AALBORG UNIVERSITET

Master's Thesis

Determinants of the startup success: What factors influence the valuation of unicorn startups in the U.S. fintech sector? An empirical analysis.

Małgorzata Antosz (20211712)
Simone Tedeschi (20212030)

4th Semester
MSc Economics and Business Administration - Finance

Supervisor:
Frederik Steen Lundtofte

Place and date of Submission:

Aalborg, 31st May 2023

Number of characters (including spaces): 176.232

Abstract

This paper examines through the use of multiple linear regression analysis which factors have significant impacts on the valuation of unicorn startups from the U.S. fintech industry. The dataset created for the purpose of this study is composed of 129 companies and cross-sectional data including both numerical variables as well as categorical. The correlation tests proved that there is a significant positive correlation between the valuation of the company and variables such as funding amount, above-average revenue, number of funding rounds, number of investors and number of employees. The linear regression model made it possible to examine the relationship between valuation and the aforementioned data. The study allowed us to draw meaningful conclusions and the results suggest that (i) funding amount is the most important factor affecting company valuation, (ii) other variables of high significance are the number of investors and activities in sub-sectors such as cryptocurrency and blockchain technologies, and (iii) the number of employees may have a moderate impact on the valuation. We conclude that as for the numerical variables the number of investors and acquired funding amount are the most important factors affecting the post-money valuation of unicorn startup companies from the fintech sector in the U.S. In addition, due to high significance of subsectors such as cryptocurrency and blockchain technologies it was possible to conclude that investors value the companies' future potential and how involved in developing new technologies and solutions they are, more than the revenues generated by these companies.

Keywords: startup, unicorn startup, fintech, valuation, linear regression

Table of Contents

List of Figures

List of Tables

I. Introduction

In recent years, the global startup environment has seen a significant increase in the formation of very successful and influential companies known as "unicorns". These unicorns, or private startup companies valued at more than $1 billion (Lee, 2013), have disrupted traditional industries and revolutionized the way business is conducted. Among the many industries, the fintech business has undergone an exceptional growth surge, characterized by ground-breaking innovations and significant financial investments (Feyen, Natarajan, & Saal, 2023). Understanding the determinants of startup success and the factors that contribute to the exceptional valuations of unicorn startups is of great interest to entrepreneurs, investors, policymakers, and researchers. The valuation of a business is an important indicator that reflects growth potential, a market position as well as perceived value. Unicorns attract substantial attention and resources which allows them to accelerate their growth, attract talent, and challenge existing market participants. However, the drivers of these high valuations remain relatively unknown, demanding thorough research.

This thesis aims to answer the research question: “What factors influence the valuation of unicorn startups in the U.S. fintech sector?”. This empirical analysis will examine a range of potential determinants that may contribute to the valuation of unicorn startups in the U.S. fintech sector. The study will explore various factors, including but not limited to, funding amount, number of investors, number of funding rounds, type of business, subsector that the company operates in, or location. The results of this work are not only a contribution to the field but can be useful for entrepreneurs and future startup founders.

All unicorns started as startups, and funding plays a crucial role in shaping their development, from creation to maturity, with different stages marking their growth journey. While age does not define a startup, growth and scalability are crucial indicators. Startup incubators and ecosystems play a vital role in supporting and nurturing these ventures, providing resources, funding, mentoring, and networking opportunities, therefore the strategic location of the business and access to such resources could be potential indicators of its success (Diggival, 2022). The startup journey consists of stages like idea generation, analysis, funding rounds (such as pre-seed, seed, series A, B, C, D, and IPO), growth, expansion, and potential exit (Lassala & Ribeiro-Navarette, 2022). Within those, funding is crucial at each stage, with various sources like angel investors, crowdfunding, venture capitalists, and private equity firms. Consequently, as the startup grows, both in terms of profitability and employee, it is able to attract more investors. Furthermore, the number of employees may have some correlation with company valuation, but it is not a strong determinant (Krch, 2018).

As of January 2023, there are 1205 unicorns globally. In 2022 (CB Insights, 2022), North America held the highest share of unicorn valuations at 46% (Statista, 2022). The fintech sector accounted for the highest number of unicorns and had record-high funding of $106.1 billion in 2021 (Eckert, 2022). However, the number of unicorns experienced a slowdown in growth from the second quarter of 2021 (CB Insights, 2022), particularly in the U.S. and Asia. The decrease in unicorn numbers can be attributed to macroeconomic factors such as inflation, increasing interest rates, and geopolitical instability, which are impacting public markets. Consequently, the funding environment has become more challenging, making it difficult for startups to achieve valuations over $1 billion (CB Insights, 2022). Thus an interesting factor to analyze is whether the year of the foundation has an impact on unicorn valuation. Some experts such as Paul Graham (2012) who invested in over 1000 startup companies, emphasize growth rather than age and size when it comes to defining what is a startup. Moreover, the research found that the relationship between funding size and firm's valuation has a U-shaped (convex) pattern, indicating a positive effect initially, and then a negative one after reaching a peak (Nazir & Tbaishat, 2023).

This thesis sheds light on the potential failure factors that may jeopardize the success and development of a young company. Any startup failures can be attributed to factors that also affect well-established unicorns (Pride, 2018). For example lack of funding, founders/investors disharmony and even overfunding. Furthermore, it has been found that about 50% of unicorns are overvalued (Gornall & Strebulaev, 2017). This might be due to the fact that startup valuations are often based on growth potential and expected development rather than financial success or fundamental data. The success and valuation of unicorns are also highly influenced by the environment in which they operate. The U.S. is home to key startup hubs like Silicon Valley, Boston, and Austin, known for their favourable startup ecosystem providing high availability of funding, a skilled workforce, and growth-oriented business culture (Stephens, 2019).

Methods for examining the relationship between the valuation of the fintech unicorn in the U.S. and tested variables include analysis of the Multiple Linear Regression using the OLS method and Correlation Tests. The model created for the study included valuation as a dependent variable and various internal factors, both numerical and categorical (expressed with dummy variables) as independent variables. This model was subject to transformation, diagnostics, hypothesis testing and model selection in order to obtain the most significant variables affecting the valuation of the company. The empirical results of this study are based on a sample of 129 U.S. unicorns from the U.S. fintech sector featured in the World Unicorn Club dataset of Unicorns by CB Insights (2023). Through the empirical analysis, we aim to uncover the key drivers behind these exceptional valuations and their implications for the entrepreneurial ecosystem. By gaining a deeper understanding of these factors, we can foster an environment conducive to innovation, growth, and sustainable economic development.

Our study makes a contribution to the literature on the factors influencing the valuation of fintech unicorn startups in the United States as we name the most significant variables influencing the valuation of the companies in the sample.

The paper is structured as follows. In section II we discuss the literature and background of startup development, funding, ecosystem, trends, threats and problems regarding startup valuation. Section III focuses on the data collection process, data characteristics and structure. The methodology of the study is outlined in Section IV. Section V presents our empirical results and regression output. In Section VI we discuss the limitations and further research on the topic. We discuss our findings and conclude in Section VII.

II. Theory background and literature review

i. Startup company: definitions and startup ecosystem

It is crucial to first have a thorough understanding of what a startup is in order to adequately discuss the concept of unicorns. Eric Ries the author of the book The Lean Startup: *How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses* (2011) describes a startup as a human institution designed to create a new product or service under conditions of extreme uncertainty. The concept of a startup became very popular during the 1990s when the number of technology companies promptly boosted due to technological development. However, the word start-up in the meaning we know today has been first introduced in Forbes magazine in 1976 and quickly became a buzzword in the discussion.

Predominantly startups are young companies that are just getting started in the early stages of their development. They normally have one to three founders that focus on designing a practical and feasible platform, service, or product that aims to introduce something entirely new to the market. The goal is to make the product or a service irreplaceable for the users. Blank and Dorf (2012) recognize a startup as a temporary organization in search of a scalable, repeatable, profitable business model. It must be remembered that startups are not just small versions of big companies, and they cannot use traditional tools appropriate for known businesses as startups are all about the unknowns, therefore entrepreneurs must find a new way to manage such a business (Blank & Dorf, 2012).

Generally, startups have a common goal which is going public, allowing early investors to support fresh, risky businesses and gain adequate rewards (Baldridge & Curry, 2022). In order to create a scalable business model ready for quick expansion, entrepreneurs must have access to relatively substantial external ventures to fund their operations. Startups often start with operational costs and expenditures that exceed their monthly revenue and budget; therefore they seek funding opportunities. Types of funding opportunities will be discussed in the following subchapters of the thesis. Although startups make up a small percentage of entrepreneurs in certain areas such as Silicon Valley or New York, they manage to attract potential investors by offering significant returns for the greater risk taken (Blank & Dorf, 2012).

Startups tend to be understood as young, recently established businesses, though that is not necessarily true. Paul Graham (2012) who invested in over 1000 startup companies, emphasizes growth rather than age and size when it comes to defining what is a startup. Thousands of young businesses may be appearing on the market, but if their strategy is not focused on growth and scalability, they simply do not have what it takes to be a successful startup on the market. But how do we define a young company?

For one company a year could be enough to develop a product and offer it on the market, for the other ten years may still not be enough to finish the testing and development stage. Therefore, the age of the company does not define its stage or growth.

To help entrepreneurs enter the market successfully and to provide assistance at the early stages, some more experienced entrepreneurs or academics created non-profit organizations called startup incubators. Incubators are often linked to universities and business schools - others are formed by governments or successful businesspersons. They help businesses to grow by providing workspaces, seed funding, mentoring and training, creating networking opportunities, accounting services, and helping with acquiring funds through bank loans, grants, venture capital or angel investors (Willson, 2022).

Startup incubators are part of a startup ecosystem that aims to create new startup companies. It is a network of businesses, entrepreneurs, universities, funding organisations, investors and mentors. It is a complex and interdependent system where one entity cannot function without the other systems as they rely on each other and are mutually beneficial (Diggival, 2022). Members of such an ecosystem are motivated to bring innovative solutions that will improve the living of the local communities, create jobs and make use of the available resources. Figure 1 shows the Startup Ecosystem model created by Taleghani and Azizi (2022).

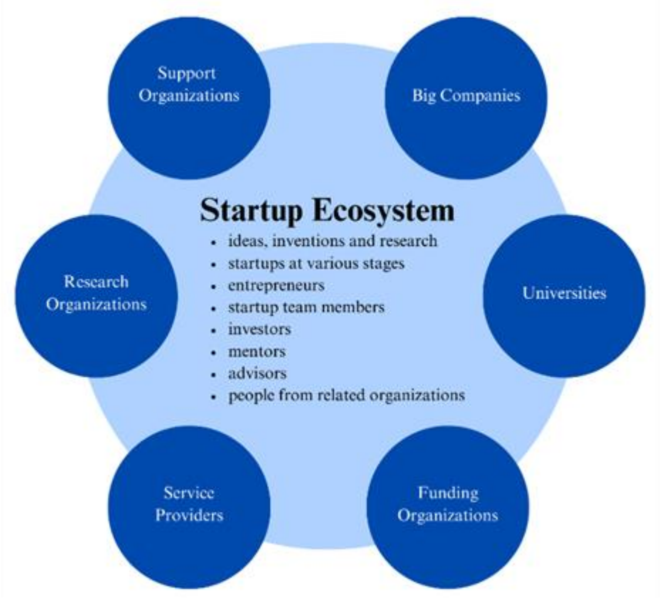

Figure 1 Startup Ecosystem model

Source: own illustration based on Teleghani, A., Teleghani, M, & Azizi, J. (2022). Startup Ecosystem.

They found out that a strong startup ecosystem can attract and retain more foreign investors. Such ecosystems are affected by external factors, therefore different startup ecosystems will work differently across different geographies due to cultural differences, different resources and knowledge pools (Startup Commons, 2019). Internal factors influencing the ecosystem include people who contribute to it, their skills and resources. Therefore universities play an important role in the ecosystem as they provide opportunities such as entrepreneurship courses, startup organizations, talents, knowledge, research and mentorship (Diggival, 2022).

ii. Startup funding and development stages

Founding a startup starts with an idea, which must be scalable. Scalable ideas can be expanded depending on the demand or other factors impacting the situation in the market. With this phase, no money is needed normally, what is necessary is plenty of time and dedication. Along with the idea, a team must be found, not a too large group of people that can share the vision of the common project and be committed to developing the product. According to Merzlova (2022), the size of the starting team is up to 4 people.

The overall journey of a startup to achieving the status of unicorn is composed of seven stages, which are idea, analysis & pre-seed funding, seed stage, market entry & series A funding, growth stage, expansion, and ending with either exit or obtaining the status of a unicorn (Merzlova, 2022). Startup companies often start their operations with high costs that exceed their initial budget or available funds, therefore they seek funds that will allow the development and growth of the venture. Funds are required at every stage of development to allow expansion, hire new team members or for product development. Funding is especially important at the early stages of development as a company cannot afford to fund its activities by the revenue it generates. Different stages require different financing strategies and funding sizes. Andrew Metrick and Ayako Yasuda (2011) characterise four stages of growth of companies in their book *Venture Capital & the Finance of Innovation:* seed stage, early stage, expansion and late stage.

These growth stages include funding stages that enable growth. There are seven funding stages (Pree-Seed, Seed, Series A, Series B, Series C, Series D, IPO) during the growth stages that include different types of investors along the way (Lassala & Ribeiro-Navarette, 2022):

1) ***Pre-Seed Funding*** - This is a stage of market research and analysis, competitor analysis, product-market fit and development of the MVP (minimum viable product). A stage where the expenses are the lowest but so are the possibilities to raise funds. It is when the first financial plan outline is created including fixed and variable costs as well as revenue sources. Is important to note that at this development phase there is no formal organization, just a research team evaluating the

feasibility of the initial concept and establishing clear, quantifiable objectives. This stage is also known as bootstrapping where the startup owners invest their resources and gather funds from friends and family. All funding is acquired in exchange for partial ownership of the company, therefore the owner's percentage of ownership decreases; this process is called a share dilution. At this stage, entrepreneurs usually manage to gather 10-50 thousand USD. The company in this phase can be worth from $10K up to $100K.

2) ***Seed Funding*** **-** the idea has been transformed into a working business at this stage. This is a part of the seed development phase where companies enter the proper funding stage. This is where the team needs to convince a potential external investor to invest in the idea. The team must therefore show the potential investors the viability of their product/service, or rather that there is market demand. The minimum viable product has been launched getting the company its first 100 active users. The product is being developed to fit into the current market with the help of new employees and systems. The approximate amount of funding needed is three million USD. That amount of money is acquired through crowdfunding, startup accelerators & incubators, and from business professionals called angel investors. At this stage company's value reaches approximately 3-6M of USD.

Equity crowdfunding is the process of financing a project by multiple investors through investment platforms (Lassala & Ribeiro-Navarette, 2022). Individuals are not required to invest a large amount of money as the whole idea of the platform is their joint investment. As a result, less experienced investors might try out several types of investments and end up owning some of the company's share capital. Startups with great potential are able to raise funds from Venture Capitals at this stage.

Business Angel (BA) investors are private individuals who invest their own capital and have knowledge ad experience in startup investment. They are willing to fund smaller operations than VCs and are willing to be more flexible (Cremades, 2019). Angel investors provide funds and mentorship in exchange for the company's equity, therefore shares are being further diluted.

The first round of financing where the VC are present is called *Series A* phase. It is the beginning of the scalability of the business model through maximizing profits and expanding business by entering new markets. Depending on the industry during the Series A funding round a company is capable to raise from $2 million to $15 million. Well-established venture capital firms that take part in Series A funding involve Sequoia Capital, 500 Startups, Google Ventures, and Intel Capital (Fundz, 2022). At Series A funding, companies can be valued at $10 million up to $30 million.

If the startups reached this stage it means that has successfully entered the market. Here the focus is on maintaining steady growth, analyzing customer and user personas, enhancing the product, optimizing marketing strategies, establishing sales processes, establishing company culture, and recruiting new team members.

3) ***Growth Phase/ Expansion*** - one of the most important stages of a startup's life cycle. At this point, the startup is no longer a high-risk venture, therefore can attract a wider group of investors such as investment banks, hedge funds and private enquiry firms (series C, D fundings etc). Here the size exceeds two hundred people but stays at less than one thousand (Merzlova, 2022). The business often operates at the break-even point, has a competitive cost structure, and is profitable (Lassala & Ribeiro-Navarette, 2022). This is the stage where Venture Capital funds start investing vastly in startups.

Venture Capital funds (VCs) spend money gathered by professional individual investors in a fund. They expect high returns for the money they invest when the business is ready to scale. They do not contribute to the company in the same way that a business angel would, but they often want a position on the board of directors to act in a supervisory role (Lassala & Ribeiro-Navarette, 2022). Once the Series A funding stage is completed, a company can enter Series B funding. According to Tomasz Tunguz, Polish globally-known Venture Capitalist at Redpoint, Series B funding is the most challenging round for a company. During Series B, startups can raise around $30 million to $50 million. According to Lassala and Ribeiro-Navarette (2022), the dilution of existing investors in each of the financing rounds oscillates between 10 and 20% depending on the stage. Due to the lack of data on the size of the current financing round, the data collection and analysis part of this thesis will solely concentrate on the overall total funding raised by the company since its first funding round.

Series B funding can be followed by Series C Funding. The general motivation for a Series C Funding Round occurs to make the startup appealing for acquisition or to support an upcoming public offering. The funding size increases as investors are more willing to fund successful startups therefore total funding can reach an average of $55 million as of 2022, while for the Series C funding, it is $73 million (Fundz, 2022). Potential investors interested in Series C include late-stage VCs, Private Equity Firms, Hedge Funds and Banks. A startup at this stage is on average valued at $100 million - $120 million. Figure 2 presents the development of financing rounds over time.

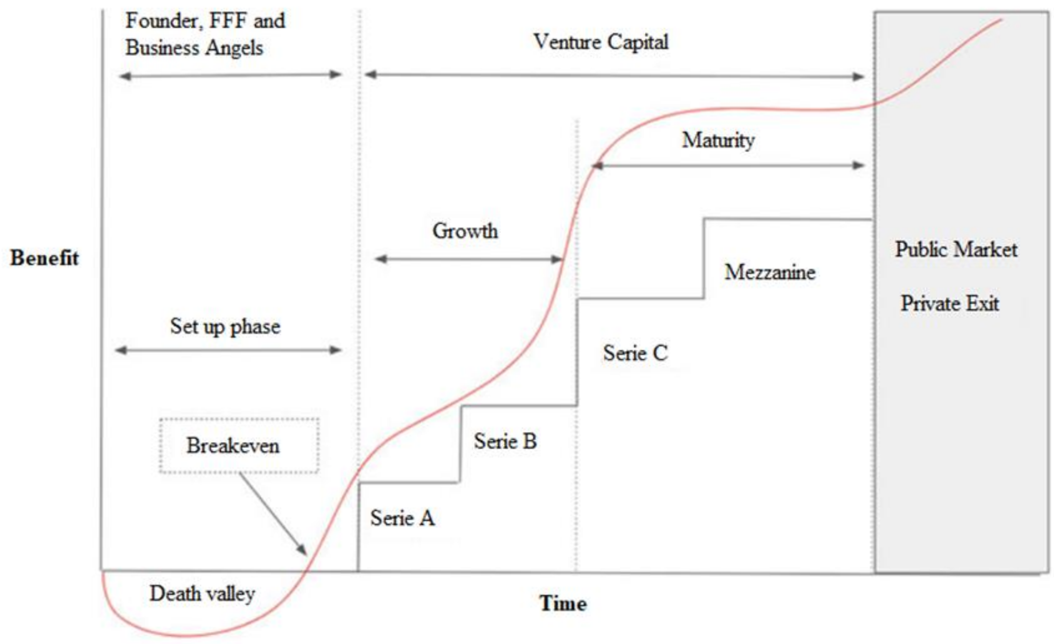

Figure 2 Startup Financing rounds

Source: Lassala & Ribeiro-Navarrete. (2022)

4) Late Stage/ Exit: Reaching the final point the startup has three possibilities, going through an exit either by being bought by a larger company for financial gains, resulting in the end of its independence, or going public through an IPO. Alternatively, the startup can continue operating as a unicorn. At this stage, the people forming the company are more the one thousand (Merzlova, 2022). IPO provides access to an additional source of financing, increases the company's creditability and in some countries may even lead to tax reduction.

However, many startup owners are not ready for the IPO or are not considering going public at all and want to keep their companies private. What is more, companies in the U.S. are staying private for longer and go public only after 8 years on average (Erdogan et al., 2016). They can enter late-stage funding series that can continue into Series D Funding. Series E Funding, Series F and G funding, private equity funding rounds or mergers and acquisitions. Unicorn Startup Companies tend to wait longer with IPO offerings as staying public brings them several benefits (Farandea, 2021). Without going public, the company does not have to be fully transparent, which can bring a competitive advantage in the innovative sector as competitors cannot acquire certain information.

A curious aspect to analyse is whether the number of employees which is expected to increase as the startup enters a new stage is actually a key determinant of the valuation of a company. Therefore, companies with a higher number of employees have a higher valuation than the ones with a lower headcount and consequently helping them to more easily access the status of unicorns.

A study from Přemysl Krch 2018, *Relationship between the company size and the value: empirical evidence* investigated with the use of linear regression and correlations the relation between the levels of valuation multiples and eight criteria of company size, including the number of employees from German companies. The paper found that there is a certain dependence on the price-to-earnings valuation multiple on most of the tested size criteria, as well as the number of employees. However, the dependence was indicated to be very low and significant only for the price-to-earnings ratio. The other tested valuation multiples, including enterprise value to sales, did not show a statistically significant level of correlation or dependence. Therefore, while there may be some positive correlation between the number of employees and company valuation, the relationship is not strong and may not be a key determinant of company valuation, and suggested further research on that (Krch, 2018).

iii. Unicorns: definition and origin

Let's start by defining what the term "Unicorn" means, a unicorn is simply a privately held company that reached the value of $1 Billion (Lee, 2013). The credit for associating this concept of business with a mythological creature goes to Aileen Lee (Rodriguez, 2015), who first came up with this name in 2013, to emphasise the rareness of the success of these types of companies.
There are other two terms used to describe even more valuable and rare companies that are not publicly traded on stock markets, the first one is "Decacorn" which is a private company with a valuation between $10 billion and $100 billion and the second one is "Hectocorn" where the valuation goes over $100 billion. In March 2020, the Chinese firm ByteDance became the first private company to reach a valuation of $140B and was the only Hectocorn until October 2021, when SpaceX and Shein joined the ranks. In total, 4.2% of unicorns, or rather, 51 private companies, fall within the Decacorn category, with valuations between $10 Billion and $100 Billion. Moreover, 22.5% of companies in the global unicorn club are valued at exactly $1 Billion. (CB Insights, 2023)

In 2013 when the unicorn club was created, Aileen Lee counted 39 of them (Lee, 2013). As of January 2023, the number of firms that reached the status of unicorns is 1205, which is 242 more new entries in comparison to January 2022 (CB Insights, 2022). This led many to reflect on whether that is not something so uncommon anymore. Nevertheless, although the number increased, the probability of a business becoming one is still extremely low, or rather 0,00006%, with an average of 7 years' time frame for a startup to reach unicorn status. (Embroker, 2023).

Geographic, sector distribution and trends

As shown in Figure 3, in the first half of 2022, North America (46%) is the region that owns the highest share of unicorns' valuation, right after there are Asia (43%) and Europe (12%).

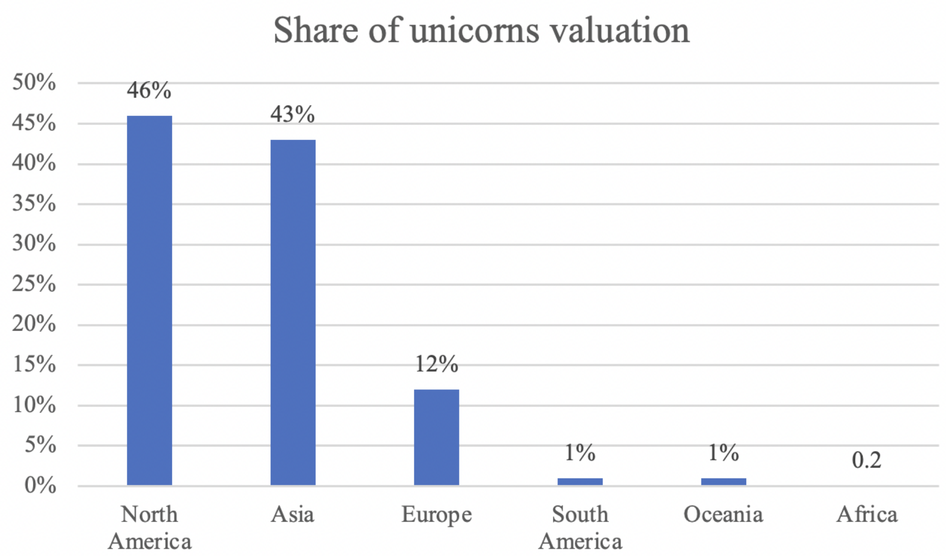

Figure 3 Distribution of unicorns' valuation by region 2022

Source: own creation, Data Statista (2022)

According to the data from CB Insights *List of unicorn startups & markets: CB insights from 2023*, the number of unicorns and the countries, the U.S. has always kept the record for the highest percentage of unicorns, in October 2022 it was 53.9%, followed by the second largest economy, China (14.3%), then India (5.7%). Lastly, the fourth place goes to the UK with 4,2%. In total 47 countries and regions are represented in the total number of unicorns globally (CB Insights, 2023).

A report delivered by Crunchbase suggests that between 2005 and 2010 only 14 firms achieved the status of unicorns (Teare, 2020). Things started changing from 2013, when until 2020 the number of unicorns had quite a constant growth, and then in 2021, the number skyrocketed. For instance, between 2016, and June 30, 2021, 869 businesses reached the 1$ Billion valuations (Eckert, 2022). One of the main factors that led to this incredible growth is technology, the development of smartphone and their apps, for instance, personal finance transactions that used to be costly, complicated or impossible can now be done with just a few taps on a mobile phone. One of the areas where most of the unicorns are concentrated indeed, the fintech sector, is also the sector with the highest amount of investment. In fact, Fintech, has record-high funds of $106.1 Billion in 2021 (Eckert, 2022) and is the first sector for the number of unicorns, accounting for 21.05% of the overall share, as shown in figure 4. Just two percentage points below there are internet software & services, and right after e-commerce and health.

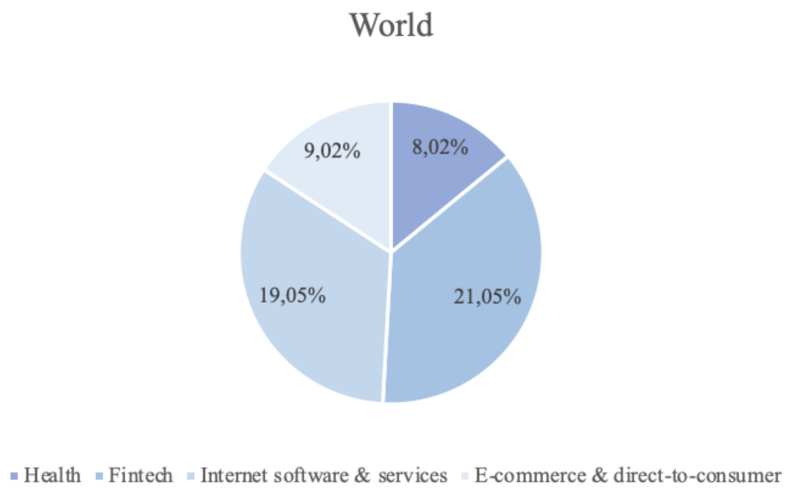

Figure 4 Top four unicorn industries

Source: own creation, Data: CB Insights (2022)

Figure 5 shows the world number of unicorns found each year. It is important to note that the data used is from 31st December 2022, therefore is from the present pool of unicorns, thus firms who lost their unicorn status (for example through an exit) are not considered. The same trend is followed by U.S. and Europe, with a progressive growth from 2013, with a soar in 2021, where just in that year the number of new unicorns was more than five hundred, more than all the previous years combined. All this until the end of 2022 when the number plummeted.

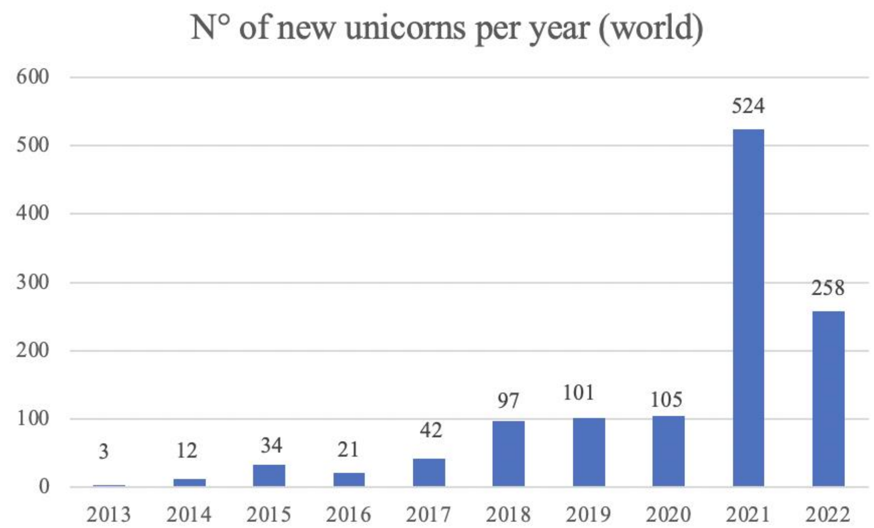

Figure 5 Number of new unicorns per year

Source: own creation, Data: CB Insights (2022)

In order to understand what happened it is important to look at when the trend started within the year, therefore, in which quarters the drop occurred.

iv. Unicorns' slowdown

Looking at the U.S. example, Figure 6 shows when the U.S. startups joined the unicorn club, within each quarter of the year from 2020 to the end of 2022. From the second quarter of 2021, the massive growth slowed down until the first quarter of 2022 when the number of new unicorns fell sharply. The same trend is present also in the world number of unicorns in Figure 6.

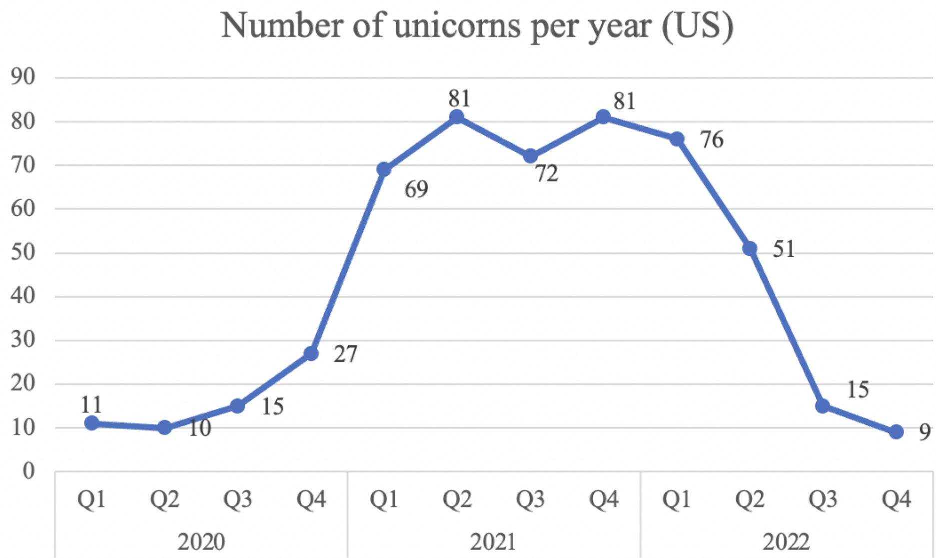

Figure 6 Number of unicorns per year 2020-2022

Source: own creation, Data: CB Insights (2022)

The unicorn's world downtrend is still affecting mainly the U.S. and Asia, on the other hand, in Europe, after a steep fall at the beginning of 2022, the number began to rise again starting from the third quarter of the year (CB Insights, 2022). According to the CB Insights report, the reason for Europe's resilience can be attributed to several reasons: first, the interest rates in the U.S. were raised earlier than those in Europe, leading to a quicker decline in the U.S. venture market. Secondly, non-VC investors like Tiger Global and SoftBank, who played a significant role in the large late-stage funding rounds in 2021, have traditionally focused more on startups in the U.S. and Asia than in Europe. Fintech experienced the biggest drop in new unicorn formation among all sectors analysed, declining by 44% quarter-over-quarter (QoQ) and 58% year-over-year (YoY). Fluctuations in the public markets caused by a tumultuous macro environment have caused a decrease in the worth of the most valuable private companies, leading to a decrease in valuations and causing investors to shy away from large investments in later stages (CB Insights, 2022). Generally, in the evaluation of a firm are considered the forecasted cash flows, which are then discounted to get the value of the company, these are fundamental factors that participate in the evaluation of a company. Therefore a considering part of the evaluation is based on the future predicted earnings. These forecasted earnings are lower in periods when the economy is facing challenges since consumer demand can easily shrink and inflation makes everything more expensive, thus the overall value of the company decreases. Consequently, the base reason for this rapid drop in the number of unicorns is due to macroeconomic factors such as inflation, increasing interest rates, and geopolitical instability affecting the public markets.

To understand how the world got to this unstable macro situation, we have to consider what happened in the last two years and understand the causes of the current situation. Firstly, in 2020, the COVID-19 pandemic resulted in a global economic slowdown and caused widespread panic among governments. In an effort to mitigate the impact of the pandemic, many governments significantly increased their public spending. Additionally, the rapid and unexpected easing of COVID-19, brought about by the effectiveness of the vaccines, has led to a surge in demand for goods and services, almost overcoming pre-pandemic levels. The combination of these factors has caused inflation to rise in 2021. Along with this, on 24 February 2022, the Russian Federation invaded Ukraine, leading to widespread destruction and loss of life, also causing a massive energy crisis. Consequently, inflation increased even more, and central banks raised interest rates which along with investors' pessimism caused venture funding to startups to decrease. The reason for this is that during a bear market, investors may shift their focus to more stable and secure investments, such as bonds and cash, rather than riskier assets like stocks or ventures.

This last paragraph about the roller coaster of unicorn births leads us to think about the importance of the year foundation of unicorns. Having the overall macroeconomic and geopolitical factors in favour when a startup is founded can be a key starting point for the business and its overall future outlook. Consequently, the year of foundation will be present in the regression analysis carried out afterwards in order to assess its significance and relationship nature with unicorn valuations.

The current funding environment has become more challenging, with investors being less aggressive in securing investment opportunities. As a result, there will be fewer high-value deals for startups, making it difficult for companies to reach a valuation of over $1 billion (CB Insights, 2022). According to Douglas Cumming and Na Dai's studies, there is a positive correlation between companies' valuation and funding size, with a U-shape (convex) relationship. Therefore, as fund size increases, firm valuations first increase, but then begin to decline after reaching a peak.

Nevertheless, the paper uses data from over 9,000 financing rounds with valuation data between 1991 and 2006, and just from VC firms (Cumming & Dai, 2011). Another study carried out on startups in the United Arab Emirates, finds also that the amount of capital raised, and post-money valuation have a U-shaped relationship, indicating a positive effect on market valuation, moreover, raising too much capital can have a negative impact (Nazir & Tbaishat, 2023). Moreover, according to the paper, an amount of at least USD 357.5 million is enough to generate a significant positive effect on the post-money valuation of the firm (Nazir & Tbaishat, 2023).

Consequently, it might be interesting to see if the same outcome can be seen also from a specific sector like Fintech and from a different country like the US.

The average time it takes to become a unicorn

The average amount of time it took for privately owned startups to reach the 1$ billion-dollar valuation, in the U.S. is seven years and three months (Ahmad, 2022). For what regards the European countries who are doing best, Finland owns the podium, with just three years and seven months, and Estonia is not far behind, with four years and nine months (Ahmad, 2022). However, is important to consider that these countries have only a few unicorns. Having way more unicorns than the previously mentioned Nordic and Baltic countries, it holds an average of six years, beating the U.S. with more than one year. Lastly, the United Kingdom had taken eight years and two months on average for unicorns to reach the 1$ billion-dollar benchmark (Ahmad, 2022).
As for the average time it takes to become a unicorn per sector, AI is the second fastest industry for unicorn startups and is a rapidly growing technology with multiple applications in various fields. Hardware and e-commerce come right after, taking a bit less than five years. Fintech is in the octave position, with seven years and nine months (Ahmad, 2022).

The trend of becoming a unicorn startup has exploded in recent years, as more and more startup companies aim to achieve billion-dollar valuations in record time. This fashion is often linked to the growth strategy of blitzscaling, which prioritizes rapid expansion over profitability to capture market share and establish dominance (Kuratko et al., 2020). Blitzscaling strategy is often associated with Silicon Valley which is home to the biggest hi-tech companies in the world. While blitzscaling has been successful for many companies, it can also be a high-risk approach that requires careful planning and execution.

Before such a strong development of new technologies, companies reached USD 1 million valuations after 10 to 30 years. For example, it took around 15 years for Target Corporation to reach a $1 billion valuation since the IPO while it took Nike 24 years to achieve that milestone. In the case of companies using new technologies, this process has been significantly shortened. Google took 8 years to reach a $1 billion valuation, Facebook took 6 years and Uber achieved it in just 3 years. Blitzscalling allows modern companies to monopolize the market and achieve a $1 billion valuation even within a year. Blitzscaling has numerous downsides, one of them including a high level of uncertainty described by Donald Kuratko (2020) who stated that in the face of uncertainty, blitzscaling deliberately favours speed over efficiency.

Factors and characteristics that facilitate the growth of unicorns

Before diving into the common features that characterise the success of unicorns, it is relevant to mention what are some factors that caused the growth of those 1$ billion-dollar startups during the last decade. From 2013 to 2021 the world experienced massive investor spending from venture capitalists and investment firms looking for high returns in a low-interest rate environment. Moreover, The availability of large amounts of cash from big tech companies gives investors hope for potential buyouts and acquisitions. For example, WhatsApp was acquired by Facebook for $19 billion, and Nest was acquired by Google for $3 billion. Especially before 2022, many investors saw risky startups as their only opportunity for appetizing returns which brought a high level of investor spending of Venture capitalists and investment firms, due to low percentage of interest rates and low yields on the S&P 500 (Eckert, 2022).

All these facilitated the rise of unicorn valuations. Furthermore, private capital had a huge impact on the timing and timing and strategy of IPOs of those late-stage venture-backed companies that desire to go public. Unicorns are raising significant sums of private capital before going public, just in the first six months of 2021, there were 404 rounds that gathered $134 billion in pre-IPO financing. Those rounds are also called “mega-rounds”, where $100 million or more is raised (Howe, 2015).

Customer-focused

It is interesting to note that 62% of the unicorns are B2C companies. Furthermore, unicorns adopt a customer-centric business strategy, where the customer is considered in all the processes, from the planning, manufacturing and delivering phase of the service or product (Khushali, 2022). Having a user-friendly app and good customer service is a fundamental factor of success since people are more and more demanding about having products easy and pleasant to use, factors that can bring them to switch brands.

Growth-driven and efficient

A vision for growth is something that unicorns have in common, implemented with an actual plan and strategy to scale the company. This is the reason they are also called exponential organizations, since the strategies they employ, allow them to scale at exponential rates, which means adding revenue at a much greater rate than costs (Ron Carucci, 2021). On the other hand, small businesses have in mind a more gradual and consistent growth, while startups have a more aggressive and fast intent (Embroker, 2023).

Unicorns have a global mindset but before expanding geographically they do product expansion.

Another important characteristic is that unicorns are efficient in what they do. The use of Minimum Variable Products, and MVP is a version of a product that has enough features to be accepted and used by customers, the main goal is then to use the feedback provided by customers to make improvements (Becker 2014). Since they only include essential features, they are cost-effective and allow startups to maximise efficiency by iteratively testing and refining the MVP based on user feedback and without wasting resources on completed products. Uber is a successful example of an MVP, starting with a simple interface and three cars, gathering user feedback after each ride to improve, and eventually becoming a well-known company (Embroker, 2023).

v. Threats and failure factors

Many startups' failures are linked to unicorns' success factors. When some of those are not achieved, the business probability of crumbling rises. It is said that threats that affect unicorns can be internal or external. The internal ones are closely related to problems within the business and management, for instance, an unprofitable business model. The external factors are mainly caused by unfavourable changes within the macroeconomic environment, these can cause the business to slow down if the company can not manage to adapt and overcome these challenges. Therefore, even though internal and external are different causes of why unicorns fail, they are indeed closely connected.

Jamie Pride in *Unicorn Tears: Why startups fail and how to avoid it* (2018) highlights that it is uncommon for a startup to be outperformed by its competitors, but most failures are caused by a lack of planning, inadequate team or poor execution. According to Pride, J. there are ten main reasons why startups fail:

1. founder(s) lack capacity
2. founder(s) lack capability
3. founder disharmony
4. ran out of cash
5. too much funding
6. investor–founder disharmony
7. solving an irrelevant problem (desirability)
8. ineffective business model (viability)
9. poor execution (feasibility)
10. external threats/competition (adaptability).

Founder capacity refers to the overall readiness of a founder to handle the demands of leading a startup. Startups can also fail due to founder disharmony, or rather to conflicts among co-founders. Thus, choosing the right co-founder and learning to work well together is a critical aspect of success. Choosing the right people, such as board members and other key stakeholders, is another important factor in the success of a startup. Having a supportive and effective network can help mitigate challenges and bring new perspectives and solutions to the table. (Pride, 2018)

Simple as it is, every startup that fails, will eventually run out of cash. This can happen regardless of the underlying cause of the failure, but rather the lack of confidence from investors due to an unsuccessful business model. When a business exhausts, its cash reserves and there is no more time to raise more capital, it can be a critical moment. Interestingly, a startup can also fail due to overfunding, when too many funds cause them to lose focus and drive. Moreover, according to J. Pride's experience in VC, in overfunding cases, startups were shifting their focus to things like office renovation and business card design, instead of continuously testing and validating their customer hypotheses and value propositions (Pride, 2018).

Another cause is founder and investor disagreement, which can harm the ability of startups to succeed. Investor–founder disharmony happened when there is a conflict between investors and founders. It is crucial to choose the right investors and align their vision with the future direction of the business to avoid such conflicts (Pride, 2018). This led to the question of whether having more investors could increase the probability of having more disharmony.

One of the major causes of failures is a flawed business model. This is because the business model is the core of the startup operation, in shapes the way a company creates, delivers, and captures value. As well as all the fundamental aspects of a company's strategy to compete and succeed in the market. Consequently, how the company generates revenue and makes a profit. One of the mistakes many founders do is to give excessive importance to their ideas and neglect the importance of a solid business model. Business model failures can be grouped into four main categories:

- Lack of Desirability: this occurs when there is no market for the product and when the problem being solved is not understood well enough or the solution is not seen as relevant by the target customer (Pride, 2018). Timing has a key role (Berger-de León, et al, 2022) in launching a product or service at the right time, neither too early, nor too late when there is demand for it and favourable market conditions. In fact, according to Bill Gross's findings on 100 Idealab and 100 non-Idealab companies, timing is what can make a difference from an average firm to a billion-dollar company. Of five factors, the timing was the one that mattered most, accounting for 42% of whether the company succeeded or failed, team/execution came in second place, with 32%. Surprisingly, the idea and business model comes after, with 28% and 24%, and lastly funding with 14% (Gross, 2015).

- Lack of Viability**:** a sustainable business model requires managing revenue, growth rates, and costs to reach break-even and profitability (Pride, 2018). Successful startups demonstrate market traction through revenue growth and a clear path to profitability.

- Lack of Feasibility: a startup may struggle in execution due to poor hiring decisions, misfocused efforts like focusing on unprofitable activities, or slow progress (Pride, 2018).

- Lack of Adaptability: even though startups are disruptive and often bring change and innovation, however, they should not overlook competition. Moreover, external threats such as macroeconomic factors, and government regulations should also be monitored and considered. Having a resilient business model that easily adapts to changes can make a difference in the survival of the firm (Pride, 2018).

vi. Unicorn valuation and overvaluation problem

Highly valued companies tend to be viewed as more successful by investors, competitors and customers therefore more and more startup founders fight to get Unicorn status. Considering the huge worths of these companies, the question to be asked is what criteria are taken into account while valuing such a company? Is it possible that some of them are overvalued? This can affect the outcome of analysis conducted on unicorns' valuation.

Published in 2017 study from the National Bureau of Economic Research concludes that on average 50% of unicorn startups are overvalued (Gornall & Strebulaev, 2017). During the study, researchers from the University of British Columbia and Standford examined the sample of 135 startups valued at $1 billion or more. The study found out that 65 should be in fact valued at less than $1 billion. Because it is dependent on assessments by venture capitalists and investors who have taken part in certain phases of the funding, the value of these start-ups is debatable. The objectivity and dependability of these evaluations are thus called into doubt. According to Lee (2019), the overall valuation may have nothing to do with the startup's financial success or any other such information. What is more, some of these highly valued companies have not yet generated a profit (Lee,2019).

There are several reasons why overvaluation can be problematic. First, it can lead to a bubble in the startup market, with investors spending money on companies that may not be able to deliver on their guarantees. This can create a situation where investors are left with overvalued assets that are difficult to sell or that lose value rapidly. Secondly, overvaluation can put pressure on unicorn startup founders to meet unrealistic growth targets or to go public before they are ready. This can lead to a focus on short-term gains at the expense of long-term sustainability and can result in negative consequences for both the company and its investors.

Privately owned startups' value is obtained from valuations developed by venture capitalists and investors who took part in the financing rounds the company held. Startups' value is also based on their growth potential and expected development; therefore the generated profit or other fundamental data do not play such a big role in the valuation (CFI, 2022). What makes the startup valuation tricky is that many of those companies do something that has never been done before and bases it on a completely new business model.

Startup valuation can be determined pre-money and post-money. The worth of a firm before any external investment and before the most recent round of funding is known as the pre-money valuation. Post-money valuation represents the value of the company after receiving outside funding from the investors. To determine the post-money valuation the amount of invested dollars is divided by the per cent that the investor receives, whereas the pre-money valuation is obtained by subtracting the amount of investment from the post-money valuation. It must be remembered that the post-money value of the company does not refer to the share price but to the total value of the firm's equity. It is important to assume that all shares of a firm are worth the same when calculating its value. Nonetheless, startups typically issue eight distinct kinds of shares, with each round having a unique share price. The incorrect method of computation would thus be to blame for the overvaluation as the post-money value is often calculated by multiplying the price per share from the most recent financing round by the number of diluted common shares.

A startup's valuation process differs significantly from that of an established business in a more secure situation. For instance, start-ups often generate new classes of shares every one to two years when there is a new round of funding, but public companies typically only have one common class of shares. Depending on when they were issued, the shares grant different rights. Investors want ever greater guarantees in return for their money in the following rounds. Due to the existence of the preferred shares, which affect the vast majority of unicorns, it is difficult to apply the traditional valuation approach. Thus, using the conventional approach would result in a skewed estimation (Gornall & Strebulaev, 2017).

There are numerous examples of overvalued startup companies, including some of the most recognized ones. According to the before mentioned study by the National Bureau of Economic, Research Airbnb was publicly valued at $30 billion while the study's valuation proved that it should be valued at $26.1 billion resolution in a 15% overvaluation. Space X, the American company manufacturing space rockets, was valued at $10.5 billion in 2017, while its actual value was $6.4 billion representing an overvaluation of 65%. This may be due to the excitement and high hopes of the investors who are willing to pour large amounts of money into startups they believe in.

vii. American startup ecosystem

The environment in which unicorns are born and grow is another major factor that affects their success and their valuation. The United States are the country that generates and has the highest number of unicorns worldwide. They are also the first country for the number of unicorn relocations; in 2021 4.1% of global unicorns moved their HQ to a different ecosystem, with 72% relocating to the U.S. (Fahimi, 2022). The reason for these primacies lay in different aspects, almost all of them related to its ecosystem. Overall, the U.S. has a thriving startup ecosystem due to various factors such as the availability of funding, the flexibility of the private sector, an integrated market and a culture that encourages risk-taking and embraces failure. Additionally, the U.S. has access to a large pool of talented individuals who contribute to the success of its ecosystem and together with universities and companies created a strong network of entrepreneurs and a connected like-minded community. In the U.S., Silicon Valley, Boston, and Austin are three geographically concentrated ecosystems that have been home to most technology ventures (Stephens, 2019). The Silicon Valley region is a prime example of a successful cluster where universities, investors, innovation, and successful companies converge to create opportunities for entrepreneurs and founders.

According to Baltrusaitis, J. (2022) in *Revealed: The U.S. has twice as many unicorns as China and India combined* the main factors that help explain the reason why the US is the leading country in regard to the number of unicorns are a well-regulated environment, large funding availability, and innovation potential. All these make thighs easier for entrepreneurs that want to start a company.

Regarding regulations, for instance, the JOBS Act enacted in 2012 made some changes in the U.S. system, to increase access to capital for small businesses and startups (Howe, 2015). The law includes several provisions to make it easier for small companies to go public, raise capital, and sell securities. For instance, it extends startups from certain SEC registration requirements when it regards raising capital through crowdfunding. It also expanded the maximum investor base of private companies before having to publicly disclose financial information from 500 to 2000 (Congress, 2012). It is hard to assess how strong was the impact of this law in the soar of startups valuations, however, it surely leads to increase access to capital for startups.

In comparison, Asian countries lack this flexibility in the private sector, which normally is subjected to stricter regulations, especially China. Due to the government's fear of having fast-growing tech companies becoming too dominant, in these countries, the governments have taken action to limit the power of their growth. Consequently, venture capitalists are more sceptical about investing in these markets. Nevertheless, the attractiveness of these companies remains strong (Baltrusaitis, 2022).
Another aspect is linked to the number of M&A deals, a high activity of M&A attracts more funding (Pala, 2022), the U.S. is the country with the most M&A deals in the world, and it has more than double than the whole European ecosystem (Choy, 2021).

With regards to funding, the U.S. still hold first place in the world, in the 2022 Global Startup Ecosystem Index report by StartupBlink, North American startup get 52.3% of all global startup funding including early and late-stage investments, while 24% for Asia Pacific region and just 18.3% for Europe (StartupBlink, 2022). The U.S. is also first when it comes to venture capital investment, which accounts for 0,63% of the GDP, more than ten times higher than Germany. Moreover, venture capital investment increased by 345% from 2010 to 2019 (Fink, 2020). All these facilitate the rise of new business thanks to the easiest access to VC investment rather, which invests in riskier business. Contrary to banks, which are more reluctant to invest in risky start-ups.

The U.S. culture of failure is extremely different from the European one. One of the main underlying factors for this difference is the investors' mindset, which then affects how founders think. For instance, according to founder Emmanuel Debuyck, in France, people tend to discourage entrepreneurs and suggest reconsidering before taking a risk like funding a company due to the fear of failing, whereas, in the U.S., the public encourages to take the plunge (Pala, 2022). This can be considered one of the reasons why companies, including many unicorns from Europe, decide to relocate to the U.S. In fact, Emmanuel Debuyck decides to relocate his company Adwanted Group to the US. This was because in France it was extremely hard to get funds from banks because of the failure of his previous business during the global financial crisis, not considering that he was running it for 18 years before folding. A Danish company called Unity Software also decided to relocate its operations and headquarters to San Francisco. The main reasons according to the findings of a survey made by the Danish Chamber of Commerce are the lack of access to second-round capital and difficulty recruiting qualified staff. Sometimes American VC firms ask to move the business to the US, due also to the higher potential of return of an exit if listed in the U.S. rather than in another market (Pless, 2022), because of the number of investors and its image. All this is to point out how the location of where to run a business has a key role in affecting the development of a unicorn. Therefore, as this factor is relevant on the world scale, it might also have an impact within the region itself, thus in specific parts of the US, such as California, New York, Boston, etc.

Silicon Valley is in fact the region that still holds and produces the greatest number of unicorns (CB Insights, 2017). It is the cluster where, universities, talent, investors, innovation, and successful companies meet, creating a diversified and efficient network for entrepreneurs that want to start a business. After the California area, the East Coast with New York and Boston have the highest unicorn population (CB Insights, 2017), Table 1 shows where the main clusters of unicorns are located within the US, based on data from 1110 U.S. companies that became unicorns between 1997 and 2021, based on their headquarter location.

Table 1 Unicorn Clusters in the U.S.

Cluster in the U.S.	Number of unicorn startups in 2022
San Francisco (California)	294
Silicon Valley (California)	189
New York	147
Boston	76
Los Angeles (California)	57
Seattle	33
Washington DC	31
Chicago	30
San Diego (California)	26
Irvine (California)	15
Philadelphia	15

Source: Strebulaev (2022)

Lastly, the U.S. is a large market with the same culture and language, where laws regarding business and funding are mostly harmonized, this is very different from Europe where the regulatory frameworks usually differ between countries. Therefore, a startup founded in the U.S. has easy access to all the states and potential US-based markets from the beginning, while for European startups it is more difficult to expand their activities and introduce their solutions across borders.

viii. The fintech sector

CB Insights database (2023) contains a list of all world unicorn startups existing nowadays. In the United States, there are currently 651 privately owned startups valued above one billion USD, actively operating. Among those, 162 are operating in the *Internet Software & Services sector*, 134 operate in the *Fintech* sector, 53 deliver Artificial Intelligence solutions and 45 offer Cybersecurity services and products. Therefore, 61% of US-based startups offer IT-related products and services. According to

McKinsey Technology Trends Outlook Report (2022), technology continues to be a primary catalyst enabling fast development of businesses, and governments and contributes to humanity's well-being. As for the number of unicorn companies, Fintech is the second largest sector in the United States. Figure 7 presents the major sectors in the U.S. startup ecosystem and the number of unicorns per sector.

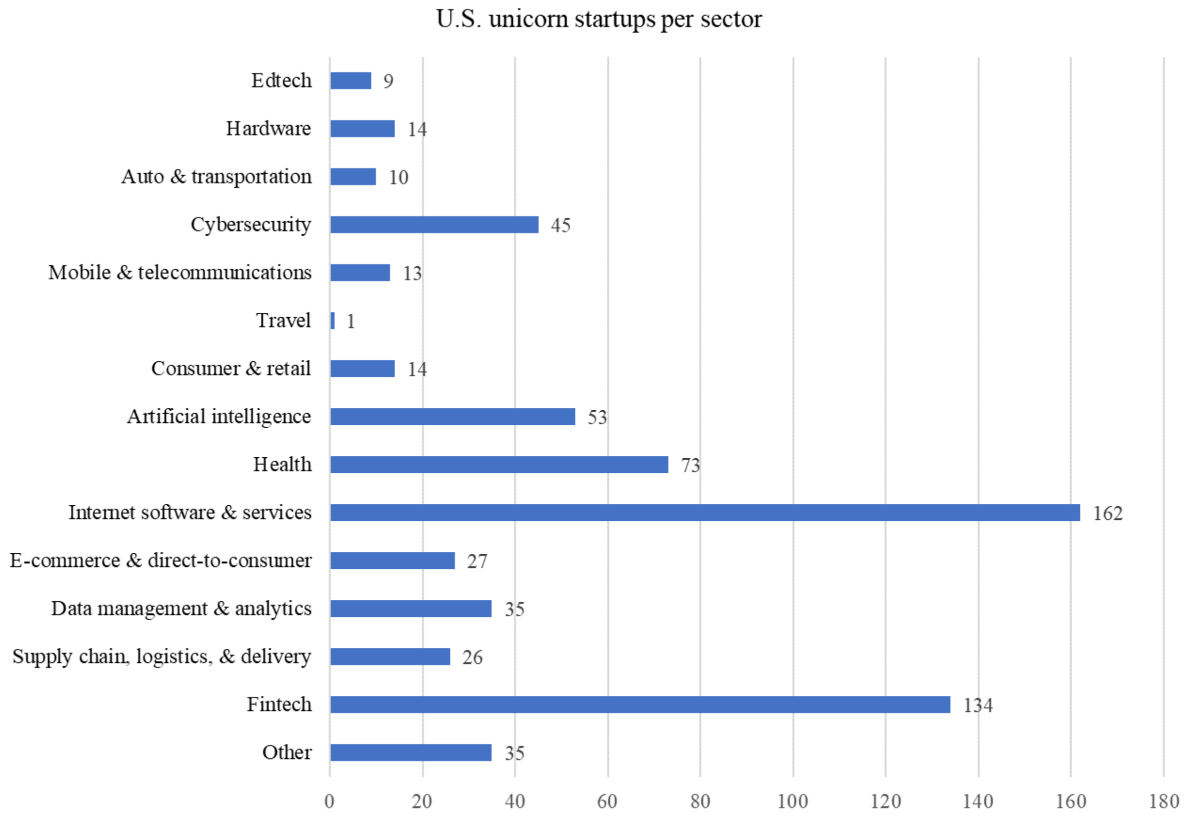

Figure 7 The U.S. unicorn startups per sector

Source: CB Insights (2023)

Fintech is short for "financial technology", and it is an umbrella term for financial software, algorithms, applications, digital banks, digital insurance solutions and investing platforms. The annual Forbes Fintech 50 list of companies includes the largest and most promising companies in the industry. The 2022 list included companies such as the established Sweden Klarna (a company that delivers online payment solutions) and the industry unicorn giant - Stripe (Point of Service solutions for B2B and B2C companies). One of the most popular Fintech companies in the world is Revolut - a British enterprise that changed digital banking and currency exchange by offering fee-free currency exchange, bank services, and cryptocurrency trading for everyone.

Fintech, short for financial technology sector accelerated during the COVID-19 pandemic. According to the report by the World Bank on *Fintech and the future of finance*, countries with a greater number of COVID-19 cases per capita had on average larger increase in their financial app downloads (both bank and non-bank apps). The report also stated that the digitalization of financial services and money helps to build more inclusive financial services and promotes economic growth (Feyen, Natarajan & Saal, 2023). What technology can do is lower costs and increase security and transparency of financial services, therefore, Fintech is said to have the "potential to transform the provision of financial services, spurring the development of new business models, applications and processes, and products" (IMF/ World Bank, 2018). Two fundamental drivers of Fintech development are universal connectivity of internet-connected devices and communication networks, data storage and computing power (Feyen, Natarajan, Saal, 2023).

The Fintech sector can be divided into subsectors including Digital Payments (Point of Service payments), Personal Banking/ Investments, Financial Management, Blockchain-based technologies and various investing platforms including NFTs and cryptocurrency trading. Fintech companies deliver goods and services on a B2B (Business to Business) and B2C (Business to Consumer) basis, or also both B2B and B2C together. In the United States alone, in 2022 the total transaction value of the Fintech sector was USD 1.77 trillion for Digital Payments and USD 1.08 trillion for Neobanking (Online Banking) and these numbers are expected to grow respectively to USD 3.53 trillion and USD 2.60 trillion in 2027 (Statista, 2023). The revenue of the Fintech sector in the United States was equal to USD 42 billion in 2021 and that number almost doubled two years later in 2023 when the total sector revenue was USD 78.45 billion This trend is expected to grow up to USD 139 billion in 2027 (Statista, 2023).

III. Data collection

This chapter will discuss the details of the data selection process needed for the cross-sectional analysis. From how the data was collected to its characteristics and composition. The data gathered was used for statistical analysis in order to find what are the significant factors determining the post-money valuation of a Fintech Unicorn startup from the United States.

To conduct the study, we created a data set composed of 129 observations of Unicorn Fintech startups from the US, collected from the *World Unicorn Club* dataset of Unicorns by CB Insights (2023). The data set contains a complete list of all world unicorns and is updated whenever a new company joins a "unicorn club". The complete list has been limited to the US-based companies from the Fintech sector only, which was possible as CB Insights delivered the list where the companies were already assigned to certain sectors and locations. Additionally, the dataset contained the current valuation of the companies.

The next step was to gather the variables that were needed to build the regression model. The variables were chosen based on the literature review and available information. As the analysed companies are privately owned they are not obliged to disclose their detailed financial data, therefore the data used mainly concerns funding information available to the public and characteristics that are known due to the companies' activities.

The quantitative and qualitative information needed to create a complete data set was manually exported from Crunchbase. Crunchbase is a platform providing business information about private and public companies. Their content includes investment and financing information, founding members and executives, mergers and acquisitions, investments and industry trends. To collect additional information about the companies' performance and activities platforms such as LinkedIn and Pitchbook were used together with the companies' official websites.

The sample contained information on all companies at a single point in time which made it cross-sectional data. The data represents the available information as of March 2023 when the sample was created.

The nature of a variable can be either quantitative or qualitative expressed with a dummy (value of zero or 1). The independent variables that were chosen for the model are shown underneath. There are studies regarding some of the qualitative variables and their relationship with the firm value. Nevertheless, for others, these are lacking, especially for funding rounds, sub-sectors and a number of investors. This is also one of the reasons we decided to include them in the regression since as mentioned from more qualitative information in the theory, the funding part of a startup has a major role in their growth. See if this is still the same as when the unicorn's status has been achieved can be an interesting finding.

However, In both cases, the results can be difficult to predict and possibly contradict some of the analytical literature especially because most of the previous quantitative studies are not about either the Fintech sector and neither unicorns:

1. Total funding amount: according to the literature the amount of funding a company get is a relevant factor when it comes to unicorns' reasons for failure. In fact, not enough funds can lead the company to run out of cash (Pride, 2018). On the contrary, overfunding can cause a false sense of stability and financial security that may lead to loose motivation and drive for efficiency (Pride, 2018). In fact, two studies, by Douglas Cumming and Na Dai, and by Nazir and Tbaishat, indicate a U-shaped relationship between fund size and firm valuations (Nazir & Tbaishat, 2023; Cumming & Dai, 2011). Both studies found that while increasing fund size initially led to an increase in firm valuations, there was a point at which further increases began to have a negative impact. Nazir and Tbaishat's study also found that raising too much capital could negatively affect market valuation, with a minimum amount of $357.5 million needed to generate significant positive effects (Nazir & Tbaishat, 2023; Cumming & Dai, 2011). The question remains whether these findings apply to specific sectors like Fintech and in different countries such as the US. Moreover, companies' funds have significant relevance in determining the company's value since they contribute to its book value and overall valuation. Thus the hypothesis here is that the higher the funding, the higher the company's overall value and success.
2. Revenue: private companies are not required to disclose information about their revenue, only the revenue range is known. Therefore, we divided the companies into 3 groups; average revenue range being USD 10-50M, above average and below average. In the model, this factor is expressed by two dummy variables - above average and below average. This particular variable is included in the model since it has an important weight when it comes to company valuation. When evaluating a company, future cash flows are predicted and these predictions are based also on the current revenue the firm is generating, even though most of the impact it has the forecasted one, which is value goes way beyond the current revenue the company is generating. A sustainable business model involves effectively managing revenue, growth rates, and costs to achieve break-even and profitability (Pride, 2018). Moreover, startups must demonstrate market traction through revenue growth and a clear path to profitability. Nevertheless, Lee (2019) suggests that the valuation of a startup may not necessarily indicate its financial success or any other relevant information. Additionally, some of these companies

with high valuations have yet to produce a profit (Lee, 2019). However, we believe that these companies are outliers.

3. Year of foundation: According to what was said previously in the literature review, the year 2021 was a boom for unicorns and startups (CB Insights, 2022). Therefore, specific periods or years can affect a company's evaluation, which is something that is connected to the timing factors as well for a product or a service within a specific market and industry (Howe, 2015). Furthermore, the hypothesis here is that older companies have an advantage as they are well-established and more mature. However, some unicorn companies achieved the status after just a year or two of existence which makes the results from this variable very exciting. Moreover, the average time to become a unicorn in the U.S. is seven years and three months and in the Fintech sector si seven years and nine months (Ahmad, 2022). Considering also that in the U.S. companies go public only after 8 years on average (Erdogan et al., 2016). Therefore, do older companies and firms that hold the status of unicorns for a long time have significantly lower valuations compared to the younger and fastest ones?
4. Funding rounds: refers to the number of funding rounds that the company has undergone. Usually, a higher number of funding rounds means that the company is promising enough to attract investors more than once (Lassala & Ribeiro-Navarrete, 2022). Moreover, the more funding rounds a startup gets, the more stages has reached, increasing its valuation and leading to the maturity stage, finally then reaching the exit (Lassala & Ribeiro-Navarrete, 2022). A high number of funding rounds does not necessarily mean high funding amount therefore this variable is predicted to deliver interesting results.
5. Number of investors: we wanted to test if a too-high number of investors could create disharmony. Normally, a higher number of investors mean also that there is more dilution for existing shareholders. Since it can be that the ownership of the company is spread among more shareholders. Moreover, disharmony between investors and founders is one of the reasons for unicorns' failure (Pride, 2018). Nevertheless, similarly to the funding rounds, a high number of investors does not necessarily mean high funding amount and high value.
6. Number of employees: normally, the number of employees increases in parallel with the growth of the company. As discussed previously there are different stages of a startup and each one of them expects an increase in the number of people working in the company (Merzlova, 2022). A study found that there is a low significance level of positive correlation between company valuation (Krch, 2018). Therefore, Including this numeric variable could help assess whether a higher number of employees significantly contribute to reaching a higher valuation. Moreover, testing it within the selected sample could lead to a different result since within the fintech sector where most companies are internet-based, this growth could also reach a peak and not be gradual anymore, assuming a U-shape curve.

7. Location: quantitative information expressed with dummy variables; a value of "1" if the Unicorn is located in California and "0" if in any other state. According to previous studies and findings, the location of the company has a significant effect on its development and growth (Blank & Dorf, 2012; Stephens, 2019). Due to the benefits or problems correlated to the business and entrepreneur environment within that location such as access to funding, network, culture, quality of labour etc. California state is well known for being the best location in the world to start a business, in fact, is the region which generates the most unicorns in the world (Gornall & Strebulaev, 2017). The question is if it is a significant advantage also for fintech unicorns.
8. Sub-sector: all the companies in the sample belong to the Fintech sector, however, they provide different goods and services to their users. We grouped them into 7 categories to see if the speciality of the company impacts the valuation. Those categories include:

    - Digital Banking
    - Personal Finance
    - Cryptocurrency (payments and trading)
    - Online payment solutions
    - Insurance/ Insurtech
    - Financial Management
    - Blockchain technologies

    Considering possible overvaluation problems (Gornall & Strebulaev, 2017) we expect at least one of them to be significant. The fintech sector has the highest number of investors (Eckert, 2022), therefore we believe this could lead to specific overvaluation problems also within each speciality. For example, fintech unicorns involved in cryptocurrency and blockchain technologies may have attracted larger funding amounts and achieved higher valuations, driven by the optimistic market outlook for these sectors, particularly until 2021. It is important to note that the valuation dynamics of these unicorns might have been influenced by fluctuations in cryptocurrency values, which experienced declines after 2021, following the bitcoin plummet (Google Finance, 2023).

    8. B2B or B2C basis: we wanted to include this variable to test whether the client base of the company impacts the valuation. We divided the companies into three categories; "B2B", "B2C" and "both". This variable was expressed with two dummy variables. A study from Dotzel and Shankar examines the impact of business-to-business service innovations (B2B-SIs) and

business-to-consumer service innovations (B2C-SIs) on firm value and found that B2B-SIs have a positive and significant effect on business value, greater than B2C-SIs (Dotzel & Shankar, 2019). Important to note that differently to B2B, more specifically, B2B-SIs are service innovations that are aimed at improving the business-to-business interactions and relationships of a company. Therefore, B2B companies in fintech offer solutions such as payment processing, risk management, and compliance to other businesses. Examples of B2B-SIs in fintech include platforms providing financial data analytics, blockchain solutions, and API integration services to other businesses in the financial industry. Stripe and Plaid are examples of B2B and B2B-SI fintech unicorns, respectively. Consequently, our results could be different from these findings.

Variables for the model were carefully selected to include those considered significant by other researchers in the literature and those that are able to provide interesting results. In addition, the number of variables has been selected so as to avoid "over-fitting", i.e. the concept in science which occurs when a statistical model fits perfectly against its training data and has very high indicators of a good fit such as an R-squared value. When this happens, the algorithm can no longer perform accurately and provide reliable results while used on other data.

IV. Data structure

The collected and analysed sample consisted of 129 companies however, in the United States there are currently 134 Fintech Unicorn startups. During the data collection process, we eliminated 5 companies due to a lack of information. This was caused either by the fact that these companies did not disclose any relevant observations or due to their absence from Crunchbase and Pitchbook databases. Some of the included in the sample companies are well-known and well-established businesses such as Stripe - an American giant operating in the digital payment solutions subsector or Deel - a company providing hiring and payments services for companies hiring international employees. The highest valuation on the list is USD 95 billion (Stripe), and the lowest valuation equals USD 1 billion (the value bar to become a unicorn company). The average valuation is USD 5.58 billion, the median is USD 2.0 billion, and the mode of valuation equals USD 1 billion. Companies with high valuations such as Stripe could be considered outliers which are observations that lie an abnormal distance from other values in a sample, in other words, “extreme values”. Before excluding them from the sample we decided to test if they will cause any problems with the regression analysis and its inference which will be further discussed in the methodology part.

Some variables used in this research represent numerical data, and others represent categorical data and had to be expressed with dummy variables. Numerical variables can be analysed with the use of popular measures which can give researchers an idea of how the data is structured. One such example in our sample is the *number of employees*. The maximum number of employees is 7 990 and is held by Stripe, while the minimum is only 4 employees at the New York-based company Unit providing financial management solutions for B2B clients. The average number of employees in the sample was 656, and the median was 416. Table 2 shows the maximum, minimum and average of the numerical variables' values.

Table 2 Numerical variables' measures

Variable	Maximum	Minimum	Average
Funding amount	USD 8.7 B	USD 35 M	USD 540.80 M
Number of funding rounds	22	1	7
Number of investors	117	1	25

Source: Dataset’s descriptive statistics

We wanted to include revenue as a numerical variable in our model, however since all the companies in the sample are privately held, they are not required to disclose their financial information to the public. Nonetheless, the Crunchbase database provides the revenue range, which we used to group companies into three categories: average revenue, above and below average revenue. The average revenue for the companies in the sample was USD 50M - 100M. That way we were able to establish that 32,56% of the companies earn USD 50M - 100M, while 18,60% have revenue above this level and 48,84% of the companies have revenue below that range which can be seen in Figure 8.

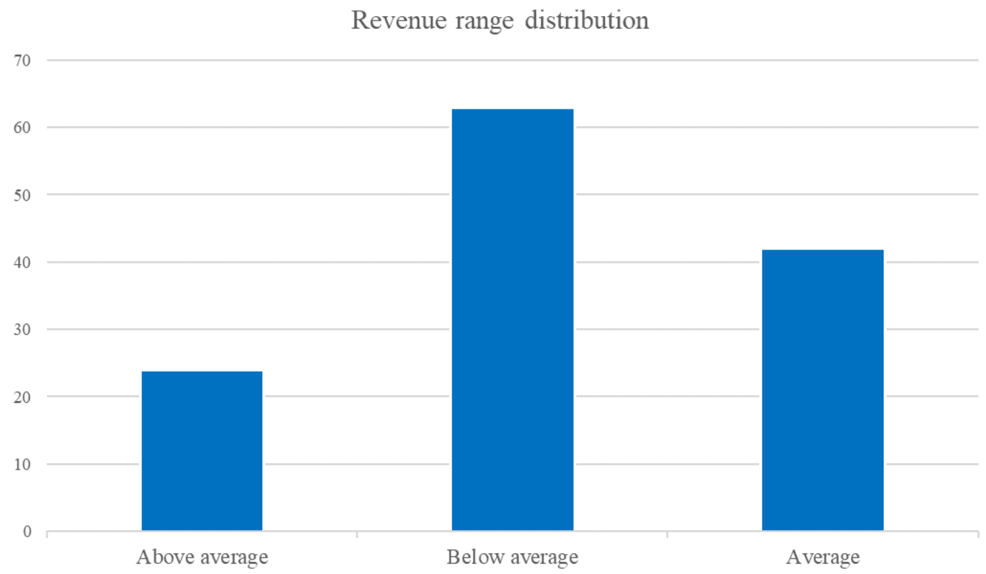

Figure 8 Revenue range distribution

S*ource: Dataset's descriptive statistics*

Although all companies come from one country and one sector, there are noticeable differences between them in terms of founders, location in the U.S., sub-sector and type of customers (B2B or B2C). Only 10 companies out of 129 from the sample were founded by women, which indicates a high gender disproportion which can be seen in Figure 9. Irfan Ahmad (2022), the founder and editor of Digital Information World says it is important to note that women are still greatly underrepresented in the billion-dollar business club. However, things are started changing, the number of female-founded unicorns has increased by over 2% annually since 2019, and in 2021 12.9% of unicorns globally had female founders. A percentage that is still too low but a sign of change, considering that from 2007 to 2013 this number was zero.

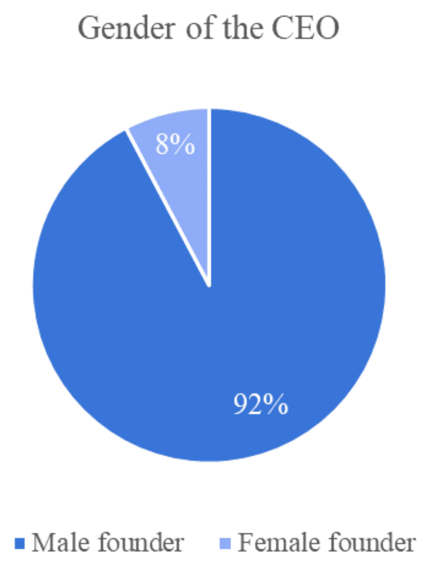

Figure 9 Gender of the CEO distribution

Source: Dataset's descriptive statistics

In terms of location, we decided to group the companies from the sample into two categories; the first one composed of the unicorns based in California state and the second composed of the companies with headquarters located in all the other states in the U.S. Figure 10 presents the geographical distribution of the Fintech unicorn startups in the U.S.

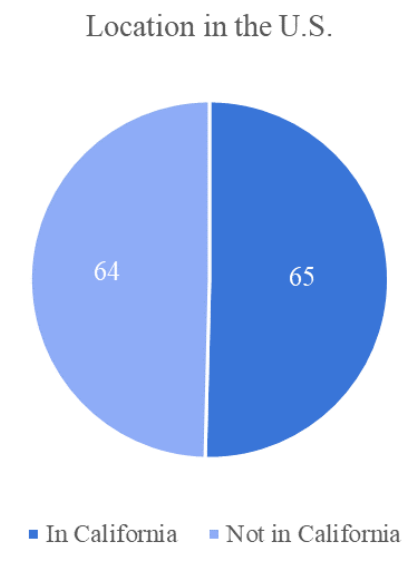

Figure 10 Location in the U.S. distribution

Source: Dataset's descriptive statistics

We wanted to test whether the basis of the operations matters, that is whether a company provides goods and services to companies or individuals. Thus, we divided companies into three categories B2B, B2C and both B2B and B2C which can be seen in Figure 11.

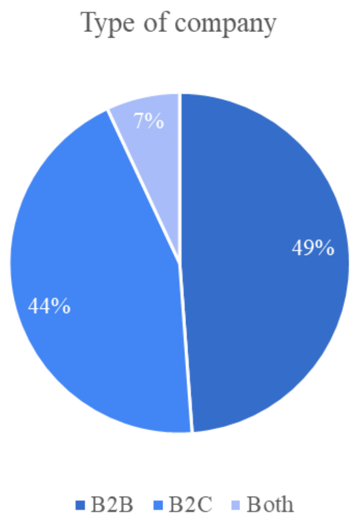

Figure 11 Company type distribution

Source: Dataset's descriptive statistics

As it can be observed, most of the companies in the sample provide goods and services to other businesses, however, the difference between the number of B2B and B2C companies is not substantial.

Considering all information discussed above, it can be seen that although the companies come from the same sector in one country, there are significant differences between them that can directly affect the company's valuation.

V. Methodology

This section provides the reader with the methods used to conduct the research and answer the research question. Therefore, explains how the general approach used in the study, fits the research design and what kind of statistical analysis is applied to analyse the data. The aim of this study was to determine the key, most significant factors that influence the final valuation of Fintech Unicorn startups in the United States. After gathering appropriate data, the next step was to detect possible model deficiencies and issues in the dataset, including those resulting from the specificity of the data such as outliers, and then fix them through adequate techniques. With the improved data a multiple regression model was created and estimated using OLS (Ordinary Least Squares method) which helped to understand which variables have the highest significance indicating how the changes in the independent variables correlate with shifts in the dependent variable that is company valuation. Quantitative data analysis, calculations, and graphic statistics were carried out in the R language environment for statistical computing and graphic through the integrated development environment called R Studio.

i. Research paradigm

According to Uba and Lincoln (1994), a paradigm is a collection of fundamental beliefs or metaphysical principles that address ultimate or first principles. Paradigm shapes our understanding and approach to researching the world around us. Furthermore, is the basis of research since involves studying and understanding a certain phenomenon. There are four essential elements of a research paradigm, namely ontology, epistemology, research methodology, and methods. It is important for researchers to take into account the objectives of their research when selecting a paradigm for investigation; otherwise, the original purpose of the research could be lost (Rehman & Alharthi, 2016).

Ontology is knowns as the concept related to the nature of reality, or rather what constitutes it. Does reality differ from person to person due to their own understanding and experience of the world or there is a true single reality that is the same for everyone? This paper considers objective ontology, where reality exists independently of our knowledge of it, and that this reality can be discovered through objective means (Hofweber, 2020).

Epistemology deals with the theory of knowledge and learning, how people build their knowledge, and thus, how what is known is created and acquired (Steup & Neta, 2020). Since the research question was assessed using quantitative analysis with the use of quantitative data and qualitative data in form of dummies. We believe a mixed-methods epistemology may be more appropriate to integrate these different data sources. Therefore, using pragmatism as an epistemology framework. The reason is that pragmatism considers both methods valid to answer the research question (Legg & Hookway, 2021), contrary to positivism where the truth of knowledge comes from tested and objective conceptualization, while interpretivism believes that knowledge is subjective to the single individual and constructed through different interpretations (Alharahsheh & Pius, 2020).

The research design of the paper consists of quantitative analysis carried out in the form of multiple linear regression on a sample composed of cross-sectional data. As mentioned before the dataset contains information which is both quantitative and qualitative which are expressed in the form of dummy variables.

Consequently, since we did not start the research with a well-known theory and hypothesis, but rather developed an explanation after sample data is gathered and analysed (Goddard & Melville, 2001; Saunders & Thornhill, 2019), the research approach of this paper is inductive. starting with the observation of Unicorns in the U.S. and raising the research question, "Determinants of the startup success: What factors influence the valuation of unicorn startups in the U.S. fintech sector? An empirical analysis." From there, the process of gathering empirical data and trying to identify patterns started, subsequently generating hypotheses and a possible theory.

ii. Cross-sectional analysis

The analysis was carried out on a cross-sectional data set. In this type of sample, each unit is given at a specific point in time (Wooldridge, 2018). Our data set consisted of a sample of US-unicorn startups from a fintech sector and a variety of other units discussed in the Data Collection part of this thesis.

Multiple linear regression was carried out on the collected sample in order to study the relationship between the dependent variable, in this case, the post-money valuation of the unicorns and the independent variables which are going to be different factors that characterize the selected businesses. The equation of the described regression is

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + u$$

Where:

- y: is the post-money valuation of the unicorn (dependent variable)
- β_0 is the intercept (the slope of the regression line)
- $\beta_1 \ldots \beta_n$ are the slope parameters of the regression line, $dE[y|x_i]/dx_i$ which measures the expected change in y with respect to x_i , holding other factors fixed
- $x_1 - x_n$ are the independent variables (regressors) that denote the firms' attributes
- u is the error term (or random error) which represents the effects of other unobserved variables that influence y. Therefore is the unpredictable part of y (Wooldridge, 2018).

When dealing with qualitative information in a regression model, dummy variables are used to express this kind of non-numerical data. Some examples are gender, industry sector, etc. The only values a dummy variable can have are either zero or one. The dummy variable to include in the model when representing binary information is only one, nevertheless, they can be more if the goal is to represent different binary information (Wooldridge, 2018). The reason for the omission of one dummy variable is to dispose of perfect collinearity meaning that one dummy would be a perfect linear function of the other dummy variable.

The steps to follow when building a statistical model are generally four. The first step is the identification which includes the formulation of the theoretical and statistical model as well as the preliminary analysis or testing. Secondly, there is an estimation, thus in this case, running the regression model in R of the sample with the use of OLS estimation. Then comes model validation/diagnostic which includes tests to validate the statistical model like the F-Test, t-test, test for heteroscedasticity, and Breusch-Godfrey test for autocorrelation. The last step is model selection where tests such as the RESET test are run to check which one is the model that best fits the data. (Wooldridge, 2018)

Identification

Assumptions on the error terms moments need to be made to set the parameters $\beta_0, \beta_1, \ldots \beta_n$. Thus when building a regression model, it is necessary to state the following assumptions:

- u, y, x_i are random variables. In our case, since we have a dataset of companies and multiple variables associated with each company, each independent variable is represented as a vector. At the same time, the dependent variable is also represented as a vector, with each element in the vector corresponding to the valuation of a specific company, the same as for the error terms.
- $E[u] = 0$ hence, $E[u|x_i] = \text{cov}(u, x_i) = 0$ i = 1, … n This implies that the error terms must have unconditional mean equal to 0 and u must be uncorrelated with any regressors (x_i). Therefore the error terms are also uncorrelated with any linear combination of the variables.
- $E[y|x] = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$

(Wooldridge, 2018)

iii. OLS estimation

The real parameter $\beta_0, \beta_1, \ldots \beta_n$ of the actual population generated by the DGP (Data Generating Process) are definite and unique, yet unknown. The term "DGP" refers to the mechanism that generates real world data. The DGP represents the true underlying process that generates the dependent variable 'y' in a regression model. In OLS estimation, the DGP produces the values of both the dependent variable

and the independent variables, which constitute the observed data in our sample. Nevertheless, the true parameters ($\beta0$, $\beta1$, ..., βn) of the DGP, which capture the relationships between the independent variables and the dependent variable, remain unknown. Therefore, these parameters are estimated by analysing a sample and making inferences within the Data Generating Process (DGP). To achieve this, the principle of Ordinary Least Squares (OLS) is applied. As the name suggests, OLS aims to estimate the relationship between the dependent variable 'y' and independent variable(s) 'x' by minimizing the squared error terms (u) derived from each individual equation (Wooldridge, 2018).
Consequently, the estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n$ are chosen simultaneously to minimize the equation:

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_n x_{in})^2$$

(Wooldridge, 2018)

It is important to note that the real parameters $\beta_0, \beta_1, \ldots \beta_n$ are not random variables. They are fixed constants. On the other hand, the estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots \hat{\beta}_n$ are random variables with a distribution called a sampling distribution.

The fitted values are: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_n x_n$
The residuals are not estimates of the error terms. The formula for the residuals for the i observation is: $\hat{u}_i = +y_i - \hat{y}_i$

When an estimator is unbiased, the expected value of its probability distribution equals the parameter it is meant to estimate.
For the estimators to be unbiased, the following assumptions of the OLS must be satisfied:

1. Linear in parameters: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + u$
   The model does not have to be linear between x and y but it must be linear on β_0 ,β_i and u through linear specification on parameters.

2. Correct specification: y is related to the independent variable, x, and the error u, as
   $$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + u$$
   The only variable with systematic effect is x. Therefore, error terms are unsystematic: Thus they must be random since they have no obvious pattern.

3. No perfect collinearity: which states that in the sample there are no constant independent variables and there are no exact linear relationships among them.

4. Zero means of *u*: all elements of *u* are uncorrelated with X

$$E[u|X] = 0 \text{ , which implies}$$

$$E[u|x_i] = 0$$

it implies that the error term is uncorrelated with the entire set of independent variables as a group.

Under the assumption of unbiasedness, OLS is a consistent estimator. Consistency means that as the sample size increases indefinitely, the sequence of estimated values converges to the true parameter value β. In simpler terms, as the sample size grows larger, the estimator becomes increasingly accurate and approaches the true value of the parameter (Davidson & Mackinnon, 2003).

An important thing to note about hypothesis four is that if instead of E[u|X] = 0 there was just $E[u|x_i] = 0$ meaning that the error term is uncorrelated with each explanatory variable individually, not as a group, as for the fourth assumption. Then OLS is still consistent, but it does not imply unbiasedness (Davidson & Mackinnon, 2003).

Since the variance of OLS estimators has a critical role in confidence intervals and hypothesis testing, another assumption will be that the errors are homoscedastic (they have the same variance) and uncorrelated. Therefore, the error has the same conditional variance given any value of the independent variables. This assumption together with the four above makes OLS an efficient and unbiased estimator (any $\hat{\beta}_i$). It is important to highlight that with just the four previously mentioned assumptions, OLS is unbiased but not efficient. The assumption states:

$$var[u_i|x_1,\ldots,x_n] = \sigma^2$$

$$cov[u_i,u_j] = 0 \qquad for\ all\ i \neq j$$

Consequently, the error terms are random. In this case, if there is uncorrelation, it might be because some regressor is missing from the model (Wooldridge, 2018). Uncorrelation can also be a result of irrelevant explanatory variables included in the model. This phenomenon is called over-specification and may affect the variance of OLS; however, OLS is still unbiased.

The case when two or more variables have a high correlation between them is called multicollinearity. This phenomenon can be an issue when the correlation is particularly high, for instance, higher than 0.7 or 0.8. This can indicate a problem in the data construction but if the value is not too high it might just be a characteristic of the data set. The consequence of strong multicollinearity is that is going to be difficult to separate the effects of their parameters increase and the estimated variance between these variables (Wooldridge, 2018).

In order to test for multicollinearity, we used the Variance Inflation Factor (VIF) which quantifies how much the variance of a regression coefficient increased as a result of collinearity; the existence of correlation among the regressors in the model (Corporate Finance Institute, 2022). Therefore the variance inflation factor for the *ith* independent variable is:

$$VIF_i = \frac{1}{1 - R_i^2}$$

In the case where R_i^2 is zero, it is not possible to predict the variance of the other independent variables from the ith independent variable. Thus, a VIF or value of 1 indicates that the ith independent variable has no correlation with the other variables, indicating the absence of multicollinearity in the regression model, while the value above 5 indicates that there is a string multicollinearity present (Corporate Finance Institute, 2022).

When carrying out hypothesis testing it can be made another assumption, specifically on the distribution, to gather more information on the mean and the variance of the error terms. The normality assumption on the error term states: The error term u is independent of the explanatory variable $x_1, x_2, \ldots, x_2$ and is normally distributed with zero mean and variance σ^2, i.e. $u_i \sim N(0, \sigma^2)$ for all i = 1,..,N. In summary, the error terms are homoscedastic, uncorrelated and follow a multivariate normal distribution (Wooldridge, 2018).

Unbiasedness and normality assumptions of OLS are called exact since these properties hold regardless of the sample size (Wooldridge, 2018).

In the case the normality assumptions of the error terms are not satisfied, thus the error is not normally distributed, then the distribution of a t statistic is not exactly t and an F statistic does not have an exact F distribution for any sample size. Hypothesis testing can still be carried out, although they are not exact tests, but they are approximated. For instance, other assumptions that can be dropped are homoscedastic errors and uncorrelated errors (Wooldridge, 2018).

iv. Data transformation

Creating an initial regression model with raw data allows researchers to notice dataset irregularities or possible violations of the OLS rules. If such data issues are present, once detected they can be handled with the use of appropriate methods for their complete elimination or general dataset improvement. Logarithmic transformation is a very common way of handling situations where OLS assumptions are violated; where there is a non-linear relationship between the independent and dependent variables, if variables are highly skewed, if residuals are not normally distributed or if the heteroscedasticity is present (Benoit, 2011). In that transformation, the natural logarithms are used, where the base is e ≈

2.71828. Transformation can be done on dependent variables only, independent variables or all variables present in the model. In a case of a multiple regression model, the equations are:

1) Linear: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + u$
2) Linear-Log: $y = \beta_0 + \beta_1 * log(x_1) + \ldots + \beta_n * log(x_n) + u$
3) Log-Linear: $log(y) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + u$
4) Log-Log: $log(y) = \beta_0 + \beta_1 * log(x_1) + \ldots + \beta_n * log(x_n) + u$

These can be estimated directly from the OLS technique, what is more, if Y and Xs in that equation have log-normal distribution, then ln Y and ln X have a normal distribution (Kissell & Poserina, 2017). The dataset prepared for the study of this thesis contains both quantitative (numerical) information and qualitative information expressed with dummy variables. Dummy variables present in the model were not log-transformed due to logarithmic properties as:

- *log(1) = 0,*
- *log(0) is undefined.*

Therefore, the transformation was done on the numerical variables only, creating a mix between the Log-Linear model and the Linear-Log model. In the Log-Log linear regression model (model where both dependent and independent variables were transformed) the interpretation of the coefficient is the following: it is an expected percentage change in Y when X increases by some percentage (Benoit, 2011). If after the logarithmic transformation of the variables in the dataset, some of the OLS assumptions are still violated, other possible issues such as outliers can be detected and handled with the use of appropriate methods to improve the regression.

Must be noted that the interpretation of the linear regression changes after the log transformation of the variables. When no transformation is applied to a model a 1 unit increase in X1 variable is associated with an average change of β1 units in Y. However, in the case of the log-log model where both the dependent and independent variables are transformed a 1% increase in X1 increase is associated with an average change of β1% in Y which is linked to the properties of logarithms. In other words, a one per cent increase in the independent variable is associated with a $100 * (1.01^{\widehat{\beta}_i} - 1)$ per cent change in the dependent variable (Yang, 2012).

v. Outliers' detection and treatment

OLS estimation is very sensitive to outliers which can substantially change the results. Outliers are observations in a data set that are significantly different from the other observations within the same dataset. They may occur because of data entry errors making it objectively easy to deal with, nevertheless, they can be an outcome of non-random sampling or because some data are generated by a different model than most of the other data (Wooldridge, 2018). Their detection is therefore very important step of the process. Outliers can be detected by calculating summary model statistics, especially minimums and maximums. Another way to detect them is visually by creating plots of the data features of which the outliers' presence is expected. The most common plots used for outlier detection are scatter plots and box plots, both supported by the R Studio. On a scatter plot, potential outliers are points that are significantly distant from the other data points, in the case of a box plot, outliers are points located above the upper bar or below the lower bar of the box as they represent extreme values (Lütkepohl & Krätzig, 2004).

It is difficult to decide whether to keep or remove such observations from the dataset in a regression analysis, as the statistical properties of the resulting estimators are complicated (Wooldridge, 2018). Outliers can appear in the sample naturally and provide information by increasing the variation in the independent variables, however, as mentioned above, they can greatly impact OLS estimates causing misleading results. As the sample of this study is rather small (consisting of 129 observations) we decided to remove outliers as in the book by Wooldridge (2018) it is mentioned that outliers especially impact OLS estimates in small samples.

Extreme values can be removed from the data set to avoid misleading regression results. There are many methods to deal with outliers; one of them is data trimming. Trimming is done on the independent variable as well as on the dependent variables and is done symmetrically by removing the same number of observations on both ends of the dataset. Symmetric trimming on the independent variables has no effect on the regression slope β or the regression intercept α and has no effect on the mean square error of the regression, therefore, it is a safe method to fix potential issues resulting from the presence of the outliers (Lien & Balakrishnan, 2005).

vi. Hypothesis testing

In the hypothesis testing part, we first choose a significance level which determines the critical value along with degrees of freedom and alternative hypothesis.

T-test:

Any single parameter in a population regression function can be tested using the T-test, testing for a specific mean value of the parameter about the real world.

$$\frac{\hat{\beta}_i - \beta_i}{se[\hat{\beta}_i]} \sim t_{n-k-1}$$

The critical value of the T-distribution, which is based on the significance level and the n -k- 1 degrees of freedom, must be compared to the value of the T-test. Where the degrees of freedom are the final numbers of independent values that may vary in the sample, while K is the number of observations. The null hypothesis will not be rejected if the T-test value is within the confidence interval. However, if the T-test value is outside the confidence interval, the null hypothesis will be rejected (Wooldridge, 2018).

F-test

The F-test is used to test multiple hypotheses on parameters. It is not possible to use the t-test to test multiple hypotheses separately since they imply different models.

The formula of the test statistic is

$$F_\beta = \frac{(SSR_r - SSR_{ur})\,/r}{SSR_{ur}/(n-k)}$$

Where:

- SSR_r is the sum of squared residuals of the restricted model and SSR_{ur} of the unrestricted.
- The F-distribution of the statistic is $F_\beta \sim F_{r,n-k-t}$, r are the degrees of freedom of the numerator and $n-k-t$ are the degrees of freedom of the denominator.

The p-value of the F-test must be used to determine its significance. The null hypothesis is rejected if the p-value is less than the significant level of 0.05 (Wooldridge, 2018).

vii. Diagnostic

As mentioned before the diagnostic aim is to assess the model's validity through various tests. Some tests are done in order to find empirical evidence of consistency, normality, and heteroskedasticity and to assess the validity of the model and the inference based on it.

Jarque-Bera test for normality

This test is used to determine whether the error terms follow a normal distribution. The test is done on the residuals which are the approximation of the error terms of the real population.

The hypothesis is:

H_0 : skewness and kurtosis match a normal distribution (which implies normal distribution).

H_1 : skewness and kurtosis do not match a normal distribution.

The test statistic is:

$$JB = \frac{n}{6}\left(S^2 + 0.25(K-3)^2\right)$$

Where *n* is the sample size, *S* is the sample skewness and *K* is the sample kurtosis. Skewness is the measure of the asymmetry of the probability distribution, while kurtosis measures the “peakiness” of the probability distribution.

When the critical value is less than 0,05, it indicates that the value is outside the probability area, on the other hand when the p-value is greater than the critical value, it indicates that the value is inside the probability area and the null hypothesis is rejected. This means that data does not support exact normally distributed error terms. Hypothesis testing can still be carried out, but the tests are all asymptotical, thus approximated tests (Wooldridge, 2018).

Heteroskedasticity

If heteroskedasticity is present, the OLS estimator is no longer the best linear unbiased estimator. Must be noted that heteroskedasticity does not influence the model fit, but it does influence the uncertainty around it by biasing the standard error and test-statistics (Astivia & Zumbo, 2019). When the model has many regressors, is advised to use the Breusch-Pagan test to test for heteroskedasticity. The Breusch-Pagan test uses the following hypotheses:

H_0 : Homoskedasticity is present (the residuals are distributed with equal variance)

H_1 : Heteroskedasticity is present (the residuals are not distributed with equal variance)

In the case of heteroskedastic errors, which means that the error terms have different variances, the OLS estimator $\hat{\beta}_i$ is still unbiased and/or consistent. However, the estimator $var[\hat{\beta}]$ of the variance that assumes homoscedastic error is incorrect, biased, and inconsistent. This implies that all exact and asymptotic hypothesis testing is not valid (Wooldridge, 2018).

Methods to fix heteroskedasticity include log-transformation of the independent variable, which has already been applied, weighted regression or correction using heteroskedasticity-consistent standard errors (Wooldridge, 2018). The presence of heteroscedasticity gives more evidence that the model is not correctly specified as already found before since it suggests that there might be important variables that are not included in the model (Wooldridge, 2018). In order to obtain reliable standard errors in a situation where log-transformation did not bring the expected results in the context of heteroskedasticity, one can apply correction using heteroskedastic-consistent standard errors. This method does not modify the model or data, but only modifies the standard errors of the model. There are four possible HC (heteroskedastic-consistent) estimators: HC0, HC1, HC2 and HC3. White (1980) introduced the estimator HC0 into the literature on econometrics; nevertheless, for small sample sizes, the standard errors from HC0 are biased, leading to liberal results in regression models (Bera, Suprayitno, & Premaratne, 2002). Due to the fact that the sample size for this study consists of less than 150 observations, it is considered relatively small. With HC0, the bias decreases as the sample size grows. To enhance the performance in small samples, MacKinnon and White (1985) proposed the estimator types HC1, HC2, and HC3. Additionally, Long and Ervin (2000) contend that HC3 performs the best in small samples because it assigns less weight to the model's important observations. Therefore, in the remaining parts of this thesis, we focused on the HC3 estimator.

Goodness-of-Fit

To define how well the OLS regression line fits the data the R-squared can be obtained, which gives an idea of the explanatory power of a regressor. In other terms, it indicates to what extent the variance of y is explained by x. Therefore, if the R-squared of a model is 0.3, then approximately 30% of the observed variation can be explained by the model's inputs. The formula is

$$R^2 = \frac{SSE}{SST} = 1 - SSR/SST$$

Where *SSE* is the sum of squares errors, and *SST* is the total sum of squares.

Model selection

Model selection is necessary in order to drop potentially insignificant variables from the model and keep the ones that are significant, thus relevant. This process is also known as backward elimination. The selection should not be on all possible combinations, but it should be on a reasonable subset. In order to assess the relative quality of the statistical model given a sample, estimators named information criteria are used. These information criteria are namely, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). AIC gives a result assessing the amount of information lost in the model, the lower the value of AIC, the better since a good model is the one that loses less information. Similarly, also the lower the value of the BIC the better, because a higher number indicates the inclusion of additional non-relevant regressors (Wooldridge, 2018). If the goal is prediction, AIC is preferred, however, if the goal is selection, inference and interpretation, BIC is preferred. To select the best possible model for the analysis of results we used BIC and AIC criteria as guides since we created different models including a different number of variables. Given the AIC and BIC criteria, we were able to select the model with the lowest prediction error among the created models.

VI. Analysis

In this section, we present the results of the linear regression and cross-sectional analysis we carried out on the sample of 129 fintech unicorn startups in the United States. The study aimed to determine the most significant variables that influence the valuation of such companies. Firstly, by discussing the process of data preparation for regression through data transformation techniques. Secondly, bearing in mind the OLS assumptions, we tried to create the best model using the transformed data in such a way that these assumptions are not violated. The model obtained was then subjected to diagnostics, multicollinearity detection and OLS estimation. Then, in order to answer the research question, backward elimination was applied to the model containing all variables with the aim to create a model containing only significant variables.

i. Data transformation and trimming

After collecting the necessary data for the analysis, we created the first linear regression model. Looking at the model summary and graphs we noticed some problems with the data, including the fact that the residuals were not normally distributed, which we were able to determine from the histogram and the scatterplot between residuals and fitted values. Additionally, the points in the QQ plot did not lie on one line, the errors were heteroscedastic, and there was no linear relationship between the explanatory variable and the dependent variable. Therefore, OLS assumptions were violated.

Aiming at data improvement, we transformed the data using natural logarithms and detected the outliers. Our model contained both quantitative information and qualitative expressed with dummy variables. Considering the logarithmic properties discussed in part iv of the methodology we transformed only the numerical (quantitative) variables. We created a model called the *Numerical Model* which contained only numerical information (the dummies were excluded), i.e., the dependent variable being *valuation*, with independent variables being *funding amount, number of funding rounds, number of investors, year of foundation and number of employees*. With this model, we then started the transformation and analysis which helped us determine that the best type of transformation, in this case, is the logarithmic one where both dependent and independent variables were transformed. This decision was based on graphical analysis, Jarque-Bera test p-value as well as AIC and BIC of the models. Additionally, the VIF function helped us to state that there is no multicollinearity within the model containing only numerical variables. The next step was to gradually add dummy variables before transformation in order to select the best model in terms of the number of variables based on AIC and BIC, VIF, and model summary.

The best model has proven to be the model containing *valuation* as the dependent variable and the following independent variables:

- Funding amount,
- Year of foundation,
- Number of funding rounds,
- Number of investors,
- Number of employees,
- Dummies for revenue: average revenue, revenue above average and revenue below average, which was omitted due to dummy variables' properties,
- Dummies for company type: B2B, B2C and both B2B and B2C which was omitted,
- Dummies for subsector: online payment solutions, digital banking, financial management, personal investment, blockchain technologies, cryptocurrency and insurance which was omitted,
- Location dummies: located in California, and located outside California which was omitted.

The summary of the aforementioned model called the Complete Model, before any transformations is presented in Table 3.

Table 3 Linear regression summary for the Complete Model

Residuals:	Min	1Q	Median	3Q	Max
	-13.8451	-1.4976	0.4527	1.6755	10.1535
Coefficients:	Estimate	Std. Error	t value	Pr (>| t |)	
(Intercept)	-2.190e+02	1.829e+02	-1.198	0.233637	
Funding_amount	8.611e-09	5.829e-10	14.774	< 2e-16	***
Year_of_foundation	1.088e-01	9.062e-02	1.201	0.232224	
Funding_rounds	-1.371e-01	1.300e-01	-1.054	0.293986	
Number_of_investors	3.263e-02	2.084e-02	1.566	0.120105	
Number_of_employees	1.438e-03	4.680e-04	3.072	0.002670	**
Average	-8.219e-01	6.822e-01	-1.205	0.230813	
Above	-7.682e-01	9.504e-01	-0.808	0.420609	
B2B	-3.715e+00	1.346e+00	-2.760	0.006759	**
B2C	-4.962e+00	1.310e+00	-3.786	0.000247	***
Online_payment_solutions	1.140e+00	1.578e+00	0.723	0.471488	
Digital_banking	9.000e-01	1.758e+00	0.512	0.609782	
Financial_management	6.041e-01	1.332e+00	0.453	0.651088	
Personal_investment	2.001e+00	1.325e+00	1.511	0.133727	
Blockchain_technologies	3.140e+00	1.565e+00	2.007	0.047185	*
Cryptocurrency	3.759e+00	1.353e+00	2.778	0.006418	**
California	1.376e+00	6.178e-01	2.227	0.027970	*

Residual standard error:	3.275 on 112 degrees of freedom	
Multiple R-squared:	0.8793	Adjusted R-squared: 0.862
F-statistic:	50.98 on 16 and 112 DF	p-value: < 2.2e-16

Source: Own creation

The AIC value of the Complete Model before the transformation was 689,96 and the BIC value was 741,44. That model was later tested for multicollinearity using VIF (Variance Inflation Factor). The results of the VIF are presented in Table 4.

Table 4 VIF for Complete model before transformation

Funding_amount	Year_of_foundation	Funding_rounds	Number_of_investors
2.789358	1.372693	2.227838	2.052850
Above	Average	B2B	B2C
1.644526	1.228787	5.444258	5.092479
Personal_investment	Blockchain_technologies	Cryptocurrency	California
4.243970	2.105229	2.764953	1.147383
Number_of_employees	Online_payment_solutions	Digital_banking	Financial_management
2.388321	2.527588	1.648877	4.293448

Source: Own creation

There was strong multicollinearity in the case of B2B and B2C where the VIF value was higher than 5, therefore we decided to remove the B2C dummy from the model and keep only B2B. After excluding the B2C dummy variable there was no more presence of multicollinearity. QQ plot (quantile-quantile plot) created to compare distributions of residuals and fitted values for the aforementioned model is presented in Figure 12.

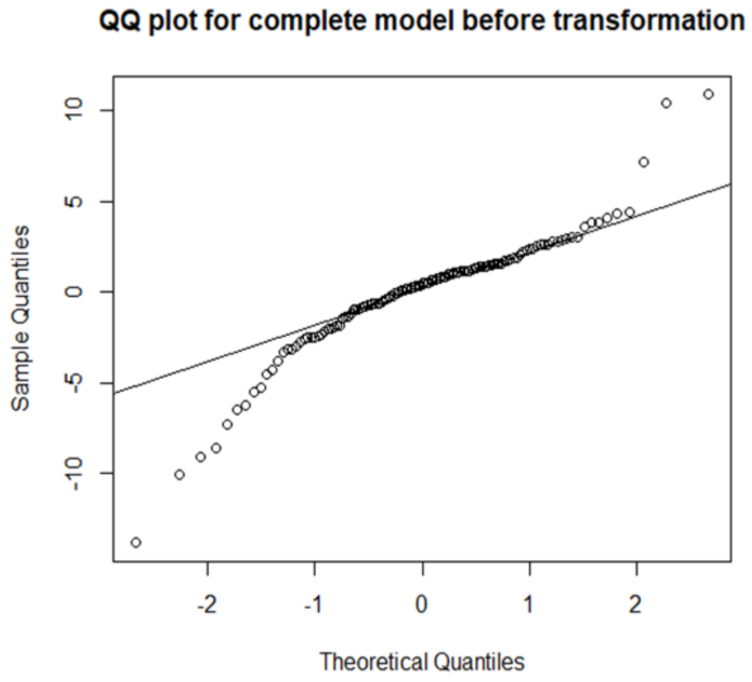

Figure 12 QQ Plot for Complete Model before transformation

Source: Own creation

When data is normally distributed, the data points in a QQ plot lie on a straight diagonal line with some minor deviations on each of the tails. Analysing the QQ plot obtained from the Complete Model before transformation we notice that it does not behave like the normally distributed one and that its tails are rather heavy. Heavy-tailed probability plot (QQ plot) indicated that error terms are not normally distributed. The following was also confirmed by the residuals' histogram shown in Figure 13 and in the scatterplot of residuals vs. fitted values shown in Figure 14.

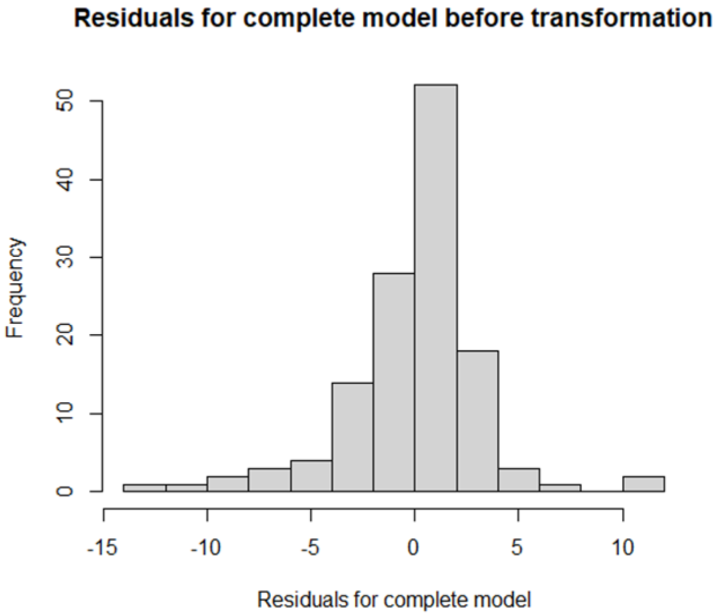

Figure 13 Residuals for Complete Model before transformation

Source: Own creation

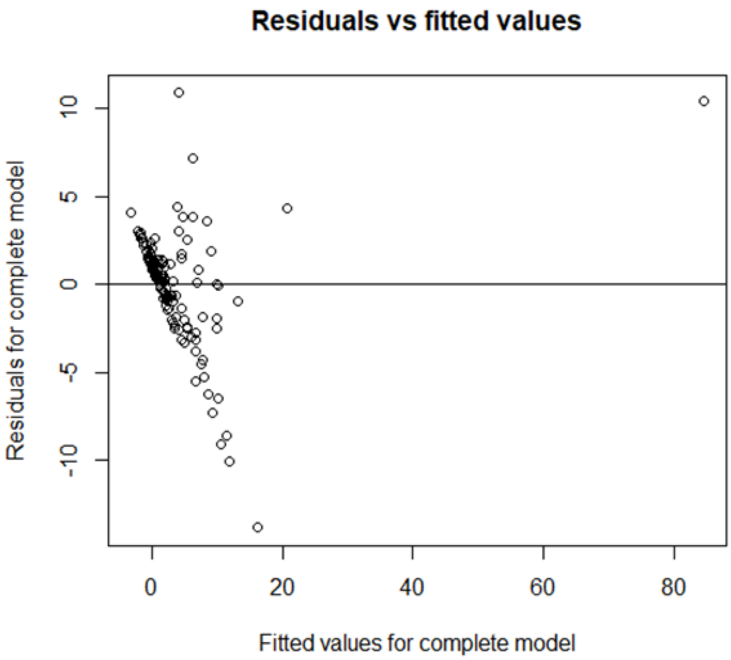

Figure 14 14 Residuals vs fitted values for Complete Model before transformation

Source: Own creation

Non-normality of the error terms was caused by the way how the sample was created and resulted from the presence of outliers. Figure 15 shows the outliers in the sample present when we create the boxplot of the dependent variable.

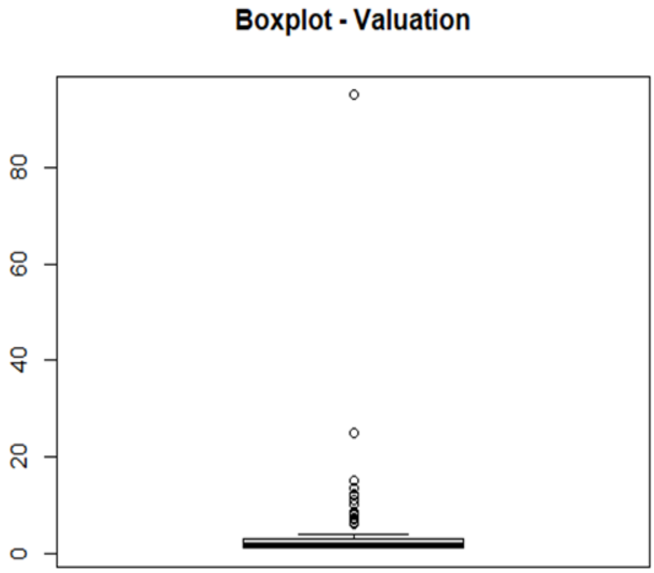

Figure 15 Boxplot of the sample's valuation values

Source: Own creation

We can see that our dataset contains extreme values. In the case of the valuation variable, the average valuation is USD 5.58 billion, the median is USD 2.0 billion, and the equals USD 1 billion however, there are companies in the sample valued at USD 8 - 25 billion and one company (Stripe, the most distant data point on the boxplot above) valued at USD 95 billion. The Complete model underwent a Log-Log transformation excluding the dummy variables which remained unchanged. Then, to remove outliers, the database was trimmed on both sides by 6.5%, which means that 8 observations were trimmed on both sides of the dataset. The data improved and outliers were not that extreme anymore however, it has to be noted that as we removed initial outliers, new ones appeared which can be seen in Figure 16.

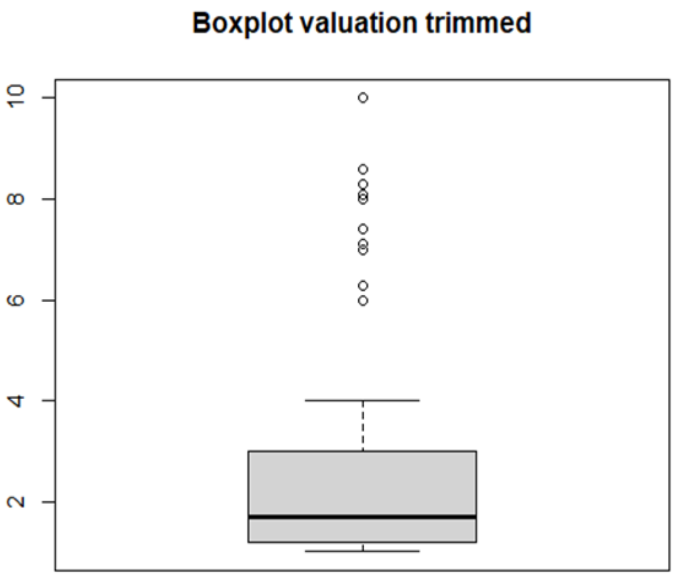

Figure 16 Boxplot of the sample's valuation values after trimming

Source: Own creation

The complete (in terms of the included variables), transformed and trimmed model was the one that further analysis was carried out. To present how the quality of the data improved and that the OLS assumptions were not violated we created the same plots and graphs as for the model before logarithmic transformation and trimming. Figure 17 shows the QQ plot of the improved model.

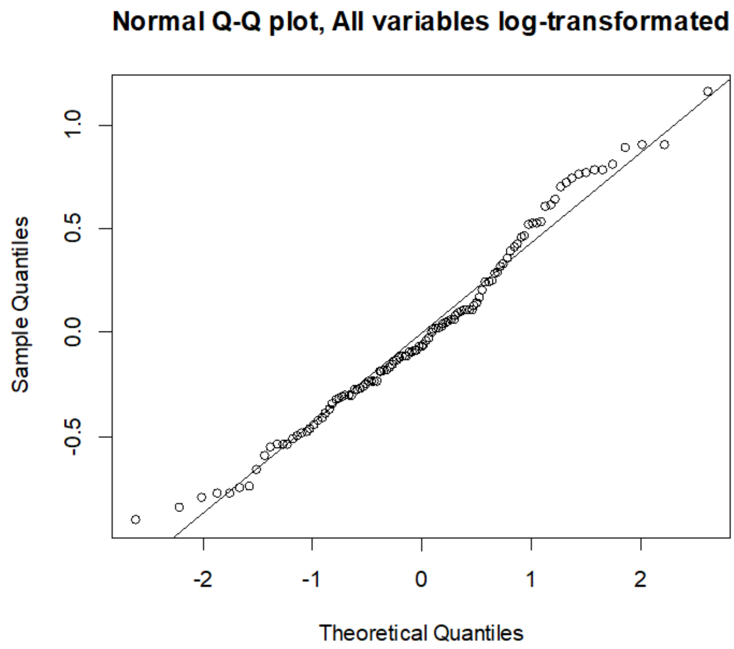

Figure 17 QQ Plot of the trimmed and transformed Complete Model

Source: Own creation

It can be observed that the tails are not that heavy anymore, and that data appears to be normally distributed (which will be confirmed in the next part with the appropriate tests), the data points in the new QQ plot lie on a straight diagonal line strongly resembling normally distributed residuals, supporting the condition that error terms are normally distributed. Figure 18 presents the histogram of residuals after log-transformation and data trimming.

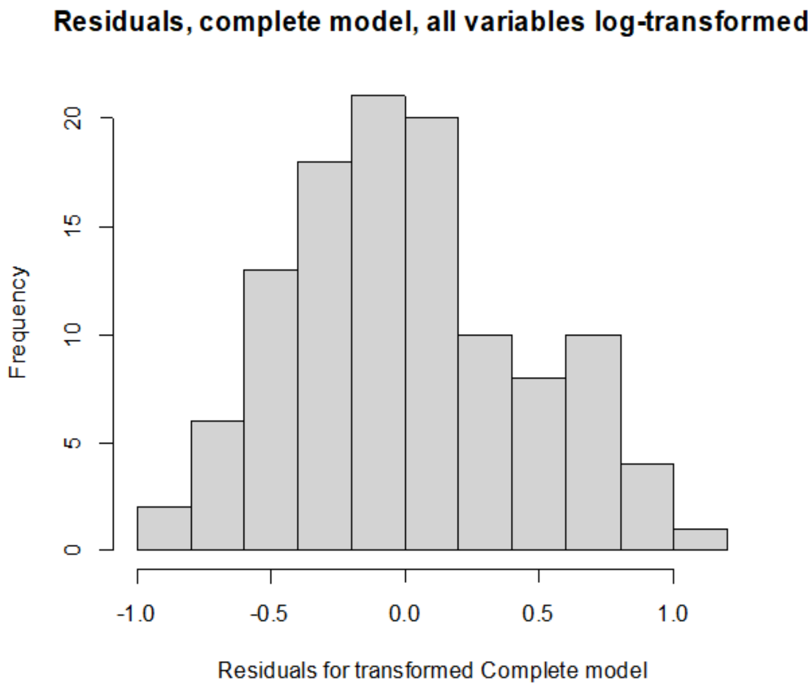

Figure 18 Residuals of the Complete Model after transformation and trimming

Source: Own creation

The last plot that was created to show that transformation and trimming improved our data was a scatterplot of residuals vs. fitted values shown in Figure 19.

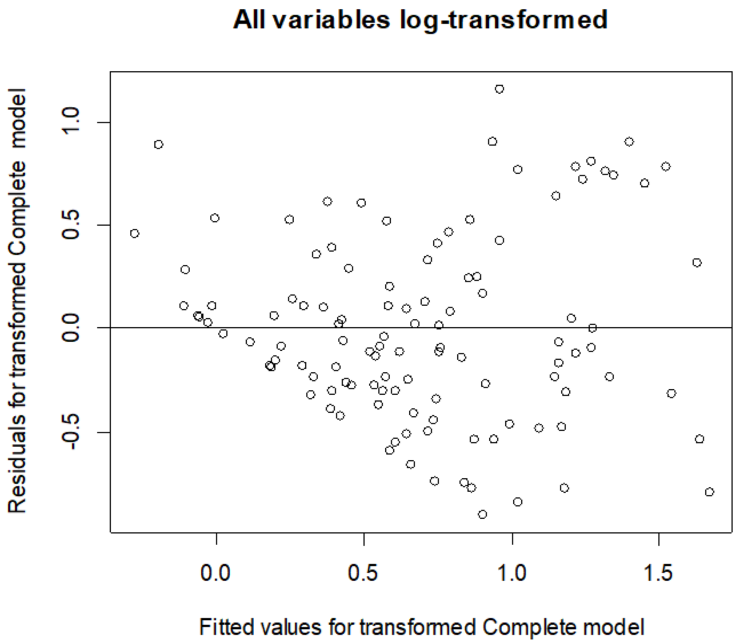

Figure 19 Residuals vs fitted values plot of the Complete Model after transformation and trimming

Source: Own creation

It can be observed that residuals are randomly distributed, instead of being clustered as they were for the model before transformation and trimming. The ideal residual plot (also called null residual plot) shows a random scatter of points distributed evenly, without any visible patterns. On the residual plot in Figure 19, residuals seem to form somewhat of a cone-shaped pattern suggesting heteroscedasticity, however, that will be verified with the use of appropriate tests in the diagnostic. Nonetheless, it must be remembered that heteroscedasticity is very common in cross-sectional studies due to data characteristics.

ii. Estimation of the model with transformed data

After improving the data by logarithmic transformations of the dependent variable and independent numerical variables, it was possible to obtain normally distributed residuals, so the next step was the estimation of the complete model. This model did not include the B2C variable due to the occurrence of multicollinearity which makes it difficult to determine the individual effect of the variable on the dependent variable. Thus, the estimated model included company *valuation* as a dependent variable and *funding amount, year of foundation, number of funding rounds, number of investors, number of employees, average revenue dummy, above average revenue dummy, location in California dummy and dummies specifying the subsector; online payment solutions, digital banking, financial management,* *personal investment, blockchain technologies*, and *cryptocurrency*. The summary of the complete model is presented in Table 5.

Table 5 Regression summary of the Complete Model after transformation and trimming

Residuals:	Min	1Q	Median	3Q	Max
	-0.90067	-0.29870	-0.0616	0.28546	1.15689

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-117.7649	229.9442	-0.512	0.609713	
log(Funding_amount)	0.45101	0.07726	5.837	7.03e-08	***
log(Year_of_foundation)	14.23827	30.19644	0.472	0.638327	
log(Funding_rounds)	-0.10411	0.12638	-0.824	0.412106	
log(Number_of_investors)	0.15272	0.08018	1.905	0.059786	.
log(Number_of_employees)	0.10845	0.05352	2.026	0.045465	*
Average	-0.11181	0.10545	-1.060	0.291635	
Above	-0.07242	0.15905	-0.455	0.649885	
B2B	0.05449	0.16065	0.339	0.735190	
Online_payment_solutions	0.22767	0.24120	0.944	0.347547	
Digital_banking	0.18214	0.29411	0.619	0.537176	
Financial_management	0.28953	0.20735	1.396	0.165813	
Personal_investment	0.26301	0.19874	1.323	0.188836	
Blockchain_technologies	0.75829	0.24608	3.082	0.002681	**
Cryptocurrency	0.73626	0.21334	3.451	0.000829	***
California	0.07031	0.09652	0.728	0.468089	

Residual standard error: 0.4806 on 97 degrees of freedom	
Multiple R-squared: 0.4871	Adjusted R-squared: 0.4078
F-statistic: 6.142 o 14 and 97 DF	p-value: 7.261e-09

Source: Own creation

It can be seen that the p-value of the F-statistic is 7.261e-09, which is highly significant. This means that, at least, one of the predictor variables is significantly related to the outcome variable. The t-statistics evaluated whether or not there is a significant relationship between the independent and dependent variables, which occurs when the beta coefficient is different from zero. In the summary table above it can be seen that change in *log(funding amount)* is significantly associated with changes in the company's valuation. A moderately significant relationship can be observed for *log(number of investors)* and *log(number of employees)*. Additionally, there seems to be a strong relationship between the company’s valuation and its operation in the *cryptocurrency* subsector of the fintech industry. Relatively high significance is also observed for *blockchain technologies*. Other variables included in the model do not seem to have a strong relationship with the dependent variable.

The AIC of the estimated model equals 171.84, and BIC equals 218.208 which is a big improvement compared to the values of AIC and BIC for the complete model without any transformations and trimming where these were respectively 689.96 and 741,44. Much lower AIC and BIC values prove that logarithmic transformation improved how well the model fits the data. The fit of the model is also assessed by examining the residual standard error (RSE) and value of R-squared. The average difference between the actual responses and the values predicted by a regression model is shown by the RSE. In this case, RSE equals 0.4871 on 97 degrees of freedom, which is also a lower value compared to the model before transformations and trimming where RSE was 3.275 on 112 degrees of freedom. The R-squared value is rather low; 0.4871 compared to 0.8793 from the complete model before transformation and trimming. However, a low R-squared value in social sciences does not necessarily indicate a negative outcome, and it is rather often in the case of cross-sectional analysis (Wooldridge, 2018). What is more, R-squared does not define if the OLS regression is valid and should not be used to compare models (Wooldridge, 2018).

iii. Correlation tests

This test was performed to find out the nature of the correlation between the dependent and each independent variable and to assess whether this effect is significant. Thus, whether those variables move together or not. It is important to note that correlation does not prove causation. Therefore, it is necessary to exercise caution in interpreting the results and to consider other potential factors or variables that may affect the dependent variable.

Table 6 shows the sample estimate of the significant correlation coefficients between the independent variables and company valuation, together with the p-value of the correlation test.

Table 6 Significant correlation between variables

Variable	log(funding amount)	above avg revenue	log(funding rounds)	log(number of investors)	log(number of employees)
Correlation coefficient	0.5656485	0.2626231	0.2708965	0.3441226	0.2987824
P-value	6.663e-11	0.004951	0.003706	0.0001899	0.001307

Source: Own creation

As can be seen in Table 6, log(*funding amount), above-average revenue, log(funding rounds), log(number of investors)* and log(*number of employees)* have a significant positive correlation with the dependent variable; company valuation. The strength of the correlation varies between the pairs, where we can see different levels of correlation, strong and moderately strong,

A strong positive correlation between the funding amount and unicorn valuation, suggests that as the funding amount of a company increases, its valuation also tends to increase. This makes sense, as companies that have raised more funding often have more resources to invest in growth and expansion, which can increase their valuation, leading also to inorganic growth.

A weaker but still significant positive correlation is present between the dummy variable containing information on companies with revenue above the average and company valuation, with a positive correlation coefficient of 0.2602127 and a p-value of 0.004951. This aligns with the theory emphasizing the significance of revenue growth (Pride, 2018; Lee, 2019).

Moderately significant positive correlations were found between the rest of the variables. As for *log(funding rounds),* the correlation test result confirms that as a company goes through multiple rounds of funding its valuation increases. It suggests that the firm has demonstrated enough potential to attract investors more than once (Lassala & Ribeiro-Navarrete, 2022). This often means that the company has reached additional stages of development, which can increase its valuation and bring it closer to the maturity stage (Lassala & Ribeiro-Navarrete, 2022).

Pride (2018) affirms in "Unicorn Tears: Why startups fail and How to avoid it", that a too high number of investors could cause disharmony among the founders and thus be seen as a reason for the failure of unicorns, even though no statistical analysis it has been used to prove this belief, but rather qualitative, based on the author noteworthy experience in the VC industry. According to our findings, this is not the case, since a significant positive relationship between the company valuation and the number of investors was founded. This could be explained by the fact that having a higher number of investors in the company also brings more expertise and diversification within the board, consequently leading to better results.

Finally, a positive correlation between the number of employees and unicorn valuation would suggest that companies with a larger workforce tend to have higher valuations. This could be true because larger companies are often seen as more established and better positioned for growth and profitability, furthermore, an increase in the number of employees as the company grow is expected (Merzlova, 2022). However, also past research (Krch, 2018) found a low significance level of positive correlation between the two variables.

It is important to keep in mind that correlation does not imply causation. Just because these variables are positively correlated with unicorn valuation does not mean that the movement of one variable causes company valuation to vary. For instance, correlation might be influenced by a third variable correlation is present due to a shared relationship with an unrelated factor (Wooldridge 2018), which in this case it might be market conditions, competitive landscape, product differentiation, and leadership team etc. Overall, these findings suggest that there is a positive relationship between these variables and unicorn valuation, but further analysis would be needed to determine the strength and direction of the causal relationship between these variables and a company's valuation.

As previously mentioned, the results discussed above regard only the correlation coefficients between variables where the test exhibited a significant–value. Below are some additional considerations regarding the results of the remaining ones.

First of all, a more precise and reliable set of results could be obtained by conducting an analysis with time series data and a larger sample size. With regard to all other variables, their correlation coefficients are near zero and exhibit p-values that are excessively high.

Despite prior research indicating that a company's location can significantly impact its growth and development (Blank & Dorf, 2012; Stephens, 2019), it is surprising that the dummy variable pertaining to whether the company is located in California or not is insignificant, given that the test yielded a negative correlation coefficient of -0.008468467. Additionally, California is widely considered the premier location in the world for starting a business (Gornall & Strebulaev, 2017). Nonetheless, these findings have neither demonstrated a significant impact nor established their dependability. Finally, as previously mentioned we believe the main constraint regarding this correlation test is the sample size. However, this is mainly caused by the size limitations of the population of U.S. unicorns in the fintech sector.

iv. Diagnostic of the Complete Model

As previously introduced, the complete model is the one where all the variables including y and all x's have been transformed into logarithms (except for the dummy variables, since log (0) is undefined) and also where the main outliers have been eliminated through data trimming. It is important to note also that the variable B2C has been excluded from this model. According to the rule, one dummy variable must be omitted from the model to avoid multicollinearity. Initially, we decided to omit the dummy for *both B2B and B2C*, however, multicollinearity still occurred as most companies are either *B2B* or *B2C*, so another variable was excluded from the model leaving only *B2B*. In this part, tests are made to assess the validity of assumptions.

Starting from testing the linear regression assumption which states that the expected value of the error terms is zero. The residuals of this model have a mean of $-2.893764e^{-18}$ which is really close to zero. This is validated also through the t-test which shows a p-value of one, meaning that the null hypothesis of zero mean of u is not rejected. The graph shown below in Figure 20 is used to give valuable information regarding the OLS assumptions of linearity and correct specification.

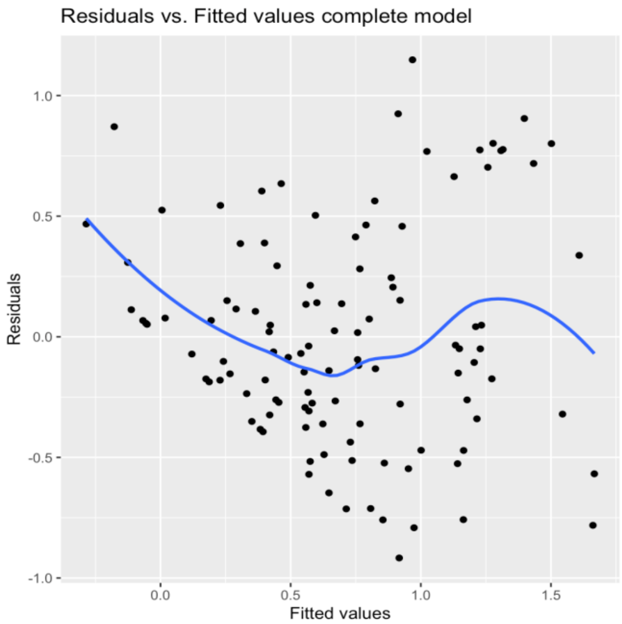

Figure 20 Residuals vs fitted values plot with a trendline for the Complete Model after transformation and trimming

Source: Own creation

As possible to see the residuals and fitted value plot doesn't follow a straight line but rather show a pattern that goes down, up and down again, this likely indicates the presence of non-linearity between the dependent variable and independent variables, as well as heteroscedasticity, which is confirmed after with the Breusch-Pagan test.

The consequences of these are upon the interpretation of the regression coefficients and the predictive performance of the model, but it doesn't necessarily mean that the model is entirely invalid. Since the plot shows a systematic effect of the residuals this can indicate that the assumption of correct specification is also not satisfied. The latter is also confirmed by a p-value of 0.009516 of the RESET tests, suggesting that there is significant evidence to reject the null hypothesis of correct functional form in favour of the alternative hypothesis that the functional form is incorrect (Wooldridge, 2018).

No perfect collinearity

Multicollinearity can cause standard errors of the regression coefficients to increase and reduce their precision, making it challenging to accurately identify the actual effects of each independent variable (Wooldridge, 2018).

By running the VIF function on R, the VIF value of each independent variable is less than the threshold of 5 which would indicate the presence of multicollinearity for that specific independent variable (Corporate Finance Institute, 2022). Table 7 shows the results of the VIF test for each independent variable in the Complete Model after transformation and trimming.

Table 7 VIF of the Complete model after transformation and trimming

log(Funding amount)	log(Year of foundation)	log(Funding rounds)
1.952290	1.568275	2.057787
log(Number of employees)	log(Number of investors)	Average
1.702262	1.871846	1.214049
Above	B2B	Online payment solutions
1.504096	3.150067	2.295600
Digital banking	Financial management	Personal investment
1.444851	4.012531	4.131162
Blockchain technologies	Cryptocurrency	California
2.171349	2.845781	1.137159

Source: Own creation

All values are above one, therefore, it is evidence that there is some correlation. As mentioned before the complete model is without the B2C variable, since their VIF results were above 5, thus over the threshold which indicates the presence of multicollinearity (Corporate Finance Institute, 2022). In Table 7 we can see that VIF does not exceed 5 for any of the variables.

Normality

The Jarque Bera test, used to assess whether the residuals are normally distributed shows a p-value of 0.1895, which is above the significance level of 0.05. Therefore, there is not enough evidence to reject the null hypothesis, thus, the error terms are normally distributed (Wooldridge, 2018). Underneath is possible to examine also the Normal Probability plot (QQplot) in Figure 21 which shows the residual points on the plot fall roughly along a straight line, meaning that the assumption of normality is likely to be met. Finding coherence with the outcome of the normality test.

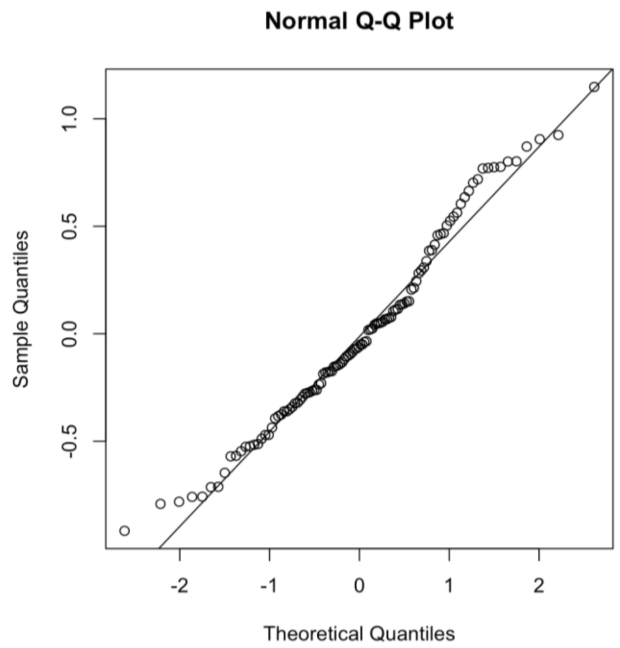

Figure 21 QQ Plot of the Complete Model after Transformation and trimming proving normality

Source: Own creation

Heteroscedasticity

The Breusch-Pagan is being used since the model has a relevant number of regressors (Wooldridge, 2018). The p-value is 0.008639, thus the null hypothesis that implies homoscedasticity is rejected, this indicates evidence of heteroscedasticity, or rather that the variance of the residuals is not constant. A consequence of this could be that the plot of the residuals might not be reliable but the estimation of the regression coefficient $\hat{\beta}_i$ should still be unbiased and consistent even in the presence of heteroscedasticity (Wooldridge, 2018). However, the variance estimator $var[\hat{\beta}]$, (which assumes homoscedastic errors) is biased and inconsistent, consequently leading to bias also on the standard errors and confidence intervals calculated for the regression coefficients. As a result, exact and asymptotic hypothesis testing based on this estimator is not valid (Wooldridge, 2018).

Log-transformation can fix heteroskedasticity, however, it did not resolve that issue in this case. To obtain trustworthy standard errors we applied appropriate correction using heteroskedastic-consistent standard errors. With the use of the HC3 estimator, we achieved homoscedastic standard errors with constant variance. Therefore, the OLS assumption of homoskedasticity is met and hypothesis testing is valid. The outcome of the correction applied to the *Complete Model* is presented in Table 8.

Table 8 The Complete Model coefficients after HC3 correction

t test or coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-117.764921	264.146581	-0.4458	0.656712	
log(Funding_amount)	0.451009	0.097511	4.6252	1.157e-05	***
log(Year_of_foundation)	14.238271	34.730636	0.4100	0.682737	
log(Funding_rounds)	-0.104107	0.121885	-0.8541	0.395129	
log(Number_of_investors)	0.152719	0.083446	1.8302	0.070298	.
log(Number_of_employees)	0.108454	0.072937	1.4870	0.140268	
Average	-0.111808	0.112710	-0.9920	0.323668	
Above	-0.072421	0.216003	-0.3353	0.738139	
B2B	0.054493	0.178732	0.3049	0.761106	
Online_payment_solutions	0.227674	0.266166	0.8554	0.394446	
Digital_banking	0.182140	0.232445	0.7836	0.435194	
Financial_management	0.289527	0.238855	1.2121	0.228401	
Personal_investment	0.263005	0.197371	1.3325	0.185803	
Blockchain_technologies	0.758287	0.257705	2.9425	0.004073	**
Cryptocurrency	0.736256	0.234260	3.1429	0.002219	**
California	0.070313	0.101663	0.6916	0.490828	

Source: Own creation

v. Backward elimination

Models containing many variables with small datasets may create misleading results or even lead to model overfitting. Backward elimination is a method used in order to improve multiple linear regression models by eliminating the least significant variables from the model one by one (Weisberg, 2014). The significance of the variable is assessed by its p-value. The regression summary of the complete model after logarithmic transformation and data trimming was presented in part ii of the methodology part of this thesis. That model was subject to backward elimination. Since our model was proven to have heteroskedasticity of the error terms, to obtain the model that meets the homoskedasticity assumption of the OLS method we applied the correction and used the *coeftest* function in R in order to obtain heteroskedasticity-consistent standard errors and trustworthy significance results. Elimination was conducted on the basis of the *coeftest* function output and started with the least significant variable with the highest p-value, which was the dummy variable for the *B2B* basis of the company operations with the p-value of 0.761106 (0.735190 before applying the correction, which was still the highest among all the variables). After eliminating one variable from the model, the summary of the new model was calculated once more and the heteroskedasticity correction was applied as well in order to detect the next least significant variable. We conducted ten rounds of such elimination in total, leading to a final model containing five explanatory variables. The most significant variables in the final model, for the given sample, proved to be: *log(funding amount)*, *log(number of investors)*, *log(number of employees)*, the dummy for the *cryptocurrency* subsector and the dummy for the *blockchain technologies* subsector. The summary of the *Final Model* is presented in Table 9.

Table 9 The regression summary of the Final Model

Residuals:	Min	1Q	Median	3Q	Max
	-0.92694	-0.35015	-0.04908	0.25466	1.27335

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-8.44550	1.14945	-7.347	4.22e-11	***
log(Funding_amount)	0.41284	0.06434	6.416	3.90e-09	***
log(Number_of_investors)	0.12778	0.06065	2.107	0.037478	*
log(Number_of_employees)	0.09108	0.04743	1.920	0.057485	.
Blockchain_technologies	0.51173	0.17395	2.942	0.004001	**
Cryptocurrency	0.50032	0.13269	3.771	0.000267	***

Residual standard error: 0.47 on 107 degrees of freedom	
Multiple R-squared: 0.4591	Adjusted R-squared: 0.4338
F-statistic: 18.16 on 5 and 107 DF	p-value: 5.173e-13

Source: Own creation

The highest p-value included in the *Final Model* equals 0.057485 and is represented by the *log(number of employees)* variable. The highest significance, therefore, and the lowest p-value is generated by the intercept of the model. The intercept represents the value of the dependent variable when all independent variables in the model are equal to zero (Wooldridge, 2018). In terms of independent variables, *log(funding amount)* has the highest significance with the p-value of 3.90e-09 which fits in the range [0, 0.001]. The R-squared value of the *Final Model* is 0.4591 which is a small decrease in comparison with the model before the backward elimination. The AIC and BIC values are respectively 157.86 and 176.96 and are the lowest values achieved so far which indicates that the *Final Model* has the best quality of all the models created for the purpose of this study.

vi. Diagnostic of the Final Model

The *Final Model* is derived from the backward elimination carried out upon the complete model which is trimmed from significant outliers and where both dependent and independent variables are transformed using the logarithmic function in order to satisfy the normality assumption of residuals. The independent variables left in the *Final Model* are *log(funding amount), log(number of investors),* *log(number of employees), blockchain technologies* and *cryptocurrency*. To verify the validity of this model, a diagnostic was carried out.

The mean of the residuals equals $4.433523e^{-17}$ and the p-value of the t-test on zero mean is one, as for the *Complete Model*. Figure 22 underneath shows the plot of the *Final Model* residuals and the fitted values.

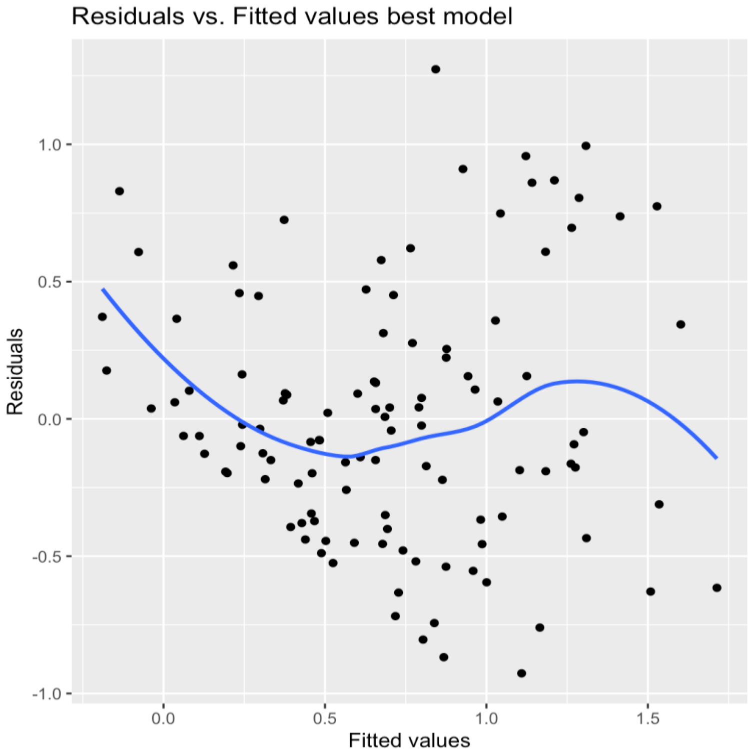

Figure 22 Residuals vs fitted values plot with trendline of the Final Model

Source: Own creation

The overall course of the line is the same as for the *Complete Model*, suggesting a possible non-linear trend in the relationship between the residuals and fitted values as well as the prospect of heteroscedasticity. The RESET test gives a p-value of 0.01449, slightly better than the one of the *Complete Model* (0.009516) but still not enough evidence in order to fail to reject the null hypothesis. Therefore, suggesting with significant evidence the presence of misspecification in the model, meaning that the linear regression model may not be a good fit for the data.

The consequences of these are upon the interpretation of the regression coefficients and the predictive performance of the model, but it does not necessarily mean that the model is entirely invalid. Since the plot shows a systematic effect of the residuals this can indicate that the assumption of correct specification is also not satisfied.

No perfect collinearity

In the model, there is no evidence of multicollinearity, since all the VIF values are below 1.4 which is considerably below five. In Table 10 below it is possible to see the outcome of the VIF and its value for each variable in the *Final Model*, which in comparison to the *Complete Model* in Table 5, they are noticeably lower.

Table 10 VIF of the Final Model

log(Funding_amount)	1.416087
log(Number_of_investors)	1.120145
log(Number_of_employees)	1.398278
Blockchain_technologies	1.134805
Cryptocurrency	1.151303

Source: Own creation

Normality

Figure 23 shows the Normal Probability plot of residuals of the *Final Model*, received as a result of backward elimination.

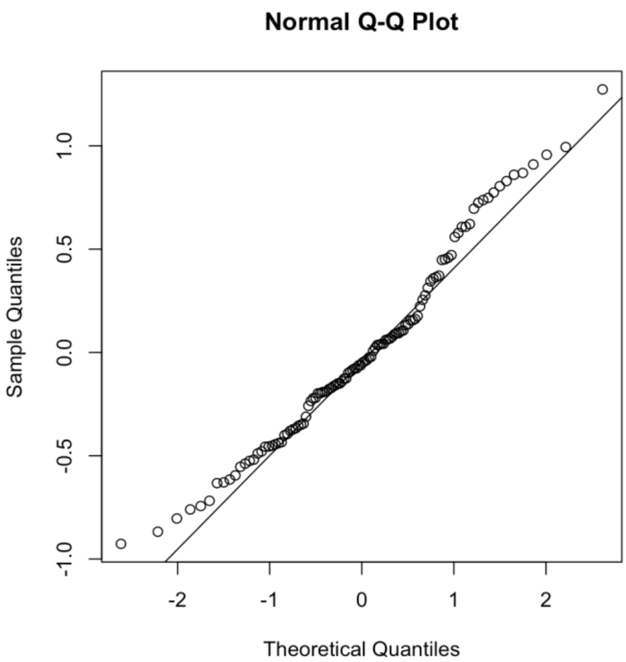

Figure 23 QQ plot of the Final Model

Source: Own creation

Most of the dots fell within the line, indicating that the residuals are normally distributed. This evidence is also proved by the Jarque Bera test with a p-value of 0.1424. Both the plot and the test allow us to confirm that the normality assumption of OLS has not been violated.

Heteroscedasticity

Despite the p-value of the Breusch-Pagan test being marginally better (0.008356) for the final model as compared to the complete model, it still falls below the conventional significance level of 0.05. This provides evidence to reject the null hypothesis that the error terms are homoscedastic. To obtain trustworthy standard errors we once again applied the heteroskedasticity-consistent standard errors correction. However, at this stage, in addition to the HC3 estimator, we also used HC0, HC1 and HC2 to compare the results, taking into account that the significance may be different for each estimator.

After applying the correction for heteroscedasticity in R using the *coeftest*, we got the following corrected coefficients presented in Table 11. Values in the brackets represent the standard errors and values above them represent the estimates.

Table 11 The Final Model's coefficients after heteroskedasticity correction

	log(Valuation)				
	OLS	HC0	HC1	HC2	HC3
log(Funding_amount)	0.413***	0.413***	0.413***	0.413***	0.413***
	(0.064)	(0.066)	(0.067)	(0.069)	(0.074)
log(Number_of_investors)	0.128**	0.128**	0.128**	0.128**	0.128**
	(0.061)	(0.055)	(0.056)	(0.058)	(0.061)
log(Number_of_employees)	0.091*	0.091*	0.091*	0.091*	0.091
	(0.047)	(0.051)	(0.053)	(0.055)	(0.059)
Blockchain_technologies	0.512***	0.512***	0.512***	0.512***	0.512***
	(0.174)	(0.155)	(0.159)	(0.167)	(0.180)
Cryptocurrency	0.500***	0.500***	0.500***	0.500***	0.500***
	(0.133)	(0.128)	(0.132)	(0.135)	(0.141)
Constant	-8.446***	-8.446***	-8.446***	-8.446***	-8.446***
	(1.149)	(1.150)	(1.182)	(1.213)	(1.282)

Observations: 113	
R-squared: 0.459	Adjusted R-squared: 0.434
Residual Std. Error: 0.470 on 107 degrees of freedom	
F Statistics: 18.162 (***) on 5 and 105 degrees of freedom	
Note: *p<0.1; **p<0.05; ***p<0.01	

Source: Own creation

vii. Results

From the results, it is possible to confirm that the variables with a strong significance are (in descending order) the *log(funding amount), cryptocurrency* and *blockchain technologies* subsectors of the fintech industry, followed by *log(number of investors)*. What is interesting, is that *log(number of employees)* had moderate significance in the case of the initial summary of the model, however after applying heteroskedasticity correction of the HC3 type, that variable was no longer significant. Therefore the variable is excluded from the final model containing only the most significant variables. There is insufficient evidence to suggest a significant impact of employee count on firm valuation, despite the proximity to significance. One possible explanation is that the fintech sector, characterized by innovation and technology intensity, does not necessarily require constant growth in employee numbers to achieve growth and results. Alternatively, there may be a threshold beyond which increasing the employee count has diminishing effects. Additionally, in the sample, there is a notable number of unicorns with high valuations despite having a relatively small number of employees, such as Carta (70 employees), Spotter (8 employees) and MobileCoin (61 employees) to name a few. In those cases, even though having a valuation between $1.7 and $10 billion, the number of employees they have is below 70.

The coefficient estimate for the variable *log(funding amount)* is 0.413. It indicates that for a 1% increase in funding amount, the estimated unicorn valuation is expected to increase by approximately 41% if all the other variables are held constant. Therefore, higher funding amounts received by fintech unicorns are greatly associated with higher valuations. The reason behind these results might suggest that unicorns that receive larger funding amounts are better perceived by the market and investors. Nevertheless, it might also be linked to the fact that funds help the company to be more competitive and cause both inorganic and organic growth. Furthermore, in the case of startup companies, the post-money valuation takes into consideration the gathered funding amount. Lastly, it must be noted that fintech was the sector with the highest funding amount among unicorns, amounting to $106.1 billion just in 2021 (Eckert, 2022).

It does not come too as a surprise that *blockchain technologies* and *cryptocurrency* subsectors are significant. In the past year, these particular subsectors of Fintech have been on the rise due to the crypto and blockchain trend that saw Bitcoin reaching a maxim evaluation of $56 thousand in November 2021 (Google Finance, 2023). The Bitcoin phenomenon like all the events where a rapid and major increase in valuation is present, brings attention and desire to speculate or invest in that particular phenomenon, generating a massive volume of new investments and possibly also causing overvaluation problems. It seems that these two particular specialities have shown a higher level of demand and growth compared to other subsectors such as digital banking or online payment solutions which were also in surge during the pandemic.

The significance of the *log(number of investors)* suggests that fintech unicorns with a larger number of investors tend to have higher valuations. The coefficient is 0.128, thus indicating that a 1% increase in the number of investors is associated with an estimated increase of approximately 12.74% in the response variable. As mentioned in the correlation test part, a plausible explanation is that a greater number of investors may contribute with additional expertise to the company, enabling them to make informed decisions and set the appropriate strategic direction, although this is just a possible explanation, and the correlations between two variables do not necessarily imply causation. A larger number of investors can also be directly related to more funds invested in the development of the company. However, this is not a rule as the amount of invested funds depends on the personal capabilities and preferences of the investors.

A surprising result is the fact that the California regressor proved not to be significant in the *Final Model*. Previous studies have shown that location in this particular area significantly influences business growth due to factors such as funding availability, networks, and labour quality (Blank & Dorf, 2012; Stephens, 2019). Despite California's reputation as the best place in the world for running a business (Gornall & Strebulaev, 2017), it may not hold the same advantage for fintech firms compared to other sectors. Moreover, the growing and more distributed presence of fintech ecosystems and favourable business environments in other locations across the country could also have an impact on the outcome of this result. Cities like Boston and New York have emerged as strong contenders with their own vibrant fintech ecosystems, offering access to specialized talent, financial institutions, and regulatory support.

Other variables, such as revenue range, year of establishment, funding rounds and B2B, also did not reach the required level of significance in the *Final Model*. The analysis also shows that the other sub-sectors included in the *Complete Model* are not significant enough to have an impact on valuation. As was previously stated, According to Lee (2019), the valuation of a startup may not accurately reflect its financial success or provide comprehensive insights. Furthermore, it is worth noting that certain high-valued companies have not yet generated profits, as highlighted by Lee (2019). Before conducting the analysis, we believed that this statement might just apply to some outliers in the world of unicorns.
These results might suggest that investors put greater emphasis on the future potential of a company, considering factors such as market opportunities, technological advancements, and competitive advantage, rather than solely focusing on the cash flows it is generating at present.

The year of foundation not being significant could indicate that the fintech sector reached maturity; this industry has experienced rapid growth and development in recent years, which could make the year of the foundation less relevant upon valuation impact. This may also be due to the fact that some companies have been able to achieve unicorn status in less than 5 years from the start of their operations, proving that disruptive innovation or blitzscaling strategy can take a company to the top faster than standard ways of development. Moreover, this result might suggest that being founded or doing business in a year where there are favourable economic conditions, does not necessarily imply the significance of this factor. However, we still believe in the importance of timing. A year of foundation may not be the right variable to verify this, since the timing factor may have a greater effect on a single company, rather than a specific sector. Nevertheless, as mentioned before a highly funded sector can significantly affect its valuation.

As mentioned in the data collection, a company that went through more funding rounds does not necessarily receive a higher amount of funding than a company that went through one to three rounds. It might be that the higher number of funding rounds means that the company is promising enough to attract more investors (Lassala & Ribeiro-Navarrete, 2022). However, according to our analysis of this specific sample characterized by U.S. fintech unicorns, there is not enough evidence to state that this has actually had an impact on their valuation, as affirmed by Lassala and Ribeiro-Navarrete in *Financing Startup*s (2022).

VII. Limitations and further research

Potential limitations of the research mainly concern the method of collecting the research sample. As mentioned in the data collection, the sample used for this study consists of no more than 150 U.S. fintech unicorn startup companies in total. Since 134 individuals make up our dataset, the sample cannot be regarded as completely random since the population may influence how generalizable the findings are. Additionally, the firms were selected from the CB Insights database and represent a single industry in a single nation, thus reducing the randomness. As mentioned above, the sample size was not large enough to ensure randomness, however, it is doubtful that the U.S. fintech sector will have many more unicorn companies in the future. Some companies will decide on IPOs, others on mergers, which will mean that they will no longer be unicorns, thus disappearing from the list provided by CB Insights. As a result, sampling and regression errors brought on by inadequate data will most certainly continue to be an issue in future studies.

In addition, during the diagnostics, evidence of model misspecification and non-linearity of parameters was observed, which not only violates the OLS assumption but also indicates potential challenges related to the accurate modelling of relationships between variables. When analysing the findings and deriving inferences from the study, these elements should be taken into consideration. Therefore, the findings should be evaluated carefully to account for any possible inaccuracies. The RESET test gave reason to believe that there was model misspecification in the study and that a non-linear model would have more power to explain the relationship between dependent and independent variables. Non-linear models are more complex than linear ones, and require more expertise from the researchers therefore due to time limitations we did not apply them in this thesis. However, with the use of non-linear RESET regression, the obtained estimates could prove to be properly specified and give trustworthy results of the test statistics.

One of the weaknesses of the data was its limited availability and reliability. Unicorns are private companies, which means that they are not required to disclose data as in the case of public companies. Therefore, variables such as the number of employees, the number of investors or the sum of funds collected found in the sources may differ from the true values. In addition, for example, the number of employees of different companies available in the sources may come from different years. This problem could be solved by using an analysis based on panel data rather than cross-sectional where values would be given over the years. However, reliable variable values would then have to be collected directly from the surveyed companies.

Despite some limitations, this study provides a comprehensive analysis of the U.S. Fintech Unicorn companies and key factors that affect their valuation.

VIII. Conclusion

In this paper we answer the question, what factors influence the valuation of unicorn startups in the U.S. fintech sector. For the purpose of this thesis, we created a data set composed of 129 observations of Unicorn Fintech startups from the US, collected from the World Unicorn Club dataset of Unicorns by CB Insights (2023). The final analysis was conducted on the sample of 113 companies as the data set contained outliers that had to be eliminated. For each company in the database, we collected data on their activities, both numerical and categorical data. The first linear regression model contained 16 variables, including 5 numerical and 11 binary dummy variables representing categorical data.

After the first estimation of the model, it was found that the residuals were not normally distributed. In order to obtain normality of residuals, 8 outliers were removed from the datasets and log transformation was applied to both dependent and independent variables excluding the dummies. Lastly, the B2C variable was excluded from the model since the VIF test proved that multicollinearity was still present. The correlation test found a strong significant positive correlation between funding amount, above-average revenue, funding rounds, number of investors and number of employees with the unicorn’s valuation.

From the diagnostic of the *Complete Model,* non-linearity of parameters, evidence of misspecification, and heteroscedasticity were observed. On the latter, the correction was applied. After backward elimination, the *Final Model* containing only the most significant variables proved to be: *log(funding amount), log(number of investors), log(number of employees), cryptocurrency,* and *blockchain technologies* subsectors. However, after applying the heteroskedasticity correction on the model *log(number of employees)* was no longer significant leading to the conclusion that the number of employees in an American fintech unicorn startup has no impact on the company's valuation. In other words, we cannot prove that having more employees leads to an increase in the unicorn’s valuation.

As expected from our hypothesis, funding amount turned out to be a strongly significant variable. Funding is of great importance for the development of the company, it is positively correlated with valuation and is additionally included in the post-money valuation formula. An interesting result is the high significance of the number of investors and the non-significance of the number of funding rounds. Before starting the study, we expected these variables would have a similar significance level, but this turned out to be an incorrect assumption. The result of linear regression suggests that the number of investors has a positive impact on the company's valuation. The fact that the company is able to attract a larger number of investors speaks very positively about it and at the same time a larger number of investors may provide more experience to the company, allowing them to make educated judgments and define the right strategic path. A greater number of investors might also be directly tied to more cash invested in the company's development.

Our study confirmed the high significance of variables representing sub-sectors of the fintech industry: cryptocurrency and blockchain technologies. That gives us evidence to believe that new technologies and future solutions for both companies and individual clients have a strong meaning and impact on the company's valuation, suggesting that companies dealing with such solutions achieve on average higher valuations. Additionally, the low significance of the revenue range may imply that investors place a larger emphasis on a company's future potential, taking into account aspects such as market prospects, technical improvements, and competitive advantage, rather than just focusing on the cash flows it generates now.

The research on the basis of this master's thesis was carried out with care for compliance with available literature and methods. Potential threats to the reliability of the regression result, such as violations of the OLS assumptions, were eliminated using appropriate methods so that the final results were as reliable as possible. Nevertheless, it should be noted that the sample was not created randomly, contains outliers, is rather small and additionally the linear assumption of OLS was violated which means that the model cannot be used for predictions and the conclusions presented in this thesis should be treated with caution. However, given the small number of similar studies regarding this sector, our focus on the relation between the valuation and quantitative and qualitative variables constitutes a new contribution to the field.

Bibliography

Ahmad, I. (2022). Which unicorn companies reached $1 billion the fastest? Digital Information World. Retrieved February 4, 2023, from https://www.digitalinformationworld.com/2022/04/how-long-does-it-take-for-startup-to.html

Alharahsheh, H.H. and Pius, A., 2020. A review of key paradigms: Positivism VS interpretivism. Global Academic Journal of Humanities and Social Sciences, 2(3), pp.39-43. Retrieved May 3, 2023, from https://gajrc.com/media/articles/GAJHSS_23_39-43_VMGJbOK.pdf

Astivia, O. L. O., & Zumbo, B. D. (2019). Heteroskedasticity in Multiple Regression Analysis: What it is, How to Detect it and How to Solve it with Applications in R and SPSS. Practical Assessment, Research, and Evaluation, 24, Article 1. https://doi.org/10.7275/q5xr-fr95

Baldridge, R., Curry, B., (2022), What is a startup? The Ultimate Guide, Forbes Advisor. Retrieved January 27, 2023, from https://www.forbes.com/advisor/business/what-is-a-startup/

Baltrusaitis, J. (2022). Revealed: The U.S. has twice as many unicorns as China and India combined. Finbold. Retrieved February 18, 2023, from https://finbold.com/revealed-the-u-s-has-twice-as-many-unicorns-as-china-and-india-combined/

Becker, R. (2014). What is a minimum viable product (MVP)? - definition from Techopedia. Techopedia.com. Retrieved January 30, 2023, from https://www.techopedia.com/definition/27809/minimum-viable-product-mvp

Benoit, K. (2011) Linear Regression Models with Logarithmic Transformations. Methodology Institute, London School of Economics

Bera, A. K., Suprayitno, T., & Premaratne, G. (2002). On some heteroskedasticity-robust estimators of variance–covariance matrix of the least-squares estimators. Journal of Statistical Planning and Inference, 108(1-2), 121-136.

Berger-de León, M., Königsfeld, J., Leypoldt, L., & Vollhardt, K. (2022, October 19). How to build a unicorn: Lessons from venture capitalists and start-ups. McKinsey & Company. Retrieved February 15, 2023, from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-unicorn-lessons-from-venture-capitalists-and-start-ups

Blank, S., Dofr, B. (2012). The Startup Owner's Manual: The Step-By-Step Guide for Building a Great Company. K & S Ranch

CB Insights (2023) The complete list of Unicorn Companies, Retrieved March 1, 2023 from https://www.cbinsights.com/research-unicorn-companies

CB Insights, The Complete List Of Unicorn Companies. (2022). Retrieved January 2023, from https://www.cbinsights.com/research-unicorn-companies.

CB Insights. (2017). The United States of Unicorns: Every US company worth $1B+ on one map. CB Insights Research. Retrieved February 21, 2023, from https://www.cbinsights.com/research/startup-unicorns-us-map/

CB Insights. (2022). New unicorns are plummeting. Here’s how volatile markets and shrinking valuations are shifting power from founders to investors. CB Insights Research. Retrieved February 5, 2023, from https://www.cbinsights.com/research/unicorn-company-slowdown/

CB Insights. (2023, January 11). List of unicorn startups & markets: CB insights. CB Insights Research. Retrieved January 27, 2023, from https://www.cbinsights.com/research/unicorn-startup-market-map/

CFI Team. (2022) Unicorn. Retrieved May 5, 2023 from https://corporatefinanceinstitute.com/resources/valuation/unicorn/

Checklist: How to Become a Unicorn Startup in 2023. (2023, January 2). Retrieved January 28, 2023, from https://www.embroker.com/blog/unicorn-startup-checklist/

Choy, L. (2021) Global M&A by the numbers: 2020 recap, S&P Global Market Intelligence. Retrieved February 20, 2023, from https://www.spglobal.com/marketintelligence/en/news-insights/blog/global-ma-by-the-numbers-2020-recap

Congress. (2012). H.R.3606 - jumpstart our business startups. Retrieved February 7, 2023, from https://www.congress.gov/bill/112th-congress/house-bill/3606

Corporate Finance Institute (2022). Variance inflation factor (VIF). Retrieved April 19, 2023, from https://corporatefinanceinstitute.com/resources/data-science/variance-inflation-factor-vif/

Cremades, A., (2019). 8 Types of Investors for Startups. Forbes Magazine, Retrieved February 2, 2023, from https://www.forbes.com/sites/alejandrocremades/2019/01/02/8-types-of-investors-for-startups/?sh=3dc035174a3e

Cumming, D. and Dai, N., 2011. Fund size, limited attention and valuation of venture capital backed firms. Journal of Empirical Finance, 18(1), pp.2-15.

Davidson, R. and MacKinnon, J.G. (2003) Econometric theory and methods. New York: Oxford University Press.

Diggival, A., (2022). Startup Ecosystem: Is it worth? BlueRose Publishers, p. 1-3

Dotzel, T. and Shankar, V., 2019. The relative effects of business-to-business (vs. business-to-consumer) service innovations on firm value and firm risk: An empirical analysis. Journal of marketing, 83(5), pp.133-152.

Eckert, V. H. (2022). Living in a world of unicorns. PwC. Retrieved February 1, 2023, from https://www.pwc.com/gx/en/about/new-ventures/unicorns-five-trends.html

Erdogan, B., Kant, R., Miller, A., Sprague, K., (2016) Grow fast or die slow: Why unicorns are staying private. McKinsey & Company, Retrieved February 2, 2023 from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/grow-fast-or-die-slow-why-unicorns-are-staying-private

Fahimi, F. (2022). Where are unicorns founded & where do they move to? Startup Genome. Retrieved February 22, 2023, from https://startupgenome.com/articles/where-are-unicorns-founded-and-where-do-they-move-to-1

Farandea, A., (2021). Why unicorn don't do IPO yet? Retrieved February 2, 2023, from https://www.linkedin.com/pulse/why-unicorn-dont-do-ipo-yet-hafizh-farandea/

Feyen, E.H.B., Natarajan, H., Saal, M., (2023) Fintech and the Future of Finance : Market and Policy Implications (English). Washington, D.C. : World Bank Group.

Fink, A. (2020). Barely any European unicorns: What can the Eu do? BARELY ANY EUROPEAN UNICORNS: WHAT CAN THE EU DO? Retrieved February 22, 2023, from https://en.irefeurope.org/publications/online-articles/article/barely-any-european-unicorns-what-can-the-eu-do/

Fundz (2022). Series A, B, C Funding: Averages, Investors, Valuations. Retrieved February 1, 2023, from https://www.fundz.net/what-is-series-a-funding-series-b-funding-and-more#intro

Goddard, W., & Melville, S. (2001). Chapter 5: Research design. In Research methodology: An introduction (2nd ed., pp. 32–34). Juta and Company Ltd.

Google Finance (2023) Bitcoin (BTC) price, real-time quote & news, Google Finance. Retrieved May 21, 2023 from: https://www.google.com/finance/quote/BTC-EUR?sa=X&ved=2ahUKEwiEg7DAjYz_AhUbgP0HHfXOAyIQ-fUHegQIBhAf

Gornall, Will and Strebulaev, Ilya A. (2017). Squaring Venture Capital Valuations with Reality . Journal of Financial Economics (JFE), Forthcoming. Retrieved March 1, 2023 from https://ssrn.com/abstract=2955455 or http://dx.doi.org/10.2139/ssrn.2955455.

Graham, P. (2012). Startup = Growth. Retrieved from http://www.paulgraham.com/growth.html

Gross, B. (2015) The single biggest reason why start-ups Succeed, Bill Gross: The single biggest reason why start-ups succeed | TED Talk. Retrieved March 1, 2023 fromhttps://www.ted.com/talks/bill_gross_the_single_biggest_reason_why_start_ups_succeed/transcript

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. Handbook of qualitative research, 2(163-194), 105.

Hofweber, T. (2020) Logic and ontology, Stanford Encyclopedia of Philosophy. Stanford University. Retrieved March 5, 2023, from https://plato.stanford.edu/entries/logic-ontology/ .

Howe, N. (2015). What's feeding the growth of the billion-dollar 'Unicorn' startups? Forbes. Retrieved January 28, 2023, from https://www.forbes.com/sites/neilhowe/2015/03/18/whats-feeding-the-growth-of-the-billion-dollar-unicorn-startups/?sh=57be08a52654

Ilya Strebulaev, Venture Capital Initiative, Stanford Graduate School of Business (2022)
IMF-World Bank., (2019). “Bali Fintech Agenda: Experience thus Far.” IMF and World Bank, Washington D.C.

Khushali, S. (2022, July 2). What is a unicorn startup and how to be one? Razorpay Learn. Retrieved January 28, 2023, from https://razorpay.com/learn/what-is-a-unicorn-startup/#Features-of-a-unicorn-startup

Kissell, R., Poserina, J. (2017) Chapter 2 - Regression Models, Optimal Sports Math, Statistics, and Fantasy, Academic Press, Pages 39-67, ISBN 9780128051634,

Krch, P., 2018. Relationship between the company size and the value: empirical evidence. In The Impact of Globalization on International Finance and Accounting: 18th Annual Conference on Finance and Accounting (ACFA) (pp. 145-152). Springer International Publishing.

Kuratko, D., Harrison,L.,& Neubert,E.(2020). Blitzscaling : The good, the bad and the ugly. Business Horizons, 63(1), pp 109-119.

Lassala, C. & Ribeiro-Navarrete, S. (2022) “Investment cycles,” in Financing startups: Understanding strategic risks, funding sources, and the impact of emerging technologies. Cham: Springer International Publishing.

Lee, A. (2013), Welcome To The Unicorn Club: Learning From Billion-Dollar Startups. Tech Crunch, Retrieved March 5, 2023, from https://techcrunch.com/2013/11/02/welcome-to-the-unicorn-club/

Lee, A. (2019). The (Unreality) of The Unicorns: UK tech sector creates one Unicorn startup per month. Newstex.

Legg, C. and Hookway, C. (2021) Pragmatism, Stanford Encyclopedia of Philosophy. Stanford University. Retrieved March 25, 2023, from https://plato.stanford.edu/entries/pragmatism/

Lien, D. and Balakrishnan, N. (2005) On Regression Analysis with Data Cleaning via Trimming, Winsorization, and Dichotomization, Communications in Statistics - Simulation and Computation, 34:4, 839-849, DOI: 10.1080/03610910500307695

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. The American Statistician, 54(3), 217-224.

Lütkepohl, H. and Krätzig, M. (eds) (2004) Applied Time Series Econometrics. Cambridge: Cambridge University Press (Themes in Modern Econometrics). doi: 10.1017/CBO9780511606885.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of econometrics, 29(3), 305-325.

McKinsey & Company (2022). McKinsey Technology Trends Outlook 2022 Report

McKinsey. (2020). Europe's start-up ecosystem: Heating up, but still facing challenges. McKinsey & Company. Retrieved February 21, 2023, from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/europes-start-up-ecosystem-heating-up-but-still-facing-challenges

Merzlova, K. (2022). Startup Development Stages: 7 Steps to Become a Unicorn. SumatoSoft. Retrieved January 30, 2023, from https://sumatosoft.com/blog/startup-development-stages-7-steps-to-become-a-unicorn

Metrick, A., & Yasuda, A. (2011). Venture Capital & the Finance of Innovatin. John Wiley & Sons, Inc.

Nazir, A. and Tbaishat, D., 2023. The Impact of Funding on Market Valuation in Technology Start-Up Firms: implication for open innovation. Journal of Open Innovation: Technology, Market, and Complexity, p.100028.

Pala, S. (2022). Why does Europe not produce enough unicorns compared to the size of its economy? LinkedIn. Retrieved February 20, 2023, from https://www.linkedin.com/pulse/why-does-europe-produce-enough-unicorns-compared-size-serhat-pala/?trk=pulse-article_more-articles_related-content-card

Pless, J. (2022). Europe births tech unicorns - only to see them leave. CEPA. Retrieved February 20, 2023, from https://cepa.org/article/europe-births-tech-unicorns-only-to-see-them-leave/

Pride, J. (2018). Unicorn tears: Why startups fail and how to avoid it. Wiley.

Rehman, A. A., & Alharthi, K. (2016). An introduction to research paradigms. International Journal of Educational Investigations, 3(8), 51-59.

Ries, E. (2011). The lean startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. Crown Business.

Rodriguez, S. (2015) “The Real Reason Everyone Calls Billion-Dollar Startups 'Unicorns',” International Business Times. Retrieved January 24, 2023 from https://www.ibtimes.com/real-reason-everyone-calls-billion-dollar-startups-unicorns-2079596

Ron Carucci(2021). Midsize companies, shouldn't confuse growth with scaling. Harvard Business Review. Retrieved January 30, 2023, from https://hbr.org/2016/07/midsize-companies-shouldnt-confuse-growth-with scaling#:~:text=Growth%20means%20adding%20revenue%20at,much%20greater%20rate%20than%20cost.

Saunders, M., Lewis, P., & Thornhill, A. (2019). Chapter 4: Understanding research philosophy and approaches to theory development. Research Methods for Business Students (8th ed., pp. 152–159). Pearson Education Limited.

Startup Commons, (2019). What Is Startup Ecosystem? Retrieved January 29, 2023, from https://www.startupcommons.org/what-is-startup-ecosystem.html, StartupBlink. (2022). An analysis of Startup Ecosystem Funding Per Region. LinkedIn. Retrieved February 22, 2023, from https://www.linkedin.com/pulse/analysis-startup-ecosystem-funding-per-region-startupblink/?trk=organization-update-content_share-article

Statista. (2022). Distribution of unicorns by region 2022. Retrieved February 1, 2023, from https://www.statista.com/statistics/1092625/share-of-unicorns-worldwide-by-region/

Statista. (2023). Fintech - United States, rRetrieved April 10, 2023 from https://www.statista.com/outlook/dmo/fintech/united-states

Stephens, B., Sibley Butler,J., Garg,R., Gibson, D. (2019) Austin, Boston, Silicon Valley, and New York: Case studies in the location choices of entrepreneurs in maintaining the Technopolis, Technological Forecasting and Social Change, Volume 146,2019, p.267-280,

Steup, M. and Neta, R. (2020) Epistemology, Stanford Encyclopedia of Philosophy. Stanford University. Retrieved March 25, 2023, from https://plato.stanford.edu/entries/epistemology/

Teare, G. (2020). Unicorns get more magical: 2020 cohort shows huge increase in valuation on IPO day. Crunchbase News. Retrieved February 2, 2023, from https://news.crunchbase.com/public/unicorns-get-more-magical-2020/

Teleghani, A., Teleghani, M, & Azizi, J. (2022). Startup Ecosystem
Weisberg, S. (2014). Applied Linear Regression (4th ed., pp. 240-243). John Wiley & Sons, Inc.

Willson, N. (2022). What is a Startup Incubator? Its Types And Services, Retrieved January 29, 2023, from https://www.topmba.com/blog/what-startup-incubator, accessed 29.01.2023

Wooldridge, J.M. (2018) Introductory econometrics: A modern approach. Boston, MA: Cengage.

Yang, J. (2012) Interpreting regression coefficients for log-transformed variables. Cornell Retrieved May 12, 2023 from https://cscu.cornell.edu/wp-content/uploads/83_logv.pdf