac{(1 + i_C)^t - 1}{t} \tag{10.58}$$

Equations (10.56), (10.57) and (10.58) can be summarized as follows:

$$(1 + i_S \cdot t) = (1 + i_C)^t \rightarrow \begin{array}{c} i_C = (1 + i_S \cdot t)^{\frac{1}{t}} - 1 \\ i_S = \frac{(1 + i_C)^t - 1}{t} \end{array}$$

On the basis of the same reasoning it is possible to define the relationship between the compound interest rate (i_C) and the continuously compound interest rate (i_{CC}), as follows:

$$M_C = M_{CC} \rightarrow C(1+i_C)^t = C \cdot e^{t \cdot i_{CC}}$$

and then:

$$(1+i_C)^t = e^{t \cdot i_{CC}} \tag{10.59}$$

On the basis of Eq. (10.59), by knowing i_C it is possible to define the *equivalent continuously compound interest rate* (i_{CC}) as follows:

$$\ln(1+i_C)^t = \ln e^{t \cdot i_{CC}} \rightarrow t \ln(1+i_C) = t \cdot i_{CC} \rightarrow i_{CC} = \ln(1+i_C)$$

and then:

$$i_{CC} = \ln(1+i_C) \tag{10.60}$$

Similarly, by knowing i_{CC} it is possible to define the *equivalent compound interest rate* (i_C) as follows:

$$\left[(1+i_C)^t\right]^{\frac{1}{t}} = \left[e^{t \cdot i_{CC}}\right]^{\frac{1}{t}} \rightarrow 1 + i_C = e^{i_{CC}} \rightarrow i_C = e^{i_{CC}} - 1$$

and then:

$$i_C = e^{i_{CC}} - 1 \tag{10.61}$$

Equations (10.59), (10.60) and (10.61) can be summarized as follows:

$$(1+i_C)^t = e^{t \cdot i_{CC}} \rightarrow \begin{array}{l} i_{CC} = \ln(1+i_C) \\ i_C = e^{i_{CC}} - 1 \end{array}$$

Equivalence Among the Interest and Discount Rates in Different Time-Periods

For periods of less than a year, relationship can be defined between the annual interest rate and the interest rates.

Generally, two interest rates are equivalent among them if they give the same result when they are applied for the same period time.

Denoting with: m the months, i the annual interest rate, $i \cdot \frac{1}{m}$ the interest rate for period less than a year (where: $i \cdot \frac{1}{2}$ is the half-year interest rate; $i \cdot \frac{1}{12}$ is the monthly interest rate; $i \cdot \frac{1}{4}$ is the quarterly interest rate, etc.), the baseline equation is the following:

$$(1+i) = \left(1 + i\frac{1}{m}\right)^m \tag{10.62}$$

On the basis of Eq. (10.62) the annual interest rate (i) can be deducted from the interest rate for a shorter period $\left(i \cdot \frac{1}{m}\right)$ and vice versa as follows:

$$i = \left(1 + i\frac{1}{m}\right)^m - 1 \tag{10.63}$$

and:

$$(1+i) = \left(1 + i\frac{1}{m}\right)^m \rightarrow (1+i)^{\frac{1}{m}} = \left(1 + i\frac{1}{m}\right)^{m \cdot \frac{1}{m}} \rightarrow i\frac{1}{m} = (1+i)^{\frac{1}{m}} - 1$$

so that:

$$i\frac{1}{m} = (1+i)^{\frac{1}{m}} - 1 \tag{10.64}$$

Therefore, these relationships can be summarized as follows:

$$(1+i) = \left(1 + i\frac{1}{m}\right)^m \rightarrow \begin{array}{l} i = \left(1 + i\frac{1}{m}\right)^m - 1 \\ i\frac{1}{m} = (1+i)^{\frac{1}{m}} - 1 \end{array}$$

On the basis of these relationships the *annual convertible interest rate m-times in a year* $J(m)$ can be defined. Denoting with m is number of time in a year and with $i \cdot \frac{1}{m}$ is the annual interest rate for a m-part of the year, $J(m)$ is equal to:

$$J(m) = m \cdot i\frac{1}{m} \tag{10.65}$$

Generally, by considering the relationship between i and $i \cdot \frac{1}{m}$ as defined in Eqs. (10.63) and (10.64), and by using $J(m)$, it generates an *annual effective interest rate* $\left(i = \left(1 + i\frac{1}{m}\right)^m - 1\right)$ higher than the annual interest rate as follows:

$$i < i = \left(1 + i\frac{1}{m}\right)^m - 1$$

Therefore, the higher m and then the higher the annual convertible interest rate m-times in a year $J(m)$, the higher the *annual effective interest rate.*

Also, by considering $i\frac{1}{m}$ as defined in Eq. (10.64), $J(m)$ is equal to:

$$J(m) = m \cdot \left[(1+i)^{\frac{1}{m}} - 1\right] \tag{10.66}$$

Therefore, the increase of m reduces the value of the $J(m)$. Generally, $J(m)$ decreases with the increase of m until reaching a defined value.

Therefore, in the presence of an annual convertible interest rate m-times in a year J(m), the increase of m has two main implications:

- first, the increase of the effective annual interest rate that is higher than the annual interest rate;
- second, the decrease of $J(m)$ until a defined value is reached.

These two effects can be showed by Table 10.3.

In the same way, it is possible to define the relationship between the annual discount rate and the discount rates for periods of less than a year.

Generally, two discount rates are equivalent among them if they give the same result when they are applied for the same period time.

Denoting with m the months, d the annual discount rate, $d \cdot \frac{1}{m}$ the discount rate for period less than a year (where: $d \cdot \frac{1}{2}$ is the half-year discount rate; $d \cdot \frac{1}{12}$ is the monthly discount rate; $d \cdot \frac{1}{4}$ is the quarterly discount rate, etc.), the baseline equation is the following:

$$(1 - d) = \left(1 - d\frac{1}{m}\right)^m \tag{10.67}$$

On the basis of Eq. (10.67) the annual discount rate (d) can be deducted by the discount rate for a shorter period $\left(d \cdot \frac{1}{m}\right)$ and vice versa as follows:

$$d = 1 - \left(1 - d\frac{1}{m}\right)^m \tag{10.68}$$

and:

$$(1 - d) = \left(1 - d\frac{1}{m}\right)^m \rightarrow (1 - d)^{\frac{1}{m}} = \left(1 - d\frac{1}{m}\right)^{m \cdot \frac{1}{m}} \rightarrow d\frac{1}{m} = 1 - (1 - d)^{\frac{1}{m}}$$

so that:

$$d\frac{1}{m} = 1 - (1 - d)^{\frac{1}{m}} \tag{10.69}$$

Therefore, these relationships can be summarized as follows:

$$(1 - d) = \left(1 - d\frac{1}{m}\right)^m \rightarrow \begin{array}{c} d = 1 - \left(1 - d\frac{1}{m}\right)^m \\ d\frac{1}{m} = 1 - (1 - d)^{\frac{1}{m}} \end{array}$$

It is possible to define the *annual convertible discount rate m-times in a year* $\mathsf{C}(m)$. Denoting with m the number of times in a year and with $d\frac{1}{m}$ the annual discount rate for a m-part of the year, $\mathsf{C}(m)$ is equal to:

Table 10.3 Annual convertible interest rate m-times in a year J(m)

i (%)	m	$i_{\frac{1}{m}}$	$J(m) = m \cdot i_{\frac{1}{m}}$	$i = \left(1 + i_{\frac{1}{m}}\right)^m - 1$	$J(m) = m \cdot \left[(1+i)^{\frac{1}{m}} - 1\right]$
12	m = 1 (1 times in a year)	$12\% \cdot \frac{1}{1} = 12\%$	$J(1) = 1 \cdot 12\%$	$i = (1 + 12\%)^1 - 1 = 12.00\%$	$J(1) = 1\left[(1 + 12\%)^{\frac{1}{1}} - 1\right] = 12.00\%$
12	m = 2 (2 times in a year)	$12\% \cdot \frac{1}{2} = 6\%$	$J(2) = 2 \cdot 6\%$	$i = (1 + 6\%)^2 - 1 = 12.36\%$	$J(2) = 2\left[(1 + 12\%)^{\frac{1}{2}} - 1\right] = 11.66\%$
12	m = 3 (3 times in a year)	$12\% \cdot \frac{1}{3} = 4\%$	$J(3) = 3 \cdot 4\%$	$i = (1 + 4\%)^3 - 1 = 12.49\%$	$J(3) = 3\left[(1 + 12\%)^{\frac{1}{3}} - 1\right] = 11.55\%$
12	m = 4 (4 times in a year)	$12\% \cdot \frac{1}{4} = 3\%$	$J(4) = 4 \cdot 3\%$	$i = (1 + 3\%)^4 - 1 = 12.55\%$	$J(4) = 4\left[(1 + 12\%)^{\frac{1}{4}} - 1\right] = 11.49\%$

$$\complement(m) = m \cdot d\frac{1}{m} \tag{10.70}$$

Generally, by considering the relationship between d and $d\frac{1}{m}$ as defined by Eqs. (10.68) and (10.69) and by using $\complement(m)$ as defined by Eq. (10.70) it provides an *annual effective discount rate* $\left(d = 1 - \left(1 - d\frac{1}{m}\right)^m\right)$ higher than the annual interest rate as follows:

$$d > d = 1 - \left(1 - d\frac{1}{m}\right)^m$$

Therefore, the higher m and then the higher the annual convertible discount rate m-times in a year $\complement(m)$, the lower the *annual effective discount rate.*

Also, by considering the $d\frac{1}{m}$ as defined in Eq. (10.69), $\complement(m)$ is equal to:

$$\complement(m) = m \cdot \left[1 - (1-d)^{\frac{1}{m}}\right] \tag{10.71}$$

Therefore, the increase of m increases the value of the $\complement(m)$. Generally, $\complement(m)$ increases according to the increase of m until a defined value is reached.

Therefore, in the presence of an annual convertible discount rate m-times in a year $\complement(m)$, the increase of m has two main implications:

- first, the decrease of the annual effective discount rate that is lower than the annual discount rate;
- second, the increase of $\complement(m)$ until a defined value is reached.

These two effects can be showed by Table 10.4.

Tables 10.3 and 10.4 show that with an increase of m, $J(m)$ decreases while $\complement(m)$ increases. In both cases until a defined level is reached.

In the case of $J(m)$, we have:

$$\lim_{m\to\infty} J(m) = \lim_{m\to\infty} m \cdot \left[(1+i)^{\frac{1}{m}} - 1\right] = \lim_{m\to\infty} \frac{\left[(1+i)^{\frac{1}{m}} - 1\right]}{\frac{1}{m}} = \frac{0}{0}$$

In order to solve the de Hôpital Theorem can be used by considering the first derivatives of numerator and denominator, as follows:

$$\begin{aligned}\lim_{m\to\infty} \frac{\left[(1+i)^{\frac{1}{m}} - 1\right]}{\frac{1}{m}} &= \lim_{m\to\infty} \frac{(1+i)^{\frac{1}{m}} \cdot \log(1+i) \cdot \left(-\frac{1}{m^2}\right)}{\left(-\frac{1}{m^2}\right)} \\ &= \lim_{m\to\infty} (1+i)^{\frac{1}{m}} \cdot \log(1+i) = \log(1+i)\end{aligned} \tag{10.72}$$

Table 10.4 Annual convertible discount rate m-times in a year $\mathtt{C}(m)$

d	m	$d\frac{1}{m}$	$\mathtt{C}(m) = m \cdot d\frac{1}{m}$	$d = 1 - \left(1 - d\frac{1}{m}\right)^m$	$\mathtt{C}(m) = m \cdot \left[1 - (1-d)^{\frac{1}{m}}\right]$
12	m = 1 (1 times in a year)	$12\% \cdot \frac{1}{1} = 12\%$	$\mathtt{C}(1) = 1 \cdot 12\%$	$d = 1 - (1 - 12\%)^1 = 12.00\%$	$\mathtt{C}(1) = 1\left[1 - (1 - 12\%)^{\frac{1}{1}}\right] = 12.00\%$
12	m = 2 (2 times in a year)	$12\% \cdot \frac{1}{2} = 6\%$	$\mathtt{C}(2) = 2 \cdot 6\%$	$d = 1 - (1 - 6\%)^2 = 11.64\%$	$\mathtt{C}(2) = 2\left[1 - (1 - 12\%)^{\frac{1}{2}}\right] = 12.38\%$
12	m = 3 (3 times in a year)	$12\% \cdot \frac{1}{3} = 4\%$	$\mathtt{C}(3) = 3 \cdot 4\%$	$d = 1 - (1 - 4\%)^3 = 11.53\%$	$\mathtt{C}(3) = 3\left[1 - (1 - 12\%)^{\frac{1}{3}}\right] = 12.51\%$
12	m = 4 (4 times in a year)	$12\% \cdot \frac{1}{4} = 3\%$	$\mathtt{C}(4) = 4 \cdot 3\%$	$d = 1 - (1 - 3\%)^4 = 11.47\%$	$\mathtt{C}(4) = 4\left[1 - (1 - 12\%)^{\frac{1}{4}}\right] = 12.58\%$

Similarly, in the case of $\mathsf{C}(m)$ we have:

$$\lim_{m\to\infty} \mathsf{C}(m) = \lim_{m\to\infty} m \cdot \left[1-(1-d)^{\frac{1}{m}}\right] = \lim_{m\to\infty} \frac{\left[1-(1-d)^{\frac{1}{m}}\right]}{\frac{1}{m}} = \frac{0}{0}$$

Also in this case it is possible to use the de Hôpital Theorem by considering the first derivatives of numerator and denominator as follows:

$$\begin{aligned} \lim_{m\to\infty} \frac{\left[1-(1-d)^{\frac{1}{m}}\right]}{\frac{1}{m}} &= \lim_{m\to\infty} \frac{-(1-d)^{\frac{1}{m}} \cdot \log(1-d) \cdot \left(-\frac{1}{m^2}\right)}{\left(-\frac{1}{m^2}\right)} \\ &= \lim_{m\to\infty} -(1-d)^{\frac{1}{m}} \cdot \log(1-d) = -\log(1-d) \end{aligned} \tag{10.73}$$

The $\log(1-d)$ can be expressed in terms of interest rate (i) rather than the discount rate (d) on the basis of Eq. (10.15) as follows:

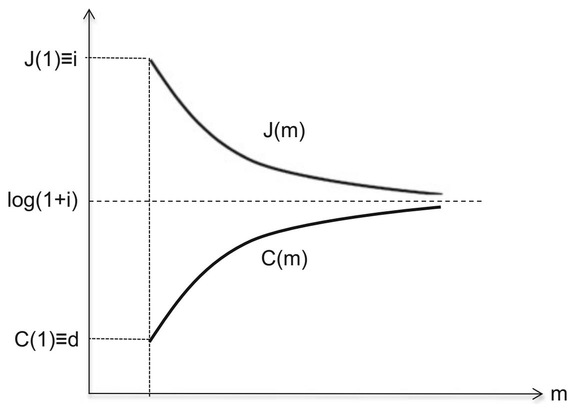
$$d = \frac{i}{1+i} \to -\log(1-d) = -\log\left(1-\frac{i}{1+i}\right) = -\log\left(\frac{1}{1+i}\right) = -\log(1+i)^{-1}$$
$$= \log(1+i)$$

The behaviour of $J(m)$ and $\mathsf{C}(m)$ to the increase of m can be represented as in Fig. 10.7.

Inflation and Interest Rates

The difference between nominal interest rate and real interest rate is due to inflation. Denote with: r the real interest rate; i the nominal interest rate; π the inflation rate; t is the time period. The relationship between real and nominal interest rates, can be defined as follows (Fisher 1930):

Fig. 10.7 Behaviour of $J(m)$ and $\mathsf{C}(m)$ to the increase of m



$$1 + r_{t+1} = \frac{1 + i_t}{1 + \pi_{t+1}} \tag{10.74}$$

By solving Eq. (10.74) for i_t, we have:

$$1 + i_t = (1 + r_{t+1})(1 + \pi_{t+1})$$
$$1 + i_t = 1 + \pi_{t+1} + r_{t+1} + r_{t+1}\pi_{t+1}$$

and then:

$$i_t = \pi_{t+1} + r_{t+1} + r_{t+1}\pi_{t+1} \tag{10.75}$$

By assuming that both real interest rates and the inflation rate are fairly small, therefore $(\pi_{t+1} + r_{t+1})$ is much larger than $(r_{t+1}\pi_{t+1})$ and so it can be dropped, so that:

$$i_t \approx \pi_{t+1} + r_{t+1} \tag{10.76}$$

Specifically, this linear approximation is given by using two 1st order Taylor expansions:

$$\frac{1}{1+x} \approx 1 - x$$

$$(1 + x)(1 + y) \approx 1 + x + y$$

Therefore, by applying this linear approximation we have:

$$1 + r = \frac{1 + i}{1 + \pi} \approx (1 + i)(1 - \pi) \approx 1 + i - \pi$$

and therefore:

$$r \approx i - \pi \tag{10.77}$$

10.2 Bond: Price and Yield

Based on the fundamentals relationship defined in the previous paragraph, the price and yield of bonds can be defined.

Governments and companies can raise the capital needed to finance their activities by issuing bonds to a public market. In this context, we will refer to the government bonds.

A bond is nothing more than a loan for which the investor (bondholder) is the *lender*, while the organization that issues the bond (borrower) is defined as *issuer*.

There are several types of bonds. However, there are some elements common to all of them. In reality, any bond can be defined on the basis of five main elements:

(a) *face value* (also known as *par value* or *principal*): it is the amount that issuer has to pay back to the investor at the bond maturity (T);
(b) *price:* it is the price of the bond that the bondholder has to pay today (t);
(c) *coupon*: it is the interest rate (in percentage) paid by the issuer. Therefore, the value of the coupon, expressed as a percentage of the face value, is the amount that the bondholder will receive as interest payments. The interest is implicit and it is equal to the difference between the capital invested and the face value and coupon paid by the issuer.
(d) *maturity date*: it is the date in the future on which the issuer has to pay back the face value of the bond. Generally, the maturity can range from as little as one day to as long as 30 years. There is an inverse relationship between maturity and interest rate: the longer the time for maturity, the higher the risk, and therefore the higher the interest rate. Generally, other variables being equal, longer term bonds may fluctuate more than a shorter term bond;
(e) *issuer*: it is the organization, government or company that issues the bonds. It is one of the most relevant variables of the bond. The issuer's stability is the main assurance for bondholder of being paid.

The rating agencies define the bond rating based on the characteristics of the issuer (based on both past and expected performance). It is a synthetic judgment about the issuer's reliability and therefore the quality of the bonds issued. Table 10.5 reports the bond rating of the three main rating agencies: Moody's, Standard & Poors (S&P), and Fitch.

There is a relevant implication: the lower the grade of the bonds, the higher the risk and therefore higher the return offered by the issuer to the investors in the bond. Therefore, not all bonds are inherently safer than stocks. Certain types of bonds can be just as risky, if not riskier in certain conditions, than stocks.

Note that it is normal to measure the free-risk rate on the basis of the government bonds. Indeed, the default risk of the governments tend to be small (mostly for the

Table 10.5 Agencies rating on bonds

Moody's	S&P	Fitch	Grade	Risk
Aaa	AAA	AAA	Investment	Highest quality
Aa	AA	AA	Investment	High quality
A	A	A	Investment	Strong
Baa	BBB	BBB	Investment	Medium grade
Ba, B	BB, B	BB, B	Junk	Speculative
Caa/Ca/C	CCC/CC/C	CCC/CC/C	Junk	Highly speculative
C	D	D	Junk	In default

developed countries). It is because the government will always be able (or should be able) to bring in future revenues through taxation. On the other hand, companies must be able to generate profit in order to survive and face their debt obligations. The difference in risk between government and corporate bonds implies that the corporate bonds must offer a higher yield than government bonds.

It is necessary to introduce a distinction between the primary market and secondary market.

The bonds are issued in the primary market and underwritten by bondholders. Bonds can be held by bondholders until their maturity and in this case all of the considerations made up until now are valid and sufficient.

But the bondholder can decide to sell the bonds owned to another investor in bonds at any time. In this second case, bonds are not held until the maturity date but sold before in the secondary market or open market. In the open market, the price of bond changes on a daily basis. Therefore, in the open market the bond's price can fluctuate just like any other traded security.

The face value is not the price of the bond. It is the amount of monetary unit that will be paid at the maturity date (*T*). It is conventionally defined in 100 monetary units. The price is the amount paid by the investor to buy the bond at the start-time (*t*). Therefore, the price P with reference to the time period $\tau = T - t$ and then $P(t, T)$ can be defined as the monetary units that the investor has to pay today in time t in order to receive 100 monetary units (face value) at the maturity date of bond in time T, as follows:

$$P(t,T) = \frac{FV}{(1+i)^{\tau}} = \frac{100}{(1+i)^{\tau}} \tag{10.78}$$

The face value is conventionally defined as equal to 100. Therefore, having defined the price at time t, the interest rate is obtained indirectly. Also, τ defines the time-period and it is equal to the difference $(\tau = T - t)$.

By knowing the price of the bond at the start of the period (*t*) and having defined the face value equal to 100 units in money, it is possible to define the interest rate that the bond pays at maturity (*T*) as follows:

$$(1+i)^{\tau} = \frac{FV}{P(t,T)} \rightarrow [(1+i)^{\tau}]^{\frac{1}{\tau}} = \left(\frac{FV}{P(t,T)}\right)^{\frac{1}{\tau}} \rightarrow 1+i = \left(\frac{FV}{P(t,T)}\right)^{\frac{1}{\tau}}$$

and therefore, we have:

$$i = \left(\frac{FV}{P(t,T)}\right)^{\frac{1}{\tau}} - 1 \tag{10.79}$$

Therefore, the price and the interest rate provide the same information: having defined the one it provides the other, as follows:

$$P(t,T) = \frac{FV}{(1+i)^{\tau}} \leftrightarrow i = \left(\frac{FV}{P(t,T)}\right)^{\frac{1}{\tau}} - 1 \tag{10.80}$$

It is can be rewritten by using the operator present value PV_t, as follows (Cesari 2012):

$$P(t,T) = PV_t[100(T)] \tag{10.81}$$

Based on the linear proprieties of the present value, any capital amount can be defined. For example a capital equal to 250,000€, it can be defined as 2500 times 100 monetary units:

$$PV_t[25{,}0000(T)] = 2500 PV_t[100(T)] = 25{,}0000 PV_t[1(T)]$$

A bond's price fluctuates throughout its life in function of many variables. Generally, when a bond is traded at the price:

- *above* its face value, it is sold at a premium;
- *below* its face value, it is sold at a discount.

It is relevant to note that there is an inverse relationship between price on the one hand and interest rate and time on the other:

- the higher the interest rate, the lower the price; the lower the interest rate, the higher the price;
- the higher the maturity period, the lower the price; the lower the maturity period, the higher the price.

There are three main categories of bonds (Cesari 2012):

(a) *bonds with a defined face value and fixed coupon*. These bonds are defined as fixed-income securities because the exact amount of cash that bondholders get back if they keep the security until maturity is known.

The coupon can be constant or variable during the bond's maturity (*step-up*, if the coupon increases over time, or *step-down*, if the coupon decreases over time) but, however, well defined at the time of emission of the bond.

There are two main types:

- *fixed-coupon bond* (Coupon Bond (CB))[1]: it has a medium-long term maturity. The maturity goes from 3, 5, 10, 15 and 30 years. Coupons are normally constant and they are paid twice a year (every six months).

[1]They are: Bonds in USA; Gilt in UK; Bund in Germany; OAT in France; BTP (Buoni del Tesoro Poliennali) in Italy.

– *zero-coupon bond (ZCB)* [2]: it has a short-term maturity. The maturity goes from 3, 6 to 12 months. They are reimbursed in one solution at maturity;

(b) *bonds with defined face value and coupon "unknown"*. These bonds are characterized by a coupon's amount unknown until the maturity. The coupon of these bonds is function of a specific market parameter as defined by the yield of other bonds. The amount of the coupon will be known at the maturity of the bond or at least at a later time following emission.

These bonds are defined *financial index bond* because they are function of an interest rates. There are two main types:

– *floater*[3]: the coupon is direct function of another bond's yield. It can take on this form:

$$Coupon = 3\% + ZCB\ yield\ to\ 12\ months$$

Therefore, there is a proportional relationship between interest rates and coupon: if the interest rates increase, the Coupon increases.

– *reverse floater*: the coupon is inverse function of another bond's yield. It can take on this form:

$$Coupon = 8\% - ZCB\ yield\ to\ 3\ months$$

Therefore, there is an inverse relationship between interest rates and coupon: if the interest rates increase, the Coupon decreases. However, the coupon cannot be negative. Therefore, the right form is:

$$Coupon = Max(0;\ 8\% - ZCB\ yield\ to\ 3\ months)$$

(c) *bonds with face value and coupon "unknown"*. These bonds are characterized by the unknown value of the coupon and/or the reimbursement value. There are two main types:

– *Real index bonds*: the coupon and/or the bond reimbursement value are function of a specific market parameter different from the interest rate. In this sense, the market parameter could be the inflation rate. In this case, the bond could take on the following form:

$$Coupon\ and\ reimbursment\ value = 3\% \cdot 100 \cdot (1 + Inflation\ Rate)$$

[2] They are the Treasury Bills in USA. In Italy they are the BOT (Buoni Ordinari del Tesoro) or the CTZ (Certificati del Tesoro Zero-Coupon) with a maturity of between 18 and 24 months.

[3] In Italy it is the CCT (Certificato di Credito del Tesoro), with a maturity of between 3 and 7 years, and index to the BOT yield.

– *Structural Bonds*: they are based on the other types of assets. They are usually issued by the company. The most common are:

- *Callable Bond*: it is structured on the basis of the Coupon Bond and the acquisition of a call option by the issuer on the coupon bond as follows:

$$Callable\ Bond = CB - Call\ Option\ (CB)$$

- *Puttable Bond*: it is structured on the basis of the Coupon Bond and the acquisition of a put option by bondholder on the coupon bond as follows:

$$Puttable\ Bond = CB + Put\ Option\ (CB)$$

- *Convertible Bond*: it is structured on the basis of the Coupon Bond and the acquisition of an option to convert the bond into stock by bondholder as follows:

$$Convertible\ Bond = CB + Call\ Option\ (CB, Stock)$$

The *yield to maturity* indicates what the debt holder earns if he keeps the debt until maturity by receiving all of the payments as promised.

The bond yield is defined with regards to a time-period (τ) equal to the difference between the start of the period t and the end of the time-period T $(\tau = T - t)$.

The yield can be defined ex-post in T-time and therefore at the end of the period (ex-post perspective) or ex-ante in t-time and therefore at the start of the period (ex-ante perspective) (Cesari 2012).

In the *ex-post perspective*, and then in T-time, the problem is simple. The aim is to measure the yield with regards to the period between T and t and therefore in the time period $\tau = T - t$. In this case the price paid in in t-time (P_t) and the price on market value in T-time (P_T) is known.

Therefore, in order to define the yield, the interest rule must be defined. For a period higher than one year, the compound interest rule is normally used. In this case, the yield and therefore the return on investment (R) it is equal to the compound interest rate, as follows (Cesari 2012):

$$i_{ex-post} = \left(\frac{P_T}{P_t}\right)^{\frac{1}{T-t}} - 1 \tag{10.82}$$

Note that if the bond pays coupons in the period $\tau = T - t$, their value must be considered in P_T. Indeed, in this case the P_T is equal to the price of the bond at T-time plus the sum of all payments received in the time-period τ.

If the coupons paid during the time-period τ are re-invested until the T-time on the basis of a compound interest rate, we have:

$$P_T = c_{t1}(1+i)^{T-t_1} + c_{t2}(1+i)^{T-t_2} + \cdots + c_T(1+i)^{T-T=1} + P \quad (10.83)$$

where c_t are the payments received in a time between the start of the period (t) and the end of the period (T) so that $t \leq t_n \leq T$ and i the interest rate used in the compound interest rule.

Otherwise, if the coupons paid are not invested in the time-period (τ) only the payments received are considered as follows:

$$P_T = c_{t1} + c_{t2} + \cdots + c_T + P \quad (10.84)$$

In the *ex-ante perspective* it is necessary to make assumptions with regards to duration, price and interest rate.

In this case the *current yield* (or *Coupon Yield*) can be calculated. The easiest criteria to calculate the bond's yield is the ratio between coupon amount and price as follows (Cesari 2012):

$$Coupon\ Yield\ (\%) = \frac{C}{P_t} \cdot 100 \quad (10.85)$$

where C is the coupon amount on an annual basis and P_t is the bond price at the start of the period (t).

Whenever the coupon amount is in percentage of the bond's face value (FV) we have:

$$Yield\ (\%) = \frac{C(\%) \cdot FV}{P_t} \quad (10.86)$$

Therefore, if the bond is bought at a par value (face value), yield is equal to the interest rate. Otherwise, if the price changes, the yield changes. This happens because the guarantee of the bond (coupon and face value) does not change. Consequently, there is an inverse relationship between price and yield, so that when:

- price decreases, yield increases;
- price increases, yield decreases.

The coupon yield is a part of the total yield. It requires calculation at the end of the period (T). Usually, the investors refer to *Yield to Maturity* (*YTM*). The YTM defines the total return that the investor will receive if he keeps the bond until maturity. It is equal to all of the interest payments that the investor will receive (by assuming that he will reinvest the interest payment at the same rates as the current yield on the bond) plus the capital gain (if purchased at a discount) or loss (if purchased at a premium), as follows (Cesari 2012):

$$YTM(\%) = \frac{C}{P_t} + \frac{\overline{P_T} - P_t}{P_t} \cdot \frac{1}{\tau} \quad (10.87)$$

where C is the coupon amount on annual basis; P_t is the bond price at the start of the period (t); $\overline{P_T}$ is the expected bond price to maturity at the end of the period (T); τ is the time-period of the bond maturity $(\tau = T - t)$.

Generally, the factor that influences a bond more than any other is the level of prevailing interest rates in the economy. When:

- the interest rates rise, the prices of bonds in the market fall, thereby raising the yield of the older bonds and putting them in line with newer bonds being issued with higher coupons;
- the interest rates fall, the prices of bonds in the market rise, thereby lowering the yield of the older bonds and putting them in line with newer bond being issued with lower coupons.

10.3 Term Structure of Interest Rate

The *term structure of interest rates (TSI)* defines the *Yield Curve* and it indicates the structure relationship between interest rates at different maturities (Alexander 1980; Altman 1987; Asquith et al. 1989; Balduzzi et al. 2001; Black and Cox 1976; Black and Scholes 1973; Brennan and Schwartz 1977, 1980, 1982; Chance 1990; Cox et al. 1980; Elton et al. 2001; Fama 1984a, b; Fraine and Mills 1961; Johnson 1967; Malkiel 1966; Rao 1982; Smith and Warner 1979; Sundaresan 1983; Zwick 1980). Specifically, the term structure of interest rates defines, at a defined time t, the relationship between the prices (or interest rates, as obtained indirectly on the price basis) of the bonds on markets and their maturity and therefore the end time-period T, as well as their time-period $\tau = T - t$. In other words, the term structure of interest rates illustrates the relationship between the interest rates when only the maturity changes in a defined both market and time: the different interest rates refer to the different maturity that define the curve of the term structure of interest rates, the Yield Curve. It is worth noting that in a different time-period and market, the structure of interest rates is different and then the yield curve is different.

In order to define the term structure, it is necessary that for each bonds all other variables must be equal. The objective is to highlight the term structure only.

The zero-coupon bonds lie at the basis of the term structure of interest rate. They are elementary bonds and they are very important for finance.

The zero-coupon bond gives the right to receive 100 units monetary at the maturity date. Therefore, the price of a zero-coupon bonds with face value equal to 100 units of money is equal to:

$$P(t,T) = \frac{100}{(1 + i(t,T))^{T-t}} \tag{10.88}$$

On the basis of Eq. (10.88) it is possible to calculate the interest rate as follows (Cesari 2012):

$$i(t,T) = \left(\frac{100}{P(t,T)}\right)^{\frac{1}{T-t}} - 1 \tag{10.89}$$

Assume a zero-coupon bond with face value equal to 100 units in money and time-period equal to 1 year $(T - t = 1)$. If the interest rates are different with regards to the different time-period, for the same bond it will pay a different price due to the change of the interest rates according to its maturity date.

Assume different interest rates for different maturity of the same bond in the same market as in Table 10.6 (adapted from Cesari 2012).

Table 10.6. shows that by changing the maturity of the bond, it changes the interest rate and therefore it changes the price of the bond. Therefore, the same bond has different prices with regards to the different maturity due to the changes of the interest rates for different time periods. The different interest rates refer to the different maturities defining the yield curve. Therefore, the term structure of interest rates, and therefore the yield curve, defines the relationship between interest rates when the maturity changes only, in the same market at a defined time. This last condition implies that the interest rates for the different maturities are defined at a defined date. Therefore, in a different date on the same market the interest rates referring to the maturities can change and therefore change the yield curve. The yield curve can undertake different shapes: increasing, decreasing, hump, etc.

Note that Table 10.6. shows a relevant relationship: price and interest rate provide the same information. Indeed, given the one it is possible to calculate the other on the basis of a mathematical relationship.

The term structure of interest rates and therefore the yield curve is fundamental for the determination of the price of all other assets.

In the same analysis, the yield curve is assumed flat. In this case the same interest rate for different maturities is assumed. This assumption is hardly noted in reality, but it has the advantage of simplifying the analysis.

It is relevant to know that the TSI's are the Internal Rate of Return (IRR) of the zero-coupon bond. They are the interest rates that can be achieved at the maturity date. For the zero-coupon bond there is no problem regarding reinvestment of the coupon.

By considering the coupons, it is possible to define the relationship between the TSI and IRR for coupon bonds.

Assume that the P_B is the bond's price with coupon c. By considering the IRR, we have (Cesari 2012):

Table 10.6 Relationship between interest rate, price and maturity

PRICE	Maturity Time-period (T-t)									
Interest rate i(t,T) (%)	1	2	3	4	5	6	7	8	9	10
3	97.09	94.26	91.51	88.85	86.26	83.75	81.31	78.94	76.64	74.41
4	96.15	92.46	88.90	85.48	82.19	79.03	75.99	73.07	70.26	67.56
5	95.24	90.70	86.38	82.27	78.35	74.62	71.07	67.68	64.46	61.39
6	94.34	89.00	83.96	79.21	74.73	70.50	66.51	62.74	59.19	55.84
7	93.46	87.39	81.63	76.29	71.30	66.63	62.27	58.20	54.39	50.83
8	92.59	85.73	79.38	73.50	68.06	63.02	58.35	54.03	50.02	46.32
9	91.74	84.17	77.22	70.84	64.99	59.63	54.70	50.19	46.04	42.24
10	90.91	82.64	75.13	68.30	62.09	56.45	52.32	46.65	42.41	38.55

$$P_B = \frac{c}{(1+i)} + \frac{c}{(1+i)^2} + \cdots + \frac{c+100}{(1+i)^n}$$

By considering the TSI, we have:

$$P_B = \frac{c}{(1+R_1)} + \frac{c}{(1+R_2)^2} + \cdots + \frac{c+100}{(1+R_n)^n}$$

Therefore, the IRR is a function of the TSI and others elements such as duration n, the coupon c, the nature and solidity of the bond offered, etc.

By considering the n rate (R_1, R_2, R_n) of the TSI, the IRR can be defined as follows:

$$f(i, i, i, \ldots, i) = f(x_1, x_2, x_3, \ldots, x_n) \equiv \frac{c}{(1+x_1)} + \frac{c}{(1+x_2)^2} + \cdots + \frac{c+100}{(1+x_n)^n} \tag{10.90}$$

It is possible to analyse this relation by considering three main relationships (Cesari 2012):

(a) IRR and the bond offered;
(b) IRR and the bond duration;
(c) IRR and the coupon bond;
(d) Duration

(A) IRR and the Bond Offered

The capability of the bond offered to face debt obligations has important effects on the price and then on the interest rate, with all other conditions equal. The higher the default probability of the bond offered, the lower the value of the bond: the lower price or in equivalent terms, the higher the interest rate to be offered to the bondholders in order to acquire the bond.

The price of the bond is inverse related to the default probability of the bond offer: the higher the default probability, the lower the price and subsequently the higher the interest rate. Therefore, ceteris paribus, the IRR of the bond increases with the increase of the default probability of the bond offer.

The bond offer rating measures the bond offer's ability to face debt obligations: the lower the rating, the lower the bond offer's ability to face debt obligations, then higher the default risk and then lower the bond's price and higher the interest rate to be offered ex-ante to the bondholders in order to sell the bond. The higher interest rate is required by bondholders in order to face the higher default risk of the bond offer.

Therefore, each bond offer has a specific IRR and therefore a specific TSI. The difference between the TSI of the bond offer and the TSI of the market with regards to the best rating class, defines the credit spread. Denoting with $R^e(t, T)$ the specific

TSI of the bond offer and with $R(t,T)$ the TSI of the best class rating in the market, the credit spread (CS) can be defined as follows:

$$CS = R^e(t,T) - R(t,T) \tag{10.91}$$

Generally, the IRR of issuer can be defined on the basis of four elements (Cesari 2012):

$$R^e(t,T) = p(t) + \pi(t) + term\ spread + credit\ spread \tag{10.92}$$

where $p(t)$ is the real interest rate in short-term and $\pi(t)$ is the inflation rate.

(B) IRR and the Bond Duration

Assume a bond with coupon. In this case, the relationship between the IRR (i) and the duration (n) is due to the TSI.

It is possible to show this relationship by considering three main cases (Cesari 2012). In all cases, assume two different bonds with the same coupon but different with a different duration: bond (A) with coupon equal 3% and duration of 5 years; bond (B) with coupon equal 3% and duration of 10 years.

Case 1: TSI increases over time

In case the TSI increases over time, the bond with a long duration has a higher IRR than the bond with short duration as shown in Table 10.7 (adapted from Cesari 2012).

Table 10.7 shows how the TSI's increase over time, the longer bond (bond B) is preferred to shorter bond (bond A).

Case 2: TSI decreases over time

In case the TSI decreases over time, the bond with a short duration has a higher IRR than the bond with long duration as shown in Table 10.8 (adapted from Cesari 2012).

Table 10.8 shows how the TSI decreases over time, shorter bond (bond A) is preferred to longer bond (bond B).

Case 3: TSI is flat over time

In case the TSI is flat over time, both bonds have the same IRR equal to the TSI as shown in Table 10.9 (adapted from Cesari 2012).

(C) IRR and the Coupon Bond

There is a relationship between IRR and the coupon bond. With other conditions equal, the higher the coupon, the higher the value of the bond.

The relationship between IRR and coupon is strictly related to the TSI curve. Specifically:

– if TSI increases, the bond with a lower coupon has a higher IRR;

Table 10.7 Case 1: TSI is assumed to increase over time

Years	Duration	TSI (%)	Discount rate	Coupon	Face value
0	01/01/2006				100
1	01/01/2007	3.0	0.97087	3	
2	01/01/2008	3.5	0.93351	3	
3	01/01/2009	4.0	0.88900	3	
4	01/01/2010	4.5	0.83856	3	
5	01/01/2011	5.0	0.78353	3	
6	01/01/2012	5.5	0.72525	3	
7	01/01/2013	6.0	0.66506	3	
8	01/01/2014	6.5	0.60423	3	
9	01/01/2015	7.0	0.54393	3	
10	01/01/2016	7.5	0.48519	3	

Bond A		
Price	Coupon and face value	Present value
−91.60		
	3	2.91262
	3	2.80053
	3	2.66699
	3	2.51568
	103	80.70320
IRR 4.937%		

Bond B		
Price	Coupon and face value	Present value
−70.84		
	3	2.91262
	3	2.80053
	3	2.66699
	3	2.51568
	3	2.35058
	3	2.17574
	3	1.99517
	3	1.81269
	3	1.63180
	103	49.97497
IRR 7.188%		

– if TSI decreases, the bond with a lower coupon has a lower IRR.

These relationships are based on two different elements (Cesari 2012):

– *general rule*: the higher the coupon, the higher the cash-in of investment. Therefore, the higher the coupon, the lower the time of investment return and

Table 10.8 Case 2: TSI is assumed to decrease over time

Years	Duration	TSI (%)	Discount rate	Coupon	Face value
0	01/01/2006				100
1	01/01/2007	7.5	0.93023	3	
2	01/01/2008	7.0	0.87344	3	
3	01/01/2009	6.5	0.82785	3	
4	01/01/2010	6.0	0.79209	3	
5	01/01/2011	5.5	0.76513	3	
6	01/01/2012	5.0	0.74622	3	
7	01/01/2013	4.5	0.73483	3	
8	01/01/2014	4.0	0.73069	3	
9	01/01/2015	3.5	0.73373	3	
10	01/01/2016	3.0	0.74409	3	

Bond A

Price	Coupon and face value	Present value
−89.08		
	3	2.79070
	3	2.62032
	3	2.48355
	3	2.37628
	103	78.80884

IRR 5.562%

Bond B

Price	Coupon and face value	Present value
−98.04		
	3	2.79070
	3	2.62032
	3	2.48355
	3	2.37628
	3	2.29540
	3	2.23865
	3	2.20449
	3	2.19207
	3	2.20119
	103	76.64167

IRR 3.232%

then the lower the effective duration of investment. On the other hand, the lower the coupon, the higher the time of investment return and then the higher the effective duration of investment. Therefore, regardless of nominal duration, the effective duration of the bond depends on the amount of the coupon;

Table 10.9 Case 3: TSI is assumed flat over time

Years	Duration	TSI (%)	Discount rate	Coupon	Face value
0	01/01/2006				100
1	01/01/2007	5.0	0.95238	3	
2	01/01/2008	5.0	0.90703	3	
3	01/01/2009	5.0	0.86384	3	
4	01/01/2010	5.0	0.82270	3	
5	01/01/2011	5.0	0.78353	3	
6	01/01/2012	5.0	0.74622	3	
7	01/01/2013	5.0	0.71068	3	
8	01/01/2014	5.0	0.67684	3	
9	01/01/2015	5.0	0.64461	3	
10	01/01/2016	5.0	0.61391	3	

Bond A

Price	Coupon and face value	Present value
−91.34		
	3	2.85714
	3	2.72109
	3	2.59151
	3	2.46811
	103	80.70320

IRR 5.000%

Bond B

Price	Coupon and face value	Present value
−84.56		
	3	2.85714
	3	2.72109
	3	2.59151
	3	2.46811
	3	2.35058
	3	2.23865
	3	2.13204
	3	2.03052
	3	1.93383
	103	63.23307

IRR 5.000%

– *effect of TSI*: if the TSI increases over time, the discount factor decreases over time. Therefore, its effect on present value is higher for a higher coupon than a lower one.

The combination of these two elements affects the IRR. Assume two bonds: bond (A) with a coupon of 3% and a duration of 10 years; bond (B) with a coupon

Table 10.10 Case 1: TSI increases over time

Years	Duration	TSI (%)	Discount rate	Face value
0	01/01/2006			100
1	01/01/2007	3.0	0.97087	
2	01/01/2008	3.5	0.93351	
3	01/01/2009	4.0	0.88900	
4	01/01/2010	4.5	0.83856	
5	01/01/2011	5.0	0.78353	
6	01/01/2012	5.5	0.72525	
7	01/01/2013	6.0	0.66506	
8	01/01/2014	6.5	0.60423	
9	01/01/2015	7.0	0.54393	
10	01/01/2016	7.5	0.48519	

Bond A

Price	Coupon	Coupon and face value	Present value
−70.84			
	3	3	2.91262
	3	3	2.80053
	3	3	2.66699
	3	3	2.51568
	3	3	2.35058
	3	3	2.17574
	3	3	1.99517
	3	3	1.81269
	3	3	1.63180
	3	103	49.97497

IRR 7.188%

Bond B

Price	Coupon	Coupon and face value	Present value
−85.72			
	5	5	4.85437
	5	5	4.66755
	5	5	4.44498
	5	5	4.19281
	5	5	3.91763
	5	5	3.62623
	5	5	3.32529
	5	5	3.02116
	5	5	2.71967
	5	105	50.94536

IRR 7.037%

of 5% and a duration of 10 years. Consider two cases: (i) TSI increases over time; (ii) TSI decreases over time (Cesari 2012).

Case 1: TSI increases over time

In case the TSI increases over time. Therefore, the bond (A) with lower coupon than the bond (B) is characterized by a higher IRR as shown in Table 10.10 (adapted from Cesari 2012).

Therefore, if the TSI increases, the bond with a lower coupon has a higher IRR.

Case 2: TSI decreases over time

In case the TSI decreases over time. Therefore, the bond (A) with lower coupon than the bond *B)* is characterized by a lower IRR as shown in Table 10.11 (adapted from Cesari 2012).

Therefore, if the TSI decreases, the bond with a lower coupon has a lower IRR.

(D) **Duration**

In the previous cases, the analysis shows that the higher amount of coupon reduces the effective duration of the bond. Indeed, the higher the coupon, the faster the return investment.

It is possible to define the duration as the weighted average of a single deadline by using the present values of each expected payment as weights and by using the TSI as interest rate to calculate the present value.

Denote with $c_t(c_1, c_2, c_3, \ldots c_n)$ the future expected payments for each year $(t = 1, 2, \ldots, n)$. Since it is a weight average, the denominator is the sum of the weights. But in its formal expression, it is the price of the (P) of the bond. The duration (D) is equal to (Cesari 2012):

$$D = \frac{\sum_{t=1}^{n} t \cdot c_t \cdot \frac{1}{(1+R_t)^t}}{\sum_{t=1}^{n} c_t \cdot \frac{1}{(1+R_t)^t}} = \frac{\sum_{t=1}^{n} t \cdot c_t \cdot \frac{1}{(1+R_t)^t}}{P} \tag{10.93}$$

Note that if the coupon is equal to zero, the duration is equal to the time n of the investment.

Assume two bonds: bond (A) with a coupon of 3% and a duration of 10 years; bond (B) with a coupon of 5% and a duration of 10 years. Consider two cases: (i) TSI increases over time; (ii) TSI decreases over time.

Case 1: TSI increases over time

In case the TSI increases over time, the higher coupon, the lower the effective duration as shown in Table 10.12 (adapted from Cesari 2012).

Case 2: TSI decreases over time

In case the TSI decreases over time, the higher coupon, the lower the effective duration as shown in Table 10.13 (adapted from Cesari 2012).

Table 10.11 Case 2: TSI decreases over time

Years	Duration	TSI (%)	Discount rate	Face value
0	01/01/2006			100
1	01/01/2007	7.5	0.93023	
2	01/01/2008	7.0	0.87344	
3	01/01/2009	6.5	0.82785	
4	01/01/2010	6.0	0.79209	
5	01/01/2011	5.5	0.76513	
6	01/01/2012	5.0	0.74622	
7	01/01/2013	4.5	0.73483	
8	01/01/2014	4.0	0.73069	
9	01/01/2015	3.5	0.73373	
10	01/01/2016	3.0	0.74409	

Bond A

Price	Coupon	Coupon and face value	Present value
−98.04			
	3	3	2.79070
	3	3	2.62032
	3	3	2.48355
	3	3	2.37628
	3	3	2.29540
	3	3	2.23865
	3	3	2.20449
	3	3	2.19207
	3	3	2.20119
	3	103	76.64167

IRR 3.232%

Bond B

Price	Coupon	Coupon and face value	Present value
−113.80			
	5	5	4.65116
	5	5	4.36719
	5	5	4.13925
	5	5	3.96047
	5	5	3.82567
	5	5	3.73108
	5	5	3.67414
	5	5	3.65345
	5	5	3.66865
	5	105	78.12986

IRR 3.353%

Table 10.12 Case 1: TSI increases over time

Duration in years	TSI (%)	Discount rate	Face value
			100
1	3.0	0.97087	
2	3.5	0.93351	
3	4.0	0.88900	
4	4.5	0.83856	
5	5.0	0.78353	
6	5.5	0.72525	
7	6.0	0.66506	
8	6.5	0.60423	
9	7.0	0.54393	
10	7.5	0.48519	

Bond A

Price	Coupon	Coupon and face value	Present value	Weights (%)	Duration effective
70.84					
	3	3	2.91262	4	0.04
	3	3	2.80053	4	0.08
	3	3	2.66699	4	0.11
	3	3	2.51568	4	0.14
	3	3	2.35058	3	0.17
	3	3	2.17574	3	0.18
	3	3	1.99517	3	0.20
	3	3	1.81269	3	0.20
	3	3	1.63180	2	0.21
	3	103	49.97497	71	7.05
				100.00	8.39

Bond B

Price	Coupon	Coupon and face value	Present value	Weights (%)	Duration effective
85.72					
	5	5	4.85437	6	0.06
	5	5	4.66755	5	0.11
	5	5	4.44498	5	0.16
	5	5	4.19281	5	0.20
	5	5	3.91763	5	0.23
	5	5	3.62623	4	0.25
	5	5	3.32529	4	0.27
	5	5	3.02116	4	0.28
	5	5	2.71967	3	0.29
	5	105	50.94536	59	5.94
				100.00	7.78

Table 10.13 Case 2: TSI decreases over time

Duration in years	TSI (%)	Discount rate	Face value
			100
1	7.5	0.93023	
2	7.0	0.87344	
3	6.5	0.82785	
4	6.0	0.79209	
5	5.5	0.76513	
6	5.0	0.74622	
7	4.5	0.73483	
8	4.0	0.73069	
9	3.5	0.73373	
10	3.0	0.74409	

Bond A

Price	Coupon	Coupon and face value	Present value	Weights (%)	Duration effective
98.04					
	3	3	2.79070	3	0.03
	3	3	2.62032	3	0.05
	3	3	2.48355	3	0.08
	3	3	2.37628	2	0.10
	3	3	2.29540	2	0.12
	3	3	2.23865	2	0.14
	3	3	2.20449	2	0.16
	3	3	2.19207	2	0.18
	3	3	2.20119	2	0.20
	3	103	76.64167	78	7.82
				100.00	8.86

Bond B

Price	Coupon	Coupon and face value	Present value	Weights (%)	Duration effective
113.80					
	5	5	4.65116	4	0.04
	5	5	4.36719	4	0.08
	5	5	4.13925	4	0.11
	5	5	3.96047	3	0.14
	5	5	3.82567	3	0.17
	5	5	3.73108	3	0.20
	5	5	3.67414	3	0.23
	5	5	3.65345	3	0.26
	5	5	3.66865	3	0.29
	5	105	78.12986	69	6.87
				100.00	8.37

In both cases the duration of the bond with a lower coupon is higher than the bond with a higher coupon. However, in both cases the duration is lower if the TSI increases.

In the analysis on duration, IRR coupon, the TSI shows that it is not possible to define criteria capable of identifying ex-ante the best bond. Indeed, if admitting the existence of this criteria, it implies the existence of arbitrage opportunities. All criteria used is based on the expectation of the interest rates.

10.4 Expectation Theory of Term Structure Interest Rate

A theory about TSI can be defined as a theoretical model capable of explaining the curve of TSI and its movements over time. There are several models to estimate the curve of TSI (McCulloch 1971, 1975; Nelson and Siegel 1985; Svensson 1994; Merton 1974; Vasicek 1977; Cox et al. 1985; Ho and Lee 1986; Hull and White 1990; Heath et al. 1992; Brace et al. 1997; Chan and Thomson 1988; Campbell 1986; Carleton and Cooper 1976; Chambers et al. 1984; Fama 1976, 1984a, b; Modigliani and Sutch 1966; Roll 1970; Van Horne 1965, 1966). It is outside of this present work. The aim of this paragraph is to provide a brief introduction on the expectation theory of TSI.

The term structure of interest rates (TSI) are spot rates and relative prices $P(t,T)$ where t is the current time and T is the end-time.

Denote with: T and S two end-times where $S < T$; $R(t,T)$ the long-term interest rate; $R(t,S)$ the short-term interest rate; $R(t,S,T)$ the forward interest rate applied in time between short-term (S) and long-term (T). In a condition of non-arbitrage, we have (Cesari 2012):

$$[1+R(t,T)]^{(T-t)} = [1+R(t,S)]^{(S-t)}[1+R(t,S,T)]^{(T-S)} \tag{10.94}$$

Equation (10.94) shows that in a condition of non-arbitrage the long-term interest rate $(R(t,T))$ is equal to the short-term interest rate $(R(t,S))$ combined with the forward interest rate $(R(t,S,T))$ in the time period between short-term (S) and long-term (T).

This Eq. (10.94) cannot be considered a theory of long-term interest rate. However, its application avoids arbitrage in financial markets and therefore theories capable of admitting arbitrage.

One of the most relevant theories on the long-term interest rates is the *Expectation Theory* (also known as *Unbiased Expectations Theory*) whose fathers are Fisher (1930) and Hicks (1939).

For further understanding assume a context of no-uncertainty (Cesari 2012). Assume two different time periods: time S and time T with $S < T$. Assume two different strategies:

- *perfect matching*: to buy an asset with maturity in time S at price $P(t,S)$ with earning equal to $\frac{100}{P(t,S)}$;
- *yield curve riding*: to buy an asset with maturity in time T and sell it in time S by earnings equal to $\frac{P(S,T)}{P(t,T)}$.

In a condition of no-uncertainty, the no-arbitrage condition implies that the results of these two strategies must be equal as follows:

$$\frac{P(S,T)}{P(t,T)} = \frac{100}{P(t,S)}$$

Therefore, the total rate of return in the entire period $\left(\frac{P(S,T)}{P(t,T)}\right)$ is equal to rate of return to maturity $\left(\frac{100}{P(t,S)}\right)$. Based on this equivalence, the price in the future time S can be achieved, as follows (Cesari 2012):

$$P(S,T) = \frac{P(t,T)}{P(t,S)}100 = Q(t,S,T) \tag{10.95}$$

Equation (10.95) shows that in a condition of no-uncertainty and no-arbitrage, the forward price is equal to the future price.

By introducing the uncertainty, it is possible to maintain the structure of the reasoning by saying that the forward price is not equal to the future price but it is equal to the *expectations* on future price. By using the operator expected value $E[.]$, Eq. (10.95) becomes (Cesari 2012):

$$Q(t,S,T) = E_t[P(S,T)] \rightarrow Q(t,S,T) - E_t[P(S,T)] = 0 \tag{10.96}$$

It is relevant to note that the $E_t[.]$ is the expected value according to the information available to the current time t. This is the pure version of the expectation theory.

Unlike the pure expectation theory, Keynes hypothesized negative differences between the forward price and future expected price called *normal backwardation*. In this case Eq. (10.96) becomes (Cesari 2012):

$$Q(t,S,T) < E_t[P(S,T)] \rightarrow Q(t,S,T) - E_t[P(S,T)] < 0 \tag{10.97}$$

This happens because hedgers tend to be short-sellers and then they have to push the speculators to assume a long-position. It implies a forward price lower than the future expected price resulting in a positive profit expectation on long-positions. Therefore, the forward price tends to increase over time between time t and time S because the forward price must be equal to the spot price.

Similarly if the hedgers assume a long-position. In this case, the relation is inverted and there is a positive difference between the forward price and the future expected price as follows (Cesari 2012):

$$Q(t,S,T) > E_t[P(S,T)] \rightarrow Q(t,S,T) - E_t[P(S,T)] > 0 \tag{10.98}$$

This happens because hedgers tend to be long-position and then they have to push the speculators to assume short-position. It implies a forward price higher than the future expected price resulting in a positive profit expectation on short-positions. Therefore, the forward price tends to decrease over time between time t and time S because the forward price must be equal to the spot price.

The pure version of the *Expectation Theory* can be applied on interest rates. The theory can be formulated by saying that the forward interest rate is the expected value of the spot interest rate in the future time. On the basis of Eq. (10.96) we have (Cesari 2012):

$$R(t,S,T) = E_t[R(S,T)] \rightarrow R(t,S,T) - E_t[R(S,T)] = 0 \tag{10.99}$$

In a condition of no-arbitrage, Eq. (10.94) can be rewritten as follows:

$$[1+R(t,T)]^{(T-t)} = [1+R(t,S)]^{(S-t)}[1+E_t[R(S,T)]]^{(T-S)} \tag{10.100}$$

A general formulation can be obtained by considering a special case of a single period interest rate (Cesari 2012). Assume: $S = t+1; T-S = 1; T = t+2$.

In a condition of no-arbitrage, by applying Eq. (10.94) we have:

$$[1+R(t,t+2)]^{(t+2-t)} = [1+R(t,t+1)]^{(t+1-t)}[1+R(t,t+1,t+2)]^{(t+2-t-1)}$$

and then:

$$[1+R(t,t+2)]^2 = [1+R(t,t+1)][1+R(t,t+1,t+2)]$$

By using the linear approximation, we have:

$$R(t,t+2) = \frac{R(t,t+1)+R(t,t+1,t+2)}{2}$$

It shows that the interest rate at 2 years is equal to the arithmetical average of the one-year spot rate and one-year forward rate in a year.

Based on the pure expectation theory, applying Eq. (10.99) we have:

$$R(t,t+1,t+2) = E_t[R(t+1,t+2)]$$

and:

$$R(t,t+1) = E_t[R(t,t+1)]$$

and then:

$$R(t,t+2) = \frac{1}{2}\sum_{h=1}^{2} E_t[R(t+h-1,t+h)]$$

Generalizing we have (Cesari 2012):

$$R(t,t+\tau) = \frac{1}{\tau}\sum_{h=1}^{\tau} E_t[R(t+h-1,t+h)] \tag{10.101}$$

Equation (10.101) shows on that the spot rate for the period of τ years $(R(t,\tau))$ is an arithmetical average of the τ future expected single period rates.

Therefore, the return investment on the long-term is equal to the return of investment in a short-term repeated from year to year (roll-over mechanism). On the basis of this mechanism, the pure version of the *Expectation Theory* can be used to explain the dynamic of the TSI curve. Specifically (Cesari 2012):

- if the curve of TSI has a positive slope, it implies a long-term rate $(R(t,t+\tau))$ higher than short rate $(R(t,t+1))$. It can be explained with expectations of future spot rates higher than current spot rates;
- if the curve of TSI has a negative slope, it implies a long-term rate $(R(t,t+\tau))$ lower than short rate $(R(t,t+1))$. It can be explained with expectations of future spot rates lower than current spot rates;
- if the curve of TSI is flat, it implies a long-term rate $(R(t,t+\tau))$ equal to the short rate $(R(t,t+1))$. It can be explained with expectations of future spot rates equal to the current spot rates.

A general formulation of the *Expectation Theory* that also considers the *normal backwardation*, can be achieved by introducing a *liquidity premium* (or *term premium*) (δ). In this case, Eq. (10.99) becomes (Cesari 2012):

$$R(t,t+\tau-1,t+\tau) = E_t[R(t+\tau-1,t+\tau)] + \delta(\tau) \quad with \;\; \delta(\tau) > 0 \tag{10.102}$$

The liquidity premium $\delta(\tau)$ is positive and it represents the difference between the forward rate and the future expected rate.

Based on this relationship, Eq. (10.101) can be rewritten as follows (Cesari 2012):

$$R(t,t+\tau) = \frac{1}{\tau}\sum_{h=1}^{\tau} E_t[R(t+h-1,t+h)] + \pi(\tau) \tag{10.103}$$

The term $\pi(\tau)$ in Eq. (10.103) is a premium linkage to the uncertainty of the future rates. Based on:

- *Liquidity preference theory*, the premium is positive $(\pi(\tau) > 0)$ because it is the premium required by investors to invest in the long-term renouncing to liquidity.

– *Preferred habitat theory*, the premium is negative $(\pi(\tau) < 0)$ because the premium is required by investors to invest in a short-term while they prefer the long-term;
– *Pure version of the Expectation Theory* requires no premium $(\pi(\tau) = 0)$.

References

Alexander GJ (1980) Applying the market model to long-term corporate bonds. J Financ Quant Anal XV(5):1063–1080
Altman EI (1987) Investing in junk bonds: inside the high yield debt market. Wiley, New York
Asquith P, Mullins D, Wolff E (1989) Original issue high yield bonds: aging analysis of defaults, exchanges and calls. J Finance 44:923–953
Balduzzi P, Elton EJ, Green TC (2001) Economic news and the yield curve: evidence from the US treasury market. J Financ Quant Anal 36:523–543
Black F, Cox JC (1976) Valuing corporate securities: some effects of bond indenture provisions. J Finance 31:351–367
Black F, Scholes M (1973) The pricing options and corporate liabilities. J Polit Econ 81:637–654
Brace A, Gatarek D, Musiela M (1997) The market model of interest rate dynamics. Math Finance 7(2):127–147
Brennan MJ, Schwartz ES (1977) Convertible bonds: valuation and optimal strategies for call and conversion. J Finance 32:1699–1715
Brennan MJ, Schwartz ES (1980) Conditional predictions of bond prices and returns. J Finance 35(2):405–416
Brennan MJ, Schwartz ES (1982) Bond pricing and market efficiency. Financ Anal J 38(5):49–56
Campbell JY (1986) A defense of traditional hypotheses about the term structure of interest rates. J Finance 41(1):183–193
Carleton WT, Cooper IA (1976) Estimation and uses of the term structure of interest rates. J Finance 31:1067–1083
Cesari R (2012) Introduzione alla finanza matematica. Concetti di base, tassi e obbligazioni, 2nd edn. McGraw-Hill, New York
Chambers DR, Carleton WT, Waldman DW (1984) A new approach to estimation of the term structure of interest rates. J Financ Quant Anal 19(3):233–252
Chan A, Thomson HE (1988) Jump-diffusion and the term structure of interest rates. J Finance 43(1):155–174
Chance DM (1990) Default risk and the duration of zero coupon bonds. J Finance 45(1):265–274
Cox JC, Ingersoll JE, Ross SA (1985) A theory of the term structure of interest rates. Econometrica 53(2):385–407
Cox JC, Ingersoll JE, Ross S (1980) An analysis of variable rate loan contracts. J Finance 35(2):389–404
Elton EJ, Gruber MJ, Agrawal D, Mann C (2001) Explaining the rate spread on corporate bonds? J Finance LVI(1):247–279
Fama E (1976) Forward rates as predictors of future sport rates. J Financ Econ 3:361–377
Fama E (1984a) The information in the term structure. J Financ Econ 13(4):509–528
Fama E (1984b) Term premiums in bond returns. J Financ Econ 13(4):529–546
Fisher I (1930) The theory of interest. Macmillan, New York
Fraine HG, Mills RH (1961) The effect of default and credit deterioration on yields of corporate bonds. J Finance 16:423–433
Heath D, Jarrow R, Morton A (1992) Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. Econometrica 60(1):77–105

Hicks JR (1939) Value and capital, 2nd edn. Clarendon Press, Oxford
Ho T, Lee S (1986) Term structure movements and pricing interest rate contingent claims. J Finance 41(5):1011–1029
Hull J, White A (1990) Pricing Interest-rate derivative securities. Rev Financ Stud 3(4):573–592
Johnson R (1967) Term structure of corporate bond yields as a function of risk of default. J Finance 22:313–345
Malkiel BG (1966) The term structure of interest rates. Princeton University Press, Princeton NJ
McCulloch JH (1971) Measuring the term structure of interest rate. J Bus 44:19–31
McCulloch JH (1975) An estimate of the liquidity premium. J Polit Econ 83:95–119
Merton R (1974) On the pricing of corporate debt: the risk structure of interest rates. J Finance 29(2):449–470
Modigliani F, Sutch R (1966) Innovations in interest rate policy. Am Econ Rev LVI:178–197
Nelson CR, Siegel AF (1985) Parsimoneous modeling of yield curves for U.S. treasury bills. NBER Eorking Paper, No. 1594
Rao RKS (1982) The impact of yield changes on the systematic risk of bonds. J Financ Quant Anal XVII(1):115–128
Roll R (1970) The behavior of interest rates. Basic Books, New York
Smith C, Warner J (1979) On financial contracting: an analysis of bond covenants. J Financ Econ 7:115–161
Sundaresan M (1983) Constant absolute risk aversion preferences and constant equilibrium interest rates. J Finance 39(1):205–212
Svensson LEO (1994) Estimating and interpreting forward interest rates. NBER Working Paper, No. 4871
Van Horne J (1965) Interest-rate risk and the term structure of interest rates. J Polit Econ LXXIII:344–351
Van Horne J (1966) Interest-rate expectations, the shape of yield curve, and monetary policy. Rev Econ Statistics, XLVIII, pp 211–215
Vasicek O (1977) An equilibrium characterization of the term structure. J Financ Econ 5 (2):177–188
Zwick B (1980) Yield on privately placed corporate bonds. J Finance 35(1):23–30

Chapter 11
Conclusions

The book deals with the economics of value creation of companies and their measuring processes. In an ideal world, the investors know everything about a company and its managers. In this context it is reasonable to expect that the current share price is perfectly aligned with the intrinsic value of the company as defined on the basis of its fundamentals. Consequently, maximizing the current share price is equivalent to maximizing the value of the company over time.

Unfortunately, in the real-world investors only know something about the company and its managers. Their knowledge about the company is based on public information. Consequently, investors don't know what is really going on within a company, what decisions managers are making and their real effects on the economic and financial performance over time. It does not mean that the stock market is inefficient. On the contrary, markets do a great and fundamental job with public information. It only means that the market cannot price information that it does not have and then in the short-term a misalignment between current share price of the company and its intrinsic value based on its fundamentals is possible. However, in the long-term they tend to align.

Several approaches can be followed to deal with the economics of value creation of the company and its measuring process. This book is based on a *shareholder-oriented capitalism*. Consequently, the company can thrive only if it is able to create value for the shareholders over time.

In the period of the Great Recession as derived by the financial crisis of 2007–2008, this approach could be unpopular. Indeed, with common sense and in a part of academics the recent financial crisis is usually charged to the shareholder-oriented capitalism. Many parts requested and pushed for more regulations and for a deeper change in company governance and its focus by shifting from the shareholders' value to a set of stakeholders' value. This push is strong in Europe for economic and business culture.

In my opinion, the financial crisis cannot be attributed to a shareholder-oriented capitalism but to its distortion: the problem has been the passage from a long-term

P. De Luca, *Analytical Corporate Valuation*,
https://doi.org/10.1007/978-3-319-93551-5_11

perspective to a short-term perspective. Indeed, one of the most fundamental principles in the shareholders' value creation perspective states that the company's capability to create shareholder value in the long period is not the same as maximization of its short-term profits. Often the choice to maximize the shareholders' value in the long-term perspective is irreconcilable with the choice of maximizing the shareholders' value in the short-term period. Consequently, if the value creation in the long-term is confused with the profit in the short-term, it generates a great problem capable of damaging the shareholders' interests as well as the stakeholders' interests. Therefore, the main problem is not the shareholders' value but the short-term perspective of some managers. The best managers do not make decisions able to maximize the profit in the short-term by destroying the company's capability to create value in the long-term. Indeed, the main aim in the shareholder-oriented capitalism is to maximize a company value for a current, as well as, future shareholder. Also, the best managers known that the company's capability to create value over time in the long-term perspective implies the development of the company's fundamentals rather than accounting and financial make-up to increase the short-term profits by pushing-up the current share price. Finally, the best managers know well that the maximization of the shareholders' value implies the maximization of the stakeholders' value as well. Indeed, we should consider the three main categories of stakeholders: employees, customers, suppliers.

With regards to the employees, a company that attempts to increase the profits by reducing the employees' benefits, will have extreme difficulty in engaging and maintaining high quality employees. The lower quality of employees means products of a lower quality and therefore it results in lower customer satisfaction with lower revenues and profit for the shareholders. Today, one of the most important company's resource for its success in business is human capital. In the context characterized by strongly skilled employees, the capacity of the company to attract and to retain the best employees is often the key to success in the long-term. Therefore, in the shareholder's perspective the company's ability to attract and to retain quality employees on the basis of their satisfaction means high productivity that, in turn, results in higher value creation in the long-term. Only on the short term can the company produce profits against its employees. But it a myopic strategy that it leads to value destruction in the long-term and then to the company's failure.

With regards to customers, the company's ability to compete in the business is strictly functional of the alignment between products characteristics and customers' expectations. If the company wants to sell the current products as well as the future product generations of product, they must be in line with the customers' expectations in terms of its quality and price. Customers can be fooled once at the most. This strategy can, at the most, increase the revenues but only in the short-term. Anyway, it deeply damages the reputation of the brand with negative effects on the company's capability to create value over time. Consequently, the company's capability to create value over time for shareholders implies the value creation for customers also.

Finally, with regards to the suppliers, the main objective of the company is to create solid relationships with them in order to guarantee a high quality product.

Indeed, only in this case the supplies can be in line with product characteristics that, in turn, must be in line with customers' expectations. Whenever the company obtains profit by fooling the suppliers or by imposing harmful conditions on them, it damages the relationships with suppliers with negative and relevant effects on the quality of the product. It reduces the quality for customers with negative effects on the company's capability to create value for shareholders. Also in this case, the bad strategy to generate profits for the company by exploiting suppliers can be, in the best case, effective only in the short-term. In the long-period it destroys value for company and shareholders.

Based on these considerations, the shareholders' value creation is function of the value creation for employees, customers and also suppliers. The company's ability to create value for the current and future shareholders in the long-term implies and requires the same company's capability to create value for employees, customers and suppliers.

It is worth noting that the opposite is not true. Indeed, the maximization of employees' value (by high wages and benefits do not align with performance), customers' value (by an unjustified product price lower than its production costs), suppliers' value (by costs much more than their real value), reduce the company's ability to create profits over time and then it destroys shareholders' value.

The company's ability to create value for the shareholders over time is strictly related to the deep understanding of the business model of the company as well as the investors' behaviour in the capital markets. The company valuation can be considered one of the most relevant fields in which the classical paradigms of the company meets the paradigms of the capital markets. Indeed, the right company's valuation requires high competence in the fields of strategy, financial management, corporate finance and capital markets.

The basic equation of the value is based on a principle dating back to Alfred Marshall in 1890: company creates value if and only if, the return on capital invested exceeds its cost of capital.

The amount of value is equal to the difference between cash in-flows deriving from the investment and the cost of capital invested able to reflect the time value of money and the risk premium. Consequently, to create value over time, the company must invest the capital raised at a rate of return higher than its capital cost. Therefore, there are two main variables of value creation:

- the *return on capital invested*;
- the *cost of capital*.

In this book the company's ability to invest the capital raised by obtaining a high return, is investigated by the *analysis of the company's fundamentals* with regards to its business model and its economic and financial performances over time.

Otherwise, the company's cost of capital invested in its business is investigated through an *analysis of the risk-return profile of the company in the capital markets*, on the basis of investors' models about risk-returns.

Therefore, the models of company's fundamental analysis are integrated with the models used by investors in the capital markets to diversify risks and maximize their expected returns.

The integration between the company's fundamental analysis and the investors' risk models and returns in the capital markets is essential for the company's success over time. It is not possible to correctly understand the company's ability to create value over time and to measure this value without the simultaneous deep knowledge of these models and their integration.

The Return of Capital Invested

The return on capital invested in the business is function of the company's business model and the quantitative effects on its economic and financial dynamics. Specifically, the company's ability to create profit over time requires an analysis based on two main parts:

- the *qualitative analysis* of the business model;
- the *quantitative analysis* of the company's performance which regards to the effects of the business model choices on the economic and financial dynamics over time.

The ***qualitative analysis of the company's business model*** is the first step to further understanding the company's ability to create value over time. The qualitative analysis of the company's business model proposed in this context is called "*Company Strategic Formula*" (*CSF*).

The CSF defines the strategic profile of the company. Specifically, the CSF defines the way in which the company is organised internally and how it manages the relationships with external players for self-development over time. The CSF can be considered as the ideal conceptual place in which, on the basis of a systemic and dynamic paradigm: (i) the *ideas* are developed; (ii) the *decisions* are made; (iii) the *operations* are defined and planned. On the basis of a systemic and dynamic perspective, the CSF allows for the transformation of the "*system of ideas*" into the "*systems of operations*" by means of the "*systems of decisions*" in order to achieve and maintain economic-financial equilibrium over time.

The CSF defines the strategic profile of the company, by considering two different "*strategic fronts*":

- *internal strategic front*: it refers to the internal structure of the company;
- *external strategic front*: it refers to the structural relationships between the company and the players of its environment classified into two main groups of business players and financial players.

The *Internal Strategic Front (ISF)* refers to the internal structure of the company. It is defined from all elements, tangible and intangible, needed for the production of goods and services. The internal structure defines the company's specific characteristics by generating its uniqueness. It gives form and substance to the CSF by establishing the uniqueness of the thinking and operation of the company.

Specifically, the internal structure of the company is defined on the basis of three main elements:

- *Corporate governance*: it refers to the rules and the procedures by which the decision-making processes in the governmental area and how the managerial and operating activities of the company are defined;
- *Organizational architecture*: it refers to how the company's resources are combined and coordinated between them for company operations. There are two main levels involved: (i) the *organizational structure* of the company, with regards to both hard and soft elements that give form, substance and operation to all parts of the organization; (ii) the *operations*, with regards to the processes and procedures that cross the company vertically and horizontally;
- *Strategic resources*: it refers to the company's tangible and intangible assets and to the human skills necessary for their coordination. The company's strategic resources represent the most important way of competing in the business. Indeed, they provide the company with uniqueness and are able to protect its competitive advantage from imitation processes by generating "isolation mechanisms".

The *External Strategic Front (ESF)* refers to the structural relationships between the company and external players. Company competitiveness is due to its ability to create value for all of its players simultaneously.

The external players of the company can be classified into two main groups: (i) business players and (ii) financial players. Based on the differences in their nature, interests and behaviour, the external strategic front can be divided in two main parts:

- *Strategic Business Area (SBA)*: it refers to the real market in which the company carries out the business. The company can operate in more than one business. In any case, any strategic business area can be defined on the basis of two main elements: (i) *competitive players*, that refers to the players with which the company defines relationships. Specifically, they are customers, suppliers, competitors; (ii) *product system*, that refers to the product offered by the company with regards to its material and immaterial elements, service components and economic and non-economic terms. In each SBA the company competes by means of a defined business strategy in order satisfy customer requirements and expectations better than competitors. It allows the company to obtain a competitive advantage in the business and greater profitability than competitors.
- *Capital Market*: it refers to the financial markets in which the company looks for the capital, in equity and debt, needed for its survival and development over time. The capital market can be defined according to two main elements: (i) *financial players*, that refers to the investors in equity and debt; (ii) *financial company profile*, that refers to the risk-return profile of the company. It is function of the company's expected cash-flows on one side and investor expectations about risk and returns. In the capital market the company competes

through its financial strategy in order to acquire the capital needed, in equity and debt, at profitable conditions.

The internal and external strategic fronts are two parts of a whole. The success of the company is function of their joint quality. They are subsequently strictly connected by systemic and dynamic relationships. Therefore, the CSF must be characterized by a "*consonance*" between all structural elements of the internal and external strategic fronts. This consonance must be: (i) *Systemic*: all elements of the internal and external strategic fronts must be aligned between them; (ii) *Structural*: there must always be correspondence between the characteristics of each element of the strategic fronts, both internal and external, based on well-defined and structural bidirectional relationships; (iii) *Dynamic*: the systemic-structural relationships between elements of the internal and external strategic fronts must be dynamic over time and never static. Therefore, the CSF can be defined as "*consonant*" only if the relationships between all of its elements can be defined as *Systemic-Structural-Dynamic*.

The ***quantitative analysis of the company's performances*** requires an investigation into the economic and financial dynamics over time with regards to the past and the future. In order to simplify the comparison between the past and the future for the same company and between different companies over time, the same analytical schemes should be used. Several analytical schemes should be used. They are defined on the basis of the specific purpose of management according to the decision-making process.

In this context, the analytical schemes used are defined based on the financial approach to company assessment and they are defined in order to investigate the three main pillars:

– *Operating Income and Net Income*;
– *Capital Invested and Capital Structure*;
– *Free Cash-flow from Operations and Free Cash-flow to Equity.*

While the first defines the economic dynamic, the second and the third define the financial dynamics of the company.

Using the analytical schemes proposed in this context, the following should be borne in mind:

– they are defined with a view to the financial community rather than the accounting one. Therefore, they must not be confused with the analytical schemes used for balance sheet and income statement analysis and for definition of the classic accounting ratios. Moreover, terminology is not strictly based on the accounting rules;
– they are strictly connected between them. Therefore, the definition of each one is strictly related to the composition of each other;
– they are defined based on non-financial companies. Furthermore, they can also be used for financial companies after some changes in their structures;

– they are used to analyse the expected future economic and financial dynamics for an estimate of company value. Therefore, their application to past data is necessary to link the past and future in a coherent manner.

Furthermore, for greater understanding of the economic and financial dynamics of the company over time, past values should be aligned with expected future values. The alignment procedure between past and future values regarding Operating and Net Income, Capital Invested and Capital Structure, and Free Cash-flows from Operations and Free Cash-flows to Equity, as represented in the analytical schemes used, can be achieved by a procedure based on three main steps:

– *the first step, is the collection and recognition of past values*: the aim is to build Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity of the company in the past. For this objective, the analyses should be based on the statement of financial position, income statement and cash flows statement on the one side and on the internal management accounts of the company on the other. The combination of these two data sources allows for an analysis of the real conditions of the company. An analysis of the management accounts is necessary for three main reasons: (i) they are built to support management in the decision-making phase; (ii) they are characterized by both monetary and non-monetary quantitative data; (iii) they are well known in their composition and dynamics thanks to the technique of the variance analysis implemented constantly;
– *the second step, is the "adjustment" of past values*: the aim of this step is to obtain the "normalized" value of Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity of the company in the past. The aim of the process is to define these values in stand-alone conditions of the company. Therefore, their effects on extraordinary events in the broadest sense are not considered;
– *the third step, is an estimate process of value in the future*: the aim of this step is to build estimates on Operating and Net Income, Capital Invested and Capital Structure, Free Cash-flow from Operations and Free Cash-flow to Equity of the company in the future. A company business plan should be defined in order to achieve this objective. It is created by defining the Company Strategic Formula and by estimating its effects on future economic and financial dynamics.

Based on these three steps, the origin of the company, where it is and where it plans to go should be clear. Thanks to normalization of the past economic and financial dynamics their values can dialogue with those expected for the future. Consequently, it is easier to highlight the jumps between the past and the future and to evaluate whether or not they can be fulfilled in the future based on the strategies that will be implemented.

The qualitative and quantitative analyses are strictly related. The competitive advantage of the company, on the basis of its business model, must be reflected in the economic and financial values over time. Consequently, it is not possible to investigate into the company by only taking into consideration the analysis of its business model without considering the effects of the strategic choices on the economic and financial dynamics. At the same time, it is not possible to investigate into the company's ability to perform by considering the economic and financial dynamics without clearly understanding the source of the strategic choices.

In the analysis of these two parts jointly, three are the main caveats to keep in mind:

- *first, there must always be full consistency between the business model of the firm and its economic and financial dynamics over time*. The economic and financial analysis measures the quantitative effects of the business model on economic and financial dynamics with regards to the three dimensions of Operating and Net Income, Capital Invested and Capital Structure, and Cash-flow from Operations and Cash-flow to Equity. Therefore, while the analysis of the business model is a qualitative analysis, the analysis of economic and financial dynamics is a quantitative analysis. The two types of analysis cannot be separated and they are normally used together in the definition and assessment of the company's business planning. Consequently, an estimate of the expected economic and financial performance must be a coherent and consistent translation of the business model adopted by the company;
- *second, the future is the reference time*. The value of the company is function of its ability to generate value in the future. In an analysis of the past, with regards both to the business model adopted and the economic and financial dynamics that it has fulfilled, it is important to understand if the future expectations of the company, as defined in the business model to be implemented and in the estimation of the expected cash-flows, are really reasonable or unreasonable;
- *third, the assumptions are the key variable of the forecast*. The business model implemented and the estimation of the expected economic and financial dynamics are based on assumptions. Then, the quality of the forecast is function of the quality of the baseline assumptions.

  Generally, an analysis of the assumptions requires strict coherence or a reasonable relationship based on personal elements of the company or straightforward to acquire. In this sense, the reference assumptions should be clearly defined for each variable, also in their relationship with other assumptions. Each assumption must be individually reliable and coherent with each other.

In order to investigate the company's profitability it is necessary to investigate into the product's margin profitability. Indeed, the company's profitability is function of product's profitability. The most relevant problem concerns the right definition of the full cost of product in every moment of company life. It seems to be simple to define the product cost. Unfortunately, this simplicity is only apparent.

Indeed, the correct measurement and knowledge in each period the product cost allows us to answer some fundamental questions such as: how much did the product cost? What is the right price on the basis of production costs on the one hand, and market competition on the other? What is product margin and then its profitability?

The wrong determination of product cost is one of the main elements of the wrong prediction of product margin and, on a general level, company profitability.

The Cost of Capital

One of the more relevant problems in the valuation process is the risk. Indeed, the company's value is function of its future capability to create cash-flows higher than the cost of capital. Consequently, risk enters into a valuation process both through the cost of capital, that can be interpreted as the price of risk, and in the uncertainty future cash-flows.

The cost of capital for the company is one of the most relevant variables in the company's valuation models. It is probably one of the most relevant topics for managers and financial economists. For decades several studies focused on the relationship between capital structure, cost of capital and company value. Despite a broad experience approach in both academic and practices, it is not surprising that the method for estimation of the cost of capital is still under intensive discussion.

An estimate of the cost of capital for the company requires the investors' behaviour and expectations in the capital market. Unfortunately, it requires the knowledge of their models about the risk valuation and the expected returns estimation. The greater the managers' skill to understand the investors' behaviour and their choices, the greater the company's probability to satisfy the investors' expectations by acquiring the capital needed for its development at favourable conditions.

Consequently, the managers must define their strategies and operational processes by considering the business and industrial logics with regards to customers, suppliers, competitors as well as the financial criterion with regards to investors in equity and debt.

The theory of the choices under uncertainty leads to the decision-making process in capital markets. The aim is to analyse the behaviour of the rational investor under uncertainty. Specifically, the aim of the theory is not to define a set of criteria for the investor's preference for general validity because all investors are different from one another. Otherwise, the aim of the theory is to define a set of criteria of the decision-making process based on a few principles characterized by generality, rationality, economic significance, consistency with individual criteria, and therefore able to have a normative function.

In this regard, the theory defines the criteria by which the rational investor chooses between the real possible options, considering the restrictions, on the basis of the expected effects that could be achieved according to their nature and that can be sorted in consideration of the relative probability.

The portfolio choices (or portfolio selection) is a problem related to wealth allocation between different investment assets.

The mean-variance approach is the most widely used in the portfolio selections. The portfolio selection is based on two variables: (i) expected value of the portfolio return; (ii) variance of the expected portfolio return measuring the portfolio risk. An efficient portfolio must satisfy the Pareto optimal condition. Therefore, the investor prefers the portfolio that is capable of maximising its expected return to an equal variance or the portfolio capable of minimizing its variance to an equal expected return.

This approach simplifies the problem of portfolio selection. There are two main advantages: first, it does not require any specifications in terms of probability distribution; second, it is simple and intuitive because it is only based on the mean and variance. However, it is also true that this approach neglects a lot of relevant information about distribution probability.

The entire portfolio selection process can be simplified on the basis of two main phases of the portfolio selection process:

– *optimization phase*: the aim is to define the diversified portfolio and the efficient frontier. The definition of the diversified portfolio is based on the statistical characteristics of the assets. Specifically, the expected return of the portfolio is equal to the weighted average of the expected returns of the assets, while the portfolio variance is the function of the covariance between the assets' expected returns. The assumption refers to the investors' homogeneous expectations about the statistical characteristics of the assets implying that all investors define the same efficient frontier.
– *maximization phase*: the aim is to choose the optimal portfolio among the efficient portfolios defined on the efficient frontier. None of the efficient portfolios on the efficient frontiers can be preferred over the others by definition. The choice of the optimal portfolio among the efficient portfolios requires a clear definition of the investor's preferences about risk.

While the *optimization phase* is characterized by objectivity because it is valid for the entire market and not for the single investor, the *maximization phase* is characterized by subjectivity because it is the function of the investor's risk preferences.

Based on the portfolio selection theory, if the investor is diversified the cost of capital is function only for the risk that investor cannot diversify away. The unique risks that any company must face are not priced in the cost of capital. Consequently, the cost of capital used as discounted rate of future cash-flows of the company is not related to the risk link to the cash-flows uncertainty. The risk associated to the future expected cash-flows must be considered in their estimation.

One of the most popular models to estimate the company's cost of capital derives from the Capital Asset Pricing Model (CAPM). The CAPM is the most well-known equilibrium model in the capital market.

The standard form of CAPM provides a clear description of capital market behaviour if its basic assumptions are respected. There are two main problems.

First, it is based on strictly basic assumptions some of which are very far from conditions of reality. This is not a problem in itself. The fact that these differences from reality are irrelevant enough, they do not have a material effect on the model's explanatory power. However, during the years some different versions of the standard CAPM have been developed by changing specific basic assumptions. The aim is to understand and to explain the standard version of the CPM in greater detail, with the investor's behaviour on the one hand, and the assets price on the other hand.

Second, the CAPM describes the conditions of equilibrium about returns on the macro level. It does not describe this equilibrium on a micro level with regards to individual investor behaviour. Indeed, most investors and institutions have a risky assets portfolio different from the market portfolio.

Therefore, while the model can explain the capital markets behaviour as an entity, it is unable to explain the investors behaviour. Indeed, the investor's portfolio is usually different from the market portfolio.

On the basis of these considerations, the cost of capital of a specific company is estimated on the basis of Security Market Line (SML) as derived from the CAPM.

Discounted Cash-Flows Approach

Based on these two variables, return of capital invested and the cost of capital, the company's ability to create value over time for its shareholders is the function of the effectiveness of the Company Strategic Formula to create expected cash-flows as well as investors' models to diversify the risk and maximize expected returns in order to estimate the cost of capital.

The basic equation of value states that the company creates value if and only if, the return on capital invested exceeds its cost of capital. The explicit application of the baseline equation can be realized through several methodologies. Among them the Discounted Cash-Flows model (DCF) is the best. It is commonly used in the financial community. It is relevant since all members of the international financial community use a common criteria and language. The DCF preference by investors is based on several reasons. Among these, three are the main ones:

(1) the DCF is based on cash-flows. The cash-flows are preferred by investors who distrust of accounting measurements. Specifically, bondholders are interested in cash-flows from operations because the company's capability to pay interest on debt and to reimburse the debt depends on their amount. Otherwise, the shareholders are interested in cash-flows to equity because the company's capability to pay dividends depends on their amount and timing. Accounting earnings and value are not one and the same. Also, if earnings and cash flows are correlated, earnings do not tell the whole story of value creation. Then, focusing too much on earnings, and then earning growths, often it leads the company to stray from a value-creating path. Earnings do not drive value in their own right. Only cash-flows can be considered as the fundamental value drivers.

It is worth noting that cash-flows are the monetary dimensions of the company. In the capital markets only the monetary values can be moved over time by giving real relevance and fairness to the mechanism of actualization and capitalization. Indeed, the capital market has two main functions: first, it allows for the movement of money over time; second, it allows for the allocation of money between assets and future events. In this context the financial monetary costs as well as the relationship between risk and return are based only on the movement of the monetary value over time in order to consume and invest;

(2) the *value-creation principle* states that the company creates value only if the returns on capital invested are greater than the cost of capital. The DCF allows for rewriting of the basic equation between the return on capital invested and the cost of capital in terms of expected cash-flows over time and cost of capital used as discounted rate.
Specifically, it allows to focus attention on three main variables of the company's value: company's capability to create cash-flows (free cash-flows from operations and free cash-flows to equity) over time; the cost of capital; the time in the valuation process and then the possibility to calculate the present values. Therefore, the DCF is based on the present value rule enounced by Irving Fisher, father of the modern financial theory: the value of any asset is equal to the present value of the expected cash-flows discounted at an appropriate rate of riskiness. Consequently, the asset's value, both real and financial, is the functions of the expected cash flows that will be realized by the asset in the future, their distribution in time, and their uncertainty;

(3) the DCF is in line with *value-conservation principle* that is one of the most relevant corollary of the value-creation principle. It states that anything that does not increase cash-flows over time does not create value, regardless of whether the decisions made improved earnings or otherwise make their financial statements look stronger. Therefore, in investors' perspective anything that does not increase cash flows does not create value.
So value is conserved, or unchanged, when the company changes all also by changing accounting techniques, but it does not change the total available cash flows. Therefore, the cash flows are the key driver of the value creation and its measurement. Indeed, the value-conservation principle tells us what to look for when analysing whether the same decision or action will create value: the cash-flows impact and nothing else.

By using the DCF approach, the company value is equal to the current value of expected future cash-flows and the cost of capital is used as a discount rate. Therefore, there are three main variables:

- *Time*: the referenced time is the future. The value of the company is strictly related to future performance rather than past performance;
- *Cash-flows*: company performance is measured in cash-flows terms. Specifically, the expected future cash-flows from operations and to equity;
- *Cost of capital*: it is the cost of debt and the cost of equity and it defines the discount rate for expected future cash-flows.

The *General Equation of value* can be defined, based on these three main variables as follows:

$$W_F = \sum_{t=1}^{\infty} \frac{CF_t}{(1+k)^t} \quad (11.1)$$

where: W_F is the company's value; t is the period-time of valuation; CF_t is the expected future cash-flows for each year (t) (note that they refer to the expected value of cash-flows but in order to simplify the formalization the operator $\mathrm{E}[CF_t]$ is not used, by the meaning is the same); k is the cost of capital used as a discounted rate.

Equation (11.1) has a great theoretical relevance. It estimates the value of the company based on expected cash flows, arising from the fundamental analysis of the company and the cost of capital. Also the equation defines the relationship between company value, the expected cash-flows and the cost of capital in the time of valuation: the company's value increases together with an increase in the expected cash-flows and decreases together with an the increase in the cost of capital.

The general equation has a relevant theoretical importance but it is not applicable directly. There are two main problems to be solved before:

- the valuation time-period;
- the valuation perspective.

The first problem is definition of the *valuation time-period*. In Eq. (11.1) time goes from 1 $(t = 1)$ to infinite $(t = n)$. Therefore, Eq. (11.1) is not directly applicable. The problem can be solved by dividing the valuation time-period in two conceptual parts:

- *definite time-period*: it is the time period of *analytic valuation*. Generally, this time period is equal to 3 or 5 years on the basis of company characteristics and its market, and it defines the time period of the business plan;
- *indefinite time-period*: it is the time period of *synthetic valuation*. It goes from the end of time-period of analytic valuation to infinity by using the Terminal Value (TV). Generally, the Terminal Value measures the company's value after the analytic valuation.

By distinguishing between the analytical valuation and the synthetic valuation, Eq. (11.1) can be rewritten as follows:

$$W_F = \sum_{t=1}^{n} \frac{CF_t}{(1+K)^t} + \frac{TV_n}{(1+K)^n} \quad (11.2)$$

where TV_n indicates the Terminal Value at the end $(t = n)$ of the period of the analytical valuation.

Therefore, while the first part of the equation estimates the company value in a given time period (analytic value of the company), the second part of the equation estimates the company value in an indefinite time period by using the Terminal Value (synthetic value of the company).

The second problem is the *valuation perspective*. Application of Eq. (11.1) requires the definition of its variables: identification of the expected cash-flows to be discounted and identification of capital cost used to discount the expected cash-flows. The solution of the problem requires the definition of the valuation perspective. They could be two perspectives:

- *Equity Side perspective*: the Equity Value (W_E) of the company is estimated on the basis of Free Cash-flows to Equity (FCFE) discounted at the Cost of Equity. Use of the Cost of Equity (K_E) is due to the nature of the free cash-flows to be discounted: they are the residual cash-flows after the coverage of the company's needs and the debt obligations and destined to equity remuneration;
- *Asset Side perspective* the asset value called Enterprise Value (W_A) is estimated. The Enterprise Value is estimated based on the Free Cash-flow from Operations (FCFO) discounted to the Cost of Capital. Use of the Cost of Capital Levered, including both the cost of equity and the cost of debt, is due to the nature of the free cash flows to be discounted. In fact, these cash flows derive from the operating activities of the company and they are used in remuneration of both equity-holders and debt-holders. Therefore, the Enterprise Value is the value generated by the company's operating activities and they must be distributed among the investors in equity and debt.

It is possible to summarize as follows:

$$W_F = \sum_{t=1}^{n} \frac{CF_t}{(1+K)^t} + \frac{TV_n}{(1+K)^n} \rightarrow \begin{matrix} Equity\,Side \rightarrow W_E = \sum_{t=1}^{n} \frac{FCFE_t}{(1+K_E)^t} + \frac{TV_n}{(1+K_E)^n} \\ Asset\,Side \rightarrow W_A = \sum_{t=1}^{n} \frac{FCFO_t}{(1+K_A)^t} + \frac{TV_n}{(1+K_A)^n} \end{matrix} \tag{11.3}$$

There are several models to estimate the Equity Value and the Enterprise Value. Indeed, by changing the cash-flows aggregated several configurations of the model can be obtained. In this context, following a financial approach, the *Equity Value* is estimated by using: (i) Dividend Discounted Model; (ii) Free Cash Flow to Equity Discounted Model; (iii) Multiples on Equity Value.

Similarly, the *Enterprise Value* is estimated by using: (i) Free Cash Flow from the Operations Discounted Model based on Cost of Capital Approach and on Adjusted Present Value Approach; (ii) Discounted Economic Profit; (iii) Multiples on Enterprise Value.

In both cases, Equity Value and Enterprise Value, the models are defined on the basis of three scenarios: (i) *Constant growth (or one-period model)* that assumes a constant growth over time indefinitely; (ii) *Two-Stage Growth (or two-period model)* that assumes an initial period characterised by extraordinary growth and a second period characterised by a steady-state growth rate expected to continue indefinitely; (iii) *Three-Stage Growth (or three-period model)* that assumes a first period characterized by growth constant at the same level, a second period characterized by a changing growth from its level in the first period to a long-run steady-state level and, a third period characterized by a constant growth indefinitely. Obviously, it is possible to use others steps by passing from the three-period model to an n-period model. Generally, by moving from the one-period model to three-period model (or n-period model), more information is required in terms of quantity and complexity. Also, more variables imply more complexity in the forecast process. Otherwise, the use of few variables implies a high level of simplicity but a low level of confidence about the value estimated. The trade-off between complexity and manageability must be solved on the basis of the information available on the company and the analyst's forecasting skills.

Finally, it is relevant to note that the large use of a rigorous quantitative analysis to integrate the company's fundamental analysis and the investors' models about risk and return in the capital markets in order to the company valuation, is not to complicate the analysis but, on the contrary, to simplify the discussion. There are three main reasons:

- first, models are easier to understand if they are studied in their formal construction. The mathematical form allows us to further understand the models in their construction, assumptions, and then, in their clear capabilities and limits;
- second, the mathematical form does not allow inappropriate manipulation of the equations and, consequently, uncorrected use of the models. Every equation is the result of a rigorous formal process and their modification can be realized only by following the same rigorous formal process;
- third, the mathematical form does not allow for attribution of the equation meanings that are not supported by their strict formal derivation. Every equation acquires form and meanings strictly related to the mathematical process of derivation. The clear derivation step-by-step of each equation does not allow errors in the equations interpretation and, consequently, in the use of models.