Department of
Business and Management

Chair of Advanced Corporate Finance

Firm valuation methods: how to value a firm. Focus on Beta Equity Determinants, a derivation through Machine Learning on Python.

Supervisor:

Prof. Pierluigi Murro

Co-supervisor:

Prof. Rosella Santella

Candidate:

Michele Bigonzetti

759841

Academic Year 2023/2024

Index

Introduction 4

Chapter One: Firm Valuation Methods 7

1.1 Capital Budgeting, NPV, and Free Cash Flow 8

1.2 Dividend Discount Model (DDM) 17

1.3 Valuation Multiples Method 23

1.3.1 Comparables Method (Comps) 30

1.3.2 Comparables Precedent Transaction Method (Precedents) 34

1.4 Discounted Cash Flow Method 39

1.5 Valuation with Leverage 43

1.5.1 WACC method 46

1.5.2 Adjusted Present Value (APV) Method 48

1.5.3 Flow-to-Equity Method 52

Chapter Two: Cost of Capital, a focus on Beta Equity 55

2.1 The Cost of Debt 56

2.2 The Cost of Equity (Listed Companies) 62

2.3 The Cost of Equity (Not Listed Companies) 80

2.4 Focus on Beta Equity 85

2.4.1 Determinants of Beta Equity 86

2.4.2 Correlation between Determinants and Beta Equity 95

Chapter Three: EDA and the derivation of Beta Equity through Machine Learning on Python 104
3.1 EDA on over 25,000 stocks 106
3.1.1 Data Cleaning 107
3.1.2 Graphical Representation of the Data 114
3.1.3 Statistical Representation of the Data 120
3.2 Development and computation of a Machine Learning code 121
3.2.1 Adjusting the Data for the Machine Learning code 122
3.2.2 Modelling the Data and Splitting in Training and Testing Sets 124
3.1.3 Creating a XGBoost Classifier and fitting the Model 126
3.3 Apply the Model to Practical Examples 128
Conclusions **132**
Bibliography **136**
Sitography **138**

Introduction

When an investor, a company, an academic or whoever wonders what the value of a certain target company is, it becomes necessary to find a way of measuring its performance and its real price. This task is not so easy, and it is frequently delegated to some experts that do all the "dirty job"; still, it has to be done and, even if the process has not to be performed by anyone, it would be way more efficient if everyone interested in the company's value would be aware of how it works. In this sense, this paper is not addressed only to those ones that have to do the "dirty job" but, indeed, it wants to be appreciated by people that are genuinely interested in the subject too. Still, the topics covered will be in some cases more difficult but, in the end, the conclusion should help to better explain some key parts that are usually less intuitive.

Especially it will be demonstrated if the Beta Equity (one of the components necessary to compute the Cost of Equity) is affected by some factors, defined in the paper as "**Determinants**", that will help to further understand the company value and eventually predict Beta Equity values on privately held companies.

Additionally, the existence of those Determinants will help in better recognising risky companies from non-risky ones. In fact, as it will be deeply covered at the end of the second chapter, it will be of primary importance to demonstrate the correlation between Beta and riskiness. Such correlation permits to definitely guarantee that the Determinants are able to analyse and predict a company risk exposure.

Furtherly, the possible demonstration of the existence of Beta Equity Determinants can even help to change and affect the companies by addressing the factors towards a lower or higher Beta Equity.

In the following pages, as starting point, it will be introduced the Capital Budgeting, it serves as introduction to what will be really presented in the first chapter, **how to value a company**. Capital Budgeting, in fact, can be really useful because it is usually considered when it comes to value a project. Company Valuation shares a lot of key factors with Capital Budgeting; for this reason, one of the most important valuation methods, the Discounted Cash Flow Method, is particularly similar to Capital Budgeting.

The Valuation Methods are divided in three groups:

- Valuation methods using the dividend (**Dividend Discount** Model);
- Valuation methods using the cash flows (**Discounted Cash Flow** Method, **WACC** Method, **Adjusted Present Value** Method, and **Flow-to-Equity** Method);
- Valuation methods using the multiples (**Comparables** Method and **Comparables Precedent Transaction** Method).

After having extensively covered the Valuation Methods, the focus of the research moves to the discounting factors. These constitute the rates at which the cash flows (if it used the Discounted Cash Flow) or the dividends (if it is used the Dividend Discount Model) are discounted and, as it will be showed, their value has an incredible impact on the firm price, even a 1% change can completely transform the perception about a company.

The most important discount factor is the cost of equity which, as it can be easily understood from the importance given to the firm (equity) valuation, is the one that, when it comes necessary to discount something, is always present (the cost of debt is used when it comes to evaluate the firm as a whole, comprised the debt). Still, the second chapter will not only focus, in general, on the cost of equity but also on the Beta Equity which, along with the Free Cash Flow prediction, represents one of the elements of major uncertainty.

This paper’s research on Beta Equity, after having explained how it is usually estimated, focuses on what causes a certain company to have a high Beta Equity and what causes another company to have a low Beta Equity. After having looked at a number of academical studies it was clear that this topic, if properly managed, and furtherly explained, would have had solid grounds. Since the literature is quite fragmented and generical about this topic, the end of the second chapter will discuss how the Beta Equity is influenced by the so-called “Determinants”, the elements that cause a Beta Equity to be higher or lower (or equal) than one.

Therefore, **this paper wants to conjugate the already existing research making it possible to define a general theory about Beta Equity Determinants** and, if possible, even **reinforce it through a series of statistical and data driven analysis** that will be exhibited in the last chapter.

In the third and last chapter, because deemed necessary, it will be explained and discussed the creation of a DataFrame in Python that will be able to (i) show if there really exists a correlation between the Beta Equity and the Determinants and (ii) create a Machine Learning algorithm that, through the grouped data and the intercurrent relationships between the Beta Equity and the Determinants, should be able to describe a pattern and predict the Beta Equity of privately held companies.

Everything done during this paper, especially the last part, has to be considered as partially limited by the instruments available; in this sense, for example, the DataFrame would have been way more detailed and larger if they were used stronger computational systems.

The conclusion will try to sum up everything said during the paper.

Chapter One

Firm Valuation Methods

The purpose of this chapter is to capture the main, and most frequently used firm valuation methods. In general, valuation can refer to projects or firms, in here will be shown the different techniques that practitioners use to derive the effective price of a company. Still, it will be introduced the definition of NPV (Net Present Value) and capital budgeting applied to projects, this way it will be possible to highlight the key concepts of how financial valuation works and, most of all, of how Free Cash Flows are computed. Although it is possible to define which one of these techniques is the most used practically, it can't be proved if there is a best choice since it depends on several factors and conditions. It will be described the Dividend Discount Model and the Valuation Multiples Method (divided in Comparables Method and Comparables Precedent Transaction Method). Then, the focus will move on to the Discounted Cash Flow Method which sets its grounds on the Free Cash Flows with the hypothesis that leverage (Debt to Equity Ratio) is constant.

Finally, the chapter concludes by considering the valuation with leverage, this is necessary to describe how it can address the problem of a variable (during the years) leverage ratio. This last part permits to realize the completion of the set of possibilities that characterize a valuation; it will be evident how the amount and type of information about a company can determine a different method and how in certain situations each method can determine a different result. The computation of the discounting factor (cost of capital), even if mentioned a few times, will be covered in detail in Chapter 2.

1.1 Capital Budgeting, NPV, and Free Cash Flow

The importance of capital budgeting is evident and derives from a simple intuition: a company is identified as the sum of the initiatives, especially the projects, that it develops. Each project will have initial or ongoing costs and revenues that characterize the future of the firm, the more those projects are efficient and effective the more the company will benefit and increase its value. Since the firm can be seen as the sum of its projects, it is evident the importance of properly representing those costs and revenues adequately; it is also important to well define when those projects are profitable and when, on the contrary, they constitute a loss.

Before introducing the NPV it's compelling the definition and introduction to capital budgeting. Capital budgeting is defined as the process through which a company forecasts the earnings and costs of an investment opportunity and, through an investment rule, chooses the best option. This sort of analysis permits to understand and predict (obviously through assumptions) if the project will result in a loss or in a gain and, in case of multiple choices, opt for the most profitable one.

The common literature and practice determines a profitable project by using as an investment rule the Net Present Value[1] (NPV). The NPV sets its ground on the basics of financial mathematics, specifically the concept of discounting. It can be said that the value of one euro today is different from one euro in one year because of the effect of time on money (the so-called "time value of money"). If, for example, a bank lends money, it is usually expected that the mortgage payment is not equal to the exact value of the money lent divided by the number of periods in which it is registered a payment. On the contrary,

[1] Widely used is also the IRR which, substantially, is the same but valuates the projects from a different perspective; in fact, while the NPV focuses on determining the profitability by looking at the difference of actualized future flows, IRR focuses on understanding what is the internal rate of return, and in case of multiple projects which one has the highest return.

the borrower is expected to pay something more due to the cost of lending, differently said, the **opportunity cost** of the bank which could have invested those money on a different, fruitful, project. What is important in understanding the opportunity cost of capital is the risk and term aspect; the opportunity cost is equal to the return of an investment opportunity with similar risk and similar maturity that is offered in the market. Investors are like a bank; they invest in a company by buying shares and expecting a certain return[2], that return must be aligned with alternative opportunities with a **similar risk** (purchasing a government bond, depositing the money in a restricted bank account or investing in a company with a similar profile) and **maturity**. The idea of opportunity cost of capital becomes clear with the description of NPV through an example:

An Investor can decide to invest his money (10,000.00 Euros) in two distinct business opportunities. In both cases, the maturity is 5 years, but the inflows and discount factors are different:

- *in the first case the investor will receive each year 2,400.00 Euros with a discount factor of 5%;*
- *in the second case the investor will receive each year 2,500.00 uros with a discount factor of 10%.*

Opportunity A		
Cash Flows	2,400.00	
Maturity	5	years
Discount Factor	5%	
Cost	-10,000.00	

T	0	1	2	3	4	5
Inflows		2,400.00	2,400.00	2,400.00	2,400.00	2,400.00
Outflows	-10,000.00					
Total	**-10,000.00**	**2,400.00**	**2,400.00**	**2,400.00**	**2,400.00**	**2,400.00**
Present Value	-10,000.00	2,285.71	2,176.87	2,073.21	1,974.49	1,880.46
NPV	390.74					
IRR	6%					

Opportunity B		
Cash Flows	2,500.00	
Maturity	5	years
Discount Factor	10%	
Cost	-10,000.00	

T	0	1	2	3	4	5
Inflows		2,500.00	2,500.00	2,500.00	2,500.00	2,500.00
Outflows	-10,000.00					
Total	**-10,000.00**	**2,500.00**	**2,500.00**	**2,500.00**	**2,500.00**	**2,500.00**
Present Value	-10,000.00	2,272.73	2,066.12	1,878.29	1,707.53	1,552.30
NPV	- 523.03					
IRR	8%					

[2] The return in this case can take the form of both distribution of dividends and increase in the price of the stock (capital gain).

From this example, it is possible to capture two main conclusions[3]. First, to compute the NPV it is necessary to discount the future flows and then add the initial investment, this way it is possible to correctly value everything in the present considering the time value of money. Second, even if two investment opportunities have the same initial cost and maturity but different cash flows, it is not certain that the one with a higher payout is the best choice since the one with higher payout can have a certain level of risk that its payout is not sufficient to remunerate ("it's not worth the shot").

Now, it is possible to formalize the NPV formula:

Formula 1: Net Present Value

$$NPV = \sum_{n=1}^{N} \frac{C_n}{(1+r)^n} - Initial\ investment$$

It can be understood that, if $NPV > 0$[4] the project is profitable and has a positive value, otherwise, if $NPV < 0$, it is better to not invest because the project will result in a non-convenient choice. What plays a fundamental role is the discounting factor, and this is evident, especially looking at the example above. If the discounting factor is large, it means that the investor expects a certain return that is coherent with a greater risk of the project. This explains why in the previous example Opportunity B had a lower NPV; in that case, the discount factor is double but the inflows, even if larger, do not justify and assure a proper reward for the additional risk.

The presence of different choices with the same risk usually permits to create a portfolio in which the diversification reduces the firm-specific (idiosyncratic) risk[5], this is one of

[3] The third one refers to the calculation of IRR. Since IRR identifies the internal rate of return, its calculation determines what is the maximum discount factor, in other words, the highest possible value that the discount factor can have if the NPV must be positive. This, in fact, explains why if the discount factor is higher than the IRR, the NPV will be negative and, if the IRR is higher than the discount factor, the NPV will be positive. IRR can be used as a different decision rule in presence of multiple projects, the highest one is the best. NPV is preferred because is more intuitive and immediately can declare if a project is worth the risk.

[4] In that case, the cost of capital (or discounting factor) will be defined as the "hurdle rate" or minimum acceptable rate of return.

[5] Markowitz model. The investor can create a diversified optimal portfolio in which the overall risk is lower than the sum of the singularly taken assets. This effect is the result of the correlation among assets which determines the possibility to eliminate the firm-specific risk (not the systematic one because cannot be diversified).

the reasons that explain why an investor would have a preference among investment opportunities with similar risks. Differently, the choice among different projects with same NPV but different risk usually depends on the risk exposure: if an investor is a risk taker (embraces the risk) will opt for investment opportunities with high returns, otherwise, they will be risk averse and will choose risk-free opportunities.

Now that it is clear how NPV works, it is possible to move on to Capital Budgeting.

A capital budget lists the projects and investments that a company plans to undertake during the coming year. To determine this list, firms analyse alternative projects and decide which ones to accept through a process called Capital Budgeting[6]. The purpose of implementing capital budgeting is to analyse the effect of a project on the company's cash flows, this is done through the NPV which evaluates if the project has a positive or negative impact on them. The rule of thumb in capital budgeting is to always look at ***incremental*** inflows or outflows. This rule is in line with the purpose of capital budgeting; if the management wants to know the incremental or decremental cash flows resulting from a company project it is logical to think that will be considered only those costs or earnings which are incremental (or decremental) to the already existing company operations. Let’s consider the situation in which an automotive company is already producing a certain car in a certain segment and decides to invest in R&D to produce a new model of that car to substitute the already existing one. The revenues resulting from that project will not be considered entirely as incremental earnings since the number of sales lost if the car company is willing to maintain the current model represent a cost (or cannibalization cost). In this case, the incremental earnings are identified as the difference of earnings between selling in the following years of the new model and the earnings that would have been registered in case the model would have not been substituted. This way of dealing with costs and revenues help to understand if the project investment generates sufficient cash flows.

Along with costs and revenues during the life of the project in capital budgeting, it is necessary to consider also different other elements:

[6] Jonathan Berk, Peter DeMarzo, “*Corporate Finance, Global edition*”, Pearson Education, 2020.

- ***Capital Expenditures*** (CAPEX) are those funds used (usually initially) to undertake a project by acquiring, upgrading, and maintaining physical assets. Even if they represent a cash expense, they are not included in the calculation of earnings;
- ***Depreciation*** regards the physical consumption of CAPEX during the years and is, therefore, strictly correlated to the useful life of a good. There are several methods to compute the depreciation, three are the most used: the *straight-line method*, the *accelerated method*, and the *units-of-production method*. The most intuitive one is the **straight-line method** which divides equally the cost of an asset by its useful life. In the **accelerated method** (Modified Accelerated Recovery System or MACRS depreciation) the allocation of costs is concentrated in the first years. In the **units-of-production** method, finally, the amortization quota corresponds to the actual use of the asset in a certain period. Even if it is considered a cost, depreciation doesn't impact or represent a real cost since it is a non-cash expense; in fact, depreciation is inserted in the income statement outside COGS and operating expenses[7]. Because of the depreciation tax shield, companies prefer to opt for a depreciation method that accelerates the depreciation in the first years, this way the depreciation tax shield will be higher (the discount is way lower because most of the depreciation is discounted in the first years);
- ***Interest expenses*** are not included in capital budgeting since they arise from the corporate choices of the project financing decision. This explains why, in capital budgeting, net income is referred to as *unlevered net income*;
- ***Taxes*** represent the final cost in a company income statement. The tax rate is the marginal corporate tax rate, which is applied to the incremental dollar of pre-tax income. It is relevant to the specific case in which a company closes the year with pre-tax losses or Net Operating Losses (NOL), in that case, the taxes of the year will assume a positive sign, as to be intended as a gain or will be managed according to corporate laws in the state in which a company operates[8]. Also, in

[7] The depreciation of assets has a beneficial effect in the income statement since it is used as a subtraction from taxable income, this means that the amount of taxes will be lower. This effect is defined as "depreciation tax shield".

[8] Generally, the tax deduction is managed through the so-called "Tax Carryforwards" which implies the reduction of future positive Pre-tax Income to reduce the payment of taxes. If the NOL is higher than the following year Pre-Tax Income the remaining part of Tax Carryforward will be used as deduction in the

this case it is evident how capital budgeting intends to consider only incremental losses or gains (in this case, thanks to the loss, the company will profit from a tax deduction).

This first set of voices constitutes the heart of a capital budgeting representation, nonetheless, it is necessary to specify those other elements that contribute (indirect effects) or not to a capital budgeting decision.

- ***Opportunity Costs*** represent a missed revenue caused by the utilization of the asset in the new project. Sometimes a project involves the utilization of an asset (warehouse, machinery, land) that would have generated revenue if used in an alternative situation (for example a warehouse could be rented for money), this is inscribed as a cost. The idea is that, as anticipated at the beginning, the earnings derived from the project must account for only the incremental effect on the company, lost revenues are reduced earnings in the final computation of the overall project;
- ***Project Externalities*** are defined as opportunity costs, they represent indirect effects caused by the new project towards other, already in place, projects. Those effects can be beneficial or not and are inserted in the capital budgeting decision. A typical, negative, project externality is constituted by the eventual entrance of a new product that can reduce the sale of already existing ones[9], in that case, it will be necessary to account for the reduction in sales and, therefore, reduce the earnings of the new project by that amount;
- ***Sunk Costs*** do not imply any effect on a capital budgeting cash flow since they represent a cost unrecoverable and independent from the new project even if necessary for its success. Sunk costs are all those expenses that will occur even if the project doesn't take place, therefore, are not included in incremental earnings. The most frequent sunk costs are (i) fixed overhead expenses (only additional ones), (ii) past R&D expenditures, and (iii) unavoidable competitive effects[10].

---

following years until it will be completely consumed and it will be kept in the balance sheet through the indication "Deferred tax asset".

[9] This effect is usually referred to as "cannibalization".

[10] it happens that the entrance of a new product doesn't damage itself the sale of the older ones which would equally suffer from a decline caused by ageing and the entrance of new competitors.

All those considerations are finalized at defining the unlevered net income, necessary for the computation of the Free Cash Flow. Below is represented the simplified formula:

Formula 2: Computation of Unlevered Net Income

$$Unlevered\ Net\ Income = EBIT\ \times (1 - \tau)$$

The incremental earnings forecast, shown more broadly, is summarized in the following table:

Incremental Earnings Forecast
Sales
- Costs
EBITDA (Gross Profit)
- Selling, General and Administrative
- R&D
- Depreciation
EBIT
- Income Tax
Unlevered Net Income

So far, the analysis was limited to how it is measured the firm performance; now it is time to go a little further and determine how real profits, the cash, stream into the firm bank account. The amount of cash available to the firm is computed through the calculation of Free Cash Flow which better represents a project performance and inevitably substitutes Unlevered Net Income in the computation of a capital budgeting decision. Free Cash Flow differs from the Income Statement in two ways. First, the income statement considers some costs that do not result in cash outflows (amortization and depreciation). Second, the income statement does not consider certain expenses, especially those referring to

CAPEX and inventories. For these reasons, Free Cash Flow presents the following additional voices:

- Capital Expenditures
- Depreciation
- Net Working Capital

Capital Expenditures, which are not considered in the Income statement, in the Free Cash Flow they are immediately subtracted from the Unlevered Net Income because they constitute a cash outflow.

Depreciation, on the other hand, doesn't represent a real cash outflow but is inserted in the income statement; for this reason, it is added back into the Free Cash Flow calculation.

Net Working Capital ("**NWC**") is generally defined as the difference between current assets and current liabilities, specifically, it is composed of the following:

Formula 3: Net Working Capital

$$\begin{aligned} Net\ Working\ Capital &= Current\ Assets - Current\ Liabilities \\ &= \boldsymbol{Cash + Inventory + Receivables} \\ &+ \boldsymbol{Other\ Current\ Assets}^{11} - \boldsymbol{Payables} \\ &- \boldsymbol{Other\ Current\ Liabilities}^{12} \end{aligned}$$

By looking at its composition it becomes immediately understandable why it is so important. Two are the main reasons for which NWC is relevant:

1) **Liquidity Metric**: A positive NWC indicates that the company has more current assets than current liabilities and, in this sense, it means that it has sufficient resources to meet its short-term obligations.

[11] Among the Other Current Assets, they are recognised:
- Prepaid Expenses (Goods or Services paid but not yet received by suppliers);
- Accrued Income (Goods or Services delivered but not yet paid by customers);
- Tax Receivables.

[12] Among the Other Current Liabilities, they are recognised:
- Deferred Income (Goods or Services sold but not yet delivered to suppliers);
- Accrued Liabilities (Goods or Services received but not yet paid to suppliers);
- Tax Payables;
- Pension and Social Security Payables.

2) **Operational Efficiency Metric**: a change in NWC can be the result of a change in the company's operational efficiency. An increase in receivables could be caused by a change in operational policy that extends the expiration date for the reception of the money derived from goods or services sold. An extension of the deadline could cause huge issues in cash flow results. Similarly, an increase in inventories could be caused by production inefficiencies or the presence of excess inventory at hand.

Companies must operate by assuring a certain level of cash to cover unexpected losses, in the same way, they have to guarantee always ready inventories to satisfy consumer demand. The difference between Receivables and Payables is defined as trade credit and determines if a company is a net creditor or debtor, it is necessary to keep a certain balance between them because, otherwise, the company would be too exposed to short-term debt or would not take full advantage of payables and their useful effect on company profits and cash management. The relevant thing for Free Cash Flow Analysis is not by itself NWC, but its variation, since an increase in NWC would represent an additional investment and the reduction of available cash, practically a reduction of the cash flow. In the calculation of Free Cash Flow, an increase in NWC (i.e., current assets increasing more than current liabilities) is typically considered a use of cash because it represents an increase in working capital that ties up cash that could otherwise be used for other purposes. Conversely, a decrease in NWC (i.e., current liabilities increasing more than current assets) is considered a source of cash because it frees up cash that was previously tied up in working capital. In the end, the Free Cash Flow will be the result of the following equation:

Formula 4: Free Cash Flow

$$Free\ Cash\ Flow\ (FCF) = Unlevered\ Net\ Income + Depreciation - CapEx - \Delta NWC$$

This formula can be extended to understand the real purpose of Depreciation:

$$FCF = (Revenues - Costs - Depreciation) \times (1 - \tau) + Depreciation - CapEx - \Delta NWC = (Revenues - Costs) \times (1 - \tau) - CapEx - \Delta NWC + \boldsymbol{\tau} \times \boldsymbol{Depreciation}$$

Now it is evident that Depreciation is used as an advantage that increases the Free Cash Flow, for this reason, the last term is defined as "**Depreciation Tax Shield**".

1.2 Dividend Discount Model (DDM)

The description of the different valuation methods starts with an initial warning: for what concerns the discounting factors used in the different paragraphs it will be just defined which type is necessary without entering in in-depth considerations; the cost of capital will be fully addressed in Chapter 2.

The first valuation method is the Dividend Discount Model, in this paragraph the reader will find its main features and characteristics along with its multiple adaptations in different situations. The importance of this model can be intuitively explained through the general meaning of the Investment Decision Rule (applied to share purchases) and the Law of One Price.

When an investor decides to invest in a company it will for sure try to understand what the expectations and the performance of that company are. In general, investors seek two types of gains: ***dividends*** and ***capital gain*** (the difference between selling price and purchasing price). For this reason, when computing its decision rule, an investor decides to effectively determine the price of a stock by looking at its future return; this return can be seen, first, in a one-year perspective:

$$V_0 \leq \frac{Div_1 + P_1}{(1 + r_E)}$$

The value, in zero, is equal (at max) to the next year's sum of dividends and the price of the company discounted by the cost of equity. This value comes from the buyer's perspective, in the same way also the seller will consider that value as the minimum. The

Law of One Price[13] in this case aligns the buyer's perspective with the seller's and concludes that the firm stock price should be:

Formula 5: Firm Value according to the Law of One Price

$$V_0 = \frac{Div_1 + P_1}{(1 + r_E)}$$

This way the Law of One Price stands. From this formula it is clear the composition of the cost of equity (or total return of the stock):

$$r_E = \frac{Div_1 + P_1}{V_0} - 1 = \frac{Div_1}{V_0} + \frac{P_1 - V_0}{V_0}$$

r_E resembles the investor expectations about dividends and price movement, it is composed of Dividend Yield and Capital Gain Rate (respectively first and second term on the right side of the equation).

Formula 5 can be extended to a multi-year horizon; precedingly, it was considered the expectation of next year's results while to be more precise and better customize the firm valuation, the future expectations can be extended to two years in the future:

Formula 6: Firm Value according to the Law of One Price (two-years extension)

$$V_0 = \frac{Div_1}{(1 + r_E)} + \frac{Div_2 + P_2}{(1 + r_E)^2}$$

The pattern is clear, for each new year it must be discounted the new dividend and if it is the last year also the value of the stock in that year. Let's be clear about that $\boldsymbol{P_2}$, which is nothing more than a $\boldsymbol{P_1}$ calculated as V_0 but in year 2:

$$P_1 = \frac{Div_2 + P_2}{(1 + r_E)}$$

And therefore:

[13] The Law of One Price states that the price of identical goods in different markets (in this case from different perspectives of seller and buyer) must be equal, otherwise there will be arbitrage opportunities. The adaptation to real world cases attributes the small differences in prices to transaction costs, that's why sometimes the same asset can have two different prices.

$$V_0 = \frac{Div_1 + P_1}{(1 + r_E)} = \frac{Div_1}{(1 + r_E)} + \frac{P_1}{(1 + r_E)} = \frac{Div_1}{(1 + r_E)} + \frac{\left(\frac{Div_2 + P_2}{(1 + r_E)}\right)}{(1 + r_E)}$$

Now that it is evident how this model works, the **Dividend-Discount Model Equation** can be generalized:

Formula 7: Dividend Discount Model Equation

$$V_0 = \frac{Div_1}{(1 + r_E)} + \frac{Div_2}{(1 + r_E)^2} + \cdots + \frac{Div_N + P_N}{(1 + r_E)^N}$$

Even if it has been defined the general formula, still it is difficult to address the main problem: how it is computed the last term? The P_N value remains a question mark, even if you can project future dividends you will arrive at a certain point at which you will need to evaluate the company (the same problem as at the beginning). To address this issue, common practice has defined two different ways of calculating P_N:

- **Constant Dividend Growth Gordon Model**
- **Dividend-Discount Model with Constant Long-Term Growth**

Both models take into consideration the future growth of the company earnings and are strictly dependent on the assumptions made on it. Usually, its value is computed through a formula:

Formula 8: Earnings Growth Rate

$$Earning\ Growth\ Rate = g = Retention\ Rate^{14} \times Return\ on\ New\ Investment$$

and represents the firm's sustainable growth rate. In other cases, more practical ones, the value of the growth rate is not computed but taken as an average value of a certain sector. Small variations of the growth rate have a huge impact on the company's estimated value, and, for this reason, it has to be calculated by considering several factors and adaptations.

The importance of the growth rate is maximal when evaluating a firm through the Gordon Model, it is composed as follows:

[14] The fraction of earnings that the company doesn't distribute as a dividend. The multiplication of the retention rate by the earnings equals the new investment.

Formula 9: Gordon Model

$$V_0 = \frac{Div_1}{r_E - g}$$

This model is simple and has a high possibility of underestimating or overestimating the firm value and presents different shortcomings. First, the only factor that really can be more certain, the dividend, has the lowest impact on the overall valuation. Second, it is strange that a company grows at the same rate forever, usually it has different stages and periods (at early stages of life the growth rate is way higher while at maturity the growth is minimal). Third, this method is not adaptable to start-ups since the first years of life usually are characterized by losses. Substantially it can be said that this model is more adaptable to those companies that are at mature stages whose dividends are with constant growth and are almost certain.

To resolve these issues and open to different kinds of firms, it has been developed a different model, always based on dividends and growth rate but more precise:

Formula 10: Dividend-Discount Model with Constant Long-Term Growth

$$V_0 = \frac{Div_1}{(1 + r_E)} + \frac{Div_2}{(1 + r_E)^2} + \cdots + \frac{Div_N}{(1 + r_E)^N} + \frac{1}{(1 + r_E)^N}\left(\frac{Div_{N+1}}{r_E - g}\right)$$

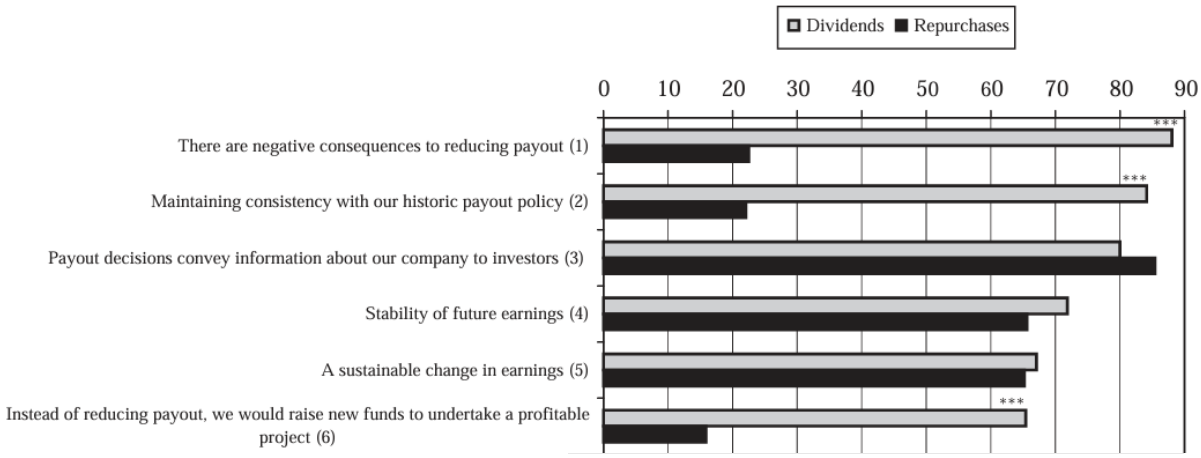
It is now possible to account for first year's losses and open this model up to start-up firms, sometimes its implementation can correspond to the utilization of multiple growth rates in the same formula to perfectly adapt to supposed various stages of company life. In addition to this, it can be said that the presence of multiple years of consideration reduces (but not too much) the terminal value $\left(\frac{Div_{N+1}}{r_E - g}\right)$ impact on the overall valuation and performs better customization to personal expectations. Still, this type of model presents a couple of limitations usually attributed to the firm future decisions about financing structure, dividend payout rate, and share count.

Since it is important to consider and analyse **payout decisions**, as they are reflected in future dividend forecasts, it becomes necessary to focus on payout policies. In a relatively recent survey[15] about payout policy adopted by financial managers it emerged that, in general, managers fear the most two things: the reduction of dividends (reduction of dividends per share if the number of shares stays the same) and the rapid increase of dividends. Both are connected because a rapid increase in dividends becomes more difficult to maintain and could result in a future possible reduction. About dividend policy it can be said: "Dividend policy is conservative"[16]. From the survey and the CFO's perspective, it is clear how there is an asymmetric reaction from the investors in interpreting dividend distribution. When the companies increase dividends the positive investor reaction is lower, and nonproportional, than the negative reaction in case of a reduction in dividends. This explains why, frequently, CFOs decide to operate in a zero-dividend policy which maintains the decisions free of limitations caused by dividends (in the survey emerged that several companies paying dividends would come back and never start). Why do managers fear the reduction in dividends? Reduction in dividends could suggest that the CFOs expects a negative future for the firm or the necessity to accrue the profits in the reserves for possible future losses. An increase in dividends usually means two things: CFOs are sure about increased future profits, and CFOs are sure about more stable inflows. This explains why investors' reactions, even if lower, push prices up.

A second, relevant, theme is the importance of **share repurchases**; in the survey became clear how, in the presence of share repurchases (a different form of distribution of dividends), managers would consider them more flexible than dividends. It is in practice proved that if a company one year decides to opt for share repurchases and the following year doesn't maintain the same decision there will be no (or generally lower) effect on prices. Their flexibility has no impact on the share price, permits to offset stock option dilution, and allow managers to decide with more freedom when to return capital to investors. For this reason, CFOs prefer this form of distribution.

Below are reported a part of the key answers taken out of the survey:

[15] Brav A., Graham J.R., Harvey C.R., Michaely R., "*Payout policy in the 21st century*", Journal of Financial Economics, 2004.
[16] Lintner key findings.



It’s clear how there is a substantial difference between share repurchases and dividends especially when considering the concerns about the future. Managers fear negative consequences in reducing payout, avoid misaligning the present with historic payout policy, and prefer to raise new funds for a project rather than risk the reduction of the payout. The perception of managers about share repurchases is different, only 20% apply the same thinking as for dividends and this underlies how share repurchases appear more flexible. Differently, when considering those two methods as a good proxy of future company performance for investors, managers feel the same way about dividends and share repurchases, this means those two instruments can be easily interchangeable.

This digression has permitted to better understand how the payout policies work, if a company is mature and expects steady flows it is more probable that the dividends will grow too because the managers are sufficiently confident about the company's future. Differently, if a company has never really paid dividends it may never start.

The Dividend Discount Model places too much confidence in expectations about dividends and their growth rate even if in their nature can become difficult to predict. The other element of uncertainty is the number of floating shares, this number can increase just like reduce in response to company necessities and structure. The utilization of share repurchases has proved to be a growing instrument used by CFOs to reduce the dilution effect and provide a harmless way of paying back investors. A reduction in shares could cause a rapid increase in dividends per share and harm the success of the valuation.

This explains why different valuation methods become necessary.

1.3 Valuation Multiples Methods

The first solution to the Dividend Discount Model problem is provided by the Valuation Multiples method. This kind of valuation still sets its grounds on the Law of One Price but, rather than look at the company's future projections, tries to determine its price by looking at comparable firms. The Law of One Price is, in this context, extended to an extreme case in which it is hypothesized that there exists a company whose characteristics are completely identical to those of the analysed company and the cash flows are identical too. In that case, it will be clear the necessary coincidence of prices due to the Law of One Price.

The problem with this is represented by the evident impossibility of assisting at two completely identical firms; multiple factors can vary starting from the management quality, the products sold (which even if similar are not identical), and the geographical location.

Because of this limitation, the Valuation Multiples Method must be dealt with adequately by considering the right risks and innumerable differences between similar companies. It is, therefore, clear that also in this case, even if the problem of forecasting dividends and share repurchases is avoided, it can be assisted at a different type of hypothesis. Also in this case the result cannot be pure science and truth.

The first difficulty that arises from the Multiples Valuation Method is to manage the differences in scale (the company dimension), this problem can be solved by considering no more balance sheet, income statement, or cash flow voices by themselves but transformed in a **ratio**. These ratios are defined as “**Valuation Multiples**” and have the characteristic of creating a sort of normalized value that can be compared among the different companies. The ratio is characterized by the presence in the numerator of a measure of market valuation and the denominator of a universal measure of financial performance. It is like creating a common (among companies) index that permits to understand, about a certain statement voice, if the stock has a higher or lower price. In this setting, if it is defined this ratio, and there is a comparable firm already listed, it will be easy for a company object of valuation to be priced by using that ratio.

Valuation Multiples can be distinguished into two macro-categories:

Equity Value Multiples: the denominator presents a financial voice whose streams are solely directed to equity holders (Net Income, EPS). This class of multiples is way more frequent for stock valuation when investors acquire minority stakes. Its wide utilization is mainly due to its more intuitive and simple nature; investors can, more easily, find the information necessary to quickly create the multiples and make an investment decision. Still, investors do not have to fall into the trap of capital structures, similar companies from the sector and size view may be different in the capital structure (different Debt to Equity). In those cases, the multiples may suffer from the so-called "gearing effect"[17] which is directly dependent on the financing structure of the comparable firms.

The Equity multiples are the following:

- the most used (and famous) multiple is the ***Price/Earnings (P/E) Ratio*** and presents at the numerator the share price while at the denominator the Earnings Per Share (EPS). The rationale is that you can intuitively understand the valuation of an analysed company by simply multiplying its EPS by this ratio; the higher the EPS the higher the equity value. The P/E ratio is a basic multiple and has different disadvantages: first, it doesn't work if there are yearly losses, second it doesn't account for company growth and, lastly, can be subject to accounting policy differences;
- an evolution of the P/E is the ***Forward P/E Ratio*** which inserts no more the EPS at year zero but at year one (the expected earnings per share at year one);

[17] Indicates the level of indebtedness which usually increases the results because the company operates through leverage (but at higher risks).

this alternative method is more precise for valuation purposes because it considers future earnings and can be furtherly developed:

$$Forward\ P/E = \frac{P_0}{EPS_1} = \frac{\frac{Div_1}{EPS_1}}{r_E - g} = \frac{Dividend\ Payout\ Rate}{r_E - g^{18}}$$

Even if earnings are not as reliable as current or historical data this doesn't mean that the utilization of future earnings doesn't permit to define a more precise valuation. Forward earnings appear as the best choice especially if looking at future valuation; two similar companies can have very different P/E, this difference could be caused by the fact that one of the two has a better future perspective or a better Forward P/E which drives the price up. In this sense it has been proved that forward multiples explain prices way better than historical multiples[19], and performance increases as the forecasted horizon increases too. The problem in using Earnings arises when there are losses, in start-up valuation it is difficult to use P/E as multiple, for this reason, there are a few different alternatives;

- the ***PEG (Price/Earnings-to-Growth) Ratio*** is frequently used to complete the P/E analysis. It's composed of, at the numerator, the P/E Ratio, while, in the denominator, the earnings growth rate (multiplied by 100). This sort of ratio is used to fully complete the information provided by the P/E alone. Since it considers the growth rate (usually the 5-years) too it can exhibit growth potential over a stock valuation. A PEG Ratio of one defines a proper stock valuation while a PEG ratio below one indicates that a stock is undervalued. So, it can happen that two similar companies, company A and company B, have a P/E of respectively 15x and 20x and a growth rate of 10% and 25%. This would mean that stock A is undervalued concerning stock B, but, if those P/E are divided by their 5-year growth rates it should be seen that the PEG ratio of A is 1.5 while that of B is 0.8. The higher growth rate of Company B completely upsets the initial conclusion and demonstrates that, if they are considered the growth rates, Company B is undervalued while Company A is

[18] Forward P/E is higher when computing high growth firms.
[19] Liu J., Nissim D., Thomas J., "Equity Valuation using Multiples", Journal of Accounting Research, 2002.

overvalued. Its strong dependence on the growth rate (which is frequently mistaken) and the absence of proper consideration of dividends distributed (mature firms may be growing slower but could still be distributing high dividends) make this multiple more adaptable when comparing high-growth stocks. Additionally, in those cases it will be necessary to properly evaluate the company risk which could be a driver of the higher growth; the PEG Ratio doesn't consider the company risk level;

- another Equity multiple is the ***Price to Book (P/Book) Ratio*** which is calculated by dividing the company's market capitalization by its last year's book value. This multiple is a good measure for capital-intensive companies in which assets are a core driver of earnings (banks, automotive, airline, etc.). In general, P/Book is used as a first analysis of a listed company, so is good for investor decision; when it is one it means that the company is valued properly while if it is more than one it is overvalued. Still, it is not for sure that a high P/Book means overvalued because a company could justify such a high value by its, equally high, ROA (Return on Assets), future growth and profits, patents, or other intangible assets. The major drawback in using P/Book is represented by its largely dependence on the fact that book value doesn't consider a lot of relevant elements in company valuation. In addition, book value is largely influenced by accounting policies that determine the way of inscribing the tangible assets (they can be inscribed at historical cost or not) and intangible assets (some accounting policies do not permit to insert intangible assets unless they have been acquired);
- the ***Price to Sales (P/Sales) Ratio*** is another ratio frequently used but, as it is evident, doesn't respect the coherence rule of the ratios. If the numerator regards only the equity part of the company, it is not correct to use in the denominator a value that is destined to both equity holders and debt holders. For this reason, the Sales are more appropriate as an Enterprise Value multiple.

Enterprise Value Multiples: the denominator presents a financial voice whose streams are directed to both equity holders and debtholders (Sales, EBITDA, EBIT). The Enterprise Value is defined as the total value of the firm, and it's computed as:

Formula 11: Enterprise Value

$$\boldsymbol{Enterprise\ Value} = Market\ value\ of\ Equity + Market\ value\ of\ Debt - Cash\&Eq$$

The market value of equity is here identified as the multiplication of the company's fully diluted shares in the money[20] by the stock's market price. This means that the computation of equity value is not limited to the shares outstanding, but it includes also in-the-money options, stock options, warrants, and convertible securities. Enterprise Value, which is different from Equity value, accounts also for the dilutive effect of certain instruments. Enterprise Value ("EV") represents the cost of buying the right to the whole of an enterprise's core cash flow[21]; by defining it as the total cost to take over the business it becomes easy to understand why is added the debt and subtracted the cash (the so-called "Net Debt") because it could be immediately used to pay the creditors.

The consideration of Enterprise Value is more frequent in Mergers & Acquisition transactions. Enterprise multiples permit to avoid the problem of capital structure that arises in Equity multiples since it is considered the financial structure of the companies used as comparable and puts all the companies at the same level. Because of this, in case of the presence of comparable with different financing structures, it is preferred to use Enterprise multiples, find the EV of the target firm, and, finally, subtract the net debt to compute the equity value.
The Enterprise multiples are the following:

- ***Enterprise to Sales (EV/Sales) Ratio***: in this case, the Sales can be used and the main advantages in considering them, rather than earnings, are zeroed exposure to accounting differences and possible utilization also in the presence of negative earnings. Still, its simplicity is the main reason why practitioners and analysts don't use this multiple, sales are rarely a direct value driver since they don't indicate any grades of performance. EV/Sales has been proven not to be so effective in determining the intrinsic company valuation, and for this reason, it is not so widely used;

[20] When the weighted average exercise price is below the current stock price.
[21] ____, UBS Global Equity Research, "Valuation Multiples: A Primer", 2001.
If you buy a company for the Enterprise price, then you have also the right of receiving interest payments.

- ***Enterprise to EBITDA (EV/EBITDA) Ratio***: this is easy to determine and its simplicity along with a sufficient similarity with cash flows makes this ratio the most popular of EV multiples. Still, since EBITDA is just a proxy for cash flow, it doesn't consider tax management (its value creation), capital expenditures, and depreciation. These drawbacks do not impede this ratio to be a perfectly valid one since, practically speaking, EBITDA represents, in the majority of cases, an optimal performance-based accounting value;
- ***Enterprise to EBITDAR[22] (EV/EBITDAR) Ratio***: this ratio is used in companies whose rental and restructuring costs have a huge impact on the operating process. By considering the EBITDAR, rather than EBITDA, analysts try to normalize capital intensity between companies operating with huge rental costs (restaurants, hotels, casinos) or that have restructured in the last year; this way it is possible to compare them without influencing their operating performance. This ratio, just like EV/EBITDA, has the advantage of not being exposed to accounting differences in the computation of depreciations and amortizations. The disadvantages are the same as EV/EBITDA with additional difficulties in determining rental costs which could not be expressly indicated in the income statement;
- ***Enterprise to EBIT (EV/EBIT) Ratio***: this ratio is not as popular as EV/EBITDA but is optimal when evaluating capital-intensive companies whose assets' depreciation represents a relevant cost. Just like the EBITDA also EBIT ignores value creation through tax management but, in addition, is susceptible to differences in depreciation policy;
- ***Enterprise to NOPAT[23] (EV/NOPAT) Ratio***: is an easily understood and accessible measure of a company's value. It provides an idea of how much you are willing to pay for each dollar of income;
- ***Enterprise to Free Cash Flow (EV/FCF) Ratio***: is a good multiple since it considers CAPEX and other elements of capital intensiveness and solves the problems of EBIT with the accounting differences since the depreciation is

[22] Earnings Before Taxes, Interests, Depreciation, Amortization and Rental costs.
[23] Net operating profit after tax, also viewed as "unlevered net income" already defined.

added back. EV/FCF is a good proxy[24] for understanding how fast you can pay back the company price since FCFs define the cash overall inflows. Unfortunately, the presence of capital expenditures quite complicates the FCF computation since they are often irregular and can be easily manipulated;

- ***Enterprise to Invested Capital***: another multiple highly used with capital-intensive companies. Its computation can be done in two ways:
  - $Invested\ Capital\ =\ Fixed\ Assets\ +$
    $Net\ Working\ Capital\ (NWC)$
  - $Invested\ Capital\ =\ Total\ Shareholders\ Equity\ +\ Total\ Debt$

  The problem with dealing with invested capital is that the book value of tangible assets is subject to accounting differences and most of the time is represented at its historical cost;
- ***Enterprise to forecasted flows (EBITDA$_1$, NOPAT$_1$,* FCF$_1$)**: obviously, also in Enterprise valuation can be considered future flows. As for Equity multiples, these become necessary when targeting a start-up or still-growing company. The process is the same but, instead of inserting equity flows, in the denominator are present forecasted enterprise flows. The importance of using forecasted values derives from the necessity to make the life cycle of the target company coincide with the life cycle of the comparable firms. In this sense, it has been proved[25] that the multiples start to coincide when the companies are in the same life cycle. If the comparable companies are all mature firms and the target one is an advanced startup, it's not sure that they will be the same but, maybe, in five years they will all be at the same level.

After having discussed the difference between Equity and Enterprise multiples, the Valuation Multiples Method, then, can be divided into two other categories:

- Comparables Method (Comps)
- Comparables Precedent Transaction Method (Compaq)

[24] It's the reciprocal of Free Cash Flow Yield, a performance metric whose purpose is to determine the ability of the company to create cash. The Levered FCF Yield considers only the cash flows destined to shareholders while the unlevered one considers the overall cash flows.
[25] ____, UBS Global Equity Research, "Valuation Multiples: A Primer", 2001.

1.3.1 Comparables Method (Comps)

Comparable Companies analysis ("comparable companies" or "Comps") is one of the primary methodologies used for valuating a given company, division, business, or collection of assets (together, or separately, defined as "**target**"). The core of this analysis involves the selection of a universe of comparable companies for the target one ("comparables universe"); these companies are benchmarked one against another. Trading multiples are then calculated for the universe, which serves as the basis for extrapolating a valuation range for the target.

This valuation range is calculated by applying the selected multiples to the target's relevant financial statistics[26].

After having defined the multiples in the previous paragraph, it is now time to understand how they are used in the Comps Method. First, it is important to set the framework and the steps of this method, it can be said that there are five, relevant, steps when performing a Comps analysis:

I.	**Select the Universe of Comparable Companies**
II.	**Locate the Necessary Financial Information**
III.	**Spread Key Statistics, Ratios, and Trading Multiples**
IV.	**Benchmark the Comparable Companies**
V.	**Determine Valuation**

[26] Joshua Pearl, Joshua Rosenbaum, "Investment Banking Valuation, Leveraged Buyouts, and Mergers and Acquisitions", Wiley, 2019.

Select the Universe of Comparable Companies: the foundation of the Comps method is represented by the selection of a set of comparable companies. This selection represents one of the most difficult and delicate things in the whole process since the comparables will be used as instruments to define the price. Too different comparables completely change the outcome and for a series of reasons it is mandatory to fully understand the target company[27]. Analysts usually start by identifying a broad range of competitors, also through a survey, which is in his turn selected as a subset of "closest comparables"[28].

Locate the Necessary Financial Information: in the next step, it has to be found all the necessary information for the definition of the trading multiples in order to have an overall picture of the comparables with ratios and key financial statistics. The information about Income statements, Balance Sheet, Cash Flow Statements, and Market data (about shares and other financial instruments issued) is quite easy to find for listed companies (usually on their websites for a law provision). Only Market data, credit ratings, and analyst projections, usually are not available for free and have to be either calculated or bought by Rating agencies.

Spread Key Statistics, Ratios, and Trading Multiples: now, all the ratios, key statistics, and multiples can be derived. This part regards the calculation of market valuation measures like equity and EV, the identification of income statement and balance sheet values and ratios necessary for the calculation of size[29], profitability[30], growth[31], financial

[27] The selection of comparables should start only after an in-depth and exaustive analysis of the target company. This is particularly difficult for private companies whose info are not easily and freely available but necessitate of further in-depth research on websites and newspapers. The exception raises in the case of M&A and takeover acquisitions in which detailed (and not publicly available) information is available.

[28] From a Business Profile it matters: the sector, the products and services, the type of Customers, distribution Channels, and Geography. From a Financial Profile: Size, Profitability, Growth Profile (EPS growth rates for mature firms, Sales or EBITDA growth rates for early stage), Return on Investment, Credit Profile, and Financial Leverage.

[29] Market Valuation, Sales, EBITDA, EBIT, Net Income, Assets

[30] ROE (ROCE), ROI (RNOA), ROIC, ROS, Asset Turnover, Gross profit margin, EBITDA and EBIT margin, Net income margin, Dividend Yield (Dividends/Price).

[31] Historical and estimated growth rates

equilibrium[32], and financial leverage[33]. The multiples discussed in the previous chapter can be added to new ones tailored to the sector-specific characteristics:

Valuation Multiple	***Sector***	***Valuation Multiple (per share)***	***Sector***
Enterprise Value /		*Equity Value (Price) /*	
Broadcast Cash Flow Or Subscribers	• Media • Telecommunication	*Book Value*	• Financial Institutions • Homebuilders
EBITDAR	• Casinos • Restaurants • Retail	*Net Asset Value*	• Financial Institutions • Mining • Real Estate
EBITDAX (Exploration expense)	• Natural Resources • Oil & Gas	*Cash Available for Distribution*	• Real Estate
Reserves	• Metals & Mining • Oil % Gas • Natural Resources	*Funds From Operations (FFO)*	• Real Estate

Benchmark the Comparable Companies: this step represents the core of the whole process and is summarized in the known word “benchmarking”. In this phase, the analysts determine the ranking of the closest comparables by looking at the calculated ratios and statistics. This is done by analysing and comparing each of the comparable companies with the target. Also, it has to be considered the presence of outliers; by defining a subset of interested companies it is possible to identify the **best** comparables for the final valuation. This process of selection is usually done in a two-step way:

I. it is given a look at the key financials and ratios seen in step III to place accordingly the comparables and identify those ones more similar to the target (best comparables and potential outliers);

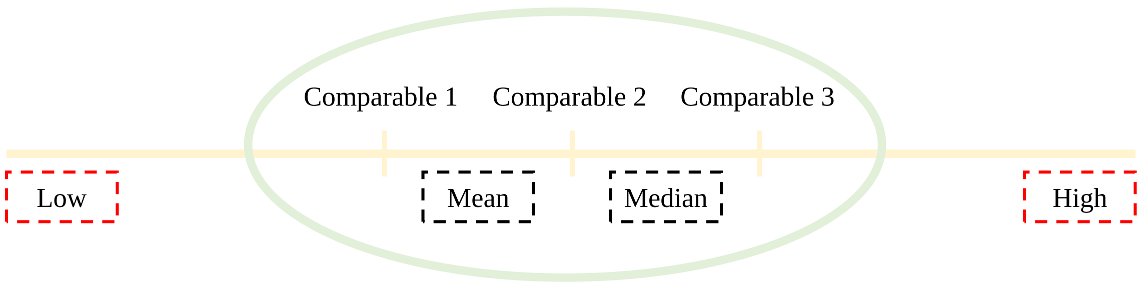
[32] Liquidity ratios: Current Ratio, Quick Ratio, and Cash Ratio
[33] Credit ratings, leverage ratios.

II. focus on the multiples by defining those that are more suitable for the sector and making a comparison with the financials and ratios to understand the outliers.

Both steps present highs, lows, mean, and median. After a first look, are frequently created subgroups of the comparable companies filtering per age, subsector, or size.

Determine Valuation: the multiples are used to create a range of possible values. The analysts usually start by taking into consideration the mean and median values of the most suitable multiples (they are chosen in the way explained in the precedent paragraph) for that company and occasion (e.g. it will not be possible to use a P/E ratio if the target company has current losses) to derive a proper range of possible values. In this part highs and lows play a necessary role in defining the **ceiling** and the **floor** of the valuation. Then, the range is further narrowed by considering the selected closest comparables and using the others as additional information. At this point, the ability and experience of seniors is necessary to have a hint for the final decision in terms of multiples which will be used to calculate the implied valuation range.

The oval below shows what would be a typical valuation range.



1.3.2 Comparables Precedent Transaction Method (Precedents)

Comparables Precedent Transaction analysis ("**Precedents**") is similar to the Comps method because it employs a multiples-based approach to derive an implied valuation range for a target. Precedents, differently, is no longer based on trading multiples of comparable listed companies but on multiples paid for comparable companies in prior M&A transactions.

The core part of this method is represented by the identification of proper precedent transactions; it is, therefore, evident the difficulty of implementing Precedents on certain occasions. Sometimes there may be no precedent transactions or the companies involved have too many differences with the target. In general, it can be said that the number of M&A transactions is way lower especially if compared to the number of companies in the market. The restricted number of transactions, along with the difficulty of digging into them (sometimes the details of mergers are not public), make the research of acquisitions involving companies similar to the target one really hard. Also, the year of the transaction makes it difficult to understand if it is a good comparable or not; generally, the recent transactions are the most relevant ones since are the result of present market trends and conditions. This means that the universe of comparables precedent transactions becomes even smaller since, except for some cases[34], older transactions have to be excluded.

As in Comps, also in Precedents the multiples play a central role. The utilization of acquisition multiples of majority stakes creates another problem: precedents provide a

[34] When the precedent transaction occurred in a similar point of the target lifecycle or during a similar macroeconomic period.

higher trading multiple which includes a premium. The premium price paid by purchasing companies is justified by two main reasons:

- buyer pays a so-called "**control premium**" when acquires the majority stake of a company, this control premium equals the monetized advantage of being a majority shareholder (which implies the right to drive decisions and operations in the target company);
- a merger frequently results in the creation of synergies between the buyer and the target, synergies permit to cut the costs of production and the reaching of economies of scale.

For the reasons just mentioned Precedents is more adaptable for M&A transactions because includes the control premium rather than the simple calculation of intrinsic value. The **pros** and **cons** of the Precedents method are listed below:

Pros	**Cons**
Market-Based: the multiples calculated are the effective valuation given by the market	*Time lag*: the difference in time can be severely relevant, especially in some fast-growing sectors
Objectivity: it avoids assumptions about the target future performance and follows an already implemented example	*Existence of Comparable Acquisitions*: in some cases, it is really difficult to find precedent transactions related to a certain sector
Current: provides a current example of an M&A transaction which reduces the uncertainty about market reaction	*Availability of Information*: information about transactions is not always public and sometimes is not even sufficient to calculate multiples
Time-Saver: reduces the steps that have to be performed to reach a proper valuation	*Acquirer's Basis for Valuation*: the acquirer usually possesses confidential information about the target, this permits to know more about future performance, this is not captured in the multiples

Precedents analysis, like Comps, can be divided into five steps:

I.	**Select the Universe of Comparable Acquisitions**
II.	**Locate the Necessary Deal-Related and Financial Information**
III.	**Spread Key Statistics, Ratios, and Trading Multiples**
IV.	**Benchmark the Comparable Companies**
V.	**Determine Valuation**

The last three phases are kind of identical to the Comps method, only the first and the second one differ since it is different the type of focus of the information. While in Comps it was necessary to deal with comparable companies in here, they must be considered more elements. The focus of this paragraph will be, therefore, on steps I and II.

Select the Universe of Comparable Acquisitions: the identification of a set of precedent comparable transactions is a difficult task that implies a strong knowledge of the target and can get even harder when considering developing sectors or businesses. In general, the first thing to do is to find and select as many relevant transactions as possible whose time, sector, size, and geographical position better suit the target. This can be done by conducting a first research on comparable companies (as seen in the Comps method) and, afterward, making a research on their history. Once it has been defined a list of comparable acquisitions it is time to dig more in depth into the story of the transaction whose understanding could help to better interpret the multiples paid. The understanding of the reasons and the story behind the acquisition can both help to determine if the transaction is a good comparable and justify an eventual premium paid or a discount received by the buyer. Among the relevant considerations are recalled:

- **Market Conditions**: the economic environment has to be intended within specific sectors and company cycles, especially relevant is the condition of capital markets. The consideration of economic bubbles (prices of a certain sector are inflated) or periods of low-rate debt financing (low-interest rate increases the amount of financing which makes it easier to purchase at higher prices) for sure have an impact on the price, in this case, it drives them up;

- **Deal Dynamics:** what was the buyer a Strategic Buyer or a Financial Buyer[35]? What were the motivations of the buyer and the seller for the transaction? The seller requires speed in the execution because needs cash or the buyer needs the target because could lead to high synergies? Still, the takeover was hostile or friendly[36]?

Locate the Necessary Deal-Related and Financial Information: this step is the hard part because, even if it was possible to identify a valuable comparable transaction, it can still be difficult to group and locate all the necessary information for the determination of the transaction multiples. In general, it can be said that, if the transaction happened with the involvement of public companies, the details have to be published and made available by law. If, on the contrary, the transaction was private the info and the multiples are way more difficult to get, and it can happen that analysts or others interested are not even able to derive the valuation multiples. Some publicly available instruments (e.g. **Capital IQ** or **Bloomberg terminal**) supply information about privately held companies, among this can figure the transaction price of a M&A deal or directly the EV paid by the purchasing company. Still, those instruments, especially Bloomberg, are quite expansive and can't be purchased by individuals alone. Another solution, at least in Italy, is to take access to one of the internet portals (e.g. Telemaco, Cerved) available that provide, at a cost, all the documents concerning a certain private company. Usually, the bylaws or another formal document contains the EV or the purchasing price; this way, after having identified a transaction whose details are not available, the desired value can be obtained through a transversal way.

The other phases are the same as those in the Comps Method; the transaction multiples, which have been defined as the target multiples for the valuation range, are selected and

[35] A Strategic Buyer guarantees a higher premium due to its expertise on the field and its capacity to create synergies. The Financial Buyer is not an expert of the sector and usually is a company, like a Private Equity firm, whose purpose is to generate high returns through a successive exit. Financial buyers usually operate through leverage which, during periods of stability of capital markets, could drive the resources up and also higher the prices.

[36] There are several reasons why a friendly takeover can be cheaper:

- Friendly takeover takes place without the utilization of defensive strategies like: Poison pills, Poison Puts, Anti-Takeover Amendments, Golden Parachutes, White Knight etc.;
- Friendly takeover is reached through a due diligence period which implies a friendly disclosure of internal information which can help better determine the price;
- Co-operation of the management.

used to derive the final value. As it has been said before, the Precedents method is a safe one if the transaction is recent, you have a valid comparable and you are evaluating a target for a M&A transaction; otherwise, the variables in the game are too many and the valuation multiples frequently comprehend premiums due to the acquisition.

For the calculation of intrinsic value both the Comps and Precedents analysis do not represent the most faithful option, especially if it is available more information about the target company, they are suggested as starting point and, specifically for the Precedents, when there is a concrete possibility of **evaluating the control premium**. The fact that both are based on market valuation implies the difficulty of deriving the intrinsic value which is usually twisted by market imperfections and biases.

To address the issue of having a limited number of comparables and transactions, and correctly derive the intrinsic value of a company without market interference, it will be introduced in the following paragraphs the Discounted Cash Flow Method with its variations and adaptations to different cases.

1.4 Discounted Cash Flow Method

The Discounted Cash Flow analysis ("DCF analysis" or the "DCF") is a fundamental valuation methodology broadly used by investment bankers, private equities and other finance professionals. It is premised on the principle that the value of a company, division, business, or collection of assets (together, or separately, defined as "target") can be derived from the present value of its projected Free Cash Flow (FCF)[37].

The Discounted Cash Flow (DCF) method represents the last step of a company valuation since it supposes a deep understanding of the target business, its weaknesses, its strengths, and future expectations about its performance. These expectations or forecasts are different from those of the other methods which are generally short-term ones. A good DCF goes beyond 2-3 years and focuses deeply into a company's prospective income and eventual projects with an average of five years projections.

DCF is different from DDM in considering no more Dividends but focuses on the cash inflows and outflows by adjusting for those expenses in the income statement that are effectively no cash outflows (depreciation and amortization) and those cash outflows that are not considered in the income statement (CAPEX and Net Working Capital). Additionally, DCF starts from the Unlevered Net Income and doesn't consider the interest payments. For this reason, Free Cash Flows (Formula 4) are forecasted and discounted to determine the Enterprise Value (Formula 11) which, as it has been said before, comprehends also the interest payments towards the creditors. The DCF will, therefore, be defined as:

[37] Joshua Pearl, Joshua Rosenbaum, "Investment Banking Valuation, Leveraged Buyouts, and Mergers and Acquisitions", Wiley, 2019.

Formula 12: Discounted Cash Flow Model

$$V_0 = PV(Future\ Free\ Cash\ Flow\ of\ the\ Firm)$$

This formula must be extended:

$$V_0 = \frac{FCF_1}{1 + r_{wacc}} + \frac{FCF_2}{(1 + r_{wacc})^2} + \cdots + \frac{FCF_N + V_N}{(1 + r_{wacc})^N}$$

The first, evident, distinction between the DCF and the other models is the discount rate. So far, it has been used the firm's equity cost of capital, r_E, which is justified by the calculation of the sole equity part of the company. Since in here it is calculated the Enterprise Value, now, the appropriate discounting factor is the (total) firm's cost of capital which is defined as Weighted Average Cost of Capital ("**WACC**") whose specifics will be analysed in-depth in the following paragraph. The distinction between cost of equity and WACC is, basically, that the cost of equity represents the opportunity cost of capital of an equity investor, while WACC represents the weighted average of the cost of equity and cost of debt (the opportunity cost of creditors and investors together). WACC and cost of equity can coincide if a company operates without debt.

The model is based on a series of assumptions:

- constant Debt to Equity Ratio[38];
- constant terminal growth rate;
- no variation in risk-free interest rate (Beta Equity is calculated by considering it).

As for Comparables Valuation, also DCF can be divided into a series of steps:

I.	**Study the Target and define Growth Perspectives**
II.	**Project Earnings and Free Cash Flows**
III.	**Calculate the Weighted Average Cost of Capital**
IV.	**Determine Terminal Value**
V.	**Calculate Present Value and Determine Valuation**

[38] The WACC formula considers the financial leverage of the company and weights the cost of debt and equity by considering its respective financing structure. WACC (after-tax), which accounts for the presence of taxes, is computed in the following way:

$$WACC = \frac{E}{E + D} r_E + \frac{D}{E + D} \times r_d \times (1 - tax\ rate)$$

Study the Target and define Growth Perspectives: this phase is characterized by a first approach to the target's business and sector with a further focus on the company's performance key drivers. These drivers are *internal*, all those internal growths like new facilities or new customer contracts, or *external*, like acquisitions or macroeconomic and customer patterns. All those patterns help to better understand if the company is effectively growing and how much it is expected to grow.

Project Earnings and Free Cash Flows: past performance (typically the prior three years) represents a valuable option in providing a starting point for the determination of future growth rates (especially if are considered mature, non-cyclical firms). Past growth rates are usually computed by considering financial statements adjusted by non-recurring elements and events (it is kind of a normalization process). Projections and growth rates are not a problem when performing a M&A transaction, in such cases, usually labelled as "Management Cases", the managers of the target company provide a set of forecasts. Those forecasts must be appropriately dealt with since they usually represent an exaggeration of future performance or the consideration of a possible Business Plan that includes future access to additional bank loans; in these cases, it is important to adjust for management's optimistic expectations and opt for a **scenario analysis** with **base**, **worst**, and **best** scenarios.

Calculate the Weighted Average Cost of Capital: WACC computation is not difficult, per se, because it is a simple weighted average (see note 38); the real difficulty is to properly manage the computation of the cost of equity and cost of debt along with the proper definition of the target capital structure (Debt / (Debt + Equity)). The cost of capital (debt and equity) will be covered in detail in the next chapter; the capital structure is usually taken as the average of the competitors.

Determine Terminal Value: another element to be considered in the computation of the Present Value of future Free Cash Flows is the so-called Terminal Value, V_N. The latter usually accounts for three-quarters of the overall company's value; because of this, its computation is crucial in the whole FCF method, and it becomes necessary that the projections stop only if the company has reached a steady state and not a cyclical high or low. Terminal value can be computed in two ways:

- **Perpetuity Growth Model**: which is defined as the value of the firm N under an assumption of constant growth rate over the remaining life of the firm:

$$V_N = \frac{FCF_{N+1}}{r_{WACC} - g_{FCF}}$$

With $FCF_{N+1} = FCF_N \times (1 + g_{FCF})$.

- **Exit Multiple Method**: works as Comps method and multiplies last year's EBITDA or EBIT by a valuation multiple of comparable companies (normalized by considering sector or economic cycles).

These methods are used in conjunction, this way they permit to avoid discrepancies or absurd results (if the two calculated terminal values are too distinct there is a problem).

Calculate Present Value and Determine Valuation: In the last step, valuation is reached through Formula 11. The latter computes the intrinsic Enterprise Value of the company; equity value will be easily derived by subtracting the Net Debt.

The length of projections depends on the type of the life stage in which the company is currently positioning, if the company is a mature firm, the cash flows will be usually projected for five years. Otherwise, if the company is in an early stage of rapid growth, it will be more appropriate to forecast cash flows until revenues reach a **steady leve**l (which can take up to or above ten years).

Even if the problems related to DCF and the multiples approach have been solved by the FCF method, still can be noticed a series of uncertainties due to the assumptions made at the beginning. The most evident one is the assumption of a constant Debt to Equity ratio which becomes hard in certain situations, especially if the target company has no intention to lock up the D/E ratio and doesn’t commit herself to such a bond (keeping the financing structure constant can result also in several costs). For this reason, it must be designed a new model that can adapt also to variations of the leverage exposure; the following paragraph exhibits the remaining methods that can help to better understand how to manage this last problem and the understanding of the impact of debt on the company valuation.

1.5 Valuation with Leverage

This paragraph outlines various valuation methods that consider the tax benefits associated with debt and permit to bring forward a company valuation involving a target with changing financing structure. These methods include:

- WACC Method
- Adjusted Present Value Method
- Flow-To-Equity Method

Before entering into the details, it is necessary to make a first, general, introduction of what is the interest tax shield and the advantages of leverage. So far, the interest expenses have not been considered, that's why Net Income was referred to as "Unlevered Net Income". Now, the attention can be directed towards the way interest expenses are managed and the consequential impact they exert on the intrinsic value of the company. An effective method to illustrate the impact of interest payments on a company's Enterprise Value is to provide a straightforward example highlighting the contrasting alterations in Net Income between a leveraged company and an unleveraged one. Let's suppose:

- Interest Expense: 300
- Corporate Tax Rate: 35%
- EBIT: 1000

Company X Income Statement		
	With Leverage	Without Leverage
EBIT	1000	1000
Interest Expense	(300)	0
Income Before Tax	700	1000
Taxes	(245)	(350)
Net Income	455	650

Upon initial examination, one might infer that interest expenses reduce the Net Income, as evidenced by the lower result presented by Company X with Leverage. However, this conclusion may be misleading if one were to adopt a broader perspective that considers the total amount available to all investors, rather than solely focusing on the amount available to equity holders. Such an approach aligns seamlessly with the definition of Enterprise Value, which encompasses both debt and equity. In a general sense, it can be demonstrated that the presence of interest expenses does not reduce the company's value, as the company's value is defined as the sum of Net Debt and Equity; rather, it augments it. Consequently, when contemplating the total amount accessible to all investors (both equity and debt holders), the dynamics of the situation undergo a transformation:

	With Leverage	Without Leverage
Interest Expense	300	0
Net Income	455	650
Total Available to all Investors	755	650

The company employing leverage generates greater wealth due to the deductibility of interest expenses. This becomes apparent when it is recognized that the difference, amounting to 105, arises solely from the product of the corporate tax rate (35%) and the deducted interest expense (300):

$$300 \times 0.35 = 105$$

The augmented worth inherent in the leveraged entity stems directly from the concept known as the "Interest Tax Shield," denoting the reduction in tax liabilities resulting from the deductibility of interest expenses. Interest Tax Shield, or the amount of avoided taxes in the presence of leverage, is defined as:

Formula 13: Interest Tax Shield

$$Interest\ Tax\ Shield = Corporate\ Tax\ Rate \times Interest\ Payments$$

It is evident how, in the presence of leverage, companies can derive a certain benefit from the utilization of debt; that's why, in general terms, a Chief Financial Officer must deal with the optimal financial structure[39].

This anticipation allows for the presentation of the aforementioned methods.

[39] This thematic will not be covered but, in substance, the getting indebted is preferred because the cost of debt is frequently lower than the cost of equity, this is generally true until the financial leverage reaches too high levels (in that case interest expenses become way too high). WACC already considers the deductibility of interests, that's why the utilization of debt could represent an additional element in favour of getting leveraged. Companies try to reach the financing mix that minimizes WACC and maximizes company's present value of FCFs; financing structure represents, therefore, a strategic choice whose decision impacts the creation of value for investors.

1.5.1 WACC Method

The WACC Method is simply an updated version of the Free Cash Flow Method because, in this case, it is inserted the possibility of computing the Debt Capacity. The after-tax Weighted Average Cost of Capital, r_{WACC} (already presented in the note n. 38) is formalized as:

Formula 14: After-Tax Weighted Average Cost of Capital

$$r_{WACC} = \frac{E}{E+D} r_E + \frac{D}{E+D} r_d (1 - Corporate\ Tax\ Rate)$$

Where:

$D\ =\ market\ value\ of\ debt\ (net\ of\ cash)$

$E\ =\ market\ value\ of\ equity$

The reduction in the cost of debt, resulting from the deduction of the Corporate Tax Rate, facilitates the integration of the interest tax shield's advantages into the calculation of the present value of future Free Cash Flows. Consequently, the fundamental principles remain consistent with those elucidated during the introduction of the Discounted Cash Flow method. In this context, the adjustment resides in the computation of the present value (Formula 12), where an after-tax Weighted Average Cost of Capital (WACC) is utilized, incorporating the deduction of interest expenses.

As evidenced from the formula, maintaining a stable Debt to Equity ratio remains highly significant. While companies may endeavour to uphold such stability, it is imperative to acknowledge that this approach could potentially impact the firm's debt levels. If a company wants to compute the amount of debt, also called "debt capacity" (D_t), necessary to keep the leverage constant, it must compute the levered continuation value, V_t at each time and multiply it by the target debt-to-value ratio, d:

Formula 15: Debt Capacity

$$D_t = d \times V_t^L$$

The debt capacity represents the amount of new debt (incremental debt) that has to be issued each year to keep the debt-to-value constant. V_t^L (with t that goes from 0 to N) represents the present value of future cash flows (for a five-year project, V_2^L is defined as the present value of future cash flows from year two and above but not those preceding it).

1.5.2 Adjusted Present Value (APV) Method

The Adjusted Present Value (APV) method, originally formulated by Steward J. Myers (1974), represents the second alternative in valuation with leverage situations. APV can directly show the advantages of choosing a financing structure with a mix of debt and equity rather than equity alone. Myers, in his 1974 paper[40], says:

> "Everyone seems to agree that there are significant interactions between corporate financing and investment decisions. The most important argument to the contrary-embodied in Modigliani and Miller's (MM's) famous Proposition I specifically assumes the absence of corporate income taxes; but their argument implies an interaction when such taxes are recognized."
>
> "Perhaps the greatest advantage of the APV concept is that it guides the corporate financial manager through various problems that turn into a can of worms when analysed by any approach relying on the cost of capital. Here are some examples:
>
> 1. APV provides a natural basis for analysis of the lease vs. buy or lease vs. borrow decision …
>
> 2. Suppose subsidized borrowing is available for certain investments (e.g., for pollution control facilities). How does this affect the investment's value? The impact is clear in the APV framework."

[40] Steward C. Myers, "Interactions of corporate financing and investment decisions – Implications for capital budgeting", The Journal of Finance, 1974.

What is typical in the APV method is how it computes the Levered Value of the target company; unlike others, APV first computes the company's Unlevered Value, and then, adds the present value of the Interest Tax Shield[41]:

Formula 16: Adjusted Present Value Method

$$V_0 = Unlevered\ Value + PV(Interest\ Tax\ Shield)$$

This way it directly identifies the advantage of recurring to debt financing, and, for this reason, this kind of method is frequently used in Leveraged Buyouts (LBOs) situations (because of the changing capital structure).

Myers declarations discussed above help to explain why APV method is preferred in those situations in which Debt to Equity ratio is destined to change, while WACC and Flow-to-Equity Methods (whose details will be described in the next section) are more suitable for constant target Debt to Equity situations. This is clear since, as it has been shown before, in the WACC method it would be necessary to change the WACC calculation every time the financing structure changes to account for each level of leverage changing[42].

Because of its structure, the APV method is summarized in three steps:

1. computation of the Unlevered Cost of Capital
2. computation of the Unlevered Value of the Firm/Project
3. computation of the Present Value of the Interest Tax Shield

Computation of the Unlevered Cost of Capital: the Unlevered Cost of Capital is simply the Pre-Tax WACC:

$$r_U = \text{Pre-tax } WACC = \frac{E}{E+D} r_E + \frac{D}{E+D} \times r_d$$

Or the WACC without the tax benefit. Pre-tax WACC is, in here, chosen because suggests that the cash flows must be discounted without considering the tax deductibility of interest

[41] Includes also other incremental value of debt financing like costs of debt issuance, financial distress costs, and subsidies to debt financing (e.g. below-market rate loan).
Stephen Ross, Randolph Westerfield, Jeffrey Jaffe, "Corporate Finance", McGraw-Hill Education, 2022.

[42] Steven N. Kaplan, Richard S. Ruback, "The valuation of cash flow forecasts: an empirical analysis", The Journal of Finance, 1995.

expense. At the same time, unlevered cost of capital permits to apply the same discounting factor to the whole company's projects since it assumes a constant Debt to Equity ratio. In here, especially, it is possible to notice that the Unlevered Cost of Capital is assumed by generalizing the current company's structure and adapting the varying levels of leverage to each project by separately calculating the present value of the Interest Tax Shield. The Unlevered Cost of Capital stays the same for all the projects involved and, for the general company valuation, it is applied the same rule. To consider Unlevered Cost of Capital as the overall, independent from leverage, expected return are necessary two assumptions:

1) the company riskiness doesn't change with the increase in leverage;
2) the size of the tax shield does not change the overall riskiness of the firm.

Computation of the Unlevered Value of the Firm/Project: Free Cash Flows are, then, discounted by the unlevered cost of capital; this way is calculated the project or company unlevered value. The unlevered value of the firm/project corresponds to all equity financed value of the firm/project. Even if the Unlevered cost of capital considers the cost of debt, it doesn't account for the interest tax shield and supposes that there will not be any change in financing structure (or the presence of any sort of tax rate), this way changes in debt exposure are not considered inside of the unlevered company/project value.

Computation of the Present Value of the Interest Tax Shield: in simple words the present value of the interest tax shield during the life of the project or the company. The tax shield is simply the result of the following multiplication:

$$Interest\ Tax\ Shield = Debt\ Outstanding\ (in\ t-1) \times r_d \times Corporate\ tax\ rate$$

In the context of project valuation with target Debt to Equity ratio the Debt Outstanding is substituted with the debt capacity (formula 14). In this sense, APV becomes more difficult if it is dealing with a target Debt to Equity ratio but, on the contrary, if it deals with changing levels of debt, it becomes the best method because it's no more necessary the computation of debt capacity (which in his turn necessitates the computation of the actualized values of forecasted free cash flows) and it can be directly used the computation of interest expenses. Interest Tax Shield is discounted by Unlevered Cost of

Capital because, as it becomes clear while computing the debt capacity, the more the NPV of the project is high the more the debt capacity is high. Since interest tax shield depends on debt capacity, and debt capacity depends on project's performance, the tax shield will share the same risk of the project or the unlevered cost of capital. This is true until the firm maintains a target leverage ratio.

When it becomes difficult for a company to maintain a constant Debt to Equity ratio, debt can be organized and properly managed in two alternative ways:

1) the **Constant Interest Coverage ratio** supposes that a good proxy of incremental interest payments can directly link interest payments to a constant fraction of Free Cash Flows, since corporate decision about debt is that it increases as earnings increase. This way:

$$Interest\ paid\ in\ Year\ t = k \times FCF_t$$

   And, because interest payments move as the project moves, the discounting factor will be again the unlevered cost of capital:

$$PV(Interest\ Tax\ Shield) = PV(tax\ rate \times k \times FCF)$$
$$= tax\ rate \times k \times PV(FCF)$$

   The present value of the FCF using the unlevered cost of capital as discounting factor is simply the unlevered value of the firm. The levered value will be:

$$V^{levered} = V^{unlevered} + tax\ rate \times k \times V^{unlevered}$$

2) Considering **Predetermined Debt levels,** the computation of interest costs is simple and corresponds to the new issuing debt (known in advance) multiplied by the cost of debt. In this case interest tax shield is not discounted by unlevered cost of capital but by cost of debt since the debt levels are independent by the project or the company.

1.5.3 Flow-to-Equity Method

The last method in here presented is the Flow-to-Equity Method which differentiates from others in evaluating the firm/project by considering the Cash Flow only available to equity holders. The Free Cash Flow to Firm (FCF) is replaced with Free Cash Flow to Equity (FCFE) and, just like WACC method, is discounted and computed the Net Present Value; so, change only the values to be discounted and the discounting factor. In this sense, Incremental Net Income substitutes Unlevered Net Income by including the subtraction of interest expenses. The Flow-to-Equity method considers only the final cash inflows or outflows directly related to equity holders or the amount of additional cash available to pay dividends or conduct share repurchases. Interest expenses are directed to debt holders and do not represent cash available to equity holders, for this reason they are subtracted from the EBIT. So, in general, three are the main differences with the WACC method:

1) Unlevered Net Income is substituted with Incremental Net Income:
$$Net\ Income = Unlevered\ Net\ Income - interest\ Expense \times (1 - tax\ rate)$$
2) Free Cash Flow to Equity is computed by adding Depreciation, substituting CAPEX, substituting increases in NWC and, **this is the difference**, adding the Net Borrowing. Net Borrowing is defined as:
$$Net\ Borrowing_t = D_t - D_{t-1}$$
$D_t = Debt\ Capacity\ in\ t$

Increases in net debt represent more cash coming into the equity holders' pockets, while a decrease usually represents a repayment of debt or a cash outflow from equity holders' pockets.

3) The Free Cash Flow to Equity is discounted by equity cost of capital (like in the WACC method the financial structure is assumed constant, otherwise also the cost of equity can change)

Free Cash Flow to Equity can be summarized in the following way:

$$FCFE = FCF - (1 - tax\ rate) \times Interest\ Expense + Net\ Borrowing$$

Still, despite the differences among the three methods, their results are all the same with equal final Net Present Value (this is theoretically necessary). They differ one from another in the way they are computed and the functionalities in the contingent situations:

- WACC Method is the most suitable for target Debt to Equity ratio situations since it represents the easiest method and mostly used.
- APV Method is the most suitable in changing leverage situations, the other ones would imply a constant recalculation of the discounting factor.
- Flow-to-Equity Method is the most suitable in those situations in which it must be calculated the equity value, a firm’s capital structure is too complex, and the market values of the other securities are not so easy to be found.

This first chapter has shown the principles of capital budgeting and valuation through an in-depth analysis of the various valuation methods, their strengths, and weaknesses. To sum up, in one hand it can be said that Comparable companies’ analysis (and Precedent) is designed to reflect “current” valuation which is also the result of market and price movements. For this reason, in many cases, it is more relevant and precise than intrinsic valuation analysis, such as Discounted Cash Flow analysis. Still, irrational investor sentiment can skew market valuation either too high or too low; market valuation is also subject to stock market cycles in which some sectors or companies can demonstrate to be overvalued. Furthermore, the difficulty in finding perfectly matching comparables, may fail to accurately capture a given company’s true value. On the other hand, intrinsic valuation methods (which presents different facets like DDM, FCF, WACC, APV, and FTE) correct some of the problems related to Comps but introduces the element of prediction inside of company pricing. Prediction and forecasts represent a key and weak spot especially regarding the growth rate in the terminal value calculation which usually accounts for the two thirds of the overall company intrinsic value. A small misalignment in one of the assumptions (even a one percent movement) can result in huge changes in price which can be calculated through sensitivity and scenario analysis. As a result, intrinsic valuation methods must always be used togheter with Comps and Precedents valuation methods, this way it is possible to get a range of values in which market interference cooperates with intrinsic elements. Market and intrinsic company details serve each other to assure and control, in one hand, the proper functioning of the market, and, on the other hand, a proper set of realizable assumptions. This first chapter permits to move on to the next topic and enter in detail with the characteristics of the cost of capital.

Chapter Two

Cost of Capital, a focus on Beta Equity

The discounting factor stands as a pivotal determinant in forecasting, particularly when analysing Free Cash Flows. In the preceding chapter, valuation methods were presented without delving into the computation of the cost of capital or its determinants. Given that even a slight variation in the discounting factor can result in significant alterations in corporate intrinsic value, it is imperative to meticulously examine how the cost of capital is derived, the various factors influencing it and their management. This chapter is anchored in the Weighted Average Cost of Capital (WACC) formula (below the unlevered formula), which is constituted of a weighted average of both the cost of debt and the cost of equity:

$$r_{WACC} = \frac{E}{E+D} r_E + \frac{D}{E+D} r_d$$

The rationale of focusing on the cost of debt and subsequently on the cost of equity becomes apparent. The latter, being notably more challenging to compute, serves as the focal point of this paper, as it entails a concentrated investigation into its attributes and the computation of Beta Equity (a crucial component for determining the cost of equity).

The chapter concludes with an exhaustive exploration of the Determinants of Beta Equity, essential for a comprehensive understanding of its impact on the calculation of the cost of equity. The significance of this final segment paves the way for the subsequent chapter, wherein a Machine Learning algorithm will be employed to capture and calculate a company's Beta Equity based on its information.

2.1 The Cost of Debt

The initial phase in understanding the Weighted Average Cost of Capital involves the computation of the cost of debt ($\boldsymbol{r_d}$). This component is fundamental to a company's operations, as it dictates the anticipated interest expenses and reflects debtholders' expectations regarding the company's debt exposure and default probability. Typically, a company's balance sheet gathers various types of debt rather than just one. However, not all forms of debt are pertinent to the computation of the cost of debt, several factors can impact a false estimation of the cost of debt:

- the **first** factor pertains to secured debts and convertible bonds. While these instruments offer a form of security in the event of default, they do not accurately represent the true cost of debt without any form of guarantee;
- the **second** factor involves the disparity between the debt yield to maturity and the cost of capital. When debt carries a risk of default, its yield may inflate to exceptionally high levels, thereby overstating the anticipated return and failing to provide an accurate value for the cost of capital. As the yield is calculated as the Internal Rate of Return (IRR) of an investment in a corporate bond, it does not incorporate the possibility of default and the potential loss of repayment. In this context, it is beneficial to present an example comparing the yield to maturity (**YTM**) under two scenarios: a Corporate Bond with no default risk and a Corporate Bond with certain default risk.

$$No\ Default\ (Risk\ free)\ Zero\ Coupon\ Bond$$

$Risk\text{-}free\ yield\ to\ maturity$: 3.00%

$Face\ Value$[43]: 10,000.00

$Maturity$: 1 $year$

$$Price = \frac{10{,}000.00}{1 + 3.00\%} = 9{,}708.74$$

$$Certain\ Default\ \ Zero\ Coupon\ Bond$$

$Risk\text{-}free\ yield\ to\ maturity$: 3.00%

$Face\ Value$: 10,000.00

$Maturity$: 1 $year$

$Expected\ default\ payback$: 9,500.00 ($risk\ free$)

$$Price = \frac{9{,}500.00}{1 + 3.00\%} = 9{,}223.30$$

$Therefore, the\ yield\ to\ maturity\ of\ the\ Corporate\ Bond\ will\ be$:

$$YTM = \frac{Face\ Value}{Price} - 1 = \frac{10{,}000.00}{9{,}223.30} - 1 = 8.41\%$$

When the Default is certain the YTM increases considerably even if the expected return (the risk-free yield to maturity in this case) stays the same; for this reason, when considering the cost of debt, it is necessary to properly manage the effects of default risk into the computation of the YTM.

- **finally**, it is plausible that certain debts of the company were contracted during a notably advantageous or disadvantageous economic period. In such instances, the yield to maturity may significantly deviate from the current rate due to the markedly altered condition of the company or its exposure to default risk.

[43] The value that has to be paid back by the debtor at maturity.

Addressing these discrepancies appropriately becomes imperative in these scenarios as well.

The computation of the cost of debt is not univocal but, as in everything that involves valuation, is subject to a series of circumstances that change drastically how it is managed and calculated. In general, the fastest way of dealing with the cost of capital is to simply divide the interest expenses by the total debt outstanding:

Formula 17: Cost of debt, first method

$$r_d = \frac{interest\ expenses}{total\ debt\ outstanding}$$

This way, the cost of debt is presumed to be the average of the different interest rates stipulated in the precedent years and of the different debt instruments in the balance sheet. Even if this method represents a fast option, it usually becomes inappropriate because generalizes the cost of debt and doesn't consider the present as the moment of valuation since, among all the debt outstanding it can figure some debt stipulated even a decade before. For this reason, two are the main pitfalls in using this technique:

1) as it has been stated before, from the balance sheet can emerge a series of different debt instruments that have nothing to do with the cost of debt interpreted as the expected return of a borrowed sum against the risk of not getting paid back;
2) most of the already in place debt instruments were stipulated in a different period or in a different company's life cycle. This means that, in the present, the cost of debt precedingly agreed upon doesn't necessarily (or most probably) represent the current financial or managerial company's condition.

A solution to the first issue is represented by the utilization of the Weighted Average Cost of Debt:

Formula 18: Cost of debt, Weighted Average Cost of Debt

$$Weighted\ Average\ Cost\ of\ Debt = \sum_{n=1}^{N} YTM\ of\ instrument\ \times weight\ of\ the\ instrument$$

This way, the debt instruments inserted into the computation of the cost of debt will be only those selected as more appropriate and coherent with the discounting purpose of the Free Cash Flow. The utilization of a weighted average cost of debt proves advantageous under circumstances where identifiable debt instruments have been recently issued, or where debt instruments were issued during a period within the company's lifecycle characterized by a default risk similar to the present conditions. In such scenarios, the current cost of debt can be accurately delineated as the weighted average of selected interest rates. Conversely, in situations different from this one, the employment of a weighted average becomes overly generalist and fails to reflect the precise contemporary cost of debt, which may diverge significantly from historical values.

For instance, during periods of high inflation, the cost of debt experiences a notable escalation. In recent years, the interest rates on bank loans surged from less than 1% in 2020 to 3.5% and higher presently. Numerous studies have illustrated a substantial decline in companies' net present values (NPVs) over the past years, primarily attributable to a rapid elevation in the cost of debt and consequently, a significant augmentation in the WACC[44]. It's clear that using past debt instrument yields to calculate the present cost of debt doesn't match today's conditions. Even if a company's rating hasn't changed, today's cost of debt has risen significantly due to recent global events. For these reasons, it is necessary to define different approaches to cost of debt computation.

There are two additional methods used in the cost of debt computation whose utilization depends upon which one of the previously explained issues can occur.

In instances in which the company has recently just issued some bonds and it is possible to singularly select them, the primary concern is to adjust the yield to maturity to consider the default risk in order to get the effective cost of debt. The most appropriate approach in this context is as follows:

Formula 19: Cost of debt considering probability of default

$$r_d = Yield\ to\ Maturity - Probability\ of\ default \times Expected\ Loss\ Rate$$

[44] ____, KPMG, "Cost of Capital Study", 2023.

In this formula, the cost of debt is derived by subtracting the product of the loss rate in the event of default and the probability of default from the yield to maturity. Consequently, the cost of debt represents the anticipated return for a debt holder, which is not solely the yield to maturity but is adjusted to reflect the expected risk of not recovering the full amount.

However, the drawback of employing this method lies in the uncertainty surrounding both the probability of default and the expected loss rate. Typically, rating agencies accompany their assessments of firms' creditworthiness with average default rates per rating category, distinguishing between estimates for recessionary and normal market conditions. Below is presented an example of Global Corporate Annual Default Rates by Rating Category from a Standard & Poor's 2021 Annual Global Corporate Default And Rating Transition Study:

(%)	AAA	AA	A	BBB	BB	B	CCC/C
2016	0.00	0.00	0.00	0.06	0.47	3.76	33.17
2017	0.00	0.00	0.00	0.00	0.08	1.00	26.67
2018	0.00	0.00	0.00	0.00	0.00	0.94	27.18
2019	0.00	0.00	0.00	0.11	0.00	1.49	29.76
2020	0.00	0.00	0.00	0.00	0.94	3.53	47.68
2021	0.00	0.00	0.00	0.00	0.00	0.52	10.99

Sources: S&P Global Ratings Research and S&P Global Market Intelligence's CreditPro®.

Annual Default rates vary considerably from a year to another, this strengthens the aforementioned concern about how much a proper default rate is difficult to predict and how much volatile would become the computation of the cost of debt.

Regarding the Expected Loss Rate, the availability of information and real-world cases varies, making it challenging to apply a generalized average loss rate. Each firm operates within its unique circumstances, rendering it difficult to rely solely on average loss rates.

Beyond the conditions mentioned above this method doesn't represent the best option as its applicability is not always assured and can increase the level of uncertainty in the prediction.

An alternative method takes the fundamentals of the Capital Asset Pricing Model (CAPM), whose specifics will be covered in detail in the next paragraph, and defines the cost of debt in the following way:

Formula 20: CAPM Cost of Debt

$$r_d = Risk\ Free\ Rate + \beta_{debt} \times (Market\ Return - Risk\ Free\ Rate)$$

The cost of debt is built as the sum of the risk-free rate, which accounts for macroeconomic turmoil and eventual inflation[45], and the product of the Beta debt by Market Risk Premium (the return of the market subtracted by the risk-free rate). The non-risk-free part of the cost of debt represents, therefore, the additional return that a debt holder is expecting for holding a certain debt instrument. The additional return is computed by looking at the correlation between the market return and the historical debt instrument return (β_{debt}). The Beta debt could be calculated individually for each company but usually this is not possible since bank loans and bonds are not always, if rarely, traded. A solution could be represented by the utilization of average debt Betas depending on the Rating, the Maturity or, also, the sector; this way the Beta debt will be simply substituted with an average value calculated thanks to debt instruments regularly traded (whose Beta debts can be properly calculated). Finally, there is another way of computing Beta debt, but this necessitates the introduction of the Option Pricing Theory whose specifics completely deviate from the purpose of this paper. For this reason, it will be just mentioned the existence of this alternative technique without entering into the details.

[45] In this case the direct presence of the risk-free rate permits to accurately account for eventual macroeconomic conditions. According to Fisher's Formula, the presence of inflation or its expectation directly increases the nominal interest rates:

$$Nominal\ interest\ rate = Real\ Interest\ Rate + Inflation\ (or\ expected\ inflation)$$

Still, Central Banks' decisions can have an additional impact on interest rates since a restrictive policy can lead to an increase of policy rates (the interest rate set for loans towards commercial banks) and, finally, an overall increase in interest rates. In both the cases this would result in an increase of risk-free interest rates.

2.2 The Cost of Equity (Listed Companies)

The equity cost of capital should represent the expected return of an instrument or investment with a risk comparable or similar to that of the target company. The purpose of this paragraph is to define a proper rule that can correctly quantify and identify this expected return.

The Capital Asset Pricing Model (CAPM) represents the most used method to derive the cost of equity, its wide application is mainly caused by its high practicality and simplicity in adapting it to all situations. The rationale behind CAPM is that all stocks' returns can be compared to the market's one, in this scenario it becomes relevant the correlation between market return and stock return. The formal equation of CAPM is the following:

Formula 21: CAPM Equation

$$r_E = r_f + \beta \times \left(E[R_{mkt}] - r_f\right)$$

$r_f = r$isk-free rate

$\beta = \text{stock correlation to market}$

$E[R_{mkt}] = \text{Expected Return of Market Portfolio}$

$E[R_{mkt}] - r_f =$ Market Risk Premium

The stock's required return, if completely correlated to the market ($\beta = 1$), will be equal to the return eventually earned if invested in the market portfolio (which is used as proxy of the market performance). If the correlation differs from one, the return will increase as the correlation increases; in this sense, the market portfolio return represents the comparable element able to quantify the company risk. If the correlation is higher than one, then the risk of the stock is higher than the market's one which results in a higher stock return. Conversely, if the stock Beta is lower than one, the risk of the stock is lower than the market's one and, therefore, a lower expected return than the market portfolio.

To gain a comprehensive understanding of the Capital Asset Pricing Model (CAPM), the subsequent sections will elucidate its foundational principles and constituent elements one by one. Firstly, it will be presented the description of how the market portfolio is computed, followed by an exploration of the **Market Risk Premium**. Finally, the discussion will culminate with the process of estimating Beta.

The **Market Portfolio** is not, per se, difficult to construct, but hides a series of uncertainties and variables that directly affect its computation. First, the market portfolio represents the best possible proxy of the market, in this sense, it would be utopic and ideal to include in it all the existing stocks. Since the market is constituted by hundreds of thousands of listed securities it becomes particularly hard to specifically identify each one of them and create a portfolio constituted by everyone. Second, each stock is not worth the same in the market, this means that for example the AAPL shares are way more numerous and, maybe, expansive than a small cap firm just listed. In these circumstances, it becomes evident that it is not possible to give for each stock the same weight. For this reason, the difficult task of identifying all the stocks present in the market becomes even more difficult if for each one it has to be computed also its weight.

Two are the possible solutions, and in both cases the market portfolio is not composed of all the available stocks but is a mimicking market portfolio:

1) **construct your own market portfolio**: because the market portfolio represents the total supply of securities, and the proportion of each security, as it has been said before, has to correspond to the proportion it represents in the total market, each security will have a weight calculated considering the market capitalization:

$$x_i = \frac{Market\ Capitalization\ i}{Total\ Market\ Capitalization\ of\ all\ Securities\ in\ the\ Portfolio}$$

The investment in each security – i – is equal to its weight calculated in the abovementioned way multiplied by the total available amount of money to invest. This type of portfolio is called a **value-weighted portfolio.** The main drawback in this method is represented by the fact that if the price of a stock changes, so does the market capitalization; this implies a constant rebalancing of the weights that have to be recalculated and have to reflect the current market capitalization proportions.

2) **Choosing Market Indexes**: When the focus of the analysis are stocks traded in specific countries with high capitalization volumes (United States is the most relevant one), a second, easier to apply, option becomes effective: to choose as market portfolio an **index**. The most famous value-weighted index is the S&P 500 which collects 500 leading companies publicly traded in the U.S. which cumulatively account for almost the 80% of all the U.S. market capitalization. It is a float-adjusted (considers only shares available to the public, doesn't consider those held by the management, governments, and employees) index whose introduction comes back to the middle of 1900 (1957). Even if the S&P 500 was introduced so long ago, and in recent times new indexes like the Wilshire 5000 (which considers all the stocks listed in US major stock exchanges) have been introduced, the results among the two are almost identical. In fact, correlation between S&P 500 and Wilshire 5000 is almost one and indicates a perfect correlation. The utilization of already existing indexes completely simplifies the whole process but restricts the analysis to stocks from the same country and doesn't permit to generalize in order to simultaneously compare different stocks from different countries.

Once the market portfolio has been constructed (or identified), the next step is to compute its return. In practice, the market portfolio is computed by firstly taking all the weekly (or monthly) data of the stocks composing the portfolio along with their market capitalization; then, once all the data is saved, the next step is to compute weekly (or monthly) returns. Those returns are finally multiplied by the weight of the stock and then they are all summed up. The market weekly (or monthly) returns are simply the sum of the weighted returns of the portfolio composing securities.

This paragraph concludes with a final part in which it will be showed a Python code whose purpose will become clearer at the end of this paragraph. In here, since this part is less understandable if not supported by a proper visualization, it will be exhibited a practical example of how, with the utilization of Python, a market portfolio is built.

Python programming language, for those unfamiliar with it, represents a useful tool especially when dealing with a huge amount of data; its widespread use is mainly justified by its simplicity and adaptability to various tasks. The utilization of a programming language, rather than directly using a program like Excel, permits to tailor the requests and the functions used. So, in a practical case like this, it is possible to create new functions and algorithms that permit to fasten up and automatize all the dataset creation process.

Another great Python resource is represented by the possibility of taking advantage of other users' or developers' work by simply importing existing libraries which are a collection of different algorithms. The libraries most used in here are for sure NumPy and Pandas for Exploratory Data Analysis (EDA) and YFinance for the download of financial information; the first two permit to exhibit the data in the so called "DataFrame" which simplify the data visualization and open to numerous other functions while YFinance is a collection of functions that get access to Yahoo Finance Databases.

The computation of the market portfolio in Python follows three steps.

First, a portfolio is built by selecting a series of securities, this is done by searching from different databases and files that collect the tickers (the names by which the different companies are listed in the stock exchanges, for example Apple Inc.'s ticker is AAPL). The first thing to do is to try looking for a collection of tickers in a .csv or .txt file, this collection must consider the highest amount possible of securities from different countries. Especially because the stocks considered are not contingent to a specific country, choosing to build your own market portfolio becomes the best choice since the available indexes do not consider the worldwide securities but indicative to the same country (in the U.S. the S&P 500, in Italy the FITSE MIB and so on). In this case it was possible to find a publicly available database published in 2018 with more than a hundred thousand tickers. Of these ones some are no more listed and other ones are over the counter securities (**OTC**). Because of this, and because the list of tickers stops in 2018, they have been considered two additional databases found online containing the list of European and U.S stocks' tickers. The process of extrapolating the tickers and searching them on Yahoo Finance to collect all the financial data and the company characteristics is always the same in all the three databases.

```python
# Importing the necessary libraries (Pandas collects also the majority of NumPy functions)
import pandas as pd
import yfinance as yf
import numpy as np
```

```python
df_new = pd.read_csv('YFtickers.csv', sep=';') # The first .csv file that has to be read in Python
tickers_new = df_new['Ticker'].values.tolist() # tickers are collected from the database
```

```python
beta_list = []
list_current = []
list_DE = []
list_gross = []
list_cap = []
list_country = []
list_exchange = []
list_currency = []
list_sector= []
list_name = []
list_industry = []
list_ticker = []

for ticker in tickers_new:
    try:
        stock_data = yf.Ticker(ticker).info # look in Yahoo Finance to see if the ticker is still existent
        beta_list.append(stock_data.get('beta')) # get the beta calculated by Yahoo Finance
        list_ticker.append(ticker)
        list_current.append(stock_data.get('currentRatio')) # get the Current ratio
        list_DE.append(stock_data.get('debtToEquity')) # get the Debt to Equity Ratio
        list_gross.append(stock_data.get('grossMargins')) # get the Gross Margin
        list_cap.append(stock_data.get('marketCap')) # get the stock's Market Capitalization
        list_country.append(stock_data.get('country')) # country in which is founded
        list_exchange.append(stock_data.get('exchange')) # stock exchange in which it is traded
        list_currency.append(stock_data.get('currency')) # currency of the stock exchange (currency of the price)
        list_industry.append(stock_data.get('industry')) # in which industry it operates
        list_sector.append(stock_data.get('sector')) # in which sector it operates
        list_name.append(stock_data.get('longName')) # the complete name
    except Exception as e: # avoids the interruction of the for loop if it is raised any error
        # Print the error and continue with the next iteration of the loop
        print(f"Error for ticker {ticker}: {e}")
        continue
```

```python
data_values=[list_ticker,list_name,list_country,list_exchange,list_currency,list_sector,
             list_industry,beta_list,list_current,list_DE,list_gross,list_cap]
transposed_data = list(map(list, zip(*data_values)))
nat_stocks = pd.DataFrame(transposed_data, columns=['Ticker','Name','Country','Exchange','Currency','Sector',
                                                    'Industry','Beta', 'Current Ratio','Debt to Equity',
                                                    'Gross Margin','Market Cap'])

new_dataframe = nat_stocks # create a copy of the just downloaded data
new_dataframe.to_csv('OneDrive\Desktop\Tesi Magistrale\FirstDatabase.csv', index=False) # save it in the PC memory
```

The data collected is saved in a specific format defined "DataFrame" which permits to view the data in an organized and more intuitive way:

```python
new_dataframe.head()
```

	Ticker	Name	Country	Exchange	Currency	Sector	Industry	Beta	Current Ratio	Debt to Equity	Gross Margin	Market Cap
0	OEDV	None	None	YHD	None	None	None	NaN	NaN	NaN	NaN	NaN
1	AAPL	Apple Inc.	United States	NMS	USD	Technology	Consumer Electronics	1.276	1.073	145.803	0.45027	2.606901e+12
2	BAC	Bank of America Corporation	United States	NYQ	USD	Financial Services	Banks - Diversified	1.388	NaN	NaN	NaN	2.913224e+11
3	AMZN	Amazon.com, Inc.	United States	NMS	USD	Consumer Cyclical	Internet Retail	1.160	1.045	80.037	0.46982	1.869732e+12
4	T	AT&T Inc.	United States	NYQ	USD	Communication Services	Telecom Services	0.701	0.713	135.038	0.59059	1.256745e+11

The same procedure is repeated for the other two databases and, after having deleted exact duplicates, the Final DataFrame is the following:

```python
filtered_df = pd.read_csv('FinalData2.csv', sep=',')
filtered_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52604 entries, 0 to 52603
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Ticker          52604 non-null  object 
 1   Name            52554 non-null  object 
 2   Country         34477 non-null  object 
 3   Exchange        52523 non-null  object 
 4   Currency        50020 non-null  object 
 5   Sector          31961 non-null  object 
 6   Industry        31961 non-null  object 
 7   Beta            33094 non-null  float64
 8   Current Ratio   29699 non-null  float64
 9   Debt to Equity  27510 non-null  float64
 10  Gross Margin    29717 non-null  float64
 11  Market Cap      35854 non-null  float64
dtypes: float64(5), object(7)
memory usage: 4.8+ MB
```

The resulting DataFrame is not yet totally ready for proceeding with the creation of the market portfolio. Before of passing to the second phase, it is necessary to do a bit of data cleaning[46]. All the functions and commands used in this DataFrame for data cleaning purposes will be exhibited in the last chapter with the Explanatory Data Analysis (EDA). In this part of the chapter the focus is on the construction of the Market Portfolio; because of this, the DataFrame will be showed already cleaned and ready to proceed to the second step.

Second, after having looked at all the factors and having cleaned all the DataFrame, the remaining tickers are only those ones that have the market capitalization available.

```python
filtered_df = filtered_df.dropna(subset=['Market Cap'])
```

The market capitalization, necessary for the calculation of the stock weights in the market portfolio, must be standardized by multiplying the exchange rate in US Dollars

[46] Represents the procedure that permits to organize the data, delete possible outliers, and avoid discrepancies among the values. Data cleaning includes various tasks:
- Handling Missing Values: hypothesize missing values through imputation (replacing missing values with estimated values) or deleting missing values;
- Handling Duplicates (as abovementioned);
- Standardizing Data: converting the data into a standardized scale when there is too much variance and the values become too high, or, in case of values considering currencies, to standardize all the data with the same currency;
- Correcting Errors: modify some recurring errors occurred in the downloading process;
- Handling Outliers: this process is usually done by considering different perspectives. For example, in this paper the columns of the DataFrame consider heterogenic parameters such as the company's country origin, the company's exchange market, the company's revenues and so on. The presence of a limited number of companies from a certain state or from a certain market exchange is not so meaningful and it could constitute unnecessary data that creates noise.

corresponding to the currency used in the market capitalization amount by the market capitalization amount.

```
from forex_python.converter import CurrencyRates # takes real time exchange rates

c = CurrencyRates()
print(c.get_rate('USD', 'GBP')) # how many pounds is worth one dollar
```

```
0.794994427934621
```

```
currencies = list(filtered_df['Currency'].value_counts().index)
exchange_dict = {}
for i in currencies: # specified exchange rates date back to 18/01/2024
    if i=='TWD':
        exchange_dict[i] = 31.55
    elif i == 'QAR':
        exchange_dict[i] = 3.64
    elif i == 'RUB':
        exchange_dict[i] = 89.43
    elif i == 'ARS':
        exchange_dict[i] = 818.69
    elif i == 'ILA':
        exchange_dict[i] = 3690
    else:
        exchange_dict[i]=c.get_rate('USD', i.upper())
```

The data expressed in foreign currency:

```
filtered_df[6451:6454][['Ticker','Name','Currency','Market Cap']]
```

	Ticker	Name	Currency	Market Cap
6451	BMO.TO	Bank of Montreal	CAD	9.497512e+10
6452	BLND.L	British Land Company PLC	GBp	3.561504e+09
6453	BIMBOA.MX	Grupo Bimbo, S.A.B. de C.V.	MXN	3.157672e+11

Will be converted and expressed in dollars:

```
list_exchange = []
for i in filtered_df['Currency']:
    if pd.isna(i):
        list_exchange.append(None)
    elif i in exchange_dict:
        list_exchange.append(exchange_dict[i])
    else:
        list_exchange.append(None)
filtered_df['Exchange Rate'] = list_exchange
filtered_df['Exchange Rate'] = filtered_df['Exchange Rate'].replace({np.nan: 1, 0: 1})
filtered_df['Market Cap'] = filtered_df['Market Cap'] / filtered_df['Exchange Rate']
```

```
filtered_df[6451:6454][['Ticker','Name','Currency','Exchange Rate','Market Cap']]
```

	Ticker	Name	Currency	Exchange Rate	Market Cap
6451	BMO.TO	Bank of Montreal	CAD	1.348254	7.044305e+10
6452	BLND.L	British Land Company PLC	GBp	0.794994	4.479911e+09
6453	BIMBOA.MX	Grupo Bimbo, S.A.B. de C.V.	MXN	17.044577	1.852596e+10

Third, the weekly prices over the last five years of the remaining (after the cleaning) tickers are downloaded and saved in a Python dictionary and then, converted in a Pandas DataFrame:

```python
prices_dict={}
for ticker_symbol in list(final_df['Ticker']):
    try:
        ticker = yf.Ticker(ticker_symbol)
        # Download historical price data
        data = ticker.history(period="5y", interval="1wk")
        prices_dict[ticker_symbol]=(list(data['Close']),filtered_df[filtered_df['Ticker'] == ticker_symbol]
                                   ['Market Cap'].iloc[0])
    except Exception as e:
        print( f'An error occurred: {str(e)}')
```

```python
prices_data=pd.DataFrame(prices_dict).T
prices_data.columns=['Prices','Market Capitalization']
prices_data[:5]
```

	Prices	Market Capitalization
A	[79.3819351196289, 74.6668472290039, 75.892173...	42410921984.0
AA	[28.22835350036621, 26.601903915405273, 26.851...	6579041792.0
AACG	[2.5, 2.6700000762939453, 2.0399999618530273, ...	33778504.0
AAL	[34.23106384277344, 33.91529083251953, 32.6226...	9009456128.0
AAMC	[8.958824157714844, 8.300000190734863, 6.91176...	9107809.0

Fourth, the total market capitalization is computed by summing all the market capitalizations and, through this value, it is possible to determine the individual stock's weight into the market portfolio:

```python
prices_data['Market Capitalization'].sum() # total Market Portfolio Capitalization
```

```
104009289664449.06
```

```python
prices_data['stock_weight'] = prices_data['Market Capitalization']/prices_data['Market Capitalization'].sum()
```

The total market capitalization amounts to approximately 104 trillion dollars, value similar to the worldwide cumulative market capitalization of 109 trillion dollars (of more than 50.000 listed companies worldwide). The stocks weight is near to zero and has almost no impact for most of the stocks; only the most famous and appreciated ones, like Apple (AAPL), have a certain relevance:

```python
prices_data[10:15]
```

	Prices	Market Capitalization	stock_weight
AAPL	[48.615089416503906, 49.834938049316406, 49.94...	2620413181952.0	0.025194
AAT	[42.174808502197266, 41.26216506958008, 42.884...	1602447488.0	0.000015
AAU	[0.5683000087738037, 0.550000011920929, 0.5550...	20102876.0	0.0
AB	[22.70430564880371, 23.08964729309082, 22.4576...	3927443456.0	0.000038
ABBV	[68.25521087646484, 66.41194152832031, 68.2355...	301931233280.0	0.002903

Apple weights approximately 2.5% on the overall market portfolio, this means that its eventual variation of weekly returns has a larger impact on the market weekly return compared to the other stocks just showed.

Finally, the stock prices are transformed into stock returns and are multiplied by their weights:

```python
def weekly_return(x):
    if isinstance(x, list) and len(x) == 261:
        return pd.Series(x).pct_change().to_numpy()
    else:
        return pd.Series([0] * 261).to_numpy()  # Return a series of zeros if the input is not valid

# Apply the function to the 'Prices' column and store the result in the 'Weekly Return' column
prices_data['Weekly Return'] = prices_data['Prices'].apply(weekly_return)
prices_data['Weighted Return'] = prices_data['Weekly Return']*prices_data['stock_weight'] # Weekly Returns are weighted
```

```python
prices_data[10:15]
```

	Prices	Market Capitalization	stock_weight	Weekly Return	Weighted Return
AAPL	[48.615089416503906, 49.834938049316406, 49.94...	2620413181952.0	0.025194	[nan, 0.025091975505004083, 0.0021582326298441...	[nan, 0.0006321679879427496, 5.437457799402348...
AAT	[42.174808502197266, 41.26216506958008, 42.884...	1602447488.0	0.000015	[nan, -0.021639539455635903, 0.039320660808464...	[nan, -3.333954664437359e-07, 6.05804484794599...
AAU	[0.5683000087738037, 0.550000011920929, 0.5550...	20102876.0	0.0	[nan, -0.03220129609422295, 0.0090909002241026...	[nan, -6.223854277919487e-09, 1.75708574227455...
AB	[22.70430564880371, 23.08964729309082, 22.4576...	3927443456.0	0.000038	[nan, 0.016972183613437775, -0.027369755795343...	[nan, 6.408782492570992e-07, -1.03349584097275...
ABBV	[68.25521087646484, 66.41194152832031, 68.2355...	301931233280.0	0.002903	[nan, -0.02700554762742824, 0.0274588138470668...	[nan, -7.839509650394436e-05, 7.97108946325649...

By summing the weighted returns is calculated the market portfolio weekly return:

```python
market_portfolio_return = prices_data['Weighted Return'].sum() # sum the weighted returns

formatted_arr = [f'{val * 100:.2f}%' for val in market_portfolio_return] # format in % the market portfolio return
formatted_str = ' '.join(formatted_arr)

# Print the formatted string
print(formatted_str)
```

```
nan% 0.43% 0.58% 0.37% -2.53% -0.83% -1.11% -1.45% 3.13% 0.06% 2.46% 0.33% 1.31% 0.31% -0.45% 1.22% -2.79% -0.82% -0.69%
-0.38% 2.30% 1.87% 1.42% -0.14% -0.87% -2.25% 1.16% 0.73% 1.60% 1.20% 1.25% 0.37% -0.31% 0.95% 0.19% 1.31% 1.72% 0.64% -0.
51% 0.98% 1.61% -0.87% -2.44% 3.06% 1.72% -2.93% -9.48% 0.42% -8.80% -10.91% 9.61% -1.91% 10.26% 2.53% -0.84% 2.42% 2.52%
-2.05% 3.19% 3.15% 6.34% -3.67% 2.07% -1.82% 2.98% 2.61% 1.20% -0.08% 0.46% 2.44% 1.24% 0.43% 2.42% -1.76% -1.20% 0.34%
-1.65% 1.40% 1.38% -0.09% 0.19% -4.78% 6.59% 3.54% 0.86% 2.37% 1.47% -0.26% 1.40% -0.15% 1.30% 4.77% -0.45% 1.38% -3.18%
4.43% 1.68% 0.11% -1.87% 0.95% 2.55% -0.43% 0.68% 1.39% -2.32% 1.32% -0.26% 0.10% 1.42% -1.10% 0.41% 1.57% 1.25% 1.01% -0.
97% 2.14% 0.60% -0.01% -0.44% 1.33% -0.10% 1.20% 0.49% -2.57% 2.19% 1.18% -0.92% -0.39% 0.40% -1.56% -1.14% 1.87% 0.95% 0.
64% 2.14% 0.40% -0.15% -2.40% -1.02% 2.91% -1.32% 1.81% 0.74% -2.51% 0.31% -3.85% -0.42% 1.37% -0.23% -1.03% -0.17% -1.57%
-1.42% 5.47% 1.10% 0.44% -1.00% -1.37% -2.31% -1.71% -1.30% -1.46% -1.06% 4.34% -0.28% -3.83% -5.01% 3.92% -1.90% 1.96%
-1.12% 2.93% 3.15% 0.58% 2.42% -1.01% -2.44% -3.04% 3.36% -3.43% -4.07% -2.19% 1.77% -1.50% 3.57% 2.43% -0.60% 5.79% -0.4
1% 1.65% 1.29% -2.18% -1.72% -0.29% 0.04% 2.08% 3.00% 0.26% 2.42% 1.42% -1.01% 0.11% -1.75% 1.87% -3.34% 0.28% 1.02% 2.99%
-0.13% 1.25% -0.05% 0.52% -0.36% 0.03% 1.57% 0.34% 1.38% 0.58% 2.06% -1.65% 1.93% -1.06% 2.44% 0.59% 1.41% -1.70% -0.35%
-2.03% 0.69% 2.42% -0.92% 0.49% -2.22% -0.56% -0.34% 0.44% -2.26% -1.48% 4.97% 0.58% 2.46% 0.64% 0.81% 0.42% 2.14% 0.76%
0.65% -1.04% 1.26% 0.21% 1.64% 0.94% 1.19% 0.56% 1.47% 0.83% 0.44% -0.17% 1.72% 0.48% -0.41%
```

The Expected Market Return $E[R_{mkt}]$ is simply the average of the series of returns (annualized):

```python
exp_market_return_w = market_portfolio_return[1:].mean() # starting from the second element because first one is NaN
exp_market_return_w
```

```
0.003475722591814486
```

Almost a **weekly** 0.35% return, in annual terms:

$$anunual\ compund\ return = (1 + weekly\ return)^{52} - 1$$

The Expected Market Return is almost 20% (annual).

```python
exp_market_return = (1 + exp_market_return_w)**52-1
exp_market_return
```

```
0.1977253350772481
```

The definition of Expected Market Return is necessary for the explanation of the **Market Risk Premium** that was previously defined as:

$$Market\ Risk\ Premium = E[R_{mkt}] - r_f$$

The market risk premium represents the additional return benchmark used by the investors when comparing a risky asset with a risk-free asset. The missing term of this formula is the risk-free rate that, so far, has not been defined properly. In practice, the risk-free rate is the rate at which an investor can invest (save) with the certainty of receiving back the money.

When analysing the **risk-free rate**, an investor must consider three distinct aspects:

1) the **place** in which the company object of the analysis has its core business (if the company has the fiscal headquarter in Italy but the goods and services are sold for the majority in Europe, then the focus is not exclusive to Italy);
2) which **type** of risk-free instrument to use; almost the totality of times it is used **a government bond of the most virtuous state involved**. If, from the example above, the states involved are different and it has to be chosen the most virtuous one, then, the German Bond would be the best option;
3) the **maturity** depends on the life of the project. If it is analysed a short-term project, then it will be considered a bond with short-term maturity; this way the maturity is consistent with the life of the project. If, on the contrary, the life of the project has a large or indefinite horizon – as it happens in stock valuation – then it will be considered a long-term bond (10- or 30-year bond).

The **Beta ($\boldsymbol{\beta}$) estimation** corresponds to the last step of the CAPM and cost of equity computation. In a formal definition, Beta is the measure of sensitivity of a security's returns to the returns of the market. In practice, Beta represents the level of correlation between a security and the market; this correlation can assume different values:

- $Beta < 0$ the asset tends to move in the opposite direction of the market. Those assets are often used to hedge against market downturns, for this reason they are often referred to as "**defensive**" or "**contrarian**". An example of defensive assets are the Real Estate Investment Trust (**REIT**) stocks that are known for their stable income streams from rental properties whose results are often derived from property specific factors;

- $Beta = 0$ the asset has no correlation with the market. An example of zero Beta assets are **healthcare** companies whose performance does not depend upon macroeconomic factors since the population necessitates it;
- $0 < Beta < 1$ the asset is less volatile than the market;
- $Beta = 1$ the asset moves like the market. The typical assets that have almost unitary Beta are usually those companies whose performance moves along with macroeconomic factors; the **energy** and **technology** sectors are known examples of frequently unitary Betas;
- $Beta > 1$ the asset is more volatile than the market. In this section are frequently recognized companies providing goods and services in the **Consumer Discretionary** sector. Luxury goods, travel services or entertainment are frequently exposed to macroeconomic trends such as changes in fuel prices, geopolitical events, or economic conditions. For example, during periods of economic expansion or robust travel demand, travel services companies may experience higher Beta values as their stock prices become more sensitive to changes in consumer behaviour and industry-specific factors. Conversely, during economic downturns or periods of reduced travel activity, Beta values for travel services companies may decrease as their stock prices exhibit less volatility relative to the broader market.

High Beta stocks are characterized by higher volatility while near to zero ones (not the negatives because could be noisy too but on the opposite) are less volatile. The volatility is used as a measure of a stock’s risk because implies the presence of frequently changing (increasing and decreasing) results with high range of values (can change dramatically). Each stock contains two types of risk:

1) **Idiosyncratic Risk** (or company-specific);
2) **Systematic Risk** (or market-related).

The first one is caused by company related factors like a company management team or other microeconomic (inside the company) elements, it is also known as “diversifiable risk” because in portfolios, through diversification, can be completely eliminated. The latter (systematic risk) is the risk inherent in the overall market and is characterized by macroeconomic factors like interest rates or inflation. The particularity of market-related risk is that it cannot be diversified away, this means that in a well-diversified portfolio it will be present only the systematic risk and not the idiosyncratic one. The Beta measures the level of systematic risk of a particular security but does not consider idiosyncratic risk; the reason is simple since, among practitioners, the firm-specific risk is considered irrelevant because it can be simply eliminated through diversification. From this perspective, in the CAPM it is considered only the systematic risk, calculated through the Beta, leaving unconcerned the idiosyncratic risk (deemed irrelevant).

After having understood what Beta represents and implies, it becomes compelling the definition of the steps necessary to derive it:

1) **select historical data of the target company and the market**. Even if ideally the Beta should be predicted, the first real bias is generated by the discrepancy between the expectancy about Beta and the data used. It is clear that – as in general happens when making a forecast – the utilization of historical data does not assure a precise prediction since, to guarantee so, the Beta should be stable over time and never vary. The second issue regards the **time span** of the data collected; many data sources that provide Beta estimates use the past two to five years data. The five years rule of thumb is frequently supported by different studies[47] that demonstrate how five years of data are sufficient to identify a company's Beta. Still, there are several doubts about this issue especially when the market is characterized by unprecedent events that completely wipe out general certainties. For example, in the recent years the world has faced a severe pandemic that completely changed the rules of the game; stocks commonly known for being defensive or of low volatility have totally changed. The five years rule of thumb is convenient but should be carefully considered before of opting for it;
2) **compute weekly or monthly returns (depends on what type of data has been downloaded)**. In the following step, after having checked if both the market's and the stock prices' dates coincide, it starts the real phase about Beta calculation.

[47] Nicolas Groenewold, Patricia Fraser, "Forecasting Beta: How Well Does the 'Five-Year Rule of Thumb' Do?", Journal of Business Finance & Accounting, 2000.

Since prices are not normalized and an increase of both prices is not so well read by a correlation analysis, the best possible option is to derive both market and stock (weekly or monthly depending on the price frequency) returns. This way it is possible to concretely understand the price variations along with the amount of the variation. If, for example, stock A has a price of 50 at date *x* and becomes 75 at date *x+1*, while the market has a price of 100 at date *x* which becomes 125 in *x+1*, their change is equal in terms of amount but in terms of growth are completely different. Stock A return is 50% while market return is 25%. Both stocks move in the same direction but, even if apparently seem to move in the same way, stock A has a return double than the market. This kind of information can only be derived through the return computation. The return is simply calculated by dividing the next week/month price by the current week/month price and then subtracting one, this is done for all the N available prices (the weekly/monthly returns will be $N-1$):

$$return(week\ or\ month) = \frac{P_{t+1}}{P_t} - 1$$

3) **identifying the best fitting line**. Now that both the series of returns are available, it is possible to compute the Beta. Sometimes, before of proceeding, practitioners tend to use the excess returns (returns minus risk-free rate) rather than the normal returns. This does not change the result because in practice it is subtracted the same value from both the series and this, in statistical terms, has no effective change. This passage is far more useful if it is possible to have the daily market changes of the risk-free rate to derive the excess market return in a more reasonable and correct way. Since in this case the purpose is to compute the Beta, there is no problem in proceeding with the current data.

   Beta can now be identified in two ways:

   1. as the **slope** of the Best Fitting Line equation, computed by using a statistical technique defined as Linear Regression:

      $$R_i = \alpha_i + \beta_i(R_m) + \epsilon_i$$

      The best fitting line equation represents the line that minimizes the sum of the squared deviations from the line; practically speaking is the line that permits to have, on average, ϵ_i equal to zero. ϵ_i is the error or residual term, the deviation from the best fitting line of each return from the equation (below it will be made a graphical example). It is also defined as the idiosyncratic risk (see above) because represents the over-under performance from the best fitting line and is independent from market movements. ϵ_i must be, on average, equal to zero because, otherwise, the best fitting line could be improved. Since the purpose of the whole CAPM is to derive the **expected** equity return, they are taken the expectations of both sides in the best fitting line formula:

$$E[R_i] = \alpha_i + \beta_i(E[R_m]) + E[\epsilon_i]$$

Which becomes:

$$E[R_i] = \alpha_i + \beta_i(E[R_m])$$

Because $E[R_i] = 0$ by definition.

Beta and alpha (α_i) are two constant values, so they stay the same. α_i is the frequent subject of several studies since, its existence, if proved, demonstrates the inconsistency of the CAPM. In fact, in the CAPM, alpha is supposed to be zero or near to zero; alpha is graphically representing the interception of the best fitting line with the y-axis above or below zero. If alpha is equal to 2%, the best fitting line will not pass through the origin of the axes but above zero, precisely on 0.02 on the y-axis. Alpha is frequently referred to as the "Holy Grail all active managers seek[48]", the additional return above and beyond the market exposure due to the hedge fund's trading skills. It is the additional return besides market risk compensation and its existence is still debated[49]. In this paper it will not be detailly explained the debate but, in a few words, it will be described the situation. On one side, there are those ones that believe in the CAPM and consider as only return the market return and justify the presence of the alpha as something easily superable through a multi factor model (like Fama-French[50]). On the other side, there are the hedge funds and traders

[48] Lasse Heje Pedersen, "Efficiently Inefficient how smart money investments & market prices are determined", Princeton University Press, 2015.
[49] The debate verges around the existence of additional risk premia that reduce to zero the alpha.
[50] Eugene F. Fama, Kenneth R. French, "Common Risk Factors in the Returns on Stocks and Bonds", Journal of Financial Economics, 1993.

that seek the alpha and desire it trying to beat the market and gaining additional return. To conclude, if the formula is taken as expectation, alpha is approximated to zero, and instead of the overall return it is computed the best fitting line of the excess return (over the risk-free rate), the equation becomes:

$$E[R_i] - r_f = \beta_i(E[R_m] - r_f)$$

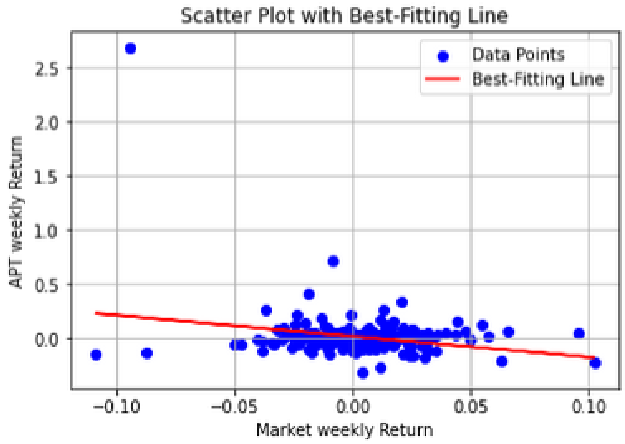
Or:

$$E[R_i] = r_f + \beta_i(E[R_m] - r_f)$$

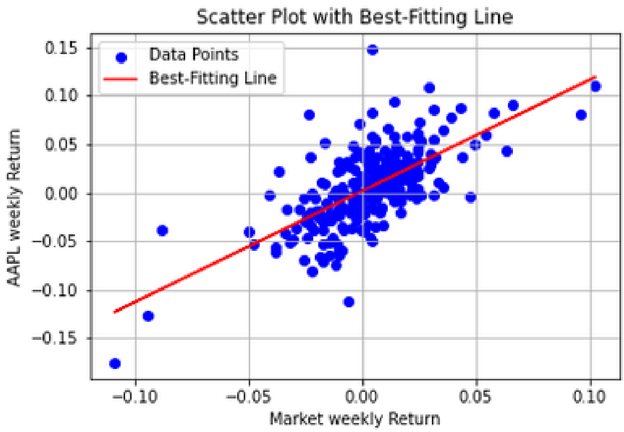
The CAPM.

Below are represented Python practical examples of the best fitting line with a comparison of a positively correlated (Beta higher than zero) stock with a negatively correlated (Beta lower than zero) stock:

```python
import numpy as np
import matplotlib.pyplot as plt

# Sample data
x = market_portfolio_return[1:]
y = prices_data.loc['AAPL']['Weekly Return'][1:]

# Calculate the slope (m) and y-intercept (b) of the best-fitting line
m, b = np.polyfit(x, y, 1)

# Format the equation of the best-fitting line
equation = f'y = {m:.2f}x + {b:.2f}'  # Example format: y = 0.80x + 0.20

# Plot the scattered data points
plt.scatter(x, y, color='blue', label='Data Points')

# Plot the best-fitting line
plt.plot(x, m*x + b, color='red', label='Best-Fitting Line')

# Add labels, legend, and the equation
plt.xlabel('Market weekly Return')
plt.ylabel('AAPL weekly Return')
plt.title('Scatter Plot with Best-Fitting Line')
plt.legend()
plt.text(0.5, 5, equation, fontsize=12, ha='center')  # Adjust the position of the equation text

# Show the plot
plt.grid(True)
plt.show()

# Print the equation of the best-fitting line
print("Equation of the best-fitting line:", equation)
```

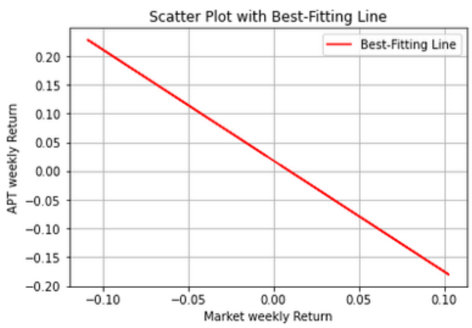



```
Equation of the best-fitting line: y = -1.93x + 0.02
```

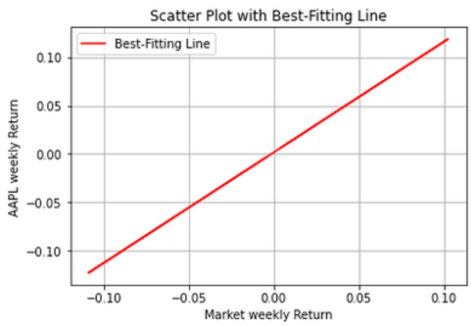



```
Equation of the best-fitting line: y = 1.14x + 0.00
```

APT returns tend to represent a strong negative correlation with the market. Rather than noticing a strong presence of positive returns when the market presents positive returns and negative returns when the market presents negative returns (as it frequently happens in AAPL scattered plot), in APT scattered plot frequently figure divergent results (positive market returns are accompanied by negative APT returns and vice versa). In APT, alpha is greater than AAPL's, the graphical difference is clearer if it is eliminated the scatter plot that presents outlier values:




```
Equation of the best-fitting line: y = -1.93x + 0.02
```




```
Equation of the best-fitting line: y = 1.14x + 0.00
```

AAPL line intersection coincides with the origin of the axes (alpha is approximately zero), APT line, on the contrary, meets the y-axis precisely at 0.02 (alpha).

2. As the result of the following formula:

$$\beta = \frac{Covariance(R_m, R_i)}{Variance(R_m)}$$

In the Python tickers database showed before, the Beta was computed by using this formula:

```
def calculate_beta(weekly_return):
    # Calculate covariance
    cov = np.cov(market_portfolio_return[1:], weekly_return[1:])[0, 1]
    # Calculate variance of market return
    var_market = np.var(market_portfolio_return[1:])
    # Calculate beta
    beta = cov / var_market
    return beta

prices_data['Beta Calculated'] = prices_data['Weekly Return'].apply(calculate_beta)
```

Obviously, the results are the same as those showed before:

```
print("AAPL Beta is:  ",prices_data.loc["AAPL"]["Beta Calculated"])
print("APT  Beta is:  ",prices_data.loc["APT"]["Beta Calculated"])
```

```
AAPL Beta is:   1.1479904574222013
APT  Beta is:   -1.9381806459079598
```

After having defined the passages necessary to compute the Beta, the results obtained can be analysed, especially looking for a consistency with the assumptions made before about a stock's native industry. The first ticker searched is JNJ, Johnson & Johnson, the pharmaceutical company:

```
resulting.loc['JNJ']["Beta Calculated"]
```

```
0.5024389532996792
```

It presents a low Beta that confirms the first assumptions about the pharmaceutical industry that has usually revealed itself as a low Beta industry. This low value should be apparent to all the stocks operating in the pharmaceutical industry, still this does not happen:

```
resulting.loc['MRNA']["Beta Calculated"]
```

```
1.601
```

The stock just showed, Moderna, one of the other ordinarily known pharmaceutical companies, presents a particularly high Beta that completely change the observations about the pharmaceutical industry trends.

To address one of the main issues about Beta, the unpredictability about its future changes, it has been introduced an additional distinction between Raw Beta (the Beta calculated so far with historical data prices) and **Adjusted Beta**. The latter is the estimation of a security's future Beta computed by starting from the Raw Beta and adjusting to future predictions. In general, according to the Efficient Market Theory, all stocks tend towards the unity (or the perfect market's portfolio Beta); in fact, adjusted Beta is the result of a simple correction:

$$\beta_{adjusted} = \frac{2}{3} \times \beta_{raw} + \frac{1}{3} \times 1$$

This way the Beta adjusted will be in every case closer to one and will represent the expectation about a future change of Beta.

Overall, even if some companies operate in the same industry and are exposed to the same macroeconomic risks (systematic risk), the Betas can assume completely distinct values and deviate substantially from one to another. As showed above, it can happen to look at some healthcare companies with above than one Beta or look at tech companies with low Beta values. It becomes apparent that there is something more that influences the value of the Beta and it is not solely the sector in which a company operates, or the products offered. In this paper it will be demonstrated that there are different other "Determinants" that characterize the Beta Equity.

2.3 The Cost of Equity (Not Listed Companies)

The precedent paragraph ended with a final consideration about the nature of Beta. So far, they have been described the different methods used to analyse and compute the cost of capital of publicly held companies (especially for the computation of the cost of equity). The CAPM, and especially Beta, is computed because it is possible to extract the data about historical price of the target company that are transformed in returns comparable with the market. But what is done when the company is not listed, and the Beta is not directly predictable through linear regression?

Since it is not possible to directly compute the Beta because there are no available data on privately held companies, it is necessary a different approach that permits to derive the Beta (also defined as "**Beta Equity**" or "**Levered Beta**") without possessing historical price data of the target. The most widely used method claims its founding grounds on the considerations made about the Comps method. In this sense, when computing the Beta of the target company, the most straightforward approach would be to directly define a list of comparable companies (usually grouped by the sector in which they operate) and then derive the average Beta. The latter would represent the Beta of the target company.

This way of approaching presents different shortcomings, the most evident one is represented by the total indifference in considering the different levels of leverage. If the target company has a D/E of 25 or a D/E of 0.5 it would be the same according to this first method. For this reason, the literature and the common practice have developed a solution to this problem that permits to **unlever** the Beta (from now on defined as "unlevered Beta" or "Beta asset") and **relever** it to adapt it at the leverage level that is preferred (the leverage level of the target company).

This concept is applied to the cost of capital too:

Consider a listed company whose cost of equity ($r_E = 12\%$*) and cost of debt (*$r_d = 5\%$*) are well known in a certain capital structure (*$\frac{D}{(D+E)} = 0.2$*). Reminding the formula of the unlevered cost of capital –* $r_U = \text{"}Pre - tax\text{"}\ WACC = \frac{E}{D+E} r_E + \frac{D}{D+E} \times r_d$ *– it is now possible to dig deep into the meaning of it.*

The unlevered cost of capital, along with what it has been already said about not considering the tax benefit, represents the cost of capital freed of whatever level of indebtedness affects the company. In this sense, the unlevered cost of capital doesn't change if the D/(D+E) is 0.2 or 0.5 because it is totally absent the "leveraging effect" of the tax benefit; instead, what changes is the equity cost of capital that is directly correlated to the debt exposure. Therefore, it is possible to compute the unlevered cost of capital by using the current capital structure and adapting the equity cost of capital to the different debt levels.

$$r_U = 0.8 \times 12\% + 0.2 \times 5\% = 10.6\%$$

The value obtained stays the same whatever the debt is, the cost of equity will change as the debt level changes:

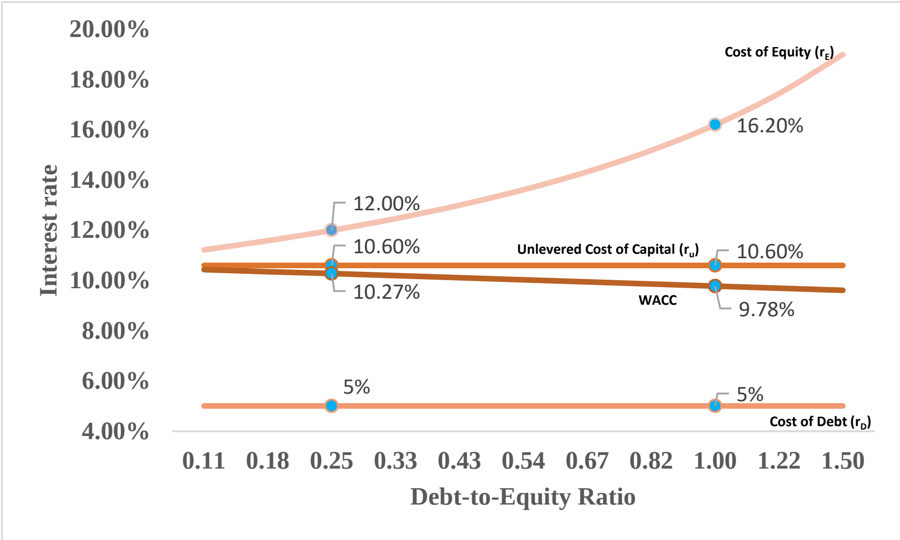
$$r_E = r_A + (r_A - r_D) \times \frac{D}{E}$$

Where $\frac{D}{E} = \frac{\frac{D}{(E+D)}}{1 - \frac{D}{(E+D)}}$

So, if the D/(D+E) ratio changes in 0.5, the cost of equity will be:

$$r_E = 10.6\% + (10.6\% - 5\%) \times \frac{0.5}{1 - 0.5} = 10.6\% + 5.6\% \times 1 = 16.2\%$$

The cost of equity increased of 4.1% (relevering to adapt the cost of equity at the new debt). The changes of the cost of capital can be summarized in the following plot:



The idea is to apply the same concept used for the cost of capital to the Beta; it is, in fact, possible to distinguish the Beta in two types:

- Beta Equity (or "Beta Levered")
- Beta Asset (or "Beta Unlevered")

This distinction was introduced by Robert Amada when he was professor of finance at the University of Chicago Booth School of Business with the publication of the later called Hamada's Equation[51]:

$$\beta_L = \beta_U \times \left[1 + \frac{D}{E} \times (1 - t)\right]$$

The above formula permits to compute the levered Beta by inserting the unlevered Beta and considering the tax benefit (*t*). By computing the inverted formula, it is possible – if it is known the Beta levered – to derive the Beta unlevered:

$$\beta_U = \frac{\beta_L}{\left[1 + \frac{D}{E} \times (1 - t)\right]}$$

Hamada's Equation supposes a Beta debt equal to zero, this means that it can only be used in those situations in which the debt exposure of the company is not so relevant that a possible change would not affect the riskiness of the debt. The situation gets more

[51] Robert S. Hamada, The Effect of the Firm's Capital Structure on the Systematic Risk of Common Stocks, The Journal of Finance, 1972.

complicated as soon as the riskiness of debt plays a significant role in determining the overall risk of the firm. In that case, the Hamada's equation is no longer adapt and necessitates of a correction:

$$\beta_L = \beta_U + (\beta_U - \beta_D) \times \frac{D}{E}$$

To determine a suitable Beta for the evaluation of a private company, it is necessary to carry out two different adjustment operations:

- the neutralization of the leverage effect in the comparables' Beta observed on the market by unlevering their Betas;
- the readjustment of the Beta unlevered (the average of the comparable companies' Beta unlevered) according to the debt level of the evaluated company in order to adapt the Beta to the leverage condition of the target.

Financial databases (e.g. Bloomberg, etc.) can help with the analysis and provide an estimate of the Beta unlevered for the companies included in the sample. As an alternative, it is also possible to opt for third-party databases which have already estimated Betas grouped by industry (e.g. Prof. Damodaran's "Levered and Unlevered Betas by Industry"). This option represents the last resort when it is not possible to identify a large enough number of listed companies or whenever the latter are in markets that are very different from the reference one (for example, if there are available only listed companies in the Chinese market and the analysis is about an American corporation).

To make everything clear below it is provided an example of how it is usually carried out a Beta Equity estimation:

A certain company C is privately held and no information about its Beta is provided; the only available information is about the financing structure (Net Debt[52] = 40, Equity = 100), the industry (Drugs, Biotechnology) and the tax rate (30%). Additional information is provided by the bank that would easily continue to lend debt without any impact on the interest rate required. Since the company is not listed it is necessary to compute the Beta

[52] In general, when it is referred to D, the proper measure is the Net Debt.

Equity in the artificial way by firstly grouping the listed comparable companies and then extracting their financing structure along with the levered Beta.

At this point the steps are the following:

1) *unlever the Beta Equity for all the comparable companies by using the inverse formula (that uses the Beta asset and the financing structure):*

$$\beta_U = \frac{\beta_L}{\left[1 + \frac{D}{E} \times (1 - t)\right]}$$

2) *take the average of the Beta Assets just derived from the comparable companies, "$\overline{\boldsymbol{\beta_U}}$".*

The resulting Beta asset will be again levered at the Company C's debt level in order to obtain its Beta Equity:

$$\beta_L = \overline{\beta_U} \times \left[1 + \frac{40}{100} \times (1 - 30\%)\right]$$

For sure this example is simplistic and does not account for all the complexities of determining the best comparable companies. In this optic, it is now time to delve into the real purpose of this paper and introduce the Determinants of the Beta Equity, the characteristics that a comparable company must have in order to really be similar, the elements that matter and affect the different levels of Beta Equity that the companies assume.

2.4 Focus on Beta Equity

All the descriptive analysis taken on so far, along with all the methodologies used in order to derive a company's intrinsic value always remind how, at the end, what really impacts the valuation is the cost of capital. Its importance is evident and, especially in some methods (especially in the DCF), becomes so relevant that even a slight change could lead to opposite or extreme results.

The cost of equity represents the most difficult element to compute; in fact, despite the cost of debt has to be equally obtained through a series of considerations, the cost of equity remains the toughest to be computed (the cost of debt can be obtained through past interest loans or computed through the rating category). For this reason, as it has been evidenced in the previous paragraph, the cost of equity necessitates a deep understanding and specific focus.

Despite the interest in the cost of equity as a whole, the only element that changes for every target company is the Beta. The latter has been previously analysed with its strengths and weaknesses, and it has been said how, even if the general definition of Beta regards the stock correlation with the market, it becomes apparent that there are several factors that impact its value. Not only the sector in which a company operates can be the driver of the resulting Beta but different other measures impact on the Beta result too.

The purpose of the following pages is to look for these other factors, defined as "Beta Determinants", with the additional help provided by past research on the matter.

2.4.1 Determinants of Beta Equity

In the past years practitioners and professors frequently tried to study the behaviour of stocks and the feature of Beta. A deep understanding of Beta would certainly help to understand the risk component of the stocks without always referring to it in market risk terms. As it has been said before, the only message Beta provides is about its correlation with the market and, therefore, a stock risk measure and default risk are usually linked with the market through market determined interactions (like price change volatility). The purpose of this paper is to really understand Beta and derive the exogenous drivers (non-price data) that have an impact on Beta result. An understanding of these exogenous measures permits to get an idea about non listed stocks too which have never been directly defined in terms of default risk because it is necessary the presence of price data.

Beaver, Kettler and Scholes, through their work[53], tried to look for some sort of evidence that could enable to demonstrate a direct correlation (association) between the Beta or the stock prices Variance (Market Determined Risk Measures) and a series of exogenous factors (Accounting Determined Risk Measures). As in the paper is initially underlined:

> *"Previous research would suggest that financial statement ratios can be used as measures of default risk, but little is known of their association with the concept of risk as defined in portfolio theory.*
>
> ...

[53] William Beaver, Paul Kettler, Myron Scholes, "The Association Between Market Determined and Accounting Determined Risk Measures", The Accounting Review, 1970.

An issue of paramount concern to the accounting profession is – what is the relationship between the accounting determined and market determined measures of risk?

...

In particular, this study will examine the contemporaneous association between the accounting determined and market determined measures of risk. By doing this, we intend to determine what accounting data are impounded in the market price data, such as to give rise to a given level of risk. If an association is observed, the evidence supports the joint hypothesis that accounting data reflect the underlying events that determine securities' riskiness and that such events are also reflected *in the market prices of securities."*

The demonstration of existence of the abovementioned correlation would have an impact under three different circumstances:

1) **it would complete the knowledge of risk determination by introducing accounting variables as additional instrument to assess a company's riskiness**;
2) **it would lead to an improvement in decision making since accounting data would be useful to the investor in forecasting the riskiness of securities, such that he can select the portfolio which maximizes his utility;**
3) **it would enable the possibility of determining the Beta of a privately held company through the consideration of not only the sector in which it operates, but also by the other exogenous factors.**

In order to proceed with the demonstration, the literature necessary to get a conclusion comprehends a first approach to the portfolio model of Markowitz theory[54] and to CAPM model. Markowitz introduced the concept of portfolio risk with a new metric: the portfolio variance (the dispersion of the values around the mean). The latter is computed through the sum of two values:

$$\sigma^2(R_p) = \frac{1}{N} average\ variance + \left(\frac{N-1}{N}\right) average\ covariance$$

[54] Markowitz theory, also known as Modern Portfolio Theory (MPT), is a framework developed by Harry Markowitz in 1952.

It becomes evident that, as the number N of stocks in the portfolio increases and tends to infinite, the first term tends to zero and remains only the second one. In this sense, for a well-diversified portfolio the only measure of risk that really matters is the average covariance.

The covariance represents, similarly to what has already been said about correlation and Beta, the extent by which two sets of values (in here two sets of stock returns) move together. In this sense, differently from correlation, a positive covariance happens when, for example, if one stock tends to move above its mean also the other stock moves above its mean. Conversely, a negative covariance presents itself when, for example, if one stock moves above its mean the other one moves below its mean.

A stock with high variance but low covariance (presumably negative) is not a risky stock to be inserted into a portfolio; on the contrary, its negative covariance can contribute to the reduction of the portfolio's overall variance (in this sense, frequently, diversification reduces the risk).

The most evident limitation of Markowitz model is the enormous amount of parameter estimation (in terms also of necessary available past data) required to assess a portfolio's variance. This drawback has encouraged other literates and researchers to find a different solution. To address the issue, William Sharpe introduced the Diagonal Model[55] whose properties would have helped to reduce drastically the amount of data necessary to compute a portfolio's variance:

$$R_i = \alpha_i + \beta_i(R_m) + \epsilon_i$$

This way, the security's return is decomposed in the two parts previously introduced, the systematic component ($\beta_i R_M$) and the individualistic component ($\alpha_i + \varepsilon_i$).

The utilization of the diagonal model changes the variance of portfolio return in the following way:

$$\sigma^2(R_p) = \frac{1}{N}\overline{\sigma^2(\varepsilon_\iota)} + (\bar{\beta})^2 \sigma^2(R_M)$$

[55] William F. Sharpe, "A Simplified Model for Portfolio Analysis", Management Science, 1963.

As the number N of securities increases, the first component goes to zero and remains solely the second factor. Hence, the individual **stock's contribution to the portfolio riskiness is measured solely by its Beta**. The concept of Beta and Covariance is strictly related, as testified by the computation formula presented before[56]:

$$\beta = \frac{Covariance(R_m, R_i)}{Variance(R_m)}$$

The direct relationship between Beta and two stocks' covariance (an increase of the covariance procures an increase of Beta) permits to conclude that, **since the covariance is demonstrated to be a measure of security riskiness, also the Beta can be referred to in this way**.

The properties and the work done on portfolio models by William Sharpe are extended to Capital Asset Pricing Models (by Sharpe, Lintner[57] and Mossin[58]) in order to determine equilibrium prices of all securities in the market. In such a framework (showed before, in here represented again), capital assets will be priced in equilibrium such that:

$$E[R_i] = r_f + \beta_i\left(E[R_m] - r_f\right)$$

> *"The capital asset pricing model states that the only variable which determines differential expected returns among securities is the systematic risk coefficient, β_i. The model further asserts there is a linear relationship between β_i and expected return, such that the greater the risk the higher the expected return.*
>
> ...
>
> *The variability of the individualistic component of return does not enter into the pricing of capital assets, since that component can be eliminated through diversification."*

Once it is possible to accept Beta as a measure of a stock's riskiness, the next step is to demonstrate the presence of correlation between the individualistic (accounting) measures and the systematic measures (Beta). In fact:

56 The formula is completely precise only if the returns are normally distributed.
57 John Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", Review of Economics and Statistics, 1965.
58 Jan Mossin, "Equilibrium in a Capital Asset Market", Econometrica, 1966.

> *"If the systematic and individualistic components* are positively correlated (at the extreme, perfectly correlated), then it is ***reasonable to view the accounting measures as surrogates for systematic risk as well****."*

If the demonstration results effective (the correlation is evident), **the exogenous variables can be used to derive a plausible expectation about a company's Beta**. Obviously, before of trying to determine such a correlation, that will be fully covered in paragraph 2.4.2, it is necessary to introduce the exogenous variables.

Both practitioners and academics have recognized various accounting risk metrics, whose attributes often imply a connection to a company's risk profile. Those accounting measures are:

- ***Dividend payout***: common practice considers **low dividend payout** (i.e. Cash Dividends/Net Income) **as positively associated to risk**. Managers and investors frequently perceive that a policy of low distribution level of dividends is a signal for a high volatility in earnings. This rationale is the result of an inverse way of reasoning. If in general a company follows a policy of dividends stabilization (maintaining stable the dividend level since a cut back could appear as a pessimistic forecast), a stable and always producing earnings company can permit itself to set a higher threshold because can count on its ability to produce earnings and distribute dividends. On the contrary, a company setting a low payout policy threshold intuitively suggests that a company is not sure about its future earnings and wants to be cautious. In this sense such a company is perceived as a company with a high uncertainty about its earnings, therefore riskier. Still, payout ratio presents two different shortcomings. The **first one** refers to the possibility of assisting to a zero or approximate to zero earnings period; in this case the payout ratio will be presumably very high. Since such an event is usually circumstantial to a single year, not even computing an average of the past payout ratios solves the problem since the single year could dominate the sequence. A *solution* could be the computation of a weighted payout ratio average depending on the earnings, this way the singular event would be weighted by the specific year earnings. The **second issue** regards the possible presence of negative earnings; this possibility represents an additional possible obstacle in the proper consideration of the payout ratio.

- ***Growth:*** consistent growth rates are usually associated to, at least, two factors: (i) **above average earnings opportunities**, or differently said, particularly high expected returns are positively associated to risk since the higher is the return, the higher is the risk incurred and (ii) a **payout policy characterized by a high retention rate**[59]. The latter refers to a case equal (but defined in term of retained earnings) to a company with a low payout ratio, also in this case a high growth rate is positively associated with risk since a high retention rate suggests a low dividend distribution and, therefore, an expectation of non-steady cash flows (results that can vary considerably during the years). So, eventually, it can be stated that, by taking into consideration the reasoning, **high growth rates are positively associated with high risk**.
- ***Leverage:*** the most obvious factor, as in fact it is, **is positively associated with risk**, as additionally testified in the process of unlevering and relevering phase of the cost of equity by using the leverage ($Net\ Debt/(Net\ Debt\ +\ Equity)$) of the target company. Additional debt increases drastically the bankruptcy and insolvency risk since even a singular negative result could affect the whole repayment plan and causing the company the financial distress.
- ***Liquidity:*** in logical terms, the presence of numerous liquid assets asserts a stable financial condition since the company's liabilities are covered by assets with similar maturity. Nevertheless, it appears **no association between liquidity measures** (the current ratio[60] is the most used one) **and market risk measures**.
- ***Asset Size:*** **small firms are believed to be way riskier than larger firms**. This is explained by several factors: (a) **large firms** usually have **a business already settled** and more **stable cash flows**, (b) large firms can count on their **market dominant position** and (c) large firms can benefit from **diversification and economies of scale**. For these reasons in there lies a **negative correlation**

[59] An important reminder is the relationship between growth rate and payout policy (where the Retention Rate represents the fraction of earnings not distributed):

$$g(growth\ rate) = Retention\ Rate \times Return\ on\ new\ investment$$

[60] Represents the ratio between current assets and current liabilities.

between size measures and risk (the greater the firm the lower the risk). The best way of determining a company size is through the total assets amount.

- ***Variability in earnings:*** a different measure is the direct computation of the earnings variability (standard deviation, the square root of the volatility). The importance of this measure is testified by the fact that all the previous ones indirectly tried to suggest the presence of the variability of earnings. The variability in earnings could be represented through a earnings to price ratio variability, still this measure would suppose the utilization of the market capitalization. The purpose of this paper is to look for determinants valid for privately held companies too, for this reason the variability in earnings will be represented differently.
- ***Accounting Beta (regression of company individual earnings with market):*** accounting Beta measures the sensitivity of a company's earnings (or other accounting measures like the ROE or the ROA) to changes in the overall market returns. It helps investors and analysts understand how a company's financial performance correlates with broader market movements. Such a measure is particularly interesting since represents a valid alternative to classical Beta when companies are not listed. Different past research (also conducted by taking into consideration Italian firms[61]) has not converged towards a unique conclusion but in some cases – among which the Italian research that presented positive results – the accounting Beta appeared as a good substitute of the classic Beta. Still, even if the substitution between accounting Beta and normal Beta is not certain, the accounting Beta represents a good measure of risk and can present a good **positive association with the market risk measures**.

Another curious variable, introduced by Chincarini et al. (2020)[62], is represented by the "**Firm Age**". This factor, even if it is not an accounting one, represents an additional element to account for the riskiness of a company. The measured Beta declines over the

[61] Intrisano, Carmelo, Giovanni Palomba, Loris Di Nallo, Anna Maria Calce, "Accounting Beta: Which Measure Is the Best? Findings from Italian Market", European Journal of Economics, Finance and Administrative Sciences, 2017.
[62] Ludwig B. Chincarini, Daehwan Kim, Fabio Moneta, "Beta and firm age", Journal of Empirical Finance, 2020.

age of a firm, therefore it is noticed **a negative association between the Beta and the firm age**. There are four possible explanations for this pattern.

The **first** explanation is that the determinants described above change as the age of the firm changes and firm becomes more mature (earnings variability generally lowers, leverage generally lowers and the payout usually increases).

The **second** explanation derives from the conception of companies' life cycles; each company follows a different series of phases which usually tend to be associated with different stages of risk and return. For example, a start-up company that has a year or two of existence presents completely different future expectations if compared to a mature company. In fact, even if a start-up has a higher probability to generate incredibly high returns (as testified by venture capital funds) it can as well result in a complete disaster because the business is too young, and it could not acquire market share. On the other hand, a mature firm with already defined customers and a business already settled does not assure incredible returns but is almost impossible to suddenly go bankrupt.

The **third** explanation regards the amount of available information pending upon a certain company. It is commonly known that as the amount of available information about a certain company increases the riskiness about that company lowers. Specifically, it has been demonstrated by Lambert et al. (2007)[63] that, if there are improvements in the information quality and amount, the cost of capital and the Beta are affected. Usually, as the firm gets older, the quality and the amount of information increases in such a way that the perceived risk of the company lowers.

The **fourth**, and last, explanation substantially depends upon the third one. The presence of information about a company can assume two different consequences: (i) the low amount of information can cause the estimation risk and (ii) holding information for a long time (not the amount) increases the familiarity with a company. Both cases' worst scenarios, the reduced number of information and the small life of a company, can cause the presence of a risk premium derived from the higher uncertainty about the target company. The absence of information in the past years (low firm age) causes **unfamiliarity**. The reward for the unfamiliarity is defined as "unfamiliarity premium"

[63] Richard A. Lambert, Christian Leunz, Robert E. Verrecchia, "Accounting information, disclosure, and cost of capital", Journal of Accounting Research, 2007.

and increases as the company age is lower. In this sense, if the unfamiliarity premium is low, the Beta and the expected return will be lower, and the price will be higher.

As previously introduced, the most practical way of demonstrating the legitimacy of defining the previous factors as Beta Determinants is to directly analyse a sample of stocks and determine if a correlation between the Determinants and the Beta exists. This is exactly what it will be done in the next paragraph.

2.4.2 Correlation between Determinants and Beta Equity

The purpose of the following analysis is to determine if, effectively, the relationship between the Determinants and the Beta Equity follows the intuitive way of reasoning and exists a real correlation among the variables towards the Beta Equity. To do so, it will be exhibited a correlation analysis factor by factor with a numerical and a graphical representation. Again, it will be used Python to address all those tasks. Specifically, it will be used a Dataset containing more than 50,000 stocks:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52604 entries, 0 to 52603
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Ticker          52604 non-null  object
 1   Name            52554 non-null  object
 2   Country         34477 non-null  object
 3   Exchange        52523 non-null  object
 4   Currency        50020 non-null  object
 5   Sector          31961 non-null  object
 6   Industry        31961 non-null  object
 7   Beta            33094 non-null  float64
 8   Current Ratio   29699 non-null  float64
 9   Debt to Equity  27510 non-null  float64
 10  Gross Margin    29717 non-null  float64
 11  Market Cap      35854 non-null  float64
```

Not all of those stocks are useful so, after a data cleaning process (the process will be presented in the final chapter with all the explanation of the various steps in an EDA – Exploratory Data Analysis) the remaining stocks are about 15,000.

The data sample used in this case considers a **restricted number of stocks** that are all traded in the various **US stock exchanges** (NYSE, New York Stock Exchange, AMEX, American Stock Exchange, or NASDAQ, National Association of Securities Dealers Automated Quotations and others). The sample size starts with almost 1,400 stocks. Not all the columns of the DataFrame are necessary since the analysis is limited to the factors previously listed. Still, three factors are not present into the DataFrame columns, these are:

- **Growth**, the growth is computed as the compound growth rate of the total assets of a company:

$$Growth = \left(\frac{Last\ Year\ value\ of\ Assets}{First\ Year\ value\ of\ Assets}\right)^{\frac{1}{Total\ number\ of\ years}} - 1$$

- **Variability in Earnings**, the purpose is to derive the variability in earnings without accessing to the usually proposed Net Income to Price Ratio that for the purpose of this paper is not adapt since the final scope is to use the model on privately held companies without listed prices. To accomplish that, it is necessary a measure of variability in earnings that accounts for differences in scale:

```
list(yf.Ticker('AAPL').financials.loc['Net Income'])
```

```
[96995000000.0, 99803000000.0, 94680000000.0, 57411000000.0]
```

```
np.std(list(yf.Ticker('AAPL').financials.loc['Net Income']))
```

```
17306866951.805576
```

```
# The absolute variation does not really reflect properly the variability of data, in the example
# below the variability is identical with the only difference of being in different scales
print(np.std([1,2,3,4,5]),'     |     ',np.std([100,200,300,400,500]))
```

```
1.4142135623730951      |      141.4213562373095
```

In this example it is evident how differences in scale affect the standard deviation. Eventually, Apple standard deviation seems to be abysmal, but it does not represent the reality of the facts. This detail becomes even more evident if comparing the standard deviations of the other two lists of values: the first and the second one are completely identical in terms of variability but the difference in scale suggests that the second list is more volatile. A solution is represented by a statistical measure defined as "**Coefficient of Variation**" that takes the standard deviation of a set of values and divides it by its mean. This way, the measure seems to work well:

Coefficient of Variation

```
# Divide the standard deviation by the mean in order to eliminate differences in scale
print(np.std([1,2,3,4,5])/np.mean([1,2,3,4,5]), '     |     ',np.std([100,200,300,400,500])/np.mean([100,200,300,400,500]))
```

```
0.47140452079103173      |      0.4714045207910317
```

```
# If the variability increases, the result is reflected in the coefficient of variation
np.std([100,200,300,400,600])/np.mean([100,200,300,400,600])
```

```
0.5376453291901642
```

```
np.std(list(yf.Ticker('AAPL').financials.loc['Net Income']))/np.mean(list(yf.Ticker('AAPL').financials.loc['Net Income']))
```

```
0.1984226152364285
```

The previous two lists now have identical coefficient of variation and, as evidenced by the second example, the coefficient increases as the variability increases. Apple's variability of earnings now seems to be more natural.

- **Accounting Beta** (Covariability in Earnings), even if it represents the most promising factor, its calculation becomes difficult since it is based in yearly variations and are necessary multiple years. Because the data of the DataFrame is composed of downloaded stocks from Yahoo Finance, the available years are at maximum four, not sufficient to compute this value. Therefore, **this will be the only factor excluded from the analysis**.

The Growth and the Variability in Earnings are included into the DataFrame:

The earnings are downloaded by YFinance library for each stock present in the subgroup of USA Stocks.

```
list_earnings = []
list_growth_asset = []

for i in list(usa_stocks['Ticker']):
    try:
            stock_data = yf.Ticker(i)
            col = stock_data.balance_sheet.columns
            if len(list(stock_data.financials.loc['Net Income']))>=2:
                list_earnings.append(np.std(list(stock_data.financials.loc['Net Income']))/np.mean(list(stock_data.financials.loc['Net Income'])))
            if len(stock_data.balance_sheet.columns)>=2:
                list_growth_asset.append(((list(stock_data.balance_sheet.loc['Total Assets'])[0]/
                                           list(stock_data.balance_sheet.loc['Total Assets'])[-1])**(1/len(col))-1))
    except Exception as e:
            list_earnings.append(np.nan)
            list_growth_asset.append(np.nan)
            # Print the error and continue with the next iteration of the loop
            print(f"Error for ticker {i}: {e}")
            continue

usa_stocks['Earnings variability'] = list_earnings
usa_stocks['Growth'] = list_growth_asset
```

```
# It is interesting the absolute value of Growth and Earnings Variability
usa_stocks['Earnings variability'] = usa_stocks['Earnings variability'].abs()
usa_stocks['Growth'] = usa_stocks['Growth'].abs()
```

The two other factors are inserted into the DataFrame. The latter, after a first elimination of the outliers and other corrections (the process will be presented in the final chapter with all the explanation of the various steps in an EDA – Exploratory Data Analysis) results in approximately 700 values.

```
usa_stocks[['Beta', 'Debt to Equity','Growth', 'Asset', 'Av Div','Year','Earnings variability','Current Ratio']].describe()
```

	Beta	Debt to Equity	Growth	Asset	Av Div	Year	Earnings variability	Current Ratio
count	703.000000	703.000000	703.000000	7.030000e+02	703.000000	703.000000	699.000000	703.000000
mean	1.049232	87.370435	0.058932	2.819938e+10	0.469786	1958.025605	0.939824	2.195373
std	0.443195	81.086219	0.050707	5.047946e+10	0.476269	48.620888	2.267908	2.344172
min	-0.066000	0.066000	0.000096	1.224370e+08	0.005100	1810.000000	0.043813	0.321000
25%	0.719500	29.086500	0.020100	2.481398e+09	0.196150	1920.000000	0.194794	1.061500
50%	1.034000	64.182000	0.044610	7.634474e+09	0.340900	1968.000000	0.344173	1.592000
75%	1.345000	123.709500	0.082120	3.018594e+10	0.587400	1999.000000	0.681532	2.463000

From this first statistical analysis the key takeaway is the average **value of Beta**; its result **equals almost 1** indicating that the sample is representative of the market. It is now possible to proceed with the correlation analysis, the determinants will be exhibited following the order of description of the previous paragraph.

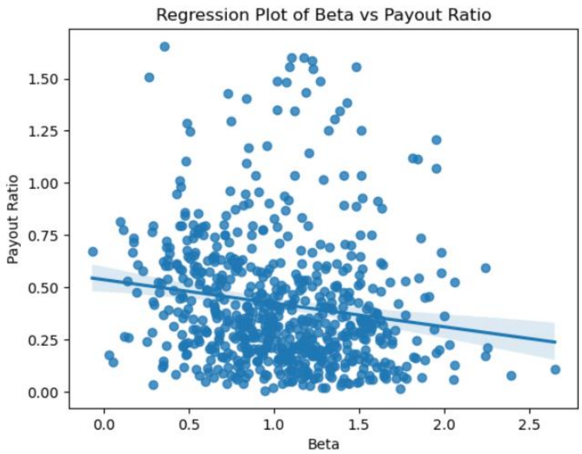
The **Dividend Payout** is the first Determinant, the hypothesis is that a low payout is accompanied by high risk. This suggests a negative correlation between Dividend Payout and Beta Equity:

```
usa_stocks[['Beta','Av Div']].corr()
```

	Beta	Av Div
Beta	1.0000	-0.1551
Av Div	-0.1551	1.0000

```
x = usa_stocks['Beta']
y = usa_stocks['Av Div']
```

```
sns.regplot(x=x, y=y)
plt.xlabel('Beta')
plt.ylabel('Payout Ratio')
plt.title('Regression Plot of Beta vs Payout Ratio')
plt.show()
```



The ***negative correlation*** is confirmed by the data that highlights a strong negative correlation (-0.15) between the Beta Equity and the Payout Ratio. This means that, by rigor of logic, to higher Payout Ratios are usually accompanied lower Beta Equity.

The **Growth,** second Determinant, *supposes a positive correlation with the Beta Equity*, a high growth rate should identify a riskier company:

```
usa_stocks[['Beta','Growth']].corr()
```

	Beta	Growth
Beta	1.000000	0.152511
Growth	0.152511	1.000000

```
x = usa_stocks['Beta']
y = usa_stocks['Growth']
```

```
sns.regplot(x=x, y=y)
plt.xlabel('Beta')
plt.ylabel('Growth')
plt.title('Regression Plot of Beta vs Growth')
plt.show()
```

The ***positive correlation*** is testified by the data; even in this case the correlation is consistent (0.15) and confirms the premises done before: the higher the growth rate, the higher the Beta Equity.

The **Leverage**, represented by the Debt to Equity Ratio, is the third determinant. Its positive correlation with the risk is commonly known and is correctly associated to the default exposure of the debt. Representative is the average value of the Debt to Equity Ratio in the US stocks sample. As it can be noticed, the average value is 87 (that has to be divided by 100 because Yahoo Finance represents it in percentage terms).

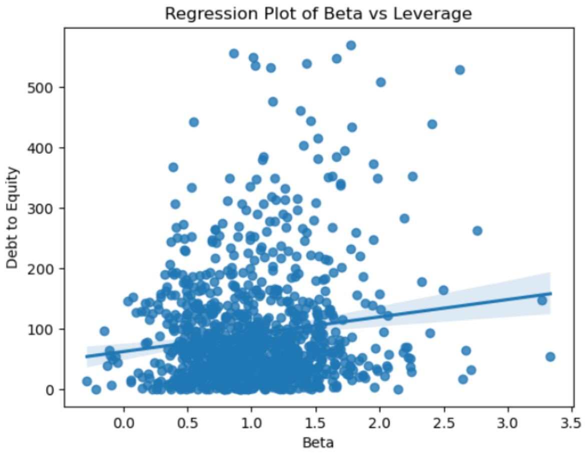
This analysis confirmed the commonly known trend:

```
usa_stocks[['Beta','Debt to Equity']].corr()
```

	Beta	Debt to Equity
Beta	1.000000	0.144549
Debt to Equity	0.144549	1.000000

```
x = usa_stocks['Beta']
y = usa_stocks['Debt to Equity']
```

```
sns.regplot(x=x, y=y)
plt.xlabel('Beta')
plt.ylabel('Debt to Equity')
plt.title('Regression Plot of Beta vs Leverage')
plt.show()
```



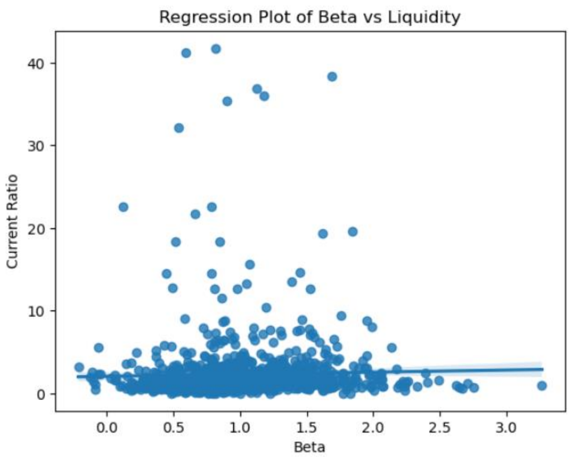
The presence of ***a positive correlation*** (almost 0.15) again demonstrates the association between Debt and Risk.

The **Liquidity**, measured through the Current Ratio, is the fourth Determinant. By logic, a high current ratio should indicate lower risk because the company can count on a good level of hedging and a good financial structure (since the assets and the liabilities durations are well balanced). Nevertheless, prior research has demonstrated how the Current Ratio does not really affect the Beta Equity. The analysis' results are the following:

```
usa_stocks[['Beta','Current Ratio']].corr()
```

	Beta	Current Ratio
Beta	1.00000	0.03226
Current Ratio	0.03226	1.00000

```
sns.regplot(x=x, y=y)
plt.xlabel('Beta')
plt.ylabel('Current Ratio')
plt.title('Regression Plot of Beta vs Liquidity')
plt.show()
```



In this case too, the data analysed displayed ***no substantial correlation*** (it even came up a low positive correlation) indicating again an alignment to past research.

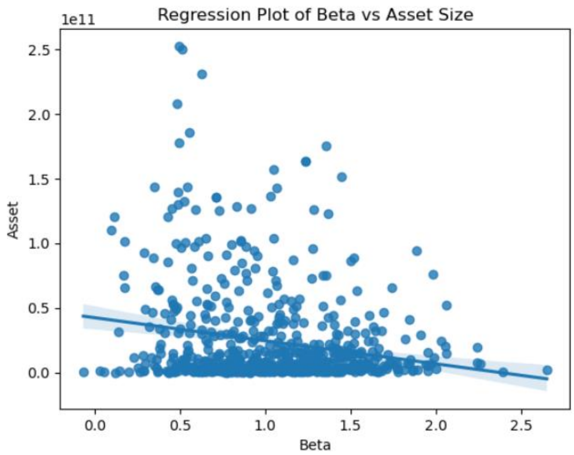
The **Size** of the company represents the fifth Determinant, its correlation with the risk has always been affirmed, representative is the Fama-French three factor model that adjusts the CAPM to include two additional elements of risk premium: (i) the "**SMB**" – small minus big – and (ii) the "**HML**" – high minus low. While the second element is not of primary interest[64], the first refers to the differential risk between a small size company and a big size company. As it has been said before, bigger companies benefit in different ways of their size and should be less risky compared to smaller ones. In this analysis it will be used the total asset size rather than market capitalization[65]:

```
usa_stocks[['Beta','Asset']].corr()
```

	Beta	Asset
Beta	1.000000	-0.204633
Asset	-0.204633	1.000000

```
x = usa_stocks['Beta']
y = usa_stocks['Asset']
```

```
sns.regplot(x=x, y=y)
plt.xlabel('Beta')
plt.ylabel('Asset')
plt.title('Regression Plot of Beta vs Asset Size')
plt.show()
```



[64] It indicates the difference in risk between a stock with a high book value to market value ratio and another with low ratio. In this paper such a correlation would even be interesting but completely depends on the market valuation of a stock. Such a ratio is not possible to be derived for a private company.

[65] The reason, again, is because the final purpose is to use these correlations on a privately held company that has no market capitalization but, instead, has the asset size.

The analysis has confirmed the assumptions made before and suggests a ***strong negative correlation*** between the firm Asset Size and the Beta, a small size firm has a higher probability of having bigger Beta.

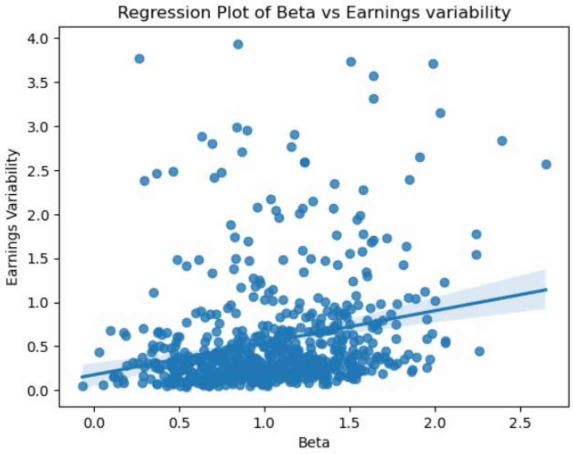
The **Earnings Variability** represents the sixth Determinant, its way of being computed has already been described in the previous pages and should represent the absolute value of earnings volatility (it does not interest if the earnings go severely up or down it is relevant if there is a change). The utilization of the coefficient of variation as the measurement method substitutes the Earnings to Price Ratio that is widely used. This analysis tries to look for a new correlation with the coefficient of variation:

```
usa_stocks[['Beta','Earnings variability']].corr()
```

	Beta	Earnings variability
Beta	1.000000	0.249567
Earnings variability	0.249567	1.000000

```
x = usa_stocks['Beta']
y = usa_stocks['Earnings variability']
```

```
sns.regplot(x=x, y=y)
plt.xlabel('Beta')
plt.ylabel('Earnings Variability')
plt.title('Regression Plot of Beta vs Earnings variability')
plt.show()
```



The result is straightforward and underlines a **strong positive correlation** between the Earnings Variability and the Beta.

The **Age** of the firm is the last Determinant, its correlation with the Beta has not been so much debated as the others but recent research has raised some questions and developed some answers. In this analysis, the result is the following:

```
usa_stocks[['Beta','Year']].corr()
```

	Beta	Year
Beta	1.000000	0.099477
Year	0.099477	1.000000

```
x = usa_stocks['Beta']
y = usa_stocks['Year']
```

```
sns.regplot(x=x, y=y)
plt.xlabel('Beta')
plt.ylabel('Year')
plt.title('Regression Plot of Beta vs Founding Year')
plt.show()
```

Again, the results suggest that it is effectively present a ***positive*** (even if not so consistent) ***correlation*** (negative if it is considered the age) between the founding year and the Beta. This means that, the sooner the companies have been founded the more is probable that they present high Betas.

The results obtained so far permitted to finally confirm the assumptions and the hypothesis made on the Beta Determinants. The existence of factors that affect the value of Beta permit to advance predictions on the possible value of Beta on a privately held company. The purpose of this paper, so far in part already obtained, will be completed in the next chapter, through the preparation of a **Machine Learning** (from now on "**ML**") algorithm that, by using the Determinants' correlations with Beta, permits to predict a possible Beta Equity value (not the specific value) of a privately held company. In the next chapter will be, in fact, considered a wider sample of stocks with the Determinants and it will be conducted on them an **EDA** (Explanatory Data Analysis) to look again for the presence of a correlation with the Beta. Once the correlation is evidenced, it will be possible to conclude with the ML algorithm.

Chapter Three

EDA and the derivation of Beta Equity through Machine Learning on Python

Once the theory and the literature have been presented, it is possible, now, to move a step forward and look at the last chapter of this paper. The importance of Beta Equity along with all the studies and the literature about its Determinants claim a final, practical, explanation.

In the following chapter it will be advanced the presentation of a full EDA process aimed at exploring the whole Dataset crafted through the download of more than 25,000 stocks from Yahoo Finance (initially composed of a population of 100,000 then reduced to approximately 50,000). The purpose of this chapter is to explore the powerfulness of Python and the utility of an EDA to understand the ML code. In the end, since the perfect derivation of Beta through a ML code is not possible, the ML code will limit itself to finally demonstrate how the Beta Determinants have an impact on the prediction of Beta since it will be possible to take a guess on fictional companies.

The EDA, together with the ML code, will be defined through a series of steps completely executed on Python. It is essential to include all the necessary functions and operations of the EDA, each one of them can help to better understand the patterns and the key elements that will be used by the ML code in order to define the ML prediction.

To further explain how the EDA is structured, it is possible to find below an extract of what will follow in the next paragraphs:

The steps will be the following:

- Read the Data saved in Memory (as DataFrame)
- Data Cleaning
  - Converting Values
  - Dropping Duplicates
  - Handling Fundamental Missing Values
  - Handling Outliers
  - Deleting Non Relevant Data
  - Impute Missing Values
- Graphical Representation of the Data
- Statistical Representation of the Data (Regression)
- Converting Numerical Values in Categorical Values
- Deleting Unnecessary Columns
- Machine Learning Code and Preparation
  - Model the DataFrame to access the ML tools
  - Split the data into training and testing sets
  - Create a XGBoost Classifier
  - Fit the Model
- Result and Performance
- Practical Examples

After a first phase in which the main operations regard the cleaning and the adjusting of the Dataset, it follows a second phase in which it is presented a graphical and statistical representation of the data. The Python code concludes with the definition of the ML algorithm and the representation of its performance along with a few practical examples to effectively test the ML algorithm.

Final considerations will conclude this chapter.

3.1 EDA on over 25,000 stocks

After having downloaded the data from Yahoo Finance into a DataFrame ("**DF**"), it must be saved in memory and opened again in order to ascertain that the whole downloading process has not to be done again (the whole process requires at least a day). The majority of the DF columns represent one of the Determinants.

```
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Ticker               28544 non-null  object
 1   Name                 28542 non-null  object
 2   Country              28544 non-null  object
 3   Exchange             28542 non-null  object
 4   Currency             28456 non-null  object
 5   Sector               27578 non-null  object
 6   Industry             27578 non-null  object
 7   Beta                 28544 non-null  float64
 8   Current Ratio        24754 non-null  float64
 9   Debt to Equity       23776 non-null  float64
 10  Market Cap           28394 non-null  float64
 11  Growth               28310 non-null  float64
 12  Year                 28544 non-null  float64
 13  Revenues             26914 non-null  float64
 14  Av Div               15210 non-null  float64
 15  Asset                28544 non-null  float64
 16  Earnings variability 27854 non-null  float64
```

The stocks in the DF are more than 28,000 and represent all those stocks (whose ticker name was available[66]) whose Beta was existent on Yahoo Finance (in the 2.4.2 paragraph the stocks were 50,000 because it didn't eliminate the rows without Beta values). These data must be properly cleaned and have to follow a series of steps in order to ascertain that they constitute a proper and well-presented DF. The whole EDA process will be divided into:

- **Data Cleaning** (composed of the different steps that go from Converting the Values to Imputing Missing Values);
- **Graphical Representation** of the data;
- **Statistical Representation** of the data (regression).

[66] Recall that the Dataset started from more than 100,000 ticker names sourced from different lists stocks available online. As said before, the stocks used in this DF coincide with the stocks whose Beta was available; the stocks without the Beta have no functionality in this paper.

3.1.1 Data Cleaning

The whole Data Cleaning process is summarized in six steps, the description of each one will coincide with the practical representation of the Python code.

1) **Converting Values**: a Data Cleaning process always starts with the conversion of non-adequate values into data that can be used for modelling and analysis. The conversion of data can be distinguished in (i) **data type conversion** and (ii) **data integrity conversion**. While in the first case it implies a simple conversion of the Python data type through a direct function, in the second case it is necessary to adjust the data in a way that can be meaningful for the purpose of the analysis. In this DF it is important to analyse the single columns (or Determinants); in particular, the columns that are subject to the majority of mistakes are those ones characterized by the presence of numerical values, these ones can usually lack some corrections that become necessary in this context.

   In fact:

   The **Year** column, that identifies the founding year of the companies, necessitates a type conversion since the years are expressed as strings and not integer values.

The Years precedently extrapolated from the yahoo finance summary have to be converted into integers

```
def remove_nonnum(k):
    l=[]
    for i in k:
        try:
            l.append(int(i))
        except Exception as e:
            l.append(np.nan)
            continue
    return l

year=remove_nonnum(filtered_df['Year'].to_list())
filtered_df['Year']=year
```

Earnings Variability and **Assets' Growth** rate columns necessitate a numerical conversion since, as it can be intuitively understood, those values are representative in absolute terms (i.e. the magnitude of the values rather than the direction) and, inter alia, become easier to be understood and analysed.

Earnings Variability and Assets' Growth are interesting as absolute values (is the dimension not the direction)

```
filtered_df['Earnings variability'] = filtered_df['Earnings variability'].abs()
filtered_df['Growth'] = filtered_df['Growth'].abs()
```

Divide by 100 the Debt to Equity Ratio because in Yahoo Finance is intended as %

```
filtered_df['Debt to Equity'] = filtered_df['Debt to Equity']/100
```

The **Asset** column needs, indeed, a numerical conversion by standardizing all the asset values into USD through the same procedure as it has been seen before.

Convert all Asset Size in USD

```
from currency_converter import CurrencyConverter # takes real time exchange rates

c = CurrencyConverter()
print(c.convert(1,'USD', 'GBP')) # how many pounds is worth one dollar
```

```
0.7863804247460758
```

```
currencies = list(filtered_df['Currency'].value_counts().index)
exchange_dict = {} # contains all the existent currencies in the DataFrame with the corresponding exchange rate
for i in currencies: # specified exchange rates date back to 02/06/2024
    if i=='TWD':
        exchange_dict[i] = 32.48
    elif i == 'QAR':
        exchange_dict[i] = 3.64
    elif i == 'RUB':
        exchange_dict[i] = 90.43
    elif i == 'ARS':
        exchange_dict[i] = 893.82
    elif i == 'ILA':
        exchange_dict[i] = 5517
    else:
        exchange_dict[i]=c.convert(1,'USD', i.upper())

index = ['Exchange Rate']
exchange_frame=pd.DataFrame(exchange_dict,index=index)
exchange_frame
```

	USD	EUR	INR	TWD	HKD	CAD	KRW	AUD	CNY	MYR	...	CHF	ILA	RUB	ARS	DKK	NZD	QAR	
Exchange Rate	1.0	0.923361	83.299631	32.48	7.80591	1.367498	1367.451524	1.50397	7.240443	4.692521	...	0.914589	5517	90.43	893.82	6.889843	1.636565	3.64	1

2) **Dropping Duplicates**: this step is fundamental and, as showed below, has a real impact on the DF. Even if the procedure is quite straightforward and fast enough to handle it in just a command, it is always necessary to pay the proper attention to the way duplicates are handled. In these cases, specifically regarding stocks, it is quite frequent to notice multiple duplicates because it is, indeed, quite frequent to list the same stock on multiple Stock Exchanges – the so-called **cross-listing**[67] – in different countries (e.g. HSBC is simultaneously traded at the London Stock Exchange, at the Hong Kong Stock Exchange and the New York Stock

[67] _____, Investopedia, "what is cross-listing?", 2022.

Exchange). Again, it is possible to find stocks traded in just one stock exchange (e.g Apple is listed only in the New York Stock Exchange) but its ticker can still be found on local stock exchanges (AAPL.MI, AAPL.VI etc.) due to a main mechanism that is the Global Depository Receipts (GDR). This mechanism allows foreign companies to list their shares into domestic stock exchanges through a depository receipt agreement with the domestic depository banks that prepare a package and issue the shares to their respective stock exchanges[68]. On one hand it would not be legit to maintain the stocks with GDRs because they refer all to the same original stock and the same market capitalization, on the other hand cross-listed stocks should be considered since they present different market capitalizations. The best option is to delete all the stocks with the same name even if this implies the elimination of cross-listed ones:

```
filtered_df=filtered_df.drop_duplicates(subset='Name')
filtered_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 19323 entries, 0 to 28542
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Ticker                19323 non-null  object
 1   Name                  19322 non-null  object
 2   Country               19323 non-null  object
 3   Exchange              19321 non-null  object
 4   Currency              19315 non-null  object
 5   Sector                18740 non-null  object
 6   Industry              18740 non-null  object
 7   Beta                  19323 non-null  float64
 8   Current Ratio         17065 non-null  float64
 9   Debt to Equity        15935 non-null  float64
 10  Market Cap            19296 non-null  float64
 11  Growth                19129 non-null  float64
 12  Year                  19323 non-null  int64
 13  Revenues              18232 non-null  float64
 14  Av Div                9605 non-null   float64
 15  Asset                 19323 non-null  float64
 16  Earnings variability  18855 non-null  float64
 17  Exchange Rate         19323 non-null  float64
dtypes: float64(10), int64(1), object(7)
memory usage: 2.8+ MB
```

The number of stocks is reduced by almost 10,000.

3) **Handling Fundamental Missing Values**: this step can be done at the beginning too, since it implies the elimination of the rows that present the absence of necessary values like Beta or other ones like Growth rate or Market Cap that, even if is not useful in this paper, it serves as a good proxy for stocks with many missing values. This step permits to keep only those stocks that have a minimum amount of information about the Determinants.

[68] _____, Investopedia, “Global Depositary Receipt (GDR) Definition and Example”, 2024.

```
filtered_df=filtered_df.dropna(subset=['Beta','Year','Growth','Market Cap'])
filtered_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 19102 entries, 0 to 28542
Data columns (total 18 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Ticker               19102 non-null  object 
 1   Name                 19101 non-null  object 
 2   Country              19102 non-null  object 
 3   Exchange             19102 non-null  object 
 4   Currency             19102 non-null  object 
 5   Sector               18561 non-null  object 
 6   Industry             18561 non-null  object 
 7   Beta                 19102 non-null  float64
 8   Current Ratio        16861 non-null  float64
 9   Debt to Equity       15791 non-null  float64
 10  Market Cap           19102 non-null  float64
 11  Growth               19102 non-null  float64
 12  Year                 19102 non-null  int64  
 13  Revenues             18038 non-null  float64
 14  Av Div               9581 non-null   float64
 15  Asset                19102 non-null  float64
 16  Earnings variability 18799 non-null  float64
 17  Exchange Rate        19102 non-null  float64
dtypes: float64(10), int64(1), object(7)
memory usage: 2.8+ MB
```

4) **Handling Outliers**: this step is crucial in order to optimize the ML algorithm. When the purpose is to find a pattern and a correlation, it is necessary to eliminate those values that represent the so-called "outliers" that are those values that do not appear ordinarily and that are usually extremely high or low. In this context it is possible to notice two types of outliers: (i) the values that are **unrealistic** (like the 2024 funding year that, even if possible, is too recent and will certainly consider a Beta value that is calculated by using a too short period of time and amount of data) or **irregular** (stocks that are traded on the Over the Counter market[69] will present Beta values not conforming to regular stock market rules) and (ii) the values that show **Beta values too high** or **too low**. In this case, the unrealistic funding year and the OTC stocks are immediately deleted and, just like them, the values that fall in the highest and lowest fifteenth percentile regarding the Beta and the highest and lowest fifth percentile regarding Growth, Earnings Variability and Asset size.

Deleting the founding years that appear unrealistic

```
filtered_df=filtered_df[(filtered_df['Year']>1600) & (filtered_df['Year']<=2024)]
filtered_df.info()
```

PNK refers to OTC stocks that are highly volatile and not regulated.

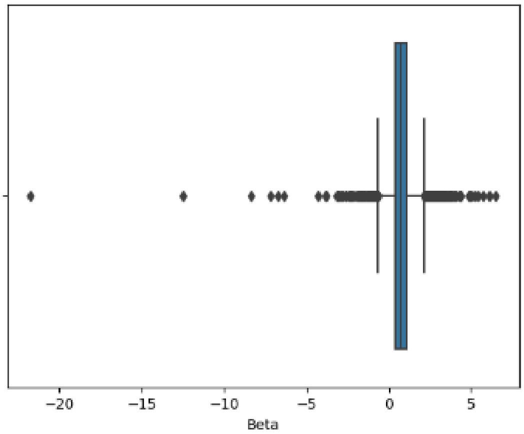
```
without_otc = filtered_df[filtered_df['Exchange'] != 'PNK']
without_otc.info()
```

[69] A Over the Conter (OTC) market trades securities not subject to any specific regulation concerning the organisation and operation of the market itself (Borsa Italiana).

```
Q1 = without_otc['Beta'].quantile(0.15)
Q3 = without_otc['Beta'].quantile(0.85)
IQR = Q3 - Q1

# Define lower and upper bounds for outliers
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

outliers = without_otc[(without_otc['Beta'] < lower_bound) | (without_otc['Beta'] > upper_bound)]

print("Potential Outliers:")
print(outliers.nlargest(5, 'Beta')['Beta'])
print(outliers.nsmallest(5, 'Beta')['Beta'])
```

```
Potential Outliers:
10966    6.492
3943     6.128
12983    5.739
1998     5.403
9238     5.227
Name: Beta, dtype: float64
16925   -21.722
20390   -12.485
8685     -8.357
11016    -7.199
15234    -6.751
Name: Beta, dtype: float64
```

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.boxplot(x='Beta', data=without_otc)
plt.show()
```




```
lower_bound, Q1, Q3, upper_bound
```

```
(-1.365, 0.264, 1.35, 2.979)
```

```
len(outliers)
```

```
114
```

Beta values falling below -1.365 and above 2.979 are eliminated; among these, for example, there was a Beta value equal to -21.722 that, certainly, represent a huge outlier.

```
def remove_outliers_iqr(df, columns):
    for col in columns:
        Q1 = df[col].quantile(0.05)
        Q3 = df[col].quantile(0.95)
        IQR = Q3 - Q1
        lower_bound = Q1 - 1.5 * IQR
        upper_bound = Q3 + 1.5 * IQR
        df = df[(df[col] >= lower_bound) & (df[col] <= upper_bound)]
    return df

clean_without = remove_outliers_iqr(without_otc, ['Growth','Earnings variability','Asset'])
clean_without.describe()
```

Fist and top 5% of Growth, Earnings Variability and Asset Size values are eliminated too.

5) **Deleting Non Relevant Data**: the last step before of concluding with the imputation is the elimination of non-relevant data that is the removal of those values whose categories (in this specific case the **Sector** of origin and the **Country** of the company) can't count a sufficient number of population (e.g., as showed below, companies that are located in Uruguay are only 7, not a sufficient number that permits to ascertain a possible correlation).

Sector

```
clean_without.groupby(clean_without['Sector'])['Beta'].count()
```

```
Sector
Basic Materials           2929
Communication Services     973
Consumer Cyclical         2927
Consumer Defensive        1419
Energy                     990
Financial                    3
Financial Services        2384
Healthcare                2039
Industrial Goods             1
Industrials               3814
Real Estate               1189
Technology                2650
Utilities                  640
Name: Beta, dtype: int64
```

Country

```
clean_without.groupby(clean_without['Country'])['Beta'].count()
```

```
Country
Argentina          55
Australia        1095
Austria            83
Azerbaijan          1
Bahamas             1
                 ...
United Kingdom    860
United States    4869
Uruguay             7
Vietnam             2
Zambia              2
Name: Beta, Length: 91, dtype: int64
```

In both cases, for Sector and Country, will be kept only those stocks whose category counts more than 50.

```
# Deleting those stocks that correspond to a sector that counts less than 50 (a number that could pontentially be significant)
fil = clean_without.groupby(clean_without['Sector'])['Beta'].count()
not_relevant = list(fil[fil<50].index)
clean_without = clean_without[~clean_without['Sector'].isin(not_relevant)]
```

```
# Deleting those stocks that correspond to a country that counts less than 50 (a number that could pontentially be significant)
fil = clean_without.groupby(clean_without['Country'])['Beta'].count()
not_relevant = list(fil[fil<50].index)
clean_without = clean_without[~clean_without['Country'].isin(not_relevant)]
```

6) **Impute Missing Values**: finally, to conclude the data cleaning process, the missing values in the Debt to Equity's, Growth's and Payout Ratio's columns will be filled through the so-called "imputation" process. The latter involves the

replacement of missing values with, in this specific case, the average value encountered on a different parameter that is usually well established and certainly correlated to the imputing column. This code is designed to impute missing values in the dataset by replacing them with the mean values of their respective industries. For example, the Debt to Equity value missing in a certain row is replaced with the average Debt to Equity corresponding to the row's specific Industry.

```
# Function to impute missing values based on industry mean
def impute_debt_to_equity(row):
    industry_mean = clean_without.groupby('Industry')['Debt to Equity'].mean()
    # Create a dictionary mapping industry to its mean
    industry_mean_dict = industry_mean.to_dict()
    if pd.isnull(row['Debt to Equity']) and not pd.isnull(row['Industry']):
        industry = row['Industry']
        return industry_mean_dict[industry]
    else:
        return row['Debt to Equity']

def impute_growth(row):
    industry_mean = clean_without.groupby('Industry')['Growth'].mean()
    industry_mean_dict = industry_mean.to_dict()
    if pd.isnull(row['Growth']) and not pd.isnull(row['Industry']):
        industry = row['Industry']
        return industry_mean_dict[industry]
    else:
        return row['Growth']

def impute_div(row):
    industry_mean = clean_without.groupby('Industry')['Av Div'].mean()
    industry_mean_dict = industry_mean.to_dict()
    if pd.isnull(row['Av Div']) and not pd.isnull(row['Industry']):
        industry = row['Industry']
        return industry_mean_dict[industry]
    else:
        return row['Av Div']

# Apply the function to create a new column 'Imputed Debt to Equity'
clean_without['Debt to Equity'] = clean_without.apply(impute_debt_to_equity, axis=1)
clean_without['Growth'] = clean_without.apply(impute_growth, axis=1)
clean_without['Av Div'] = clean_without.apply(impute_div, axis=1)
```

```
clean_without.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14758 entries, 0 to 14757
Data columns (total 18 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Ticker               14758 non-null  object
 1   Name                 14757 non-null  object
 2   Country              14758 non-null  object
 3   Exchange             14758 non-null  object
 4   Currency             14758 non-null  object
 5   Sector               14286 non-null  object
 6   Industry             14286 non-null  object
 7   Beta                 14758 non-null  float64
 8   Current Ratio        12947 non-null  float64
 9   Debt to Equity       12602 non-null  float64
 10  Market Cap           14758 non-null  float64
 11  Growth               14758 non-null  float64
 12  Year                 14758 non-null  int64
 13  Revenues             14146 non-null  float64
 14  Av Div               7446 non-null   float64
 15  Asset                14758 non-null  float64
 16  Earnings variability 14565 non-null  float64
 17  Exchange Rate        14758 non-null  float64
dtypes: float64(10), int64(1), object(7)
memory usage: 2.0+ MB
```

3.1.2 Graphical Representation of the Data

With the conclusion of the data cleaning process, now it is possible to graphically represent the final DF.

First of all, it can be attached the description of the DF after the data cleaning phase:

```
clean_without.describe()
```

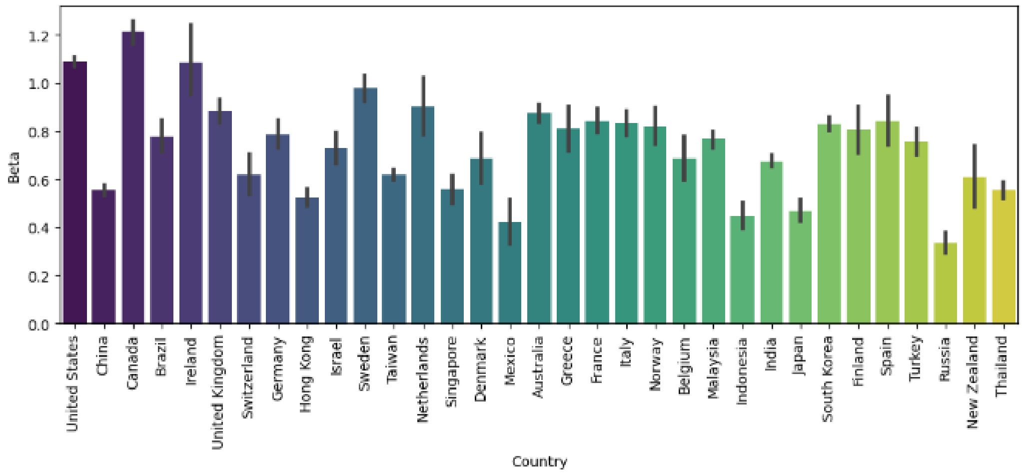
	Beta	Current Ratio	Debt to Equity	Market Cap	Growth	Year	Revenues	Av Div	Asset	Earnings variability	Exchange Rate
count	14758.000000	12947.000000	12602.000000	1.475800e+04	14758.000000	14758.000000	1.414600e+04	7446.000000	1.475800e+04	14565.000000	14758.000000
mean	0.795848	3.238697	0.812506	4.245919e+09	0.103280	1985.413335	7.441920e+09	0.658748	1.141158e+10	1.153386	463.353618
std	0.531155	6.500033	1.859845	3.928239e+10	0.144024	36.603075	5.731121e+10	0.770065	6.900941e+10	1.942063	2309.093475
min	-1.162000	0.001000	0.000010	3.313980e+02	0.000000	1609.000000	-6.977937e+08	0.000000	7.312627e-01	0.000000	0.786380
25%	0.431000	1.071000	0.103997	3.875912e+07	0.026838	1976.000000	3.226733e+07	0.244600	6.227477e+07	0.281081	1.000000
50%	0.727000	1.669000	0.382675	1.948154e+08	0.060293	1995.000000	1.695057e+08	0.453550	3.115378e+08	0.544325	4.692521
75%	1.096000	2.982000	0.889535	1.174064e+09	0.122839	2009.000000	9.797465e+08	0.779200	1.988028e+09	1.195722	32.480000
max	2.978000	107.374000	59.377950	3.156600e+12	4.051939	2023.000000	1.724146e+12	8.914600	1.439289e+12	30.074109	16040.000000

Among all the things that can be noticed it should be underlined the mean value of Beta; it is around 0.8 which makes this portfolio not a total mimicking of the market one. This could be caused by many reasons, starting from the data cleaning process and arriving to the stocks selected. Still, 0.8 is a good average value and can work for the purposes of this paper (still considering that the whole work has not been done by using virtual machines and other advanced technologies that could have helped to download an even larger list of stocks).

The purpose of the following graphs is to potentially extract information about general trends and visualize the DF in a significative way, highlighting all the fundamental variables and results paying a lot of attention to the data visualization of topics already covered during this paper. The graph representation is finalized through two factors that are not included in the paper research: (i) the company **Sector** and (ii) the company **Country** location. These two Determinants are generally recognised and represent a good tool in order to perform the graphical representation of the resulting DF. The graphs will alternate in order to follow a general reasoning and thinking that can highlight some key topics.

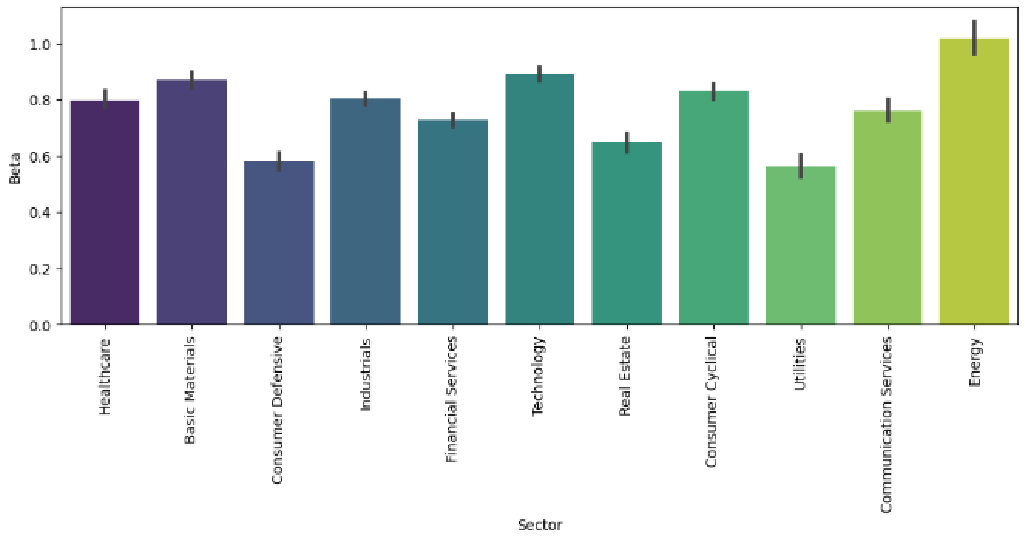
The first graph presented below shows how the average Beta changes among the countries; along with the general representation of data, the barplot presents the “error bars” (i.e. a visual representation of the variability of the data that can indicate the uncertainty or standard error around the bar's height) which provides a clear representation of the results.

```
plt.figure(figsize=(12, 4))
sns.barplot(x="Country", y="Beta", data=clean_without, palette='viridis')
plt.xticks(rotation=90)
plt.show()
```



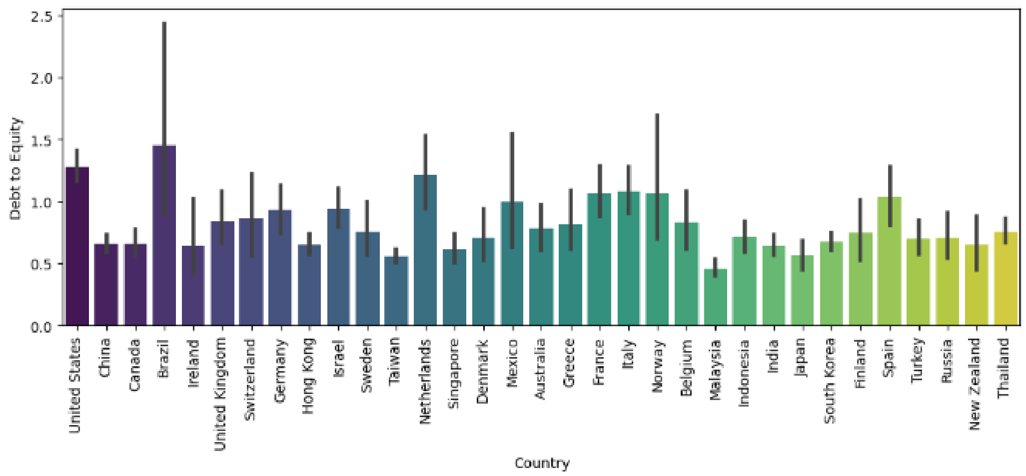
On one hand, **United States** and **Canada** present average Beta values of above 1.0 and an error bar almost minimal; they are characterized by the predominance of **Tech** and **Energy** companies that, as it can be showed below, appear as the sectors with the highest Beta.

The low Beta countries, on the other hand, can be distinguished in (i) those countries with high government intervention and influence, like **Russia** and **China**, that are characterized by state-owned companies or the direct influence in the company by the government through aids or direct purchase of the goods sold and (ii) those countries, like **Japan**, with numerous mature companies that have already a steady cash stream and a market dominance that assure low Betas. Russia and China present, additionally, a high number of companies operating in the **Utilities** and **Consumer Defensive** sectors that are distinguished by low average Betas (see the exhibit below).



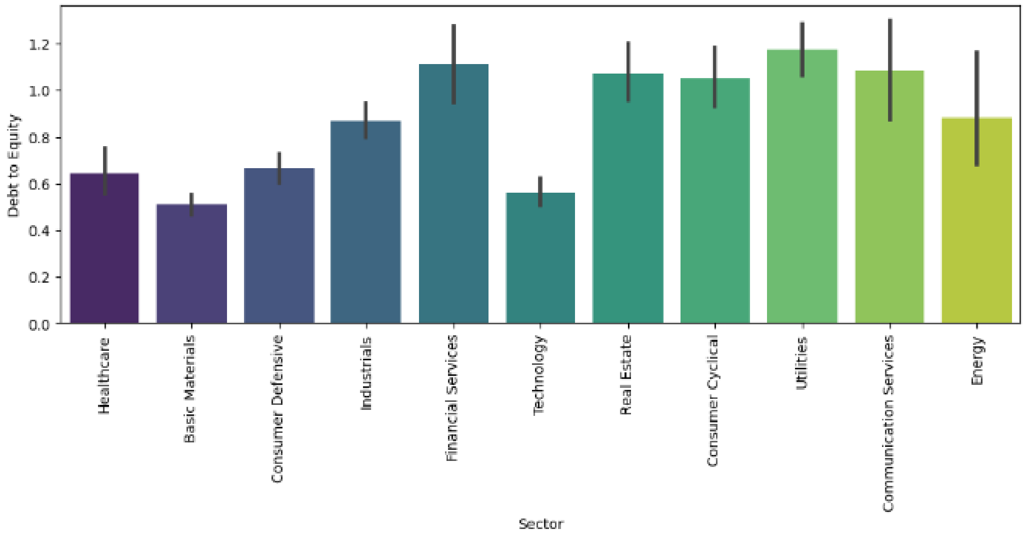
The other sectors seem to be around 0.8 which, as explained before, is the average value of this DF.

Now, it is possible to focus on those countries and sectors with highest or lowest Beta; the following graphs can help at finding a visual correlation with the highlights just showed. For example, if it is taken a look at the following chart showing the average Debt to Equity ratio per country, it is possible to notice a certain correlation with the Beta-Country graph.

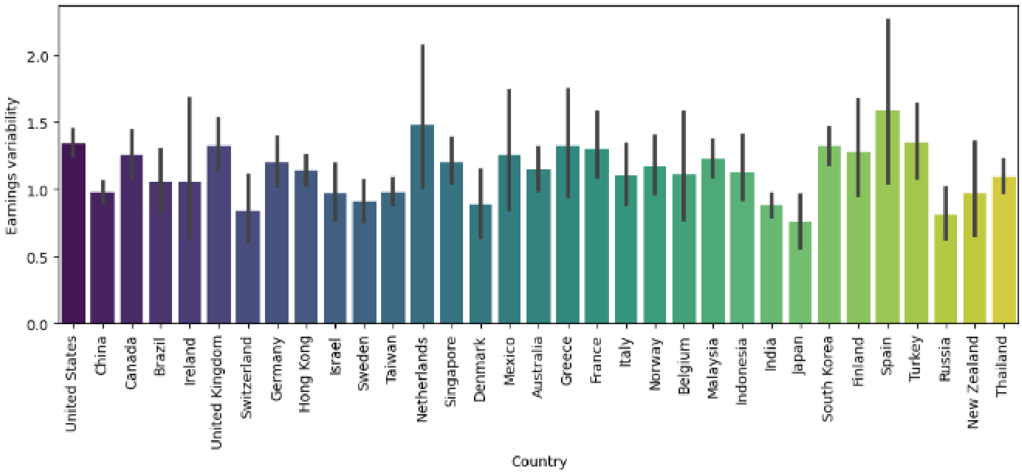


Higher than average Country Beta values (approximately 0.8) tend to coincide with higher than average Country Debt to Equity values (approximately 0.8) furtherly consolidating the paper conclusion that proves the positive correlation between Beta and Debt to Equity. Some patterns are consistent (like United States that presents the highest Debt to Equity average or Russia and China that present low Debt to Equity too), but some contradictions occur too (e.g. Canada that was a high Beta country has a low Debt to Equity). Overall, graphically appears a certain positive correlation between Debt to Equity and Beta.

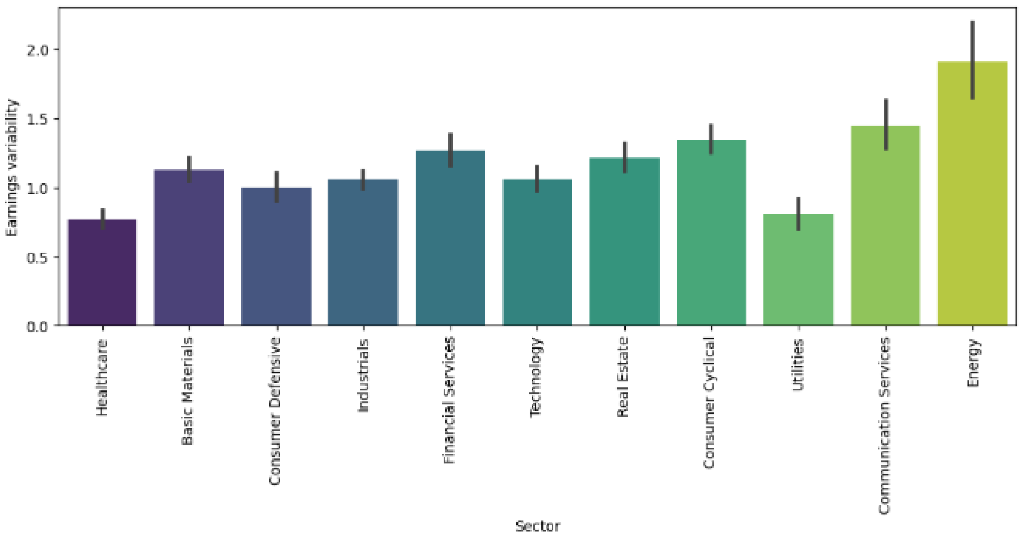
On the contrary, if it is considered the Debt to Equity-Sector graph showed below, this correlation is no more so evident and it can't be noticed any general pattern. The graphical representation can't always show optimal correlation because, otherwise, the correlation results that will be presented in the next paragraph would be way more convincing (it is kind of an anticipation, but no one would sincerely expect a perfect correlation between the Determinants and Beta).



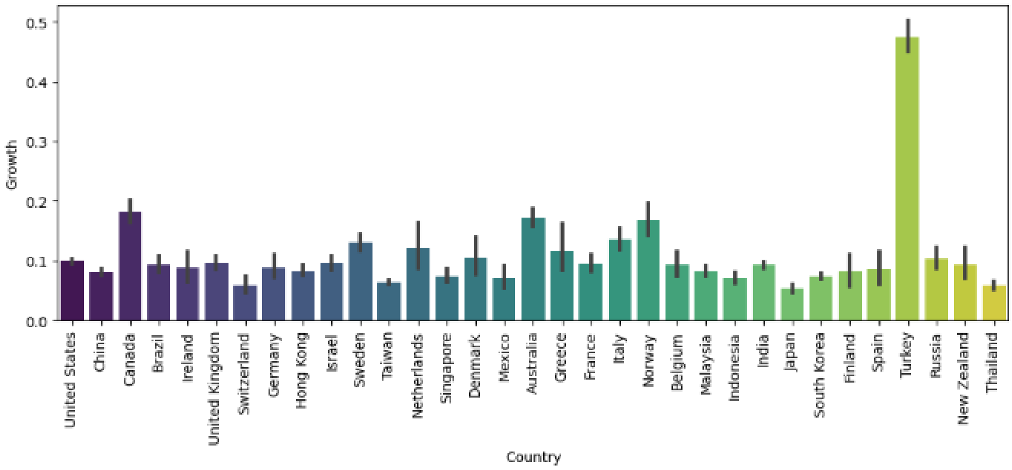
The Earnings Variability graphs can be representative too. The first one, in accordance with the Country factor, expresses the average Earnings Variability per country. United States are again characterized by high Earning Variability that according to this paper findings should imply a positive correlation with Beta. China, Russia and Japan continue to confirm the low Earning Variability values that are accompanied by a low average Beta.



With regard to the Sector-Earnings Variability graph, it becomes evident that the Energy sector continues to be distinguished by its high values related to those Determinants that have a positive correlation with Beta. The Utilities sector confirms the positive correlation of Earnings Variability with the Beta since, in here again, presents a low value.



In the end, it is showed the Growth-Country graph that, along with noticing now a consistency of Canada's higher than average Beta with Canada higher than average Growth rate, presents a curious result about Turkey. Turkey is in here an outlier that is characterized by extreme growth rates; this result, even partially caused by the selected stocks, describes a recently established trend in Turkish stocks. In fact, in the recent years, especially in the last months, Turkey distinguished itself from the other countries for the high growth stocks operating in the Tech[70] and Finance sectors. The purpose of highlighting this element is to understand a fundamental aspect of this analysis, these results depend largely on the recent market trends and macroeconomic events. For example, the last three/four years the world economy has been changed by the pandemic that in some cases totally pushed some sectors or countries and probably even changed some general established believes. Turkey is the practical example of how in a few years, or months, **everything can change**.



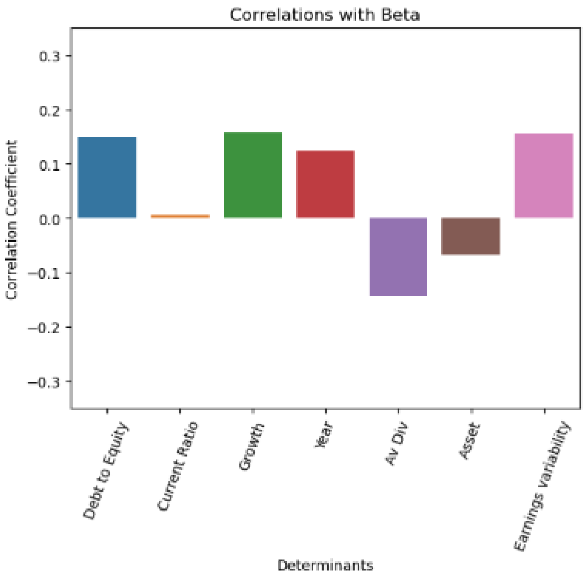
All those representations do not imply a statistical, de facto, result but try to show which factors can be helpful in order to make a first impression about a company Beta. If it is presented a company operating in the Utilities sector, in Russia, with low Debt to Equity and low Growth it should be quite sure the fact that the Beta will be below 1 (according to what the graphs suggested).

Nevertheless, it is important to notice that, even if there are some inconsistencies between the graphs (e.g. Utilities are low Beta but have high average Debt to Equity) this does not mean that the correlation doesn't exist, in the following paragraph will be exhibited the final correlations between the DF Determinants and Beta.

[70] ____, "Turkish stocks soar in world-leading rally as 'tech mania' grips market", Financial Times, 2024.

3.1.3 Statistical Representation of the Data

This brief paragraph serves as spotlight for what concerns the statistical overview of the DF. With this purpose, the different correlations with Beta are represented in a barplot graph that is showed below; it helps the reader to remember, and furtherly confirm, what the relationships between the Determinants and Beta are:



The graph confirms the correlations showed and explained in the previous chapter when it was used the U.S. Stocks sample. In here, even if with a less consistent result, the correlations are perfectly in line with the premises and analyses of this paper. Still, these results are way more important because confirm the relationships that exist between the Beta and the Determinants, just like in the U.S. Stocks sample, but applied to an even larger sample that strengthen the possibility of obtaining a properly functioning ML algorithm. The latter will be developed in the following paragraph.

3.2 Development and computation of a Machine Learning code

Now, it starts the real, computational phase of this paper that, through statistical and advanced tools provided by Python Scikit and XGBoost packages, permits to develop a fully operative ML code that will be used in the end to show how, by inserting some information about a company, it can be obtained a prediction/calculation of its Beta. Even if at the beginning of this paper the purpose was to predict the exact Beta, the results and the data didn't permit to reach such an ambitious goal. As it will be showed in the end the prediction will limit itself to identifying the imaginary company's Beta as greater or lesser than 1. This will serve as a possible way of demonstrating how, by inserting some representative values according to the correlations showed, the result will change.

The steps are the following:

- **adjusting** the data for the Machine Learning code
- **modelling** the data and splitting in **training** and **testing** Sets
- **creating** a XGBoost Classifier and **fitting** the model

Since Scikit and XGBoost work by using only categories and integer numbers all the DF has to be adjusted and modelled in order to enter into the effective creation of the algorithm.

3.2.1 Adjusting the Data for the Machine Learning code

As it was said before, in order to work with Scikit or XGBoost it is necessary to convert all values before into categorical (i.e. categories that include a certain range of values, for example the category "1-5" includes all those values that were numerical and starting from 1 to 5). This step is performed for all the Determinants with the exception of those ones that are already categories (i.e. Country and Sector Determinants).

```
# Debt to Equity
rev_category=pd.qcut(clean_without['Debt to Equity'],30)
clean_without['Debt to Equity']=rev_category

# Current Ratio
rev_category=pd.qcut(clean_without['Current Ratio'],30)
clean_without['Current Ratio']=rev_category

# Earnings Growth Rate
rev_category=pd.qcut(clean_without['Growth'],30)
clean_without['Growth']=rev_category

# Company's founding Year
rev_category=pd.qcut(clean_without['Year'],30)
clean_without['Year']=rev_category

# Company's Payout Ratio
rev_category=pd.qcut(clean_without['Av Div'],30)
clean_without['Av Div']=rev_category

# Asset Size
rev_category=pd.qcut(clean_without['Asset'],30)
clean_without['Asset']=rev_category

# Earnings Variability
rev_category=pd.qcut(clean_without['Earnings variability'],30)
clean_without['Earnings variability']=rev_category
```

The Determinants' values, that before were all float numbers, are all grouped into these categories. This way of grouping is certainly used because it helps to limit the possible values to a certain number of categories. In this case, all the columns' values were divided in 30 possible categories.

This subdivision is clear when it is showed the resulting DF that, along with the **elimination** of **unnecessary** columns that will not serve as possible determinants, takes this form and image:

```
# Deleting Ticker, Name, Exchage, Currency, Industry, Exchange Rate, Revenues and Market Cap (they do not represent a potential determinant)

clean_without = clean_without.drop(columns=['Ticker', 'Name','Exchange','Currency', 'Exchange Rate', 'Revenues', 'Market Cap'])
```

```
clean_without
```

	Country	Sector	Industry	Beta	Current Ratio	Debt to Equity	Growth	Year	Av Div	Asset	Earnings variability
0	United States	Healthcare	Diagnostics & Research	1.122	(2.552, 2.834]	(0.436, 0.489]	(0.0249, 0.0286]	(1997.0, 1999.0]	(0.205, 0.23]	(6843406803.547, 11483306359.77]	(0.185, 0.212]
1	United States	Basic Materials	Aluminum	2.446	(1.398, 1.484]	(0.337, 0.383]	(0.0102, 0.0138]	(2015.0, 2017.0]	(0.0532, 0.088]	(11483306359.77, 21647739539.314]	(2.601, 3.455]

In the resulting DF are kept only those columns that are necessary for the ML code: the Determinants that have been analysed deeply during this paper along with the Country and Sector addition.

For the ML purpose this step is crucial because, otherwise, the singular values, without any grouping, would become meaningless since it would be really difficult for the ML code to properly derive a possible pattern. This means that, if there are all singular values, some of them are maybe controversial and cause to the ML code a high level of confusion and disturbance.

ML code works in a particular way: it tries to find multiple patterns and, when it is asked to predict a value based on those patterns, evaluates the patterns found considering the effective past results; if the patterns work well (the results predicted frequently coincide with the effective ones) the ML algorithm is improved and accepts those patterns. In this case the ML code works through a "**trial and error**" mechanism. Because it works through patterns, if they are provided all singular values (i.e. in this case about 15,000) it would be like creating the same number of categories which makes it impossible to find some logical sequences (it will work only on equalities and when it does not find a past value will not give any answer). Say that it should be created this ML code leveraging the fact that there are some correlations but, for example, company A presents a Debt to Equity ratio of 1.003 with a Beta of 1.2, company B has a Debt to Equity of 1.4 with a Beta of 0.6 and all other companies with a Debt to Equity greater than one have a Beta greater than one. If it is not created a category for Debt to Equity values that go from 1 to 1.5, then it could happen that, when it is inserted the Debt to Equity ratio of the want-to-predict company and is 1.4 the ML code could use the singular case of company B and predict a Beta **below 1** when, if considered the category and the intention to leverage the correlations, the average Beta of companies with Debt to Equity greater than 1 is, indeed, **greater than 1**.

3.2.2 Modelling the Data and Splitting in Training and Testing Sets

The DF is ready to be modelled and the first thing to do is to download the libraries:

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import LabelEncoder
```

This permits to access the Scikit and XGBoost libraries that will perform the ML algorithm. The next thing to do is (i) creating, through an appropriate Python function, new columns that are called with all the new categories names and (ii) converting all rows (with the exception of the original columns and Beta column) in True-False values that are True when the column represents the characteristic of the company and False otherwise. For example, below is represented the DF with the aforementioned changes, the company in the third row presents a True value in the "Country_China" column, this means that that company is a Chinese company.

```
# Create a copy exclusively for ML purposes
ML_df=clean_without.copy()

# Categoricals are conerted into True/False columns that are necessary in ML codes
X_categorical = pd.get_dummies(ML_df[['Country','Sector','Industry','Debt to Equity','Current Ratio', 'Growth', 'Year', 'Av Div', 'Asset', 'Earnings var
ML_df=pd.concat([ML_df,X_categorical],axis=1)
ML_df=ML_df.drop(columns=['Country','Sector','Industry','Debt to Equity','Current Ratio', 'Growth', 'Year', 'Av Div', 'Asset', 'Earnings variability'])

ML_df.head()
```

	Beta	Country_Australia	Country_Belgium	Country_Brazil	Country_Canada	Country_China	Country_Denmark	Country_Finland	Country_France	Country_Germany
0	1.122	False	False	False	False	False	False	False	False	False
1	2.446	False	False	False	False	False	False	False	False	False
2	0.971	False	False	False	False	True	False	False	False	False
3	1.577	False	False	False	False	False	False	False	False	False
4	0.334	False	False	False	False	False	False	False	False	False

The DF has additionally been cleaned of all the original columns with the categories. After having separated the Beta column from the DF (which represents the target or result) the Beta values are converted in two categories: (i) **Beta greater than 1** and (ii) **Beta lower than 1**. This separation is, as it was said before, aimed at targeting future predictions towards the only forecast of Beta lower or higher than 1.

```
# Separate Beta from the rest
beta = ML_df['Beta']
ML_df=ML_df.drop(columns=['Beta'])

# Categorize Beta too
bins = [-10,1,10]
target = pd.cut(beta, bins)

pd.cut(clean_without['Beta'], bins).value_counts().sort_index()
```

```
Beta
(-10, 1]    10242
(1, 10]      4516
```

Beta greater than 1 is represented through the number "1" and Beta lower than 1 wit the number "0".

```
# Create a LabelEncoder
label_encoder = LabelEncoder()

# Fit and transform the target variable
target_encoded = label_encoder.fit_transform(target)
```

```
# Beta categories are transformed into numbers that can be read by the ML program (e.g. Beta values above 2 are encoded with number 8)
target_encoded
```

```
array([1, 1, 0, ..., 1, 0, 0])
```

The next step is to separate the data into **training** and **testing** sets, this means that the DF is, in fact, split in two parts: (i) the **training set** which represents the starting point and population that permits to Scikit to build the ML code and (ii) the **testing set** that can help the ML code to improve and understand if the resulting algorithm is well performing and predicts accordingly.

```
# Data is split in training and testing sets
X_train, X_test, y_train, y_test = train_test_split(ML_df, target_encoded, test_size=0.2, random_state=42)
```

Once the training and testing sets have been defined, it is possible to build the ML algorithm

3.2.3 Creating a XGBoost Classifier and fitting the Model

The final step of the ML algorithm preparation coincides with the creation of the XGBoost classifier. XGBoost is an efficient and scalable implementation of **gradient boosting** (powerful machine learning technique used for regression and classification tasks that builds an ensemble of *decision trees*, where each subsequent tree attempts to correct the errors of its predecessor), designed for speed and performance. The utilization of XGBoost substitutes the long and difficult creation of decisional trees that constitute the base for the majority of ML algorithms. The first step is to **create** an XGBoost Classifier:

```
clf = XGBClassifier(random_state=42)
```

And then, **fitting** the model:

```
clf.fit(X_train, y_train)
```

```
XGBClassifier
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=None,
              num_parallel_tree=None, random_state=42, ...)
```

XGBoost does all the hard work and finds the optimal patterns in order to build the ML code, the training set previously defined is fitted through XGBoost and it is created a new model.

The model built is now ready to be tested, this means that it will try to predict the results (in this case if Beta would be higher or lower than 1) of the testing set, previously separated from the training set. Once the model works on the training set it is expected to see how many of the testing values are correctly predicted (the results are obviously known because the testing set is composed of the actual Beta values). Based on how many of the tests are correctly predicted the model will be more precise and well performing. A ML code performance is measured through the so-called "**accuracy**" which, as the name itself suggests, is a metric used to evaluate the performance of a classification model. It is defined as the ratio of the number of correct predictions made by the model to the total number of predictions. Accuracy can be calculated using a *confusion matrix*, which is a table used to describe the performance of a classification model on a set of test data for which the true values are known. The confusion matrix includes four types of outcomes:

- **True Positives (TP)**: Instances where the model has correctly predicted the positive class (that the value was correct, and the model agreed with it).
- **True Negatives (TN)**: Instances where the model has correctly predicted the negative class (the value was not correct, and the model predicted it).
- **False Positives (FP)**: Instances where the model has incorrectly predicted the positive class.
- **False Negatives (FN)**: Instances where the model has incorrectly predicted the negative class.

Using the confusion matrix, accuracy can be expressed as:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

The accuracy can be easily computed in Python through a simple function, the accuracy of this model is:

```
y_pred = clf.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.2f}")
```

```
Accuracy: 0.74
```

The model has an accuracy of **0.74**, meaning it correctly predicted the outcomes for 74% of the test set instances, this result is neither excellent nor bad, it seems that the model is quite good at predicting if a certain company's Beta is higher or lower than 1 but still it's not perfect.

In the final paragraph the model is used to predict the Beta result based on some information about the Beta Determinants, let's conclude this paper with these results.

3.3 Apply the Model to Practical Examples

The whole paper ends with this final paragraph that, as it was initially intended, should have been able to predict the precise Beta value. Since the corresponding accuracy wouldn't have been acceptably high, the corresponding model was built in order to predict a Beta higher or lower than 1. Given the graphical, numerical and statistical analysis presented so far, it is now the moment of predicting a Beta given a made-up company with the Determinants' characteristics specifically chosen in order to create two distinct business case: (i) a company characterized by all those Determinants' values that suggest a low Beta and (ii) a company characterized by all those Determinants' values that suggest a high Beta. Below it is presented the code used to access the Model prediction:

```
def ml_prediction():
    global guess

    country = 'Country_' + str(input('Insert Country Name: '))
    sector = 'Sector_' + str(input('Insert Sector Name: '))
    d_e = float(input('Insert Debt to Equity value: '))
    c_r = float(input('Insert Current Ratio value: '))
    growth = float(input('Insert Earnings Growth rate: '))
    year = float(input('Insert Founding year: '))
    av_div = float(input('Insert Payout Ratio: '))
    asset = float(input('Insert Asset size: '))
    e_v = float(input('Insert Earnings variability: '))

    # Initialize the guess DataFrame with False values and the correct dtype
    guess = pd.DataFrame(data=[[False] * 400], columns=ML_df.columns)

    # Categorize inputs
    d_e = 'Debt to Equity_' + str(pd.cut([d_e], bins=clean_without['Debt to Equity'].values.categories, labels=False, include_lowest=True)[0])
    c_r = 'Current Ratio_' + str(pd.cut([c_r], bins=clean_without['Current Ratio'].values.categories, labels=False, include_lowest=True)[0])
    growth = 'Growth_' + str(pd.cut([growth], bins=clean_without['Growth'].values.categories, labels=False, include_lowest=True)[0])
    year = "Year_" + str(pd.cut([year], bins=clean_without['Year'].values.categories, labels=False, include_lowest=True)[0])
    av_div = "Av Div_" + str(pd.cut([av_div], bins=clean_without['Av Div'].values.categories, labels=False, include_lowest=True)[0])
    asset = "Asset_" + str(pd.cut([asset], bins=clean_without['Asset'].values.categories, labels=False, include_lowest=True)[0])
    e_v = 'Earnings variability_' + str(pd.cut([e_v],
                                       bins=clean_without['Earnings variability'].values.categories, labels=False, include_lowest=True)[0])

    # List of columns to set to True
    subst = [country, sector, d_e, c_r, growth, year, av_div, asset, e_v]

    # Set the respective columns to True
    for i in subst:
        guess.at[0, i.replace(']', ')')] = True

    predicted_beta = clf.predict(guess)

    if predicted_beta[0] == 1:
        print("The company's beta is higher than 1")
    else:
        print("The company's beta is lower than 1")
```

The function “ml_prediction” was built with the purpose of making it possible to directly insert all the Determinants and see how the prediction goes. It will be, in fact, very curious to see how the Model predicts differently the Beta depending on the inputs inserted.

The demonstration can start with the first example, a company characterized by the presence of the Determinants’ values that would theoretically suggest a *Beta higher than 1*, hypothetically a start-up company or a general SME. By inserting, also looking at the graphical representation of before, Canada as Country and Energy as Sector (that appeared as the factors with highest Beta average), a high Debt to Equity value (i.e. 2), a consistent Earnings Growth rate (i.e. 20%), it’s a start-up (founded in 2023), a really low Dividend Payout ratio (i.e. the 5% of net income), a particularly low Asset Size (i.e. 10,000,000.00 dollars, typical of a SME) and characterized by high Earnings Variability.

Given those inputs, the ML code should certainly predict a company’s Beta higher than 1:

```
ml_prediction()

Insert Country Name:  Canada
Insert Sector Name:  Energy
Insert Debt to Equity value:  2
Insert Current Ratio value:  1
Insert Earnings Growth rate:  0.2
Insert Founding year:  2023
Insert Payout Ratio:  0.05
Insert Asset size:  10000000
Insert Earnings variability:  3
The company's beta is higher than 1
```

The code seems to work appropriately, the company’s input values suggest that its Beta is higher than 1, **the ML code works well in this case**.

In the opposite case, the purpose is to look if, by inserting parameters that would suggest a *low Beta value*, the ML code works again appropriately. By inserting this time, Russia as Country and Consumer Defensive as Sector (that appeared as the factors with lowest Beta average from the previous graphical representation), a low Debt to Equity value (i.e. 0.3), a weak Earnings Growth rate (i.e. 2%), it’s a mature firm (founded in 1956), a really high Dividend Payout ratio (i.e. the 90% of net income), a particularly high Asset Size (i.e. 1,000,000,000.00 dollars, typical of a company with international consolidated business) and characterized by a low Earnings Variability (that again suggests how stable the cash flows are). This kind of company should exhibit a low Beta. Let’s see what the ML algorithm suggests:

```
ml_prediction()
```

```
Insert Country Name:  Russia
Insert Sector Name:  Consumer Defensive
Insert Debt to Equity value:  0.3
Insert Current Ratio value:  2
Insert Earnings Growth rate:  0.02
Insert Founding year:  1956
Insert Payout Ratio:  0.9
Insert Asset size:  1000000000
Insert Earnings variability:  0.3
The company's beta is lower than 1
```

Again, **the ML code works well and is able to well differentiate the two scenarios**. This means that, indirectly, this ML code has demonstrated that the Determinants have an effective impact on the companies' Betas. The correlation that it has been showed multiple times in this paper is accompanied by the practical example of how a ML code interprets those Determinants, their impact is doubtless and extensively demonstrated in the last chapter.

Through the following, last example, it will be showed how, as two parameters change, the results of the prediction change too. The following code has been created with the purpose of having the possibility to choose two parameters and for each combination of the two parameters see if the result changes and how:

```
def prediction_matrix():
    global guess, matrix_ml, first, second, ml_row, two_determinants_columns, f_1, f_2,row

    two_determinants = str(input('Insert the two determinants that have to be analyzed (d_e, growth, c_r, year etc.): '))
    two_determinants_columns = str(input('Insert the two determinants column names (Debt to Equity, Growth, Current Ratio, Year etc.): '))
    two_determinants = (two_determinants).split('|')
    two_determinants_columns = (two_determinants_columns).split('|')

    country = "Country_Taiwan"
    sector = "Sector_Technology"
    d_e = 'Debt to Equity_' + str(pd.cut([0.8], bins=clean_without['Debt to Equity'].values.categories, labels=False, include_lowest=True)[0])
    c_r = 'Current Ratio_' + str(pd.cut([3], bins=clean_without['Current Ratio'].values.categories, labels=False, include_lowest=True)[0])
    growth = 'Growth_' + str(pd.cut([0.01], bins=clean_without['Growth'].values.categories, labels=False, include_lowest=True)[0])
    year = "Year_" + str(pd.cut([1850], bins=clean_without['Year'].values.categories, labels=False, include_lowest=True)[0])
    av_div = "Av Div_" + str(pd.cut([0.8], bins=clean_without['Av Div'].values.categories, labels=False, include_lowest=True)[0])
    asset = "Asset_" + str(pd.cut([101000000], bins=clean_without['Asset'].values.categories, labels=False, include_lowest=True)[0])
    e_v = 'Earnings variability_' + str(pd.cut([0.5],
                                            bins=clean_without['Earnings variability'].values.categories, labels=False, include_lowest=True)[0])

    # List of columns to set to True

    dfact = {"country":country,"sector":sector,"d_e":d_e,"c_r":c_r,"growth":growth,"year":year,"av_div":av_div,"asset":asset,"e_v":e_v}

    del dfact[two_determinants[0]]
    del dfact[two_determinants[1]]

    guess = pd.DataFrame(data=[[False] * 400], columns=ML_df.columns)
    subst = list(dfact.values())
    # Set the respective columns to True
    for i in subst:
        guess.at[0, i.replace(']', ')')] = True

    def generate_dispersed_values(data, num_values=20):
        percentiles = np.linspace(0, 100, num_values)
        values = np.percentile(data, percentiles)

        formatted_values = [f"{value:.2f}" for value in values]
        # Convert values to a list

        return formatted_values
```

```
ML_example = pd.read_csv('ModeledData1.csv', sep=',')

first = generate_dispersed_values(ML_example[two_determinants_columns[0]].dropna())
second = generate_dispersed_values(ML_example[two_determinants_columns[1]].dropna())
first_str = [str(i) for i in first]
second_str = [str(i) for i in second]
prediction = 0
f_1 = 0
f_2 = 0

matrix_ml = pd.DataFrame(data=[[False] * 18],columns=first_str[1:-1])
for i in range(1,19):
    ml_row = guess
    row = [second_str[i]]
    for k in range(1,19):
        ml_row = guess
        f_1 = two_determinants_columns[0] +'_'+ str(pd.cut([float(first[k])],bins=clean_without[two_determinants_columns[0]].values.categories,
                                                    labels=False, include_lowest=True)[0]).replace(']', ')')
        f_2 = two_determinants_columns[1] +'_'+ str(pd.cut([float(second[k])], bins=clean_without[two_determinants_columns[1]].values.categories,
                                                    labels=False, include_lowest=True)[0]).replace(']', ')')
        ml_row.at[0,f_1] = True
        ml_row.at[0,f_2] = True
        prediction = clf.predict(ml_row)
        row.append(prediction[0])
    matrix_ml.loc[row[0]] = row[1:]

print('The first determinant specified is in the columns, the second in rows')
return matrix_ml
```

This code is aimed at showing how, as two Determinants change, also the prediction changes. By calling the new function "prediction_matrix" it is possible to insert the two wanted variables, that in this case are: (i) **Debt to Equity** (columns) and (ii) **Earnings Variability** (rows).

```
prediction_matrix()
```

```
Insert the two determinants that have to be analyzed (d_e, growth, c_r, year etc.):  d_e|e_v
Insert the two determinants column names (Debt to Equity, Growth, Current Ratio, Year etc.):  Debt to Equity|Earnings variability
The first determinant specified is in the columns, the second in rows
```

	0.01	0.02	0.04	0.08	0.11	0.16	0.22	0.28	0.35	0.42	0.51	0.60	0.71	0.85	1.03	1.28	1.67	2.66
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
0.16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0.20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The resulting matrix confirms definitely the impact of the Determinants on Beta, recalling that 1 (one) stands for "Beta higher than 1" and 0 (zero) stands for "Beta lower than 1" the results are perfectly in line with the paper, the literature and everything said so far. As Debt to Equity or Earnings Variability increase, the number of 1 increase too suggesting that, even in this case, **if either Debt to Equity, or Earnings Variability increases then the Beta will change and go up because they are positively correlated**.

With this last example, Chapter 3 concludes leaving the last considerations in the Conclusion (also about these last, fundamental, results).

Conclusions

The whole paper concluded with practical final examples, useful to provide a real and practical idea of everything touched and discussed during this paper. Before providing a general perspective about the last topics, it is compelling a sort of catch up about everything debated, starting from the capital budgeting and arriving to the final examples and considerations.

At the beginning, it was introduced how capital budgeting works and all its applications in real cases. Capital budgeting served as a valuable key able to introduce one of the most important elements of finance and, specifically, for corporate finance, the Free Cash Flow.

The latter, if it is possible to be derived, represents the most widespread instrument to evaluate a target company. It suggests the future cash inflows that characterize a company and distinguishes, along with the cost of capital, if a firm is more or less valuable than another. The Free Cash Flow, structurally similar to capital budgeting, practically differs from the latter in considering the inflows and outflows of the company as a whole (the capital budgeting, instead, takes into consideration the stand-alone projects brought forward). The introduction of the Free Cash Flow paves the way for the **first core element** of this paper, the company valuation.

Since the company valuation can be performed in several and distinct ways, it was possible to describe all the methods actually employed, along with a case by case specification of the suitability of a certain method to the target company characteristics. The methods discussed can be summarized as the following:

- Dividend Discount Model

- Valuation Multiples Comparables Method
- Valuation Multiples Comparables Precedent Transaction Method
- Discounted Free Cash Flow
- Adjusted Present Value Method
- Flow-to-Equity Method

The aforementioned methods all share a common factor: the utilization of a discount rate. The latter can assume the form of Cost of Equity alone or WACC (Weighted Average Cost of Capital) which considers the Cost of Equity and the Cost of Debt at the same time. The main distinction is that Cost of Equity represents the opportunity cost of capital for an equity investor, while WACC represents the weighted average of the cost of equity and cost of debt (the opportunity cost of creditors and investors together while considering simultaneously the tax benefit too). Since the common denominator of all valuation methods is the Cost of Equity it became necessary and particularly compelling to dive into its details, how it is built, its characteristics and how much its changes can affect the company valuation.

Following the analysis about the impact of the Cost of Equity on the target firm price and, along with it, the importance of properly defining it in order to not incur in absurd valuations, the components necessary to derive the Cost of Equity were detailly analysed.

In this sense, it was introduced the CAPM (Capital Asset Pricing Model) which, through a simple formula, allows to calculate the Cost of Equity. For its computation the Risk Free Rate, the Market Risk Premium and, most importantly, the Beta Equity are necessary. Beta Equity constitutes the **second core element** of this paper: its computation represents, along with the forecasted Free Cash Flows, the other factor of highest uncertainty that can take the most disparate values and substantially impact on the final valuation.

Even if the years pass by and the available data increases considerably, Beta Equity continues to represent a big question mark when it comes to valuating a firm. The literature is equally uncertain when it comes to Beta Equity computation and its characteristics; in this sense, the research appears quite poor in considering and analysing Beta Equity, in particular the elements affecting it.

For this reason, this paper moved the attention on Beta Equity; specifically, it focused on the so-called "Beta Equity Determinants", the company characteristics that have an impact on Beta Equity. The decision of choosing this subject, besides its importance, derived from its poor coverage in terms of amount of research made both in the past and, notwithstanding the wide availability of new technologies, in recent years.

In fact, even if it was possible to set a first, literature based, analysis about which factors really have an impact on Beta Equity, it appeared necessary to further examine this topic. The Determinants that came up from the various research are: (i) the company indebtedness, (ii) the company growth rate, (iii) the company founding year, (iv) the company payout ratio, (v) the company size and, finally, (vi) the company earnings variability. All of these were detailly presented and explored: specifically, it was explained why and how they have an impact on Beta Equity and which accounting variables would be the most expressive and meaningful in terms of representativeness of the corresponding Determinant (e.g. the company indebtedness is measured through the Debt to Equity ratio). This paper did not find new Determinants but, still, it was able to group all the Determinants found so far from others making it possible to define a general perspective about this topic.

Even though the explanation was solid, as well as the related academic papers, it was necessary to provide a practical example. The next step, covered in the last chapter, involved the development of a Python DataFrame with almost 25,000 stocks, of which (i) ticker names were inserted as the rows names and (ii) the Determinants (represented through ad-hoc accounting variables) were inserted as the column names. This DataFrame was used to build upon it a Machine Learning algorithm able to predict the Beta Equity by leveraging the Determinants and their correlation with Beta.

Initially, The DataFrame was subject to an EDA (Explanatory Data Analysis) finalized at showing which are the "weak spots" of the stocks, especially the correlations that are in place between the Betas and the Determinants. Then, since the correlations confirmed the hypothesis, it was possible to conclude with the Machine Learning algorithm.

The latter was designed with the purpose of showing how the prediction changed by changing the inputs (i.e. the Determinants' value). Specifically, it was evident from the examples how, if it was inserted a low Debt to Equity ratio or a high Asset Size, the Beta Equity would be predicted as lower than 1.

In the end, through this final “exhibition”, it was possible to finally arrive at the conclusion of this paper and certainly affirm that all the objectives initially anticipated in the introduction were reached. It can be said that Beta Equity is not an independent factor, but it is, as the bulk of things in finance, the result of a combination of elements. In this case, even if the combination of elements that results in the Beta Equity is not (supposedly) limited to the factors discovered in this paper, still, it can be affirmed that the correlations emerged, and the data driven results make it absolutely doubtless the existence of Determinants that have an impact on Beta Equity.

The last examples permitted to observe how much the Determinants can have an impact on the Beta Equity, this result can certainly help in different ways (already examined during the paper):

- the knowledge of the Determinants and their relationship with the Beta Equity (which was demonstrated to be directly correlated to the company riskiness) can help to better understand a company and translate such Determinants values, of that specific company, into something practical like how much the company is risky;
- the possibility of intervening on those Determinants in such a way that it is possible to address the company towards the desired direction (if it is wanted to reduce the riskiness the individual can work on some of the Determinants by, for example, reducing the indebtedness or increasing the payout ratio);
- the possibility, considering the most recent advancements in the tech sector through the introduction of Artificial Intelligence, to obtain a way more efficient ML algorithm (similar to this one) which can be able to precisely derive the Beta Equity.

Bibliography

Jonathan Berk, Peter DeMarzo, "*Corporate Finance, Global edition*", Pearson Education, 2020.

Brav A., Graham J.R., Harvey C.R., Michaely R., "*Payout policy in the 21st century*", Journal of Financial Economics, 2004

Liu J., Nissim D., Thomas J., "*Equity Valuation using Multiples*", Journal of Accounting Research, 2002.

____, UBS Global Equity Research, "*Valuation Multiples: A Primer*", 2001.

Joshua Pearl, Joshua Rosenbaum, "*Investment Banking Valuation, Leveraged Buyouts, and Mergers and Acquisitions*", Wiley, 2019.

Steward C. Myers, "*Interactions of corporate financing and investment decisions – Implications for capital budgeting*", The Journal of Finance, 1974.

Stephen Ross, Randolph Westerfield, Jeffrey Jaffe, "*Corporate Finance*", McGraw-Hill Education, 2022.

Steven N. Kaplan, Richard S. Ruback, "*The valuation of cash flow forecasts: an empirical analysis*", The Journal of Finance, 1995.

____, KPMG, "*Cost of Capital Study*", 2023.

Nicolas Groenewold, Patricia Fraser, "*Forecasting Beta: How Well Does the 'Five-Year Rule of Thumb' Do?*", Journal of Business Finance & Accounting, 2000.

Lasse Heje Pedersen, "*Efficiently Inefficient how smart money investments & market prices are determined*", Princeton University Press, 2015.

Eugene F. Fama, Kenneth R. French, "*Common Risk Factors in the Returns on Stocks and Bonds"*, Journal of Financial Economics, 1993.

Robert S. Hamada, "*The Effect of the Firm's Capital Structure on the Systematic Risk of Common Stocks*", The Journal of Finance, 1972.

William Beaver, Paul Kettler, Myron Scholes, "*The Association Between Market Determined and Accounting Determined Risk Measures*", The Accounting Review, 1970.

William F. Sharpe, "*A Simplified Model for Portfolio Analysis*", Management Science, 1963.

John Lintner, *"The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets*", Review of Economics and Statistics, 1965.

Jan Mossin, "*Equilibrium in a Capital Asset Market*", Econometrica, 1966.

Intrisano, Carmelo, Giovanni Palomba, Loris Di Nallo, Anna Maria Calce, "*Accounting Beta: Which Measure Is the Best? Findings from Italian Market*", European Journal of Economics, Finance and Administrative Sciences, 2017.

Ludwig B. Chingarini, Daehwan Kim, Fabio Moneta, *"Beta and firm age*", Journal of Empirical Finance, 2020.

Richard A. Lambert, Christian Leunz, Robert E. Verrecchia, "*Accounting information, disclosure, and cost of capital*", Journal of Accounting Research, 2007.

____, "*Turkish stocks soar in world-leading rally as 'tech mania' grips market*", Financial Times, 2024.

Allen B. Downey, *"Think Python: How to think like a computer scientist*", O'Reilly Media, 2015.

Yves Hilpisch, "*Python for Finance: Mastering data-driven finance*", O'Reilly Media, 2018.

Andreas C. Mueller, Sarah Guido, "*Introduction to Machine Learning with Python: A Guide for Data Scientists*", O'Reilly Media, 2016.

Sitography

https://www.investopedia.com/

https://it.wikipedia.org/

https://www.borsaitaliana.it/

https://www.euronext.com/

https://pages.stern.nyu.edu/~adamodar/

https://it.finance.yahoo.com/

https://www.kaggle.com/