Politecnico di Torino

Master Degree in Engineering and Management
A.a. 2022/2023
Sessione di Laurea 12/2023

Immigration and female entrepreneurship in Lombardy region

Can the presence of low-skilled immigrants, who provide home and children care, improve the chances of survival of the female-led innovative startups?

Relatrice:
Anna D'Ambrosio

Candidata:
Maria Laura Pupo

Abstract

This thesis has the scope to check if the level of low-skilled immigrants presents in the specific area of the Italian region of Lombardy has an effect on the survival's probability of the female-led startups founded in that area, starting from the idea that the low-skilled immigrants can provide house and children care services and permits the women entrepreneurs to dedicate more time to their business.

The first stage provides the study of the prior literature about the factors that can influence the startup's probability of survival, to find all variables, different from the level of immigrants, which must be considered in the statistical model's construction, with a specific focus on the factors related to the gender's difference in the entrepreneurial world.

Secondly, the statistical model is constructed, first of all the data are collected from two different databases, the Innovative Startup Register linked to AIDA, that provides the information about the startups founded on the Lombardy's territory , and the database downloaded from ISTAT, that provides the level of immigrants in each Lombardy's municipality, in particular it is taken in consideration the level of women aged between 25 and 65 with a nationality different from the Italian's one. All this information is used in the construction of the linear regression that has the duration of the startup as dependent variable and the level of immigration as independent variable, and other information about the startup is used as control variables. Different variants of the first basic linear regression are carried out to increase the level of fit of the model and to analyze the problem from different points of view.

Contents

Index of Tables and Figures

1. Introduction

The role of women in society is one of the main topics faced in these times. Without doubt the role of the woman in the society is changed a lot across the last decades, reducing the differences between the two genders under different points of view. In Italy, the territorial focus of this thesis, across the years society reshape the figure of women, giving her higher rights and independence and increasing her opportunities about her life choices. Although many steps have been taken, the road to gender equality is still long, some form, voluntary or not, of discrimination of women is still evident in society, in different fields. One field where there is still a certain level of inequality is the labor market, even it is impossible to deny that the women's participation rate to the labor force shows an increasing trend, it is also true that in the job's word demonstrates a still large discrepancy between two genders, both in terms of opportunities and remuneration. This gender gap on the workplace can be attributed to different factors related to society and the way in which it see the woman and her social role, in particular, the problem regard the way in which the society see the woman as *wife* and *mother*, and her responsibility inside home, that often, not only is disproportionated to the man's one, but can be in contrast with certain type of work careers. Different studies take to consideration and analyze deeply this concern, demonstrating how the hours dedicated to the house and children care by women are more than hours dedicated by men, and this is strongly connected to lower working hours supplied by women leading into frequent career interruptions and lower wages for women respect to men. Some researchers studied what are the factors that can change the working hours supply by women, decreasing the hours dedicated to the house and children care, and many of these found that there is a positive relationship between the level of *immigration of low skilled workers* in a certain territory and the working hours supply by women. The term "low skilled workers" refers to workers that perform jobs for which they do not need a studies' title. The basic concept is that this type of immigrant workers offers house and children care services to local women that, in turn, increase their hours dedicated to paid work. This concept was analyzed by the prior literature under different points of view, for instance, it was analyzed the measure in which it impacts and changes across different occupations, or what are the consequences in terms of gender pay gap. The scope of this thesis is to introduce that concept in a specific sector: the one of *innovative startups*, specifically, the purpose is to understand if the results of the previous research about the concern can be applicable to the word of startups. The

question to which this work tries to answer is: from the moment that there evidence that women, in particular mothers, tend to dedicate more hours then men to the house and children care, and that the low skilled immigrants tend to offer services for the house and children care, and there are also evidence that the two facts can influence each other for many occupations, it is possible that the same evidence influence the female-led startups' survival? If the women that decide to found a startup can dedicate more time to their startup, the startup has more probability of survival? And this increase in time dedicated to the startup can be reconducted to an increase in the low skilled immigrants' level in a specific territory? To carry out the analysis it is decided to focus the research on the entire territory of Lombardy, at the municipal level, this is not a random choice, but it is due to will the fact that Lombardy is the one of the Italian region with a higher number of startups, so it permits to have a higher number of available data.

The paper is divided in that way, Chapter 2 consists of a collection of the prior literature about the startups, in particular, it is provided a description of an innovative startup and then the factors that can influence the startups' survival probability are analyzed. Moreover, in this Chapter it will provide an overview about women in entrepreneurship, what is the participation rate is and what are the main reasons for which women decide to become entrepreneurs, and how these reasons can affect their startups' probability of survival. In the last part of the Chapter, it is presented an analysis of the available papers that faced the theme of the influence of the low skilled immigrants in a certain geographic area on the women's working hours supply. Chapter 3 explains what is the empirical method used for the research. Chapter 4 illustrates the empirical application, namely, the data collection, the formulation of the hypothesis and the creation of the statistical instrument. Chapter 5 reports the results obtained by the analysis. Finally, Chapter 6 and Chapter 7 show the discussion and the conclusion of the research and the limitations to resolve for a possible future implementation of that research.

2. Background Literature

2.1 Innovative Startup definition

2.1.1 General overview

To begin with, it is necessary to frame the concept of Innovative Startup. In literature does not exist a unique and completely shared definition of Innovative Startup, the world of research is divided on both the definition of Startup and the adjective Innovative, the last one is heavily related to the regulatory aspect, for this reason, it is necessary to provide a general framework about the different policy initiatives of the different countries, with a particular focus on the Italian case. Different studies attempt to give a definition with different approaches, it is concluded that there is not a fully shared definition. All definitions available have some common traits. It is possible to divide the areas taken into consideration by the literature into three macro-categories of characteristics: company characteristics, such as size, age, and growth rate, technology characteristics, linked to the product and the scope of the business, words such 'new', 'newsness' and 'innovation' frequently appear, and management characteristics, related to the team.

In addition, research on this topic often underlines a strong link between the nature of the startup and the high level of risk.

To follow a collection of some definitions.

Source	Definition
Dizionario di Economia e Finanza Treccani	Initial start-up phase of a new business, a newly created business or an enterprise that has just been listed on the stock exchange.[...]The first economic meaning mainly referred to the initial phase of a new company born in the internet or information technology sector.

Investopedia	The term startup refers to a company in the first stages of operations. Startups are founded by one or more entrepreneurs who want to develop a product or service for which they believe there is demand.
European Commission	Technological entrepreneurship, digital market, services in the field of websites and ICT
La Repubblica	1. launch, start-up, take-off of a new business, enterprise or similar; 2. newly established or newly listed undertaking.

Table 1: Startup definition

According to the conclusion of Skawinska and Zalewski (2020) it is possible to assume that a startup is a young, small, independent enterprise, that is creative, innovative, conducting research and development activity (R&D) to solve actual problems, and proposing prospective solutions, striving for talented employees, and sales growth, with an attractive business model.
Adding that it operates in a high-risk environment.

As said before, the concept of 'innovative' is strictly correlated to the regulation and the policy initiatives implemented by each country.
It is possible to identify different approaches (Colombelli et al., 2020) shown in Table 2.

Approach	Definition
New firms	New firms are innovative, assuming that entrepreneurship in general is an intrinsic source of innovation (Schumpeterian).
Self-declaration	A company is considered innovative when it declares itself basing on innovative

	character of the entrepreneurial project. The declaration can be verified by an independent operator or by the government.
Sector specific	A firm is innovative if it belongs to a certain sector, typically a high-tech sector.
Growth oriented	Growth orientation is considered synonymous with innovation.
Partner specific	Affiliation with certain partners is used to identify innovative startups

Table 2: Definition of 'innovative'

2.1.2 Italian framework

Important for the scope of this work is to identify the Italian framework.

The definition of Innovative Startup provided by Ministero delle Imprese e del Made in Italy is "The innovative startup is a young, high-tech company with strong growth potential and is therefore one of the key points of Italian industrial policy."

The Italian regulation uses a Self-declaration approach to recognize a startup as innovative, in 2012 it is introduced the Italian Startup Act, with the scope of introducing specific measures to support innovative startups during their life cycle till the reaching of maturity and stipulate a set of requirements that the company needs to meet to be considered innovative and so to have access to the benefits.

The requirements are the following (Ministero delle Imprese e del Made in Italy):

1. It is a new company or one established no more than 5 years ago.

2. Its headquarter is in Italy, or in another EU/EEA Member State (provided a production facility or a branch in Italy).

3. Annual turnover is lower than €5 million.

4. It is not listed on a regulated market or multilateral trading platform.

5. It does not distribute its profits.

6. It has as mission the development, production and commercialization of innovative products or services with a clear technological component.

7. It is not the result of a company merger, split-up, or of a business or branch transfer

Moreover, at least one of the three criteria must be met :

1. R&D expenditure corresponds to at least 15% of the higher value between turnover and annual costs.

2. The total workforce includes at least 1/3 of PhDs, PhD students or researchers, or at least 2/3 of the team holding a master's degree

3. The company is the owner or licensee of a registered patent (or it has filed an application for an industrial property right) or it owns an original registered software.

2.2 Factors that influence the survival of a startup

This section is an analysis of the factors that influence the survival of the innovative startup. With “startup survival” it means the ability of a startup to grow and pass from the status of startup to a firm.

From the literature it is possible to find four macro-categories in which different factors can be categorized, they are reported in the table.

Category	Description
Company factors	All characteristics related to the nature of the company and its founders.
Innovation sustainability factors	Components that can influence the level of innovation of the startup.
Financing factors	The financing status of the company and characteristics related to its investors.
Industry factors	Characteristics of the external environment in which the startup runs its business.

Table 3: Macrocategories of factors

It is important not to consider these categories as completely independent of each other, but they are strictly linked and influence one another, for this reason in the following examination, some factors are taken into consideration more than once and are analyzed under different points of view according to the category considered.

2.2.1 Company factors

In his intervention, Gross(2015) (youtube) shows what are the factors that have the major influence on the success of a startup, based on research on 200 companies, and he ranks the first five factors in this way:

1. Timing
2. Team
3. Idea
4. Business model
5. Funding

In literature, this is one of the broadest classifications of factors, other authors, who will be mentioned later, introduced other factors, but they can be seen as part of one of these five, moreover, the researchers often disagree about the ranking by weight of the success factors taken in consideration. Berkus (2006) identifies five key factors a bit different from Gross' ones, they are the following:

1. Idea
2. Founding Team
3. Having a Functional Prototype
4. Strategic relations
5. Invoicing

Contrary to Gross' idea, Berkus attributes the same weight to all of these.

The following table comes from the study of Sevilla-Bernardo et al. (2022), it is useful to compare the factors identified by Berkus and by Gross.

Berkus(2016)	Weight (%)	Gross(2015)	Weight (%)
Idea	20	Timing	42
Founding team	20	Team	32
Functional Prototype	20	Idea	28
Strategic Relations	20	Business Model	24

Invoicing	20	Funding	14

Table 4: Comparison between factors identified by Berkus and Gross

To understand as well as possible these factors classification it is helpful to identify five questions to answer about startup structure, they are reported in the following table.

Question	Aswer
When	Timing
Who	Human Capital characteristics
Why	Idea
Where	Industry factors
How	Financing factors

Table 5: Questions to answer about startup structure

Based on the conclusion of Gross's study, because they are the more general and more suitable to the goal of this work, and on the question subdivision in the table, the first three elements will be analyzed and in turn divided into factors that are influenced by, the last two are analyzed in section 2.2.3 and 2.2.4.

1. When

"The best way to really assess timing is to really look at whether customers are really ready for what you have to offer them." These are the words of Gross about timing. "When" the startup decides to launch its product on the market represents a key success factor for the business. It is necessary to have a good idea and a good team but it is not sufficient. It is needed to have a ready market, the supply has to be synchronized with the demand.

2. Who

"Who" drives a startup during its entire life cycle plays a key role in the success of the business.

A team of individuals that decides to launch a startup must have certain attributes to reach success, in literature different experts try to provide a set of these characteristics, in all of

these studies the focus is mainly on two players: the founder or entrepreneur and the team in general.

Soft skills

Tasnima et al.(2014) identify three main soft skills that the founder has to own to start a successful business:

1. Passion
2. Values
3. Personality

Not only Tasnima et al., but most of the researchers in the field put passion first in the list of key soft skills owned by an entrepreneur who launches a startup.

Cordon et al.(2009) define passion as "consciously accessible intense positive feelings experienced by engagement in entrepreneurial activities associated with roles that are meaningful and salient to the self-identity of the entrepreneur".

1. ***Passion***

The importance of the entrepreneurial passion for the survival and the success of a business is in contrast with its intangible and unmeasurable nature, and this makes it difficult to study its influence. Trying to provide a more tangible idea of passion, Cordon et al. (2013) proposes three main aspects that characterize it:

1. The involvement of intense positive feeling that drives the commitment to reach the entrepreneur's goals.
2. The entrepreneur turns these feelings into activities to construct his/her entrepreneurial identity.
3. Entrepreneurial passion is focused on three different fields: inventing new products, founding new firms and developing firms for success. An entrepreneur can have passion for one or more of them.

As said before, it is very difficult to measure in a mathematical and objective manner the passion, but it is still true in the face of a large number of young entrepreneurs that launch a business only a small part of these survive, the decisive factor may be the entrepreneurs' passion (Mayoshe and Nuringsih (2021)). In their work, Mayoshe and Nuringsih (2021) try to sustain this idea by tasting the hypothesis that emotional support, perceived

competence and entrepreneurship education can positively influence passion. They arrive at the result that the emotional support from the external environment increases the passion of the entrepreneur and then the commitment to reach the goals. The perceived competence increases the entrepreneur's self-confidence and permits them to take more opportunities increasing the probability of business survival. Moreover, even entrepreneurship education can stimulate passion and it can be formal education or come from previous experiences.

From these studies it is possible to conclude that *passion* of an entrepreneur who wants to launch a startup plays a key role in the survival of the business.

2. Values

The second soft skill indicated by Tasnima et al. is Values.

As a formal definition, values are personal beliefs on which an individual bases her/his behavior and her/his choices. From the point of view of entrepreneurship, values can be seen as the pillars of the strategic choices undertaken by the entrepreneur both in starting and running the business.

The literature shows that the entrepreneur's values influence his/her acts entrepreneurially (Tasnima) and as a consequence, they influence the life cycle of the startup and its survival.

By analyzing different studies it is possible to conclude a list of the four main values that can drive the startup toward success, it is the following:

1. **Mission driven**, the entrepreneur must be a person committed and focused on his/her goal.
2. **Creativity**, as it will be argued later a key role, the entrepreneur of an innovative startup needs to be able to create something that nobody has created before.
3. **Risk taking**, as said in different points in this work risk is one of the key elements that characterize a startup because of its nature, the entrepreneur needs to be available to undertake risks to pick the opportunities.
4. **Communication**, when an entrepreneur decides to launch a startup he/she has to relate to different actors, he/she has to be able to communicate with his/her team and employees, to keep them focused on the mission, with investors, to

obtain external financing and with all external individuals or entities that can influence the survival of the startup.

3. Personality

The last soft skill is Personality.

It is possible to define personality as “the enduring dispositions that cause characteristic patterns of interaction with one's environment and is related to a certain degree to inheritance and genetic factors” (Tasnima et al.(2014)). In other words, personality is a set of behavioral traits that influence the choice of an individual, but different from values, these traits have a genetic component.

From an entrepreneurial point of view, it is important to understand what kind of these traits, owned by the entrepreneur, lead to the startup's survival.

For many years researchers have supported and improved the theory that five key personality traits that make the individual's personality, are called the Big Five. In their work, Ciavarella et al. (2003) explain how the Big Five can influence the survival of a new business, and then also of a startup. They are reported and analyzed below.

1. ***1.*** **Extraversion.**

This trait is characterized by sociability, high energy level and assertiveness.

An entrepreneur who owns this characteristic is able to create a high number and high-quality level relationships, very important to creating strong social networks with external investors, customers and suppliers, so this increases the probability of a startup's survival.

2. ***2.*** **Emotional stability.**

The trait's features are: calm, self-satisfaction and maintaining optimism.

It is a very important characteristic for an entrepreneur that wants to launch a new startup, because of the high-risk and stressful environment that characterizes the life cycle of a startup. Emotional stability helps the entrepreneur to maintain calm and optimism even in critical situations by transforming them into new opportunities to use them in his or her favor, and as a consequence, by increasing the likelihood of startup survival in the long term.

3. ***3.*** **Agreeableness.**

An individual who owns this characteristic is courteous and friendly to other people. It has a key role in creating long-term and trusting relationships, in particular with customers.

4. **Conscientiousness.**

It is strictly correlated with perseverance. A conscientious entrepreneur is a goal-oriented individual, he/she is able to remain focused on the objective and hard worker. This characteristic increases the efficiency and productivity leading to the long-run survival of the startup.

5. **Openness to the experience.**

To increase the probability of startup survival, the entrepreneur needs to be able to catch new opportunities and to change the business idea according to the market needs, so the entrepreneur needs to be a smart and creative person able to discover the market and environmental changes and available to change his/her mind and strategy if it is necessary.

Leadership style

Another important aspect of the entrepreneurial personality on which studies are focused is the founder's leadership style. Many studies demonstrate the essential role played by the founder in the startup's success, she/he has the responsibility to motivate the employees and drive them towards a unique and completely shared goal (Zeach and Baldegger, 2016). In literature. this type of leadership is called entrepreneurial leadership, defined by Zeach and Baldegger (2016) as “a personal influence and directing another person towards a certain action or way of thinking in the context of startup”. Bass (1995) developed a leadership model based on three different approaches:

1. ***Transformational leadership***

   A leader who follows this management style establishes a relationship based on communication with other employees. She/he is interested in the growth of her/his subordinates and plays a mentoring role for them (Bass, 1995). This is associated with future-oriented and proactive leadership behavior (Zeach and Baldegger, 2016).

2. ***Transactional leadership***

This type of leadership is based on a reward system. The leader communicates to his/her subordinates what are the goals that he/she wants to achieve and, based on some performance criteria, gives them rewards and/or punishments (Bass, 1995).

3. ***Laissez-faire leadership***

This approach consists of a passive behavior of the leader. She/he acts and repairs the problems after they happen, there is no system of rewards and the leader does not act as a guide for the other employees (Bass, 1995).

Zeach and Baldegger (2016) in their study analyze which of these three approaches is the more efficient for the management of a startup, in terms of performance and survival probability. The results of the research show that shoe that transformational leadership is the best approach in the startup context and positively affects the performance and the survival probability, instead the other two approaches present some criticisms and show positive effects only for the larger startups, for the smaller ones instead have negative effect.

Human capital

By definition "entrepreneurs transfer skills that they have to a new venture" (BarNir, 2012), so the human capital that an entrepreneur gets involved in the new firm plays a key role in its growth and survival, particularly in the context of innovative startups. According to the human capital theory (Becker, 1975), human capital is the set of skills and capabilities that help individuals achieve a competitive advantage in the market. It is possible to identify two levels of human capital (BarNir, 2012) reported in the following table.

Level	Elements	Manifestation
Specific human capital	• Skills related to a specific company or industry. • Deep knowledge of the specific startup's domain • The capabilities to create networks • Problem solving skills	Prior experience in the industry

General human capital	• Learning ability • Adaptation ability • Capability to transfer knowledge across different industries	• Educational background • Career path • Family context

Table 6: The two levels of human capital

Prior research focuses on the impact of these two levels of human capital on the growth and survival rate of the startup. Regarding the specific human capital, Cooper and Bruno (1977) demonstrated a positive relationship between the survival rate and the previous experience of the founders, in particular showing that when the individuals that start the startup come from large corporations, the startup has a higher probability of survival. Bruderl et al. (1992) demonstrated that a high level of prior education, made by studies, prior career experience and family context, is positively related to a higher probability of survival.

Team

The characteristics mentioned till now regard the single entrepreneur, but is the startup is founded by two or more individuals they create *a team*, and, in addition to the individual features, there are elements related to the team that can influence the startup's survival. Unfortunately, the majority of the studies focus on individual characteristics rather than on team ones, and the works that exist often do not arrive at a unified conclusion but maintain different options open. Based on the work of Anna Brattstrom (2019), it is possible to identify three main categories of characteristics of the team that influence the startup's survival:

1. **Homogeneity.**

Studies show generally teams tend to be homogeneous, this is due to the fact that if someone wants to start a business search for mates from his/her relationships network, which often is made by similar people, and also because individuals tend to trust the people similar to them. So homogeneity can benefit the business and its survival, when the components of the team are similar to each other and have something in common the

performance of the startup can improve due to the friendly relationship between them. On the other hand, it is also true that heterogeneity may represent an advantage, in particular for startups with a high level of technological innovation, the different backgrounds of the team's members can help the startup' performance. The existing studies don't conclude unique and clear results about what kind of team is better than the other to improve the probability of startup survival.

2. **Change and rigid structure.**

In the startup team, they have to coexist in two opposite terms: change and stability.

To survive in the long term the startup team needs to be able to adapt to environmental changes, and if necessary, even to reshape the initial idea of the business. The team has to be able to transform its structure according to these changes and the evolution of the business. While some studies have brought to light that teams tend to have a rigid structure with clearly delineated roles in different components, it is not obvious what is the best process to select these roles is to improve the probability of startup survival.

3. **Synergy.**

It is important that, from the beginning, there is synergy between team components, they have to share the same values and same values. This will be easy in the initial phase once the startup is established. Then when the business is launched the first discordances come into play, due to the stress and the conflicts. If on one hand this is absolutely normal and can happen, on the other hand, the team needs to be able to recognize that the team synergy has a problem and it is necessary a turnover of the team. The team has to move in the same direction and to remain focused on the shared goal, if it is not the team has to change in favor of the startup's survival.

2.1.3 Why

Anyone who wants to initiate a startup has to ask him/herself “Why?”, why I'm doing so?, what is the background reason for why I want to start?

The entrepreneur must identify an idea, a concept, and try to transform it into a product or a service marketable on the market, to do so the entrepreneur has to convince the customers that they need his/her product or service. The basis to realize this step is a

“good” idea. But, what is a “good” idea? It is not easy to answer this question. Based on Sam Altman's speech (How to Succeed with a Startup, 2018) it is possible to select four main aspects of a successful idea, that lead to an increase in the probability of startup survival.

1. ***“A product (or service) so good people tell friends”***

An idea can have success if people spontaneously encourage their friends to use the product, in this way, it will create a network of people that demand the product in a “natural” manner.

2. ***“Easy to understand”***

An entrepreneur has to be able to explain in a few words his/her idea, and people who don't know it have to understand it in a clear manner.

3. ***“Idea generators”***

Due to the changing nature of the environment in which a startup lives, it is important for its survival to have a constant generation of new ideas to adapt the business to market changes.

4. ***“Speed”***

The process of generating an idea, testing it, and implementing it needs to be very rapid to obtain successful results.

2.2.2 Innovation sustainability factors

*“All production consists of combining materials and forces that are within our reach." To produce other things or the same things in a different way is to combine these things and these forces in a different way”.*These are the words of the economist Joseph Schumpeter used to describe the concept of progress. In 1939 he re-shaped the concept of entrepreneurship from a static view to a dynamic view, where a fundamental role is played by constant innovation.

A formal definition of innovation is given by the OECD (2009): “the implementation of a new or significantly improved product, process, a new marketing method, or a new

organizational method in business practices, workplace organization or external relations".

The difference between a startup and another SME is the high level of innovation that characterizes the first one, for this reason, it is of vital importance for the startup's survival, studies show that businesses that are not able to invest in research and development, are not able to survive in long-term. The innovation should be seen as an evolutionary, non-linear and interactive process, requiring intensive communication and collaboration between firms and organizations such as universities, financial institutions and government agencies (Hudson and Khazragui 2013), this leads to a complex system where the startup has to face the Death-Valley phase. The Valley of Death is a phase of the startup life cycle, typically when the basic idea is transformed into a commercialized product. In this phase a high number of businesses do not survive mainly due to a bad performance of one or more of the above mentioned collaborations. The literature explains the different elements that influence the innovation of a startup, in particular, from previous studies, it is possible to select five of these.

1. Financing and Innovation

There is a strong linkage between financing and innovation, because a startup to carry out innovation has to invest a lot of its capital in R&D, and to invest in R&D it has to have enough funds. This is one of the main problems faced by entrepreneurs and often it is difficult to to close the gap between innovation and financing. Often the R&D investments are subject to credit rationing due to different reasons related to the nature of this type of investments that give them a high level of risk from the point of view of the financing provider, In particular, they are characterized by asymmetric information between the startup and the investors, the type of knowledge generated by these investments is embedded in the human capital and for this reason it is difficult to transfer, moreover, they are influenced by the appropriate regime in which the firm operates and the availability and the enforceability of the intellectual property (this topic will be argued later).

The importance of the source of financing and the high-risk level of the investment makes the research of funds one of the hardest challenges for entrepreneurs.

An important role is played by the banks, the literature shows that the banking system can be seen as the foundation of the development process for an industrial system, then the basis for innovation, at the local level. However on the other hand the banking system shows a weak interest in investments in innovation activities. This lack of interest is mainly due to the nature of assets that innovative firms offer as collateral, in general, innovative firms have a high number of intangible assets but a low number of tangible assets, the banks value the first type lower than the second ones and for this reason, are less likely to grant credit. In recent years, after Basilea 2, the banks are improving their technology and instruments to evaluate innovative projects and show an increasing interest in innovation financing.

2. Turnover and Innovation

Different studies take into consideration how turnover can affect the decisions regarding the innovation activities of a startup. These studies show that positive turnover positively affects the startup's survival (Okrah et al. (2012)), in particular, positive turnover has two main positive effects on the startup's survival. A good turnover allows the entrepreneur to use internal funds to finance innovation from the moment that, as it is said before, it is not so easy to obtain funds from external investors. Moreover, positive turnover improves the reputation of the firm helping the startup to survive and, in this case, to realize a successful IPO.

3. Openness of market and Innovation

How the market's characteristics influence the survival of the startup is argued later, but now it is explained in particular the linkage, that emerged from past studies, between the degree of openness of a market and the innovation of a startup.

The study by Okrah et al. (2018) shows the level of openness in the internal market in which the startup operates plays a key role in its survival. A startup that runs its business in an open market has the possibility to attract external investors and so can obtain more funds, moreover, the openness increases the competitiveness in the market, which, as a consequence, motivates startups to improve innovation to outperform their rivals. Zilgalvis (2014) treated the UK case, showing the damage suffered by the innovation in the economic system, when the country withdrew from the European Union, passing from

an open to a more isolated economy. On the other hand, it is important to underline that closed economies of emerging countries, in particular India and China, have a high level of innovation (Banerjee and Defluo, 2019).

4. Government and Innovation

As it is easy to imagine the intervention of the government can drive the direction of the innovation.

The tools used and the measure in which the government can influence the life cycle of a startup are treated later in the section “Industry and Government factors”, in this section it is shown the aspect strictly correlated to innovation. The government can undertake different types of initiatives to create an innovation-friendly environment.

Regarding financing, the government can intervene in different manners, it can directly finance the R&D investments and act as a co-investor of the startup, it can also introduce fiscal incentives for the investors that invest in innovation activities and can guarantee loans. The literature shows that these types of intervention by the government can have two opposite effects, they can have an additional effect, so public intervention stimulates and increases the private sector investments, and can provide the financed startup with a positive reputational effect. But public intervention can have a substitution effect, the public investments replace and substitute, completely or in part, private investments.

However, the actions undertaken by the government can have a different nature from the financial ones. According to the study of Dolfsma and Seo (2013) the policies applied by the government to sustain innovation need to be strictly correlated to the type of technology taken into consideration in order to be effective. The study classifies the technologies respect two characteristics, the first is if the technological knowledge is developed in a discrete manner, so the development is independent of previous developments, or in a cumulative manner, so the improvement of technology is based on the previous improvements, the second one is if the network effect plays a key role in the demand of the technological product, so if the demand of the product of a customer depends heavily on which and how much other customers purchased the product or it doesn't. Based on these two characteristic and their combination Dolfsman and Seo propose different public actions, for example in the case of technologies that grow in a discrete manner and do have not a network effect, the government can improve innovation

by stimulating individual entrepreneurs by supporting the research center, instead in a situation in which the previous knowledge is essential for the improvement of the technology the state's policies have to guarantee the easy access to intellectual property by third parties, or, moreover, if the technology's market is characterized a cumulative improvement and a very important network effect, there is the risk that the market is totally controlled by the big players and this slow down the innovation's run, so the government can intervene with anti-trust law to prevent small players from being excluded.

5. Protection of innovation

Elements that help or hinder innovation have been analyzed so far, but equally important, for the startups' survival, is their ability to protect what they invented.

There are different ways in which a startup can protect the value of its inventions, one of these is the Intellectual Property, in particular, the most suitable IP to protect the innovation is the patent. The number of patents owned by a startup represents an important factor that influences its survival. A startup that has a large number of patents enjoys many benefits, first of all, patents can be a source of internal financing, and startups can exploit patents financially and make profits from them, by licensing or selling, according to its needs, on the other hand patents give to the company a competitive advantage respect to its competitors, from the moment that it has the control on the technology. Moreover, patents play a key role in the valuation of the startup, a high number of patents demonstrates the effort of the startup in the R&D activities and the ability to protect the results, increasing the valuation and the reputation, so the startup will be seen more attractive to the investors' eyes and that's not all, as it is said in before, in the recent years banks increased their interest toward innovation activities, they improved it techniques to evaluate technological innovation, so they analyze and evaluate more accurately firms' intangible assets, moving in this direction patents are becoming valuable to obtain loans from banks, a very important point for the startups that, due to their risk nature, have difficulty to obtain external financing.

Generally, for the innovation of a sector, the patents represent an important boost, because, to avoid litigation, competitors search for different ways to work around the technology protected by patent, improving the innovation.

2.2.3 Financing factors

Find the source of financing is one of the main challenges faced by new startups, because of the capital market imperfection in which they operate (Colombo and Piva (2018)) caused by the high level of information asymmetry between entrepreneurs and investors, in turn it is transformed in riskiness for investors, so this creates a funding gap for new firms, that if it is not filled lead to the startup's death.

In the literature, there are different studies that analyze the impact of the different kinds of investors on the startups that they sustain. Different types of funding sources are related to the different stages of the startup's life cycle (Salamzadeh et al. (2015)). The following table shows the three main stages and the associated source of financing (Salamzadeh et al. (2015)).

Stage	Source of financing
Founding stage	• Personal, Family and Friends funds • Business Angels
Seed stage	• Incubators
Development stage	• Venture Capital

Table 7: Three main stages of startup life cycle and the associated source of financing

The table doesn't include the traditional bank's loan, it is already argued in relation to the innovation.

Bypassing the detailed characteristics of each source of financing, what is in the interest of the work is their impact on the startup's survival, this is analyzed below for each source.

1. ***Personal, Family and Friends funds.***

The first stage of the startup life cycle is also called “Bootstrapping”, it represents the phase in which the founder relies on his/her strengths, by supporting the business with internal funds, and in this case, with the help of family and friends found, so not formal external investors.

The activity of bootstrapping has both positive and negative influences on the startup's survival. It increases the ability of the entrepreneur to find different strategies to sustain the business, in particular, the founder can carry out different methods of financial bootstrapping (Winborg & Landstrom, 2001), for example, he/she can choose a "Relationship-oriented bootstrapping", in this situation the entrepreneur creates a network of relationships that permits him/her to decrease the business' costs, another example is the "Delaying bootstrapping", that provides the delay of payments of the suppliers, all these methods allow better management of the startup financial resources and than increase the probability of startup survival. Moreover, this kind of activity increases the creativity of the founder (Backer and Nelson (2005)), as seen in the paragraph Company factors, it is one of the values that characterizes a successful entrepreneur, and then a successful startup. In addition, bootstrapping avoids the use of debt, which at this stage of the startup life cycle is difficult to obtain and guarantees higher flexibility to the firm, an important characteristic for startup success given the need for startups to adapt to the changing environment. On the other hand, bootstrapping also has a negative effect, it has a high opportunity cost, it involves the use of the valuable time of the entrepreneur to reach the best strategy, and the increases the financial risk over the entrepreneur that can lead to loss of good opportunities.

2. Business Angels

"A business angel is a private individual, often with a high net worth, and usually with business experience, who directly invests part of their assets in new and growing private businesses. Business angels can invest individually or as part of a syndicate where one angel typically takes the lead role." Just to introduce it, this is a formal definition of business angel provided by the European Commission.

As said in the definition, Business Angels, usually, are individuals with entrepreneurial experience and skills, so the benefits that they provide to the startup are of two different natures: the financial one and the experienced one, in the majority of cases, BAs sit directly in the Board of the financed firm. (Wiltbank (2009)) by participating in the strategic decision. It has been demonstrated that business angel financing improves the probability of survival for startups financed (Kerr et al. (2014)). The work of Croce, Guerini and Ughetto (2018), demonstrates what are the elements of BA activities that drive the performance of startups. In particular, they refer to two main characteristics of

business angels investments: BA capabilities and investment behavior. Regarding the first feature they confirmed the hypothesis that there is a positive relationship between the BA experience in the early stage investments and the improvement of startup performance, due to the fact that the BA-backed startups analyzed, demonstrate a better perspective of growth after the initial stage, and this attracts the future funding by the Venture Capitals that invest in the subsequent stage. From their analysis emerged that, the characteristic of the investment behavior of co-localization, namely the fact that BAs prefer to invest in startups that are in their same country, has a positive relationship with better startup's performance. Moreover, it is brought to light that in many cases BAs tend to co-invest with Venture Capital, or that the VCs investments are subsequential to BAs investments, and in both cases there is a positive link with better startup performance and its probability of survival.

3. Incubators

According to the National Business Incubator Association (NBIA), an incubator is *"a business support process that accelerates the successful development of startup and fledging companies by providing entrepreneurs with an array of targeted resources and services".* In other words, the incubator is a physical or virtual place in which a newly established startup can take advantage of different types of benefits, from physical offices and laboratories to financial and intellectual support.

The literature on the impact of incubators on startups is contrasting. Alongside the recognized many benefits that the incubators provide to the incubated startups, such as technological support, business assistance, financial support, mentoring, improvement of startups' reputation, and creation of a contact network (European Commission (2002)), not all studies agree about the positive relationship between the incubators' activity and the improvement of the survival rate of the startups. Schwartz (2010), taking into consideration his analysis of five German incubators, compared the long-term survival rate of 371 incubated firms and of 371 comparable non-incubated firms, he concluded that the incubators' intervention did not increase the chances of long-term startup survival. In a more recent study, Hackett and Dilts (2014), developed a theory of business incubators that demonstrates that business incubators increase the probability of start-up survival. This discrepancy from the literature can be attributed to the fact that if on the one hand, the incubators improve the performance of the incubated startup and lay the

foundation for faster growth, on the other hand there is the possibility that the startup relies totally on the incubators sustain and when this is no longer there, the startup is not able to survive (Lukes et al. (2019)).

4. Venture Capital

"Venture capital is a form of medium- to long-term investment in unlisted companies with high growth and development potential (high-growth companies) that are in the start-up phase. [...]Venture capital activity does not only involve the contribution of risk capital, but also a series of activities, [...] it participates in the company's strategic decisions by contributing his professional knowledge and experience, leaving the entrepreneur and management with the operational management" this is the definition of Venture Capital provided by Borsa Italiana. Venture Capital is similar to the already mentioned Business Angels, with some differences in the investment stage and behavior (Mason and Stark (2004)).

It is clear that Venture Capital takes to the backed firms, in particular, VCs represent an alternative channel of financing to the banks' loan, which as it is said, before is difficult to obtain for a newly established startup, as said in the definition VCs seat directly in the board by monitoring actively the management activities and they play a role of mentoring for the portfolio firms' entrepreneur, moreover, they help the startups from the point of view of the relationship by increasing their social network, of suppliers, customers and investors, and by improving their reputation (Megginson and Weiss (1997).By matching these VCs' functions and what is said till now in the work, it is easy to think that Venture Capitalists can increase the probability of survival for a startup, to this scope it is important to underline that VC investments are characterized by two main features: *selection,* VCs have very sophisticated methods to "pick" the firms to sustain and are able to select the startups with higher growth prospects (Sorensen (2007)), and *influence,* by the activity if mentoring and monitoring. Different studies analyzed if the success of the VC backed firm is due to the intervention of the VC or because the VC selected a successful startup already before the VC's help. It is demonstrated that both characteristics can influence the ability of the startup to the change in the external environment (Alperovych and Hubner (2011)) and improve the probability of realizing a successful IPO (Sorensen (2007)) by increasing the probability of the startup's survival.

2.2.4 Industry factors

From empirical studies it is evident that the characteristics of the industry in which the startup operates play a key role in the survival probability of the startup. (Bhide (2000)). The literature highlights the main industry characteristics that positively affect the survival rate of startups.

1. ***Entry barriers***, in its work Giarratanta (2004), demonstrate that the likelihood of survival for a startup is higher in an industry with low entry barriers for new entrants.
2. ***Industry novelty*** (Kim et al. (2023); Giarratana (2004)), the novelty of an industry is characterized by its date of birth in the market, which demonstrates that the more industry is new, the higher the probability of success.
3. ***Technological alliance*** (Giarratanta (2004)), in a situation of high technology in which a startup leaves, it is almost impossible to own all the necessary competencies to develop the product/service, a network of collaborative ventures in the industry to improve the performance and the survival rate.
4. ***Geographic clustering and social capital*** (Bandera and Thomas (2019)), the geographic proximity of highly innovative startups and social capital lead to benefits in the startups' performance, but they are transformed into an increase in survival rate only if they are utilized and exploited by the entrepreneurs.
5. ***Market growth perspective*** (Sam Altam (video)), the large number of startups that have had success in the past, are ones that started the business in a small market but that have grown very quickly.

2.3 Women and entrepreneurship

2.3.1 An Overview of gender gap in the entrepreneurship world

The participation rate of women in the entrepreneurship world is increasing in the last year, even the growth rate is different by country and industry (Vossenberg, 2013). The GEM (Global Entrepreneurship Monitor) in the *2021/22 Women's Entrepreneurship Report* claims *"Globally, women represent about one in three high-growth entrepreneurs and one in three innovation entrepreneurs that are focused on national and international markets."* It is shared by scholars that women play a key role in the countries' economic development (Sarfaraz et al. 2014), some studies demonstrate that higher level of female entrepreneurial activity rates are shown in countries with a higher level of total entrepreneurial activity rate (Verheul et al. 2004). This evidence attracts the attention of the government and policy makers, in particular in developing countries (Minniti and Naudè, 2010). Although the gender gap is still present in a tangible manner, women are less likely to start a business than men and their businesses grow lower and show a higher failure rate than the men's ones (Vossenberg, 2013). Prior literature tries to go inside this persistent gender gap to understand the causes and to incentivize female entrepreneurship.

2.3.2 Causes of the gender gap in the entrepreneurial participation rate

The literature points out many causes that drive the women's participation rate in entrepreneurship, in particular, they can be grouped into two categories: macro-level factors and micro-level factors. The two categories need to be considered linked from the moment that they strongly influence each other (Jamali, 2009). The macro-level factors include the causes that regard the macroeconomic and the sociopolitical environment in which women run their businesses, and depend on the country and the industry, instead at the micro-level there are the so call *push & pull factors*, that represent the motivations for which a woman decide to start her business (Butter and Moore, 1997), these motivations are influenced by the environmental factors. From prior studies, it is possible to identify three main factors that include both macro and micro factors.

1. ***Gender Role***

In their work, Eagly and Carli (2004) define Gender role as *"the descriptive and injunctive expectations associated with women and men."* In the entire world, there are different types of stereotypes and discrimination that evaluate and group together people

with some characteristics in common, by leaving them little room for maneuvering for their choices. Stereotypes are typical in the labor world, in particular regarding gender. It is usually attributed to men and women's different characteristics and abilities and to think that they are more suitable for certain works than others (Eagly and Carli, 2004). This is what happens when you talk about the leadership role, it is typical to associate the role of the leader of an enterprise with the men rather than with a woman (Schein, 2001). Society identifies the characteristics that a leader has to have in masculine behavior, and then associates this role to a gender role. Alongside this aspect it is important to underline the *theory of role congruity* (Eagly and Diekman, 2004), according to which individuals, by setting their goals in their professional life, tend to maximize their role congruity. By putting together the gender role and the theory of role congruity, it is possible to conclude that, at least in part, women, consciously or not, internalize the goals of their gender role that are not matched with the leader's features, and for this reason tend to stay away from the entrepreneurial world (Eagly and Diekman, 2004).

2. Gender Equality

With the term *Gender equality* is intended *"the state in which access to rights or opportunities is unaffected by gender."* (United Way NCA, 2022), the more a society is closed to this situation the higher is its level of gender equality. Different prior researchers tried to understand the relationship between the level of gender equality in a country and its rate of women entrepreneurs, but the results are contrary, some studies point out that gender equality does not any type of correlation with female entrepreneurship (Sarfaraz et al. 2014), others show a positive relationship (Baughn et al. 2006) and others a negative relationship (Klyver et al. 2013).

There are two main theories that explain the opposite results of the research (Vracheva and Stoyneva, 2020): the *Liberal feminist theory* (Fisher et al., 1993) and the *Preference theory* (Hakim, 2000).

The Liberal feminist theory states that there are no differences between men and women, they have the same capabilities and preferences, the discrepancies between the two sexes in the labor market and in particular in entrepreneurship, depend only on social discrimination and the social, economic and legal barriers, that women have to face in their professional life. Once these barriers fall and gender equality is reached, women and men can have access to the same opportunities and resources, the gender gap in entrepreneurship, both in terms of participation rate and business performance, is filled.

On the other hand, the Preference theory claims that women as a group are more heterogeneous than men, and present different preferences regarding their professional and private life. In particular, Hakim (2003) states that women allocate themselves on a priority scale that starts from the *"Home-centered woman"* to the *"Work-centered woman"*. According to this view, even if the two sexes can have access to the same entrepreneurship opportunities, the gender gap in entrepreneurship still exists, there is a certain number of women who prefer to dedicate their time to house care instead that to their career, and this number of women is higher respect to the one of men that make the same choice. So gender equality does not completely resolve the gender gap that characterizes entrepreneurship, both in terms of participation and business performance.

3. Institutional theory

"Institutional theory focuses on the role of social, economic and political systems in which entrepreneurs operate and their choices and behavior embedded." (Vossenberg, 2013), in other words, the Institutional theory includes all factors that contribute to creating the system in which the entrepreneurs start their business and in someway influence the women's behavior regarding the entrepreneurship. In particular Vossenberg (2013) identifies three levels of institutional context: *regulatory system*, so the set of laws and policies regarding entrepreneurship, it includes, for example, the access to capital and loans that, can be more difficult for women than men, *normative system,* this is related to the above-mentioned Gender Role, it is a set of norms that the society creates from its beliefs creating different social roles in which people are categorized, these beliefs can be linked to religious and cultural factors. The last one is the so-called *cognitive system,* which includes access to education and technology. This represents an obstacle for women who want to start a new business, because often the level of education for women is lower than the one of men.

2.3.3 How the Gender factor can influence a startup's performance and survival.

As it is said in Chapter 2.2, the success of a startup is related to the characteristics of the entrepreneur, for this reason the researchers began to wonder if even the gender factor can influence in somehow the startup's performance. Prior research based on available data suggests that female-led firms tend to be less profitable (Fairlie and Robb, 2009) and have a higher failure rate (Boyer and Blazy, 2014) with respect to male-led firms. Other studies show that, net of the size of the business and type of industry, the difference

between the survival rate of women entrepreneurs and the men's one is not so high (Perry, 2002). In any case, there is a relationship between gender and entrepreneurship (Minniti, 2009), that needs to be analyzed.

The causes of the gender gap in firms' performance and success, in particular for the startups, can be divided into two macro-categories: *motivations,* the reasons why a woman decides to start her own business, and how they differ from the ones of the men, and *entrepreneurial characteristics,* the environmental factors, that influence the entrepreneurial behavior of an individual, and in particular affect in a different way the two sexes.

1. Motivations and reasons for starting a startup

The motivations for which an individual decides to start a new business can influence the outcomes of the firm (Shaver, 1985). The motivations enclose the individual's goals in starting a new firm or startup, and for this reason, play a key role in the effort and time that the new entrepreneur employs in the creation of his/her new enterprise (Carter et al., 2001). Women and men take into consideration and prioritize their motivations in different ways (Carter et al., 2001), by analyzing these differences it is possible to understand the gender gap in startups' performance. The following table shows the six categories of reasons identified by Carter et al. (2001), that women and men rank in a different manner, plus a seventh motivation that the researchers have studied as a driver for women's entrepreneurial choices.

Category	Description
Innovation	The intention to exploit something new
Independence and work-life balance	The desire for freedom, control and flexibility
Recognition	The will to have a certain status
Role	The willingness to follow family and friendly environmental traditions
Financial Success	The desire to earn more money and achieve a certain financial situation
Self-realization	The desire to achieve self-directed goals
Gender pay gap	The difference between women's earnings and men's earnings.

Table 8: The six categories of motivation to start a startup according to Carter et al. (2001)

In this section, the different motivations and their effect on the startup success will be analyzed based on the results of the prior research. Only the first factor "*innovation"* will be analyzed in detail in section 2.3.4.

1. Independence and work-life balance

One of the advantages of self-employment is the flexibility in the labor hours and self-organization of the work-life balance, so this represents a motivation that an individual takes into consideration when deciding to start his/her business, in particular when the entrepreneur has a family with children. The care of the house and family is a role that society has always entrusted to the woman, by giving to the man less family responsibility. Even in recent years the situation has little changed and the division of labor in the house is more balanced between wife and husband (England, 2003), women show a greater involvement in the housework and rate family as an important life goal than men (Eskilson and Wiley, 1999). For this reason, women see entrepreneurship as a possibility to obtain a flexible career (Parasuraman and Simmers, 2001). This motivation to start a new firm or a startup, can influence in a negative manner the enterprise's success (Jennings and McDougald, 2007), in particular, Williams (2004), in his study conducted in eight European nations, demonstrates that the time dedicated to child-care is negatively correlated to the duration of self-employment ventures.

2. Recognition

About this motivation there are not a lot of studies, in particular, the prior literature does not analyze the relationship between the factor and the startup survival. What the research shows is that women evaluate more the recognition motivation with respect to male colleagues (Fisher et al., 1993), this can be interpreted as an opportunity for redemption for women, particularly in certain geographic areas, such as South Italia (Unicamere-Centro studi Guglielmo Tagliacarne. 2022).

3. Role

Regarding the motivation to start a new firm, studies find that the family context plays an essential role in the choice of an individual to become an entrepreneur, and that this affects to the same extent men's and women's choices. A study demonstrates that it is more likely that a woman entrepreneur has a mother entrepreneur with respect to a male entrepreneur

(Malach-Pines and Schwartz, 2008). Moreover, Caputo and Dolinsky (1998), state that for a woman to have a husband entrepreneur is more likely to become in turn an entrepreneur.

4. Financial Success

The literature states that in general entrepreneurs show a greater interest in financial success, as goals of their career, respect non-entrepreneurs, both men and women. (Carter et al., 2003), but in particular, this is a motivation felt more by men's respect for women (Carter et al., 2003). The existing studies do not examine the direct impact that this motivation to start a new business can have on effective business survival and performance, but some researchers, in particular Kasser & Ryan (1993), have studied the personal and professional outcomes of individuals that present a strong interest in financial success goals. They found that this type of individual tends to have a *control-oriented behavior,* so by focusing more on external sources rather than on interpersonal relationships. As mentioned in Chapter 2.2, the ability to create relationships and networks, and more in general interpersonal soft skills, play a key role in the entrepreneur's success. Moreover, Eagly and Johnson (1990), have investigated the difference in the leadership style between women and men and found that women tend to be more *interpersonally-oriented* with respect to their male colleagues. By putting together the results of these studies it is possible to hypothesize that when an entrepreneur is strongly driven by this motivation can influence negatively his/her business, and focusing on gender the probability that this happened is higher for male entrepreneurs.

5. Self-realization

Self-satisfaction can be identified as the will of an individual to fulfill his/her professional and career goals. Some studies report that entrepreneurs have a higher desire for self-satisfaction and respect for non-entrepreneurs (Carter et al., 2003). Gender differences regarding this motivation in literature there are conflicting opinions. Some studies state that between nascent entrepreneurs women show a higher interest in self-satisfaction (Gatewood et al., 1995). Other studies, instead, argue that men prefer this motivation more than women do, and this explains how there are more new male entrepreneurs than women, from the moment that the creation of a new startup is seen as a source of self-realization opportunities by the entrepreneurs of both genders (BarNir, 2012). There is

another study (Orban and Don Scott, 2001), that links self-satisfaction motivation with the concept of *"glass ceiling"*. By definition glass ceiling is "*a metaphorical invisible barrier that prevents certain individuals from being promoted to managerial- and executive-level positions within an organization or industry."* (Investopedia), in particular the concept is utilized in relation to a set of visible and invisible barriers that hinder the women to obtain roles at the top-level of the chain of command. This phenomenon can lead to "dissatisfaction with the salaried job" between women, who, driven by self-satisfaction, prefer to start a new business as self-employed.

6. Gender pay gap

If on one hand, the rate of female participation in the labor market is rapidly increasing, on the other hand, the gender pay gap does not follow the same speed (Cortes and Pan, 2015). Literature tries to understand how this inequality can influence female entrepreneurship, both in terms of participation rate and business success. Based on two main works, Gawel and Mroczek-Dabrowska (2021) and Cortes and Pan (2015), the reasons for the persistence of the gender gap in earnings can be explained by: *human capital theory* (Tverdostup and Paas, 2017), which states that there is a difference between men and women in the selection of human capital, *labor market segmentation* (Aidis and Weeks, 2016), states that women tend to have a preference to enter in secondary labor market with lower productivity, high returns of working long hours (Cortes and Pan, 2015), in particular among highly skilled works. Gawel and Mroczek-Dabrowska (2021) found out that the gender pay gap can work as an incentive for women to enter entrepreneurship, from the moment that is not satisfied by their employment, in particular in the industry "with relatively low barriers to entry", but it is also possible that gender pay gap is seen as an inhibitor for women to start a new business because it is part of a discrimination view that characterized some countries. The results of the research show that the gender pay gap has a positive influence on female entrepreneurship in terms of participation rate, in countries in which there is a high level of gender equality and women "do not perceive the gender pay gap as a discriminatory factor discouraging them from entrepreneurship". The study results of Cortes and Pan (2015) the gender pay gap is influenced by the high returns for overwork, and it is suffered by women from the moment that they have a "dual role in the home and labor market", and, in average, respect male colleagues can spend less time at work, this can represent an incentive for women to start a new business and become entrepreneurs, to have more flexibility in work hours.

Prior research does not directly analyze the effect of the gender pay gap on the startup's performance, but it is possible to make some hypotheses by taking into consideration different studies. As said for the motivation *Independence and work-life balance,* choosing the enterprise way can have a negative effect on the business success (Jennings and McDougald, 2007; Williams, 2004). The other effects are not directly related to the gender pay gap, but more to the *human capital theory* and *labor market segmentation* and are analyzed in the next section.

2. Entrepreneurial characteristics

The entrepreneurial characteristics are a set of factors related both to the entrepreneur's behavior and capabilities and the features of the environment in which the entrepreneur starts her/his business. These factors influence the business' performance and survival probability and their impact can be different in relation to the entrepreneur's gender.

1. Leadership and management style

As said in Chapter 2, the entrepreneur who decides to create a startup is not only an entrepreneur but also a leader, for this reason, we talk about *entrepreneurial leadership*. In particular, in his theory, Bass (1995) individuates three main leadership styles: *transformational leadership*, *transactional leadership* and *laissez-faire leadership,* that can influence the startup's performance and survival. From their literature research, Eagly and Carli (2004) found out that there is some difference in the choice of leadership style between the two sexes. They argued that women tend to be more *interpersonally oriented,* "behavior related to maintaining the interpersonal relationship", and *future orientation,* consisting in the focalization of future goals and the realization of a plan to achieve them, which are characteristics belonging to transformational leadership. On the other hand, men show a major tendency toward transactional leadership and laissez-faire leadership. There is no specific research that connects directly the female leadership style to the probability of startup survival, but as mentioned in Chapter 2.2, transformational leadership leads to a higher probability of a startup's survival respect the other two styles (Zeach and Baldegger, 2016), than it is possible to say that the female leadership style can provide an advantage to the startup in terms of better performance and likelihood of survival.

2. Financing

Financing represents a critical point for innovative startups, from moment it is not easy for them to obtain external financing, due to the high level of uncertainty and asymmetric information (Colombo and Piva, 2018), but at the same time the funds are essential to promote R&D activities (chapter 2.2), so type of investors and the ability to obtain financing can influence the performance and the survival probability of the startup. It is also important to understand what are the factors that can influence access to funds, different studies analyzed if one of these factors is gender, and what is its effect. In general, the literature shows that female-led firms acquire a lower amount of external financing than the male-led ones, particularly from banks (Fay and Williams, 1993). The explanation of this evidence comes from two points of view (Carter et al., 2007):

1. ***Supply side.***

   Prior research analyzed whether there are many differences in access to financing linked to gender if women and men have different opportunities in receiving external funds.

   Fay and Williams (1993) conducted two experiments to test if it is true that women are subjected to discrimination when they ask for a bank loan to finance their business. They found that background education is worth more for women's valuation than for men's. In particular, they argued that comparing two applicants of the opposite sex with a high school education, women have a lower probability of obtaining the loan. At the university education level, the probability gap is reduced but education is considered more important for women than for men. Moreover, some studies state that, on average, women have to pay higher interest rates or provide higher collateral than men (Coleman, 2000). This is not in line with the evidence that shows the higher risk aversion of female entrepreneurs (Watson and Robinson, 2003), but can be explained by the choice of industry where women start a business, in general, characterized by lower growth and profitability, with respect to the male entrepreneurs (Coleman, 2000; Fay and Williams, 1993).

2. ***Demand side.***

   Different studies have analyzed what kind of sources of financing women entrepreneurs prefer, and what are the differences with male entrepreneurs. Findings show that regarding the choice if financing the business with external or

internal equity, women prefer the second one (Chaganti et al., 1995), in particular, they prefer to employ personal savings or exploit strong link networks, such as family and friends (Carter et al., 2007), this type of financing is suitable for the founding stage of the startup life cycle but less effective for the next stages, so the total relying on this type of financing can lead into lower growth and survival probability for the next steps. This choice can also be explained by the fact that on average women tend to start a business in an industry with lower growth expectations and less capital-intensive respect for their male colleagues. Regarding the choice to use debt, women seem to be more risk and debt-averse with respect to men (Watson and Robinson, 2003), and spend less time and effort developing quality relationships with financial institutions (Fabowale et al., 1995). Even this evidence can be explained by the type of industries in which women run their businesses (Coleman, 2000; Fay and Williams, 1993).

3. Investment in Human Capital

The human capital involved in the creation of a new startup plays an essential role for its survival probability, so it is important to take into consideration and analyze the founder(s)' educational background, career path, and family context. These three elements can present, on average, specific different characteristics in relation to female founders and male founders, and this can drive the startup's performance and likelihood of survival. Women and men do not invest in the same way in human capital (Eagly and Carli, 2004). On the one hand respect to education, women have achieved gender parity in terms of participation rate, at least in developed countries, for example in Italy in 2022 women represented 60% of the graduated students (report of Consorzio Interunivarsitario AlmaLaurea, 2022), on the other hand is still evident the gender gap in terms of work experience and career. The greater involvement in housework and children care by women leads to a higher number of career interruptions and less time and effort dedicated to paid work with respect to men (Eagly and Carli, 2004), by reducing their specific human capital. In general, women tend to be characterized by a high general human capital, provided by their educational background and low specific human capital, due to the characteristics of their careers above mentioned (BarNir, 2012), this can represent a disadvantage for women entrepreneurs in terms of startup survival.

4. Opportunity recognition

An important characteristic of a successful startup is the basic idea on which the entrepreneur develops his/her business. The entrepreneur needs to be able to recognize good opportunities and exploit them. The literature found different processes of opportunity recognition between female entrepreneurs and male entrepreneurs, that can influence the startup's performance and success. First of all, some studies demonstrate that women and men who want to start a new business prefer different industries to each other, in particular women tend to invest in industries with lower prospects of growth and profitability (Alsos et al., 2006), and this can negatively influence the startup's survival. To find new opportunities men use financial assets and explore new markets to satisfy new needs never taken into account before, instead women prefer to exploit their prior knowledge in a certain domain and use their business to satisfy existing customers' needs that are not adequately met at the moment (DeTienne and Chandler, 2007). For this reason, male-led startups show higher growth with respect to female-led ones, which can lead to a lower probability of startup survival for the female entrepreneur, from the moment that the high growth rate of the market is considered an important element for the startups' success (Sam Altam (video)).

2.3.4 Female entrepreneurship and innovation

Until now it has analyzed the relationship between women and entrepreneurship, and how gender can affect startup success from an entrepreneurial point of view. From the moment that the focus of the work is on *innovative startups*, this section will analyze the relationship between women entrepreneurs and *innovation*. As said before women prefer industries with lower growth rates and with a low level of technology involved (Alsos et al., 2006) than men, for this reason in general women entrepreneurs tend to invest less in R&D activities (Amoroso and Link, 2018), but in this section the focus is on differences between the approaches of the two sexes toward the innovation, in the innovative startups, so in industries with the same level of innovation, and in particular what are the effects of the women approach on the startups' success.

Arcuri et al. (2023) conducted research based on a sample to understand the relationship between three elements: innovation*, gender* and *new venture growth*. Results show that innovative startups with a female leader present a higher growth rate than male-led ones. There are few studies that focus on this triple relationship, so it is not easy to understand the causes of Arcuri’s results, a possible explanation can be found in the leadership style already mentioned. The innovation activities and the R&D investment are strictly related

to the social network of the startup, to acquire knowledge to exploit in innovation, the relationship with the external plays a key role (Amoroso and Audretsch, 2022). From this point of view, women's leadership style is more oriented to an interpersonally-style (Eagly and Carli, 2004), this help women to exploit better knowledge from external sources, such as research institutions or venture capitalists (Amoroso and Audretsch, 2022).

2.4 Low-skilled immigrants and female labor supply

In the previous chapter, talking about the figure of the woman entrepreneur, it has emerged several times the role of women in the house, and how housework and child-care can influence women's decision and their entrepreneurial careers. It is possible to hypothesize that variations in housework, such as outsourcing, at least in part, can have some effects on female entrepreneurs and their firms. Starting from this point it is interesting to analyze what are the elements that can have some impact on the housework supply, and how they can drive a woman entrepreneur's career and the success of her firm. One of the elements that can impact the housework supply most studied in literature is the flow of low-skilled immigrants. The 17.2% of the total female migrants in the world are domestic workers (Cortes, 2022), this data is consistent with the ISTAT's report (2023) that states that in Italy the percentage of immigrants employed in low skilled jobs, in particular in family services, is about 18.5%, against the 1% of workers with the Italian citizenship. Starting from these statements, the purpose of this work is to investigate how the inflows of female low-skilled immigrants influence the survival of startups led to native women. Before continuing with the statistical analysis, it is necessary to report what are the results of the prior research about the influence of low-skilled immigrants on the female labor market in general. The basic idea behind the research on this topic is that there is a positive correlation between the inflow of low-skilled immigrants in a certain area and the reduction in costs faced by native workers in outsourcing housework, and consequently, this reduction in costs can influence the labor supply of high-skilled native women. It is possible to subdivide the research findings into the two steps illustrated below.

1. **Impact of low-skilled immigration inflows on low-skilled labor prices**

   Cortes (2008) carried out an analysis to understand the correlation between these two elements in different cities across the U.S. She showed that the industries in which there is a higher rate of low-skilled immigrants are the industries that include low-skilled nontraded services, such as housekeeping and babysitting, and that the native highly educated are less likely to work in these industries. Starting from these findings, Cortes moved on to study the effect of low-skilled immigration on the prices of these types of nontraded services and goods. Her results suggest that there is a positive relationship between the level of low-skilled immigrants and the reduction in prices of immigrant-intensive services. Moreover,

she explored what are the channels through which this phenomenon happens, she found that the causes are the reduction in low-skilled immigrants' wages, in particular, she reported two different possible situations, namely if the low-skilled immigrants and low-skilled natives are perfect substitutes in the same low-skilled nontraded services industries, or if they are imperfect substitutes, on the first case the effect of the immigrations have the same consequences in the reduction in wages of both immigrants and natives workers, in the second one, instead, the reduction in wages is higher for immigrants respect to natives.

2. **Impact of reduction in low-skilled labor prices and on female labor supply in terms of working hours.**

   Recalling what is said in Chapter 2.3, the hours that an entrepreneur dedicates to her/his venture play a key role in the firm success in terms of business longevity, the time dedicated to house and children care is negatively correlated to the duration of the self-employment (Williams, 2004), so after assessing the impact of the low-skilled immigration inflow in a certain area on the prices of low-skilled services, it is important to understand if this reduction in prices can have some impact on the labor supply of the native women in terms of worked hours. In the literature there are two main papers that analyze this topic (Cortes and Tessada, 2011; Cortes and Pan, 2015), in detail, these two works focus on the impact that the reduction in prices has on the labor supply of *high-skilled* women, so the analyze is conducted by taking in consideration the labor sectors which require a certain level of education, little different from the purpose of this work, that is focus on the entrepreneurship. However, these two studies can give an important contribution to the scope of this work because firstly, can provide a general idea about the relationship between the two facts, secondly, as said in Chapter 2.2 high level of education is positively correlated to the probability of startup success, so high-skilled women can be a good example representation. It is possible to subdivide Cortes and Tessada's study (2011) into the four results below illustrated.

   RESULT 1: ***"Effect in household work"***.

   The reduction in prices decreases the opportunity cost for women to dedicate time to housework, so for women who present high returns, in both monetary and non-monetary terms, to work outside the home, it is optimal to outsource the

housework, at least in part. This is verified for women who already demand house-care services and consequently increase their demand, instead, women who were not previously using these services seem to not start to do so after the price reduction.

RESULT 2: *"Effect on labor supply".*

Even in this case, not all women change their labor supply as a function of changes in prices. Only women who have a high level of return to work outside the home and who already buy house-care services increase their hours dedicated to work consequently to the price reduction. Moreover, the sensitivity of the labor supply to price changes depends on how the time dedicated to the house-care and leisure changes with the price decreasing. Leisure can be an inferior good, so it decreases with the increase in an individual's income or a normal good, so it is positively correlated with an individual's income. If the female worker considers leisure an inferior good, her labor supply will increase with the price decreasing, instead if the women consider leisure a normal good, it is not possible to determine in advance what is the outcome, but only empirically.

RESULT 3: *"Career concerns".*

As already mentioned in Chapter 2.3, the role that society attributes to women as responsible for house care, more than men, has a negative effect on women's careers that presents frequent interruptions and less time and effort dedicated to paid work respect to men (Eagly and Carli, 2004). The reduction in price allows women to choose career paths that require more working hours, by decreasing the time dedicated to the house-care. This can reduce different existing gender gaps, for example, the gaps in earnings, in human capital growth, that, in the case of entrepreneurs, can influence the success of the new firm (Cooper and Bruno, 1977).

RESULT 4: *"Children at home".*

An element that can influence housekeeping services purchasing and the labor supply by women is the presence of *young children.* The state provides free services, such as public school, but when the children are too young (0-5 years) spend a lot of hours at home. For this reason, mothers of young children are likely to

demand a higher number of hours of childcare services than non-mothers of young children. But with the decrease in prices, the statistical results show no high differences in the labor supply between mothers and non-mothers, this depends on how the mothers perceive leisure, if leisure is a normal good, the difference between mothers and non-mothers in labor supply after price reduction is not so high because mothers tend to increase their leisure with the decreasing in prices, more than non-mothers, contrary if the leisure is an inferior good, mothers tend to increase the labor supply with the prices reduction, more than non-mothers. Regarding this topic Claudia Goldin (Nobel Prize in Economics 2023) carried out different studies, in particular, in her study with Katz (2011), she explores the *cost of flexibility* in different occupations and also in self-employment. The cost of flexibility refers to all *"penalties to labor supply behavior that is more compatible with having a family"*. The cost of workplace flexibility is determined by the equilibrium between its supply (the level of workplace flexibility that firms offer to their employees, it changes by type of occupation) and its demand (the level of workplace flexibility demanded by employees), it changes by individuals. Goldin and Katz state that women ask for more flexibility at work than men and provide two events that can alter the equilibrium:

1. *Labor supply shift*

   It happens when there is an increase in the participation rate of women in a particular occupation, from the moment that women ask more flexibility, its cost in that occupation increases, then the increase the wage of the jobs with low flexibility respect to the ones with high flexibility, and men tend to move towards the jobs with lower flexibility and higher wage, increasing the gender pay gap.

2. *The cost of workplace flexibility decreases*

   In this situation more firms want to offer higher flexibility to their employees, at the current wage difference between jobs with flexibility and jobs without it. So, the cost of workplace flexibility decreases and individuals move towards jobs with higher flexibility, in general it is more likely that women do so with respect to men.

Another element that Goldin faces in her studies is the role of children in changing labor supply, particularly in the women's life (Goldin et al., 2022). The work

shows that the hours that women dedicate to paid work decrease after a birth and stay low for the first years of life of the children. The research demonstrates that women with children between 0 and 2 years old work 3.5 fewer weeks per year of respect to the father of children with the same age, and the differences in hours worked between mothers and non-mothers is higher for college graduated women than for non-graduates.

3. Empirical Methods

The scope of this thesis is to analyze whether there is any causal relationship between the presence of migrants and the duration of female-led innovative startups in Lombardy's municipalities.

The study is carried out by constructing the statistical model with the *linear regression method*, which requires three steps: the identification of the variables (dependent variable, independent variable, and control variables), the analysis of the relationship between the dependent and the independent variables and finally, the check of the robustness of the model. Before starting with the analysis, it is provided a brief explanation of how the linear regression model works.

3.1 Linear regression with a Single Regressor

3.1.1 What is it?

The *Linear Regression Model* is a statistical instrument that has the purpose of discovering whether there exists a linear relationship between two variables of interest, X and Y. in particular, the Linear Regression wants to estimate the change of the variable Y (the dependent variable), when the variable X changes to one unit (the independent variable). Mathematically, the relationship between X and Y, described by a linear regression, is represented by a strength line, with the following formula

$$Yi = \beta 0 + \beta 1 Xi + ui$$

where

- i is the index that refers to the observations, $i = 1, ..., n$;
- Y_i is the *dependent variable* or the *regressand*;
- X_i is the *independent variable*, or *regressor*;
- $\beta_0 + \beta_1 X$ is the *population regression line* or the *population regression function*;
- β_0 is the *interception* of the population regression line;
- β_1 is the *slope* of the population regression line;
- u_i is the *error term*.

As said before $\beta_0 + \beta_1 X$, represents the relationship between Y and X. This means

that once the value of X is known, it is possible to estimate the value of Y with the value of $\beta_0 + \beta_1 X$.

β_0 and β_1 are the *coefficients* of the population regression line, also called the *parameters*.

β_0 is the interception of the population regression line with the Y axis, so it is the estimated value of Y when X is 0, it does not always have a meaning in the real world, it depends on the economic application taken into consideration. β_1 is the slope of the strength line, so it represents the change in Y when the X changes by one unit, as below

$$\beta 1 = \frac{\Delta Y}{\Delta X}$$

The last term of the equation is the *error term* u_i, it includes all the other factors, different from the one represented by X, that deviates the value of the Y in the i^{th} observation from the value of Y predicted by the population regression line.

$$ui = Yi - (\beta 0 + \beta 1 Xi)$$

3.1.2 How do you estimate coefficients?

An important point of the population regression line is that, in a practical situation, the coefficients, β_0 and β_1, are unknown, for this reason, to realize the population regression line it is necessary to estimate the two parameters using a method of estimation, starting from the data available from the samples. The aim is to estimate the regression line using the estimators of β_0 and β_1, in order to represent the regression line as close as possible to the data coming from the observations, in other words, the two estimators chosen have to minimize the squared mistakes made in predicting Y using X. Supposing to represent the two coefficients β_0 and β_1 with their estimators b_0 and b_1, the regression line that predicts Y in the function of X is $b_0 + b_1 X$, and the mistake made by predicting the value of Y_i in the i^{th} observation is $u_i = Y_i - (b_0 + b_1 X_i)$, so the *sum of the squared prediction mistakes* over the *n* observations is

$$\sum_{i=1}^{n} (Yi - b0 - b1Xi)^2$$

Now the question is, what are the estimators b_0 and b_1 that minimize the *sum of the squared prediction mistakes*? The most common estimators used are the **OLS (Ordinary Lest Squares) Estimators** represented in that way $\hat{\beta}0$ (the interception) and $\hat{\beta}1$ (the solpe), they are calculated according with these formulas

$$\hat{\beta}1 = \frac{\sum_{i=1}^{n}(Xi - \bar{X})(Yi - \bar{Y})}{\sum_{i=1}^{n}(Xi - \bar{X})^2}$$

$$\hat{\beta}0 = \bar{Y} - \hat{\beta}1\bar{X}$$

where

- $\bar{Y}$ is the *least estimator* of the population mean *E(Y)*;
- $\bar{X}$ is the *least estimator* of the population mean *E(X)*

Using $\hat{\beta}0$ and $\hat{\beta}1$, the *OLS regression line* is called the **sample regression line** and is represented in that way

$$\hat{\beta}0 + \hat{\beta}1X$$

the value of Y_i predicted by the sample regression line is

$$\hat{Y}i = \hat{\beta}0 + \hat{\beta}1Xi$$

and the error term is

$$ui = Yi - \hat{Y}i$$

Thus, the final linear regression model is

$$Yi = \hat{Y}i + \hat{u}i$$

3.1.3 Measures of Fit

Once the coefficients are estimated and the OLS regression line is constructed, it is necessary to measure how well the estimated regression represents the real data. For this purpose, there are two main measures of fit to calculate and analyze: the R^2 and the *standard error of the regression (SER)*, described below.

1. The R^2

*"The **regression R^2** is the fraction of the sample variance of Y_i explained by X_i"* (Stock and Watson, 2012), so R^2 measures how much the regressor X_i is suitable to predict the value of the dependent variable Y_i. The mathematical formula to represent R^2 is the following one

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}i - \bar{Y}\right)^2}{\sum_{i=1}^{n}(Yi - \bar{Y})^2}$$

where

- The numerator is **the explained sum of squares (ESS)**, it is the sum of the squared deviations of the predicted value of Y_i, $\widehat{Yi}$ (i.e. $\hat{\beta}0 + \hat{\beta}1Xi$), from the sample average $\bar{Y}$, of all n observations.
- The denominator is **the total sum of squares (TSS)**, it is the sum of the squared deviations of Y_i from the sample average $\bar{Y}$, of all n observations.

R^2 can assume values with the interval that has as extremes 0 and 1. If it assumes a value near 1 it means that ESS = TSS, then the independent variable X_i is a good regressor to predict Y_i, if it assumes a value near 0 it means that ESS = 0, and this happens if $\hat{\beta}_1 = 0$ meaning that there is no correlation between the two variables, then in this case X_i is not a good regressor to predict Y_i.

2. Standard Error of the Regression (SER)

*"The **standard error of the regression** is an estimator of the standard deviation of the regression error u_i"* (Stock and Watson, 2012), it is measured in the same units of the dependent variable Y_i, as for u_i. In other words, it measures the average deviation of the observed data from the regression line, then by how much the value of Y_i, for each observation, deviates from its predicted value. Mathematically, it is represented with the formula

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}\, i}$$

where the term $n - 2$ is an adjustment of the bias introduced by the fact that two coefficients (β_0 and β_1) are estimated, this is called the *"degree of freedom"* correction.

3.2 Linear regression with Multiple Regressors

In the practical cases, it is rare that a variable is only influenced by another variable, in general there are different factors that are responsible for the predicted value of the dependent variable Y_i as well as the independent variable X_i. In the specifics of this thesis, the dependent variable, then the startups' survival expressed in terms of business duration, does not only depend on the independent variable, the stock of immigrants, but, as it is emerged from the background literature, there are other different factors to take in consideration, for this reason the model used will be the *linear regression with multiple regressors*.

3.2.1 Omitted Variable Bias

The omitted variable bias is a bias that affects the OLS estimators of the coefficients β_0 and β_1, when at least one of the two following conditions is verified:

1. the omitted variable is correlated with the included regressor;
2. the omitted variable is a determinant of the dependent variable.

If one of these two conditions occurs, one of the main assumptions of the least squares, that guarantee unbiased estimators, is not more valid. The assumption states that, the OLS estimators of β_0 and β_1 are unbiased when the error term u_i and the independent variable X_i are not correlated. When one of the two conditions is verified means that the error term includes some omitted variables that affect X_i, so the two terms, u_i and X_i, are correlated and the OLS estimators are inconsistent. It is necessary to eliminate the omitted variable bias by using a *multiple regression model*.

3.2.2 Multiple Regression Model and Control Variables

The multiple regression model is an extension of the single regression model that

considers more than one regressor. The model shows how to change the independent variable Y_i by changing one unit of the dependent variable X_{1i} and holding the other regressors constant.

The **population regression line**, for simplicity with two independent variables, but they can be more, is represented by the following formula

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

where

- β_0 has the same significant of the case of linear regression with one regressor, so it is the **interception** of the line with the Y axis and is the value of Y_i when X_{1i} and X_{2i} are equal to zero.
- β_1 and β_2 are respectively the **slope coefficients** of X_{1i} and X_{2i}

If the scope of the population regression line is to investigate the change of the dependent variable Y_i by changing one unit of the independent variable X_{1i} holding the independent variable X_{2i} constant, the variable X_{2i} is called **control variable**.

In accordance with what is said for the linear regression with one regressor, β_1 is the expected change in Y by changing one unit of X_1, so in this case it is possible to rewrite the formula (n formula) in the following manner

$$\beta_1 = \frac{\Delta Y}{\Delta X_1} \text{ holding } X_2 \text{ constant}$$

As for the case of regression with one regressor, the population regression line does not include all factors that influence the dependent variable *Y.* For this reason, even in this case, it is necessary to add a *"error term"* u_i that takes in consideration these factors, that has the same meaning in the case of regression with one regressor. In light of this, the **population multiple regression model is the following**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \qquad i = 1, \dots, n.$$

3.2.3 OLS estimators

As said for the linear regression with a single regression, the coefficients $\beta_0, \beta_1, \dots, \beta_k$ are unknown, so need to be estimated with the OLS method. Even in this case the OLS estimators are $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, the **population regression line** that predict the value of Y_i is

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki}$$

and the **residual error** is

$$\widehat{u}_i = Y_i - \widehat{Y}_i$$

So the **population multiple regression model** with OLS estimators is

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki} + \widehat{u}_i$$

3.2.4 Measure of Fit

As well as for the linear regression with one regressor, even in this case it is necessary to measure how well the estimated regression represents the real data, in the case of multiple regression there are three measures, below illustrated.

1. The R^2

The definition and formula of R^2 are the same as the ones in section 3.1.3

$$R^2 = \frac{ESS}{TSS}$$

R^2 can be also interpretated as 1 minus the fraction of the variance of Y_i does not explained by X_i, then the formula of R^2 can be rewritten in this way

$$R^2 = 1 - \frac{SSR}{TSS}$$

where SSR stand for **sum of squared residuals**, so it is

$$SSR = \sum_{i=1}^{n} \widehat{u_i^2}$$

In the case of multiple regression, if it is added a variable, that has an OLS estimator of its coefficient different from zero, SSR decreases and as a consequence R^2 increases. This can lead to the mistake of thinking that the fit of the multiple regression is improved, even if it is not the case. To bypass this bias it is introduced another fit measure called *Adjusted R^2*.

2. The "Adjusted" R^2

"Adjusted R^2 is a modified version of the R^2 that does not necessarily increase when a new regressor is added" (Stock and Watson, 2012). Mathematically, it is represented by the following formula

$$\overline{R^2} = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$

where

- n is the number of observations;
- k is the number of independent variables;
- $\frac{n-1}{n-k-1}$ is called the *degrees-of-freedom correction.*

As the formulas show, it is always true that $\overline{R^2} < R^2$.

3. Standard Error of the Regression (SER)

The definition of SER is the same as the one in the section 3.1.3, but there is a little difference in the formula

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

where $n - k - 1$ are the degree-of-freedoms correction.

In this case, it is divided for *n – k – 1,* rather than *n – 2*, the reason is that the denominator has the role of adjusting the downward bias introduced by estimating

the coefficients, then k indicates the number of coefficients estimated, in the case of single regressor $k = 2$, and then

$n - k - 1 = n - 2.$

When n is large enough, the effect of the degrees-of-freedom correction can be negligible.

4. Empirical Application

4.1 Data Construction

4.1.1 General description

The reference period in which the observations are carried out is between 2009 and 2019, this means that are taken in consideration the startups with a start date between 2009 and 2019. The choice of this period is backed by two reasons. Regarding the inferior limit of the interval, the reason is that it is the first year with a significant number of observations, useful to the scope of the analysis. The superior limit of the interval is chosen to mitigate the *sensoring problem*, namely the problem to consider died a startup that instead is still alive, only because it is born in a year close to the 2022 that is the last year of observation.

The construction of the *final database* provides the combination of two different datasets, the first database is the *Startups Database*, it includes all startups born between 1999 and 2022 in the territory of Lombardy, the second is the *Immigration Database*, it includes the number of women between 25 and 65 of age, residents in Lombardy, with a nationality different from the Italian one.

In the following sections the two datasets will be illustrated and finally the construction and the structure of the Final Database is shown.

4.1.2 Startups Database

The *Startups Database* is downloaded from the *Innovative Startups Register*, the register of the Ministero delle Imprese e del Made in Italy where are present all the startups on the Italian territory that match the requirements listed in Chapter 2.1. In Italy an innovative startup keeps its status for no more than five years, if after these five years is still alive it becomes a company, exiting from the register of the innovative startups and appearing on AIDA, that is an online dataset containing financial, master and commercial information of about 500.000 companies that operate in Italy. For this reason, the Innovative Startups Register, to the scope of this analysis, is linked to AIDA to keep track the startup's life even it has no more the startup status.

Database rows are the observations that are the startup, represented by an *id* associated with each year of observation of the startup, and 28 variables represented

in the 28 columns, which are explained in detail below together.

- ***id*** is the number that identifies univocally the startup.
- ***anno*** shows the year in which all the other variables are observed.
- ***anno_costituzione*** is the year in which the startup is founded.
- ***comune*** is the name of the municipality in which the startup has its headquarters.
- ***codice_istat_comune*** indicates the code provided by ISTAT, to the specific municipality.
- ***provincia*** states the name of the province to which the municipality belongs.
- ***codice_istat_provincia*** indicates the code provided by ISTAT, to the specific province;
- ***regione*** states the Italian region in which the municipality is located, in this case it is considered only the region of Lombardy as territorial scope of the analysis.
- ***stato_giuridico*** shows the legal status of the startup in the current year (2023), this information is obtained by matching the two databases, Innovative Startups Register and AIDA. This variable can assume two values: "attiva" when the startup/company is operative at the current moment, "cassata" or "in liquidazione" when the startup is died and stopped its business.
- ***data_cessazione*** represents the date in which the startup stopped its business, this data is provided by AIDA, for this reason the data is not available for startups that died before to exit from the Innovative Startups Register.
- ***ateco_2007_*** it is a code also called *code of economic activities*, provided by ISTAT, that classified the economic activities according with their industry of belonging, then these variable states the industry in which the startups operate.
- ***tot_immob_immateriali*** shows the amount of total intangible fixed assets owned by the startup for each year of observation.
- ***tot_immob_materiali*** shows the amount of total tangible fixed assets owned by the startup for each year of observation.
- ***tot_attivo*** indicates the total assets owned by the startup for each year of observation.
- ***tot_debiti*** states the total amount of debt that each startup has at each year of observation.

- ***capitale_social*** states the total amount of equity that each startup has in each year of observation.
- ***ricavi*** is the total turnover generated by the startup at the end of each year of observation.
- ***risultato_operativo*** indicates the operating income generated at the end of each year of observation for each startup.
- ***valore_aggiunto*** is the added value generated at the end of each year of observation for each startup.
- ***utile*** shows the income made by the startup at the end of each year of observation.
- ***ebitda*** for each startup at the end of each year of observation.
- ***dipendenti*** states the number of employees that work for the startup, in each year of observation.
- ***diritti_brevetto industrial*** indicates the number of patents owned by startups, this data is not available for all startups present in the database.
- ***data_iscrizione_registro_startups*** indicates the data in which the startup enters into the startups' register, it is possible that this date is not perfectly coincident with the date that indicates the startup's born, this due to bureaucratic steps that can slow down the inscription of the startup to the register.
- ***data_uscita_sezione_startup*** shows the date when a start-up ceases to be a start-up, with two possible fates; the startup became a company and exits from the startups' register to enter in AIDA, or the startup is dead.
- ***prevalenza_femminile_complessiva*** is a variable that indicates the level of women present in the Board of Directors and between shareholders. It can assume the four values reported in the following table

values	description
NO	(%women in BoD + %women in shareholders)/2 < 50%
MAGGIORITARIA	50% < (%women in BoD + %women in shareholders)/2 < 66%
FORTE	66% < (%women in BoD + %women in

	shareholders)/2 < 100%
ESCLUSIVA	(%women in BoD + %women in shareholders)/2 = 100%

Table 9: Legend of variable prevalenza_femminile_complessiva

- ***prevalenze_femminile_amm*** is a variable that indicates the level of women taking in consideration only by the Board of Directors. The roles to calculate that variable are the same of ***prevalenza_femminile_complessiva***, but as information take only %women in BoD.

4.1.3 Immigration Database

The *Immigration Database* is downloaded from ISTAT (Istituto Nazionale di Statistica), an Italian public organization that has the role to carry out different research in different public topics, one of this is the population census. The database is created in two steps the first step consists in downloading the immigrant population with the characteristics required by selecting the filter called "Cittadinanza" the label "Straniera", then the filter "Sesso" the label "Femmina", and then the selecting the range of age of interest, in this case the age is between 25 and 65, the download contains the amount of women of nationality different from the Italian one, between 25 and 65 of age for each municipality of the Italian territory, in particular for Lombardy, for each year from 2002 to 2019. The second step consists of the downloading of the data about the whole population present in each municipality across the years between 2002 and 2019. The procedure is the same as before, with the difference that for the filter "Cittadinanza" it is necessary to select "Totale" and the same for the filter "Sesso", in this case it is not necessary to filter for the age, the result is a database that contains the registered population in each Italian municipality from 2002 to 2019.

Then the two datasets are merged using STATA to obtain a database containing both the information about immigrants and the whole population. The database rows are observations composed by the name of the municipalities and the year of observation, and 6 columns representing the variables reported below.

- ***comune*** indicates the name of the municipality for consideration.
- ***anno_osservazione*** states the year in which the other variables are observed.
- ***pop_*** shows the amount of the whole population in the specific municipality

in the specific year of observation.

- ***imm_*** shows the number of immigrant women, with an age between 25 and 65, present in the specific municipality in the year of observation.
- ***codice_istat_comune*** indicates the code provided by ISTAT, to the specific municipality.

4.1.4 Final Database

The *Final Database* comes from a merger between the two previous datasets, with the addition of some new variables useful to the scope of the analysis. It shows 805 rows that corresponds to the 805 startups founded on the Lombardy territory between 2010 and 2016, and 45 columns that correspond to the 45 variables, some of which are the same as the previous two datasets, others have been added. The additional variables are explained below.

- **dprevalenza_femminile** is a dummy variable that values 0 when the variable *prevalenza_femminile*, cited above, is "NO", and values 1 when the variable *prevalenza_femminile* assumes the other values, then means that there is a presence of women on the board of administration. After the generation of this variable the row that have *dprevalenza_femminile* = 0 have been eliminated because the object of this analysis are the startups with a presence of women in the BoD.
- **data_fine** indicates the year of the date in which the startup stops its business, it is calculated in two different ways according to the availability of data. In particular, the date of startup cessation is provided by AIDA, this means that this information is available only for the startup that became companies, and then are registered in AIDA's database before dying. But it is possible that a startup stopped being active before entering AIDA's database, and for this reason the information is not available. For these startups the variable *data_fine* is imposed to be equal to the year of the date in which it is available the last balance sheet's information, plus one.
- ***duration*** is built as the difference between the variable *anno_costituzione* and *data_fine*, it represents the duration of the startup's life, and is the dependent variable of the linear regression model.
- ***ateco_2007_3d*** is the same variable as *ateco_2007_*, but it is taken with three digits, with a lower level of resolution to indicate the type of industry.

Before going on to illustrate the other variables, it is necessary to introduce them with

a general description. As already mentioned, the scope of the analysis is to understand how the stock of immigrants present in the municipality and in the year in which the startup is founded can influence its duration of survival, for this reason the information about the immigrants' stock and the amount of the whole population are taken in reference the year in which the startup is founded. For simplicity, as first step of analysis, also the other variables related to the balance sheet information of each startup are taken with reference to the year of startup foundation.

The variables *pop_* and *imm_* are split into two variables:

- ***pop_corrente*** shows the number of the whole population for each year of startup's observation.
- ***pop_iniziale*** shows the number of the whole population only for the year of startups' foundation.
- ***imm_corrente*** indicates the stock of immigrants for each year of startup's observation.
- ***imm_iniziale*** indicates the stock of immigrants only for the year of startups' foundation.
- ***imm_rate_iniziale*** indicates the rate of immigrants on the total population only for the year of startups' foundation.

In a second step, to create the third variant of linear regression, two additional variables are added to the final database:

- ***DEN_SLL_2011*** indicates the capital city of the Local Labor System (explained in detail in the next section) of which the municipality of startup's foundation is part.
- ***Imm_SLL_TOT_iniziali*** indicates the sum of immigrants present in all municipalities that are part of the Local Labor System of reference at the year of startup foundation.

4.2 Hypothesis formulation and model construction

Based on what came out from the literature review, the role of the woman in the house and family care, in particular family with young children, influences her participation in the labor market, and consequently also her career. A factor that can play an important role in this dynamic is the level of low-skilled immigration in a certain geographic area, the assumption is that low-skilled immigrants provide services for the house and family care changing the women's work's hours supply. Starting from this point, this thesis focuses the analysis on the geographical area of Lombardy's municipalities and on the female-led innovative startups, over a period of 10 years (2009-2019), to discover if there is a correlation between the amount of stock of immigrants in that area, and the startups' survival measured in terms of life duration. The analysis is conducted by creating three variations of linear regression, changing the independent variable to explore the topic under different points of view. The starting thesis tests the following hypothesis

H: *the increase in immigrants' stock in Lombardy's municipalities is positively correlated with female-led innovative startups' probability of survival, measured in terms of startups' life duration*.

Once the hypothesis is formulated, the statistical model is constructed, in this case it is used a multiple linear regression model, and it is the following

$$duration_i = \beta_0 + \beta_1 ImmStock_{jt} + \beta_2 X'_{ijt} + \varepsilon_{ijt}$$

where

- $duration_i$ is the dependent variable representing the duration of the *i* startup;;
- $ImmStock_{jt}$ is the independent variable, it represents the amount of stock of immigrant (explained in detailed in the next section);
- X'_{ijt} is a vector that include the control variable related to the *i* startup and the *j* municipality in the year *t*, they are illustrated in the next section;
- β_1 is the coefficient of the independent variable;
- β_2 is the coefficient of the control variables' vector;

- ε_{ijt} is the error term that include factors linked to the *i* startup, the *j* municipality and the year *t*.

To test the same hypothesis, a second linear regression is created. In particular, this variant takes into consideration as independent variable the rate of immigrants on the total population, the variable is explained in detail in the next section, to analyze what change if it is taken in consideration the stock of immigrants in proportion to the level of population in the place of the absolute value. The second version of the linear regression is as follows:

$$duration_i = \beta_0 + \beta_1 Immigrant\_Rate_{jt} + \beta_2 X'_{ijt} + \varepsilon_{ijt}$$

A third linear regression is carried out with a different level of geographical observation. For the previous two linear regressions, the observations are made at municipal level, namely, the dependent variables in both cases refer to the immigration's situation in the municipality where the startup is founded. This view presents a little imperfection, it assumes that the low-skilled immigrants offer their services in the same municipality where they live and where the startup is founded, this is not always true, because, of course, individuals can move in another municipality different from where they live to work, so can be *commuters*, the third linear regression wants to take in consideration this fact. To do so it is necessary to introduce the concept of *Local Labor System* (in Italian Sistemi Locali del Lavoro: SLL). The Local Labor System is defined by ISTAT as *"Local Labour Systems (LLS) represent a territorial grid whose boundaries, regardless of the administrative articulation of the territory, are defined using the flows of daily home/work trips"*, in other words they are represented by groups of municipalities within which individuals move every day to go to work and then back home. The assumption is that low-skilled immigrants can move within the LLS to offer their services, then it is necessary to take into consideration a variable that includes the level immigrants within the LLS and no longer only in the municipality of reference. The thesis is redefined in the following way

H: the increase in immigrants' stock in the Lombardy's Local Labor Systems is positively correlated with female-led innovative startups' survival, measured in terms of startups' life duration.

The linear regression is constructed that way

$$duration_i = \beta_0 + \beta_1 LLS\_immigrant_{jt} + \beta_2 X'_{ijt} + \varepsilon_{ijt}$$

With the dependent variable, the number of immigrants in the Local Labor System *j* at the year *t*, and all other control variables are unchanged.

4.3 Variables Description

As said till now, to create the statistical model, three types of variables have been identified, the dependent variable, the independent variable and the control variables, they are described in this section.

1. Dependent variable: *duration*

The dependent variable is the object of the observation, the object of observation of that work is the survival of female-led innovative startup, as measure of the "startup's survival" it is chosen the ***duration*** of the startup's life. The duration is calculated as the subtraction between the date of cessation, namely the data in which the company ceases to exist as such, provided by AIDA, and the date in which the startup was born, namely the date in which their enters in the startups register. It is important to undelight that not all startups have become companies, some of them died when they were still startups, then it is not possible to find the date of cessation on AIDA. In these cases, as date of business' end, it is taken the date of the last available economic accounts plus one, because if the startup can provide the economic accounts of the year it means that it was still alive and is dead the next year.

2. Independent variable: *imm_iniziale*

The independent variable is the variable that we change to understand how this can impact the dependent variable. In this case the scope is to understand how to change the duration of startups by changing the level of immigrants in the specific geographic area. In particular, as explained in the chapter of Data Construction, the selected sample of immigrants, is made by all women resident in Italy with a nationality different from the Italian one, between 25 and 65 years of age, because this is the sample that best represents the group of immigrants that offers the house-care work and that catches the effects nominated till now. This data is taken from ISTAT database, in particular in the linear regression is taken into consideration the stock of immigrants, with the previously explained characteristics, present in the municipalities where is registered the headquarters of the startup , at the year of the startup's foundation.

3. Control variables

The control variables are all elements that have some effects on the dependent variable

that is not included in the independent variable, for this reason it is important to individuate and explicate them in the linear regression to avoid the *omitted variable bias* (see chapter 3). The variables are selected from the previous analysis of the prior literature and are reported below.

3.1 ricavi_iniziali

The first control variable is the turnover of the startup declared at the end of the first year of its life. As said in the chapter 2.2, a good turnover has a positive impact on the startup's survival probability.

3.2 tot_debiti_iniziali

This variable is taken into consideration to extrapolate the effect of the financing received by the startup on the dependent variable, and it represents the level of startup's debt at its first year of life. The starting idea was to distinguish the startups by the type of investors, from the moment that the literature shows some evidence about this topic, by creating a dummy variables equals 1 when the startup is backed by a Business Angels group or a Venture Capital and equals 0 when the startup is self-financed or receives loan from banks, but this is not possible because the database of the startups does not give this information. Anyway, the effect of the financing choices of the startup can be well represented by the level of debts at its first year of life.

3.3 dipendenti_iniziali

The employee number is additional data available on the register of startups that indicate the number of employees that work in the startup for each year of observation, particularly in the linear regression this data is taken for the first year of startup life. This variable measures the size of the startup, to catch the effect that the size has on the dependent variable.

*3.4 sect**

This variable is created to catch the effect that the type of industry in which the startup operates has on its survival probability and then on its life's duration. It is a dummy variable that assumes the value of 1 when the startup operates in that specific sector, and zeros otherwise. The database shows in total 104 different type

of industry for the startup in the Lombardy's territory.

3.5 pop_iniziale

This variable represents the whole population present in the specific municipality where the startup is situated, in the first year of startup's life. It catches the effect of the demand on the dependent variable, so the fact that the startups tend to be established where there is higher demand and the higher the demand the higher the probability of survival and then the duration. But not only, this variable catches also the effect that the level of population has on the immigrants' influx, assuming that immigrants tend to go in geographic areas where there is higher jobs opportunity, and then in most populated areas.

Regarding the second variant of the linear regression the variables taken into consideration change a bit respect the first one. The control variables stay the same except for the variable *pop_iniziale*, it is no longer taken into consideration as control variable but it is included in the independent variable.

4. Independent variale: *imm_rate_iniziale*

The variable is constructed as the ratio between the number of the women immigrants between 25 and 65 age, at the year of startup duration and the total population at the same time, in the following way:

$$\frac{imm\_iniziale}{pop\_iniziale}$$

The expectation of this variable is that it can catch the effect that immigrants tend to go in area where already exist a higher presence of immigrants.

For the third variant of the linear regression it is considered another independent variable:

5. Independent variable: *imm_SLL_TOT_iniziale*

This variable is the number of women between 25 and 65 age with a nationality different from the Italian's one presents in the specific Local Labor System. It is not an information directly available, it has been built in the following way:

First, it is downloaded from ISTAT the database that associate each municipality to the respective Local Labor System of belonging, represented by the name of the capital city of the LLS, then for each LLS, it is made the sum of the immigrants present in each municipality in the LLS, so the amount come from the sum is the amount of immigrants women between 25 and 65 years in the specific LLS. Once the information is obtained it is added to the final database with all other variables. In particular, the information is taken during the year of startup's foundation as for other information already cited.

4.4 Descriptive Statistics

This chapter will summarize the most important variables of the database that are used in the linear regression model.

First, it will summarize the variable that gives an idea of the distribution of the startups' born across the time and the territory and then it will summarize the most important continuous variables.

1. anno_costituzione

anno_costituzione	Freq.	Percent	Cum.
2009	2	0.25	0.25
2010	3	0.37	0.62
2011	17	2.11	2.73
2012	22	2.73	5.47
2013	56	6.96	12.42
2014	71	8.82	21.24
2015	96	11.93	33.17
2016	106	13.17	46.34
2017	114	14.16	60.50
2018	155	19.25	79.75
2019	163	20.25	100.00
Total	805	100.00	

Figure 1: Frequencies of observations in each year

The variable *anno_costituzione*, is the year in which the startup is founded. The interpretation of the frequency of this variable represents the number of startups founded in that specific year, in detail, the snapshot shows a very low starting number for the first two years (2009-2010) to increase in continuous manner each following year.

2. comune

Once the distribution across the years of the startups' foundation is summarized, it is interesting to exhibit the territorial distribution of the startups across the

Lombardy territory. The description of the variable *comune* is the following

comune	Freq.	Percent	Cum.
Abbiategrasso	1	0.12	0.12
Agrate Brianza	2	0.25	0.37
Annone di Brianza	1	0.12	0.50
Asola	1	0.12	0.62
Assago	3	0.37	0.99
Azzano Mella	1	0.12	1.12
Azzano San Paolo	1	0.12	1.24
Bagnolo San Vito	1	0.12	1.37
Bareggio	2	0.25	1.61
Bellinzago Lombardo	1	0.12	1.74
Bergamo	30	3.73	5.47
Biassono	1	0.12	5.59
Bodio Lomnago	1	0.12	5.71
Bollate	1	0.12	5.84
Bovezzo	1	0.12	5.96
Brembate di Sopra	1	0.12	6.09
Brescia	25	3.11	9.19
Bresso	2	0.25	9.44
Brusaporto	1	0.12	9.57
Buccinasco	1	0.12	9.69
Busto Arsizio	4	0.50	10.19
Calcio	1	0.12	10.31
Candia Lomellina	1	0.12	10.43
Cantù	3	0.37	10.81
Cappella Cantone	1	0.12	10.93
Capriate San Gervasio	1	0.12	11.06
Carbonate	1	0.12	11.18
Casalmaggiore	1	0.12	11.30
Cassina de' Pecchi	1	0.12	11.43
Castenedolo	1	0.12	11.55
Ceresara	1	0.12	11.68
Cermenate	1	0.12	11.80
Cernusco sul Naviglio	1	0.12	11.93
Cesano Maderno	3	0.37	12.30
Chiari	1	0.12	12.42
Cinisello Balsamo	1	0.12	12.55
Cislago	1	0.12	12.67
Cividate al Piano	1	0.12	12.80
Cologno Monzese	1	0.12	12.92
Comezzano-Cizzago	1	0.12	13.04
Como	7	0.87	13.91
Cormano	2	0.25	14.16
Corsico	2	0.25	14.41
Costa Volpino	1	0.12	14.53
Crema	4	0.50	15.03
Cremona	3	0.37	15.40
Cura Carpignano	3	0.37	15.78
Curno	1	0.12	15.90
Cusago	2	0.25	16.15
Dalmine	1	0.12	16.27
Daverio	1	0.12	16.40
Desenzano del Garda	4	0.50	16.89
Erbusco	2	0.25	17.14
Esino Lario	2	0.25	17.39
Faloppio	1	0.12	17.52
Fino Mornasco	1	0.12	17.64
Foresto Sparso	1	0.12	17.76
Gallarate	5	0.62	18.39
Garbagnate Milanese	1	0.12	18.51
Gavirate	1	0.12	18.63
Germignaga	1	0.12	18.76
Gorgonzola	1	0.12	18.88

Grassobbio	1	0.12	19.01
Iseo	1	0.12	19.13
Landriano	1	0.12	19.25
Lecco	6	0.75	20.00
Legnano	4	0.50	20.50
Leno	2	0.25	20.75
Lentate sul Seveso	1	0.12	20.87
Liscate	1	0.12	20.99
Lissone	2	0.25	21.24
Lodi	3	0.37	21.61
Lomazzo	2	0.25	21.86
Lonate Ceppino	1	0.12	21.99
Lumezzane	1	0.12	22.11
Magenta	1	0.12	22.24
Malnate	3	0.37	22.61
Mantova	3	0.37	22.98
Meda	1	0.12	23.11
Merate	1	0.12	23.23
Milano	521	64.72	87.95
Missaglia	1	0.12	88.07
Monza	12	1.49	89.57
Mornico al Serio	1	0.12	89.69
Muggiò	1	0.12	89.81
Nembro	2	0.25	90.06
Offanengo	1	0.12	90.19
Opera	1	0.12	90.31
Osio Sotto	1	0.12	90.43
Ostiglia	1	0.12	90.56
Palazzolo sull'Oglio	1	0.12	90.68
Palosco	1	0.12	90.81
Pandino	1	0.12	90.93
Pavia	7	0.87	91.80
Pioltello	1	0.12	91.93
Prevalle	1	0.12	92.05
Quistello	1	0.12	92.17
Rho	2	0.25	92.42
Robbiate	1	0.12	92.55
Roncadelle	1	0.12	92.67
Rozzano	1	0.12	92.80
Saltrio	1	0.12	92.92
Salò	1	0.12	93.04
Samarate	1	0.12	93.17
San Giuliano Milanese	1	0.12	93.29
San Martino Siccomario	1	0.12	93.42
San Vittore Olona	1	0.12	93.54
San Zenone al Po	1	0.12	93.66
Sarnico	1	0.12	93.79
Saronno	5	0.62	94.41
Segrate	3	0.37	94.78
Seregno	3	0.37	95.16
Sesto San Giovanni	2	0.25	95.40
Settala	1	0.12	95.53
Seveso	1	0.12	95.65
Sirmione	1	0.12	95.78
Solaro	1	0.12	95.90
Sospiro	1	0.12	96.02
Spino d'Adda	1	0.12	96.15
Stezzano	2	0.25	96.40
Stradella	1	0.12	96.52
Sulbiate	1	0.12	96.65
Temù	1	0.12	96.77
Travagliato	3	0.37	97.14
Treviglio	12	1.49	98.63

Varedo	1	0.12	98.76
Varese	2	0.25	99.01
Varzi	1	0.12	99.13
Vedano Olona	1	0.12	99.25
Vergiate	1	0.12	99.38
Vigevano	1	0.12	99.50
Villasanta	1	0.12	99.63
Zanica	2	0.25	99.88
Zelo Buon Persico	1	0.12	100.00
Total	805	100.00	

Figure 2: Frequencies of observations for each municipality

From the snapshot it is possible to see how the location of startups' foundation is quite distributed across the region, it is possible to see this from the fact that the table shows a high number of municipalities in which the startups are founded, with a low frequency of observations for each, so a low number of startups founded in the municipality of reference. As might be expected, there is a high difference in frequency of observations between Milan and all other municipalities, Milan shows a number of observations equals to 521 startups, followed by Bergamo and Brescia with respectively 30 and 25 startups.

3. duration

It is interesting also to see the summary of the dependent variable, then the *duration* variable.

duration	Freq.	Percent	Cum.
1	3	0.37	0.37
2	174	21.61	21.99
3	165	20.50	42.48
4	117	14.53	57.02
5	346	42.98	100.00
Total	805	100.00	

Figure 3: Frequencies of observation for variable duration

In this first snapshot it is possible to see that the range of startups life's duration is between 1 year and 5 years. This does not mean that there is not startups that have a duration higher than 5 years, but it is part of the structure of the model, as

for the purpose of the research it is chosen to select as maximum value of duration 5 years, this is obtained by replacing the value of 5 years for the variable *duration* when the variable exceeded 5 years, this to avoid the already mentioned *sensoring problem*. By not taking into account that value, it is possible to see that the duration that presents a higher frequency is two years, with 174 startups that have a life of 2 years, followed by the value of 3 years with 165 startups.

Variable	Obs	Mean	Std. Dev.	Min	Max
duration	**805**	**3.781366**	**1.219412**	**1**	**5**

Figure 4: Variable duration summarized

This second snapshot shows the mean of duration of 3.7 years, with a relatively low standard deviation of 1.219412. This low value of standard deviation does not totally respect the reality because the maximum value is constructed for the model and also includes the value of duration higher than 5. Therefore, in reality the standard deviation can be slightly higher.

4. ricavi_iniziali

Variable	Obs	Mean	Std. Dev.	Min	Max
ricavi_ini~i	**805**	**30.67081**	**182.8951**	**0**	**3120**

Figure 5: Variable ricavi_iniziali summarized

The variable *ricavi_iniziali* indicates the amount of turnover generated by each startup at the end of its first year of life, they need to be considered in thousands. It shows a mean of about 31k euros, the standard deviation is high (about 183k). This can be explained by the values shown by min and the max indicators. There

are startups that concluded the first year with zero turnover, and others realize higher turnover for a max of 3120k euros.

5. *tot_debiti_iniziali*

Variable	Obs	Mean	Std. Dev.	Min	Max
tot_debiti~i	**805**	**45.05839**	**150.1861**	**0**	**2493**

Figure 6: Variable tot_debiti_iniziali summarized

The variable *tot_debiti_iniziali* indicates the amount of debt owned by each startup in its first year of life, even in this case the value must be considered in thousands. This variable presents a similar behavior to *ricavi_iniziali.* It has a mean of about 45k euros, and a high standard deviation (about 150k). Even in this case the reason can be attributed to the fact that some startups don't ask debt for their first year of life, instead others ask for a high value, with a max of 2493k euros.

5. *dipendenti_iniziali*

Variable	Obs	Mean	Std. Dev.	Min	Max
dipendenti~i	**802**	**.3453865**	**1.542299**	**0**	**24**

Figure 7: Variable dipendenti_iniziali summarized

The *dipendenti_iniziali* variable states the number of employees that work for the startup in its first year of life. The mean is between 0 and 1 employee and the standard deviation is low with a value of 1.5. Taking into consideration the min and the max values it is possible to see that in general the number of employees for all startups remains low, with a maximum of 24 employees, even if, observing the mean and standard deviation, it is possible to hypothesize that in the most of case the number of employees for the first year is very low.

6. *imm_iniziale*

Variable	Obs	Mean	Std. Dev.	Min	Max
imm_iniziale	**805**	**57780.08**	**40818.61**	**6**	**91089**

Figure 8: Variable imm_iniziale summarized

The variables *imm_iniziale* states the stock of women of age between 25 and 65 with a nationality different from the Italian one. The mean value is 57780, with a very high standard deviation (40818.62). The reason behind this high value is the same already mentioned for the previous cases, namely it is due to the high difference between the min and max, in this case the difference is particularly evident, because there are municipalities with a very low number of immigrants, with a minimum of six immigrants, and other municipalities with high number of immigrants, for a maximum of 91089 immigrants.

7. *pop_iniziale*

Variable	Obs	Mean	Std. Dev.	Min	Max
pop_iniziale	**805**	**897224.1**	**622676.7**	**545**	**1395980**

Figure 9: Variable pop:iniziale summarized

The variable *pop_iniziale* represents the level of population in each municipality of the database. This variable presents a similar behavior to the previous one. The mean is equal to 897224,1, and the standard deviation is 622676.7. Even in this case the very high standard deviation is due to the huge difference between the min and the max value of population.

8. imm_rate_iniziale

Variable	Obs	Mean	Std. Dev.	Min	Max
imm_rate_i~e	**805**	**.0563335**	**.0140503**	**.007947**	**.0747502**

Figure 10: Variable imm_rate_inizialie summarized

The variable *imm_rate_iniziale* represents rate of female immigrants of age between 25 and 65 in the total population present in the municipality of interest at the year of the startup's foundation. It shows a mean of 5% of population represented by women immigrants of age between 25 and 65, and a low standard deviation of about 0.014. The low standard deviation can be explained by the not so high difference between the min and the max of variable's values that are respectively about 0.07% and 7%.

As already said in the previous sections, to create the third variant of linear regression two variables have been added to the model, they are summarized and discussed below.

9. DEN_SLL_2011

The variable *DEN_SLL_2011* shows the name of the capital city of the Local Labor System of reference, the municipalities that belong to the same LLS show the same value of this variable. In the operation to addition that variables the number of observations decrease from 805 to 791, this because, when the database of the LLS is added to the Final Database, 14 municipalities' name of the Final Database did not match with the municipalities' name of the LLS database, this is not so strange because there are municipalities that have names that can change a bit from a database to another. Anyway, from the moment the 14 observations are not so relevant to the final goal of the analysis they are delayed. Moreover, it is important to undelight that look at the summary of the variable DEN_SLL_2011, the Local Labor System present in the Final Database are 29 instead than the 56 total LLS present in Lombardy, the reason is that in the period between 2009 and 2019 in only 29 LLS are born startups.

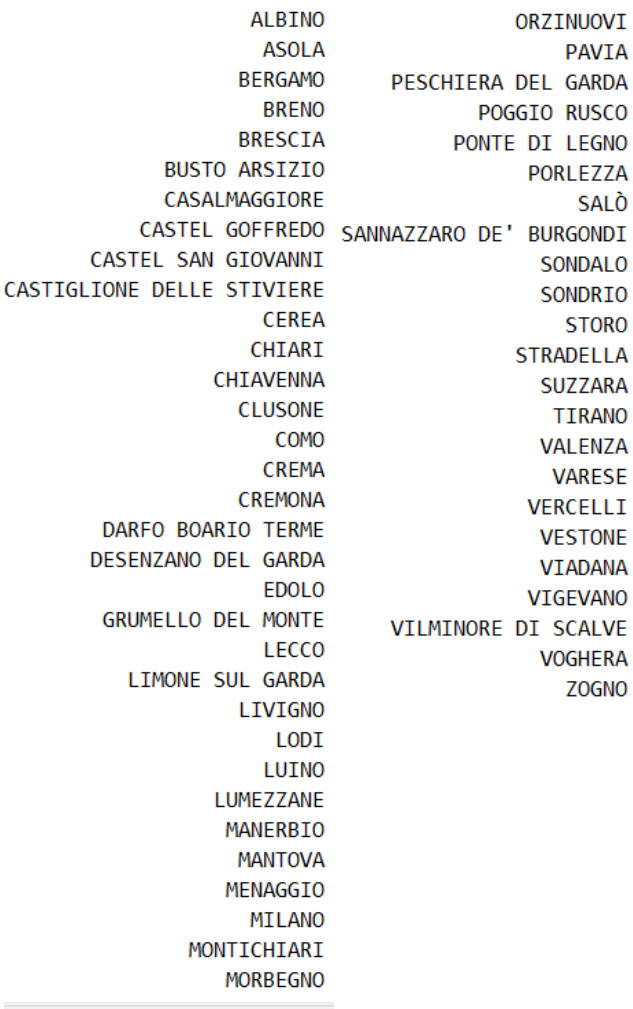

Figure 11: LLS of Lombardy in 2011

The figure 10 shows the 29 Local Labor Systems present in the Final Database, with the respective frequency of observation, namely the number of startups founded in the specific LLS between 2010 and 2019. As it was possible to image, the startups founded in Milan's LLS are significantly larger in number than all other LLSs.

DEN_SLL_2011	Freq.	Percent	Cum.
ALBINO	2	0.25	0.25
ASOLA	1	0.13	0.38
BERGAMO	53	6.70	7.08
BRESCIA	32	4.05	11.13
BUSTO ARSIZIO	16	2.02	13.15
CASALMAGGIORE	1	0.13	13.27
CASTEL GOFFREDO	1	0.13	13.40
CHIARI	6	0.76	14.16
COMO	15	1.90	16.06
CREMA	7	0.88	16.94
CREMONA	4	0.51	17.45
DARFO BOARIO TERME	1	0.13	17.57
DESENZANO DEL GARDA	1	0.13	17.70
GRUMELLO DEL MONTE	3	0.38	18.08
LECCO	11	1.39	19.47
LODI	3	0.38	19.85
LUINO	1	0.13	19.97
LUMEZZANE	1	0.13	20.10
MANERBIO	2	0.25	20.35
MANTOVA	4	0.51	20.86
MILANO	593	74.97	95.83
PAVIA	11	1.39	97.22
POGGIO RUSCO	2	0.25	97.47
PONTE DI LEGNO	1	0.13	97.60
SALÒ	2	0.25	97.85
STRADELLA	1	0.13	97.98
VARESE	12	1.52	99.49
VIGEVANO	3	0.38	99.87
VOGHERA	1	0.13	100.00
Total	791	100.00	

Figure 12: LLS in the Finale Database

10. imm_SLL_TOT_iniziali

Variable	Obs	Mean	Std. Dev.	Min	Max
imm_SLL_TO~i	791	134460.2	69052.23	100	181933

Figure 13: Variable imm_SLL_TOT_iniziali summarized

This variable shows the number of immigrants present in each Local Labor System during the year of startups' foundations. As it was possible to predict, its behavior is similar to the variable *imm_iniziale*'s one, from the moment

that, at it is explained in Chapter 4.3, the variable *Imm_SLL_TOT_iniziali* is constructed as the sum of the variable *imm_iniziale*. Therefore, it presents a very high standard deviation due to the high difference between the higher value (181933) and the lower value (100).

5. Results

This chapter will be shown and commented the results obtained from the three linear regressions.

5.1 Variant 1

5.1.1 Model's Fit

Source	SS	df	MS	Number of obs	=	801
				F(108, 692)	=	1.11
Model	175.185029	108	1.6220836	Prob > F	=	0.2163
Residual	1007.30935	692	1.45564935	R-squared	=	0.1481
				Adj R-squared	=	0.0152
Total	1182.49438	800	1.47811798	Root MSE	=	1.2065

Figure 14: Analysis of model's fit Variant 1

This framework provides an overview of the fit of the model, then how well the estimated regression represents the real data. The table on the left exhibits the analysis of the variance for the regression, in particular the values of the *SS (Sum of Squares)* and the *MS (Mean Squares)*, that are used to calculate the measures of fit illustrated and analyzed below. As argued in Chapter 3, there are three measures of fitness to consider: R^2, $\overline{R^2}$ and *SER* (Standard Error of Regression). In this case R^2 is equal to 0.1481, this means that the variables chosen to explain the dependent variable, namely the duration of the startup, explain the 14.81% of its variance, and $\overline{R^2}$ is equal to 0.0152, decreasing the portion of dependent variable's variance explained by the regression's variables to 1.52%. The two values show a low fit of the model, this means that there are other variables, that are not taken in consideration in this model, that influence the startups' duration. This is not a totally surprise, as highlighted in the Chapter 2, the variables that come into play when it comes to startups, in particular when we talk about the startups' survival, are really a lot, and many of these are unquantified, such as the characteristics of the founders, for this reason it is difficult to consider them in the linear regression model.

The last measure of fit is the *SER*, which in the output is called *Root MSE*, is equal to 1.2065, this means that the standard deviation of the regression residuals is 1.2065, it is a low value, so the standard deviation of the regression residuals is low.

5.1.2 Coefficients' Analysis

Once the model's fit has been discussed, in this section the results of the linear regression will be interpreted by analyzing the coefficients of the independent and control variables. First, it is necessary to say that for this regression it is chosen to use the so called ***linear-log model***. The linear-log model is a particular regression model in which the independent and the control variables are transformed by taking its logarithm, instead the dependent variable is kept linear. Using this type of model, it is possible to analyze how much is the change of the dependent variable in terms of unit by changing the independent variable of one percent.

To understand the impact of the independent and the control variables on the dependent variable, it is necessary to look at the coefficients shown in the results of the linear-log regression reported in the following images.

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_imm_iniziale	-.6187129	.2966318	-2.09	0.037	-1.201119	-.0363066
ln_pop_iniziale	.7301652	.3397665	2.15	0.032	.0630684	1.397262
sect1	4.019263	1.758942	2.29	0.023	.565761	7.472766
sect2	3.593334	1.747857	2.06	0.040	.1615957	7.025073
sect3	2.305645	1.352861	1.70	0.089	-.3505597	4.96185
sect4	4.076291	1.754514	2.32	0.020	.6314809	7.5211
sect5	.6143894	1.751253	0.35	0.726	-2.824018	4.052796
sect6	4.036519	1.747389	2.31	0.021	.6056994	7.467338
sect7	2.882432	1.51537	1.90	0.058	-.0928419	5.857706
sect8	2.469149	1.728192	1.43	0.154	-.9239801	5.862277
sect9	1.421325	1.740677	0.82	0.414	-1.996315	4.838966
sect10	1.880072	1.393486	1.35	0.178	-.8558968	4.61604
sect11	2.047287	1.518826	1.35	0.178	-.9347731	5.029347
sect12	3.085377	1.439127	2.14	0.032	.2597985	5.910955
sect13	3.862214	1.754089	2.20	0.028	.4182387	7.30619
sect14	3.648006	1.743679	2.09	0.037	.2244694	7.071542
sect15	1.991808	1.746367	1.14	0.254	-1.437006	5.420622
sect16	1.707903	1.744792	0.98	0.328	-1.717819	5.133624
sect17	1.664397	1.387219	1.20	0.231	-1.059265	4.38806
sect18	3.707146	1.731255	2.14	0.033	.3080042	7.106288
sect19	2.014679	1.50958	1.33	0.182	-.949227	4.978585
sect20	2.221892	1.355326	1.64	0.102	-.4391529	4.882937
sect21	.7459962	1.731017	0.43	0.667	-2.65268	4.144672
sect22	4.322135	1.742812	2.48	0.013	.9003011	7.743969
sect23	3.675177	1.749365	2.10	0.036	.2404774	7.109876
sect24	2.020639	1.519104	1.33	0.184	-.9619681	5.003245
sect25	1.854761	1.73311	1.07	0.285	-1.548022	5.257545
sect26	2.492422	1.392517	1.79	0.074	-.2416423	5.226486
sect27	2.804205	1.362961	2.06	0.040	.1281697	5.480239
sect28	3.621443	1.746187	2.07	0.038	.1929829	7.049904
sect29	2.095692	1.435392	1.46	0.145	-.7225541	4.913937
sect30	2.760211	1.378621	2.00	0.046	.0534291	5.466994

sect31	3.000666	1.350478	2.22	0.027	.3491403	5.652192
sect32	3.739082	1.746945	2.14	0.033	.309133	7.169031
sect33	3.49937	1.732441	2.02	0.044	.0978994	6.90084
sect34	2.609051	1.420775	1.84	0.067	-.1804954	5.398598
sect35	2.156717	1.431832	1.51	0.132	-.6545389	4.967972
sect36	2.816624	1.364924	2.06	0.039	.1367358	5.496512
sect37	2.096451	1.369867	1.53	0.126	-.5931431	4.786044
sect38	2.346236	1.721907	1.36	0.173	-1.034554	5.727025
sect39	2.138241	1.756875	1.22	0.224	-1.311204	5.587685
sect40	1.676567	1.440034	1.16	0.245	-1.150792	4.503926
sect41	1.922344	1.51728	1.27	0.206	-1.056682	4.901369
sect42	3.338463	1.754064	1.90	0.057	-.1054628	6.782388
sect43	1.906027	1.745706	1.09	0.275	-1.521489	5.333542
sect44	2.859272	1.439939	1.99	0.047	.0320992	5.686445
sect45	2.918768	1.39285	2.10	0.036	.1840487	5.653487
sect46	3.017035	1.518352	1.99	0.047	.0359064	5.998165
sect47	3.652222	1.742293	2.10	0.036	.2314079	7.073037
sect48	.6584371	1.743457	0.38	0.706	-2.764663	4.081537
sect49	1.902291	1.754943	1.08	0.279	-1.54336	5.347942
sect50	3.661985	1.732845	2.11	0.035	.2597201	7.06425
sect51	2.923669	1.321896	2.21	0.027	.3282598	5.519077
sect52	2.491448	1.331295	1.87	0.062	-.1224141	5.10531
sect53	2.902846	1.434233	2.02	0.043	.0868765	5.718816
sect54	3.024827	1.5319	1.97	0.049	.0170972	6.032558
sect55	3.355029	1.740876	1.93	0.054	-.0630044	6.773062
sect56	2.294808	1.436342	1.60	0.111	-.5253034	5.11492
sect57	2.029409	1.275585	1.59	0.112	-.4750724	4.53389
sect58	2.70337	1.512576	1.79	0.074	-.2664182	5.673159
sect59	3.52981	1.721673	2.05	0.041	.1494809	6.910139
sect60	1.928563	1.746271	1.10	0.270	-1.500062	5.357188
sect61	2.274009	1.393681	1.63	0.103	-.4623413	5.01036
sect62	3.427597	1.740692	1.97	0.049	.0099252	6.845269
sect63	3.728209	1.730288	2.15	0.032	.3309653	7.125452
sect64	3.355029	1.740876	1.93	0.054	-.0630044	6.773062
sect65	2.688607	1.325973	2.03	0.043	.0851953	5.29202

sect66	2.514122	1.304462	1.93	0.054	-.0470554	5.0753
sect67	2.159401	1.366355	1.58	0.114	-.5232967	4.8421
sect68	0	(omitted)				
sect69	2.420896	1.740677	1.39	0.165	-.9967456	5.838538
sect70	2.243323	1.252649	1.79	0.074	-.2161256	4.702771
sect71	2.320184	1.258603	1.84	0.066	-.1509553	4.791323
sect72	2.965519	1.437046	2.06	0.039	.1440261	5.787011
sect73	3.508512	1.517429	2.31	0.021	.529196	6.487829
sect74	3.41272	1.743892	1.96	0.051	-.0112351	6.836674
sect75	3.674154	1.743921	2.11	0.035	.2501428	7.098166
sect76	.4170897	1.517269	0.27	0.783	-2.561913	3.396092
sect77	2.555884	1.271732	2.01	0.045	.058967	5.0528
sect78	2.624335	1.315733	1.99	0.046	.041028	5.207642
sect79	2.668694	1.512057	1.76	0.078	-.300076	5.637465
sect80	2.526309	1.260531	2.00	0.045	.0513852	5.001233
sect81	1.985148	1.333025	1.49	0.137	-.6321105	4.602406
sect82	2.491301	1.318183	1.89	0.059	-.0968171	5.079419
sect83	-.5634006	1.747192	-0.32	0.747	-3.993834	2.867032
sect84	1.80865	1.289874	1.40	0.161	-.7238865	4.341187
sect85	3.412173	1.740672	1.96	0.050	-.0054593	6.829805
sect86	2.929758	1.292593	2.27	0.024	.3918828	5.467633
sect87	1.899428	1.517275	1.25	0.211	-1.079587	4.878444
sect88	2.93446	1.394913	2.10	0.036	.1956909	5.673229
sect89	.5930038	1.51038	0.39	0.695	-2.372474	3.558482
sect90	3.210282	1.333967	2.41	0.016	.591173	5.829391
sect91	.5761203	1.745324	0.33	0.741	-2.850646	4.002887
sect92	2.245201	1.307494	1.72	0.086	-.3219291	4.812332
sect93	2.031205	1.326026	1.53	0.126	-.5723109	4.634722
sect94	1.60716	1.518993	1.06	0.290	-1.375228	4.589548
sect95	3.686139	1.745502	2.11	0.035	.2590248	7.113253
sect96	3.30388	1.741428	1.90	0.058	-.1152366	6.722996
sect97	2.52271	1.753837	1.44	0.151	-.9207701	5.966189
sect98	3.50481	1.740785	2.01	0.044	.0869555	6.922664
sect99	3.442078	1.424379	2.42	0.016	.6454556	6.2387
sect100	1.613923	1.728818	0.93	0.351	-1.780434	5.008281
sect101	3.368398	1.517407	2.22	0.027	.3891244	6.347671
sect102	2.009326	1.510861	1.33	0.184	-.9570956	4.975748
sect103	3.476313	1.740664	2.00	0.046	.0586971	6.893929
sect104	1.652886	1.519729	1.09	0.277	-1.330947	4.636718
ln_ricavi_iniziali	-.0477657	.0347331	-1.38	0.170	-.1159607	.0204293
ln_tot_debiti_iniziali	-.070282	.0328595	-2.14	0.033	-.1347982	-.0057657
ln_dipendenti_iniziali	.3469872	.1270463	2.73	0.006	.0975448	.5964296

Figure 15: Results of linear regression Variant 1

Starting from the independent variable *ln_imm_iniziale* the results show that, contrary to expectations, the coefficient has a negative sign with a value of -0.6187129, this mean that there is a weak negative relationship between the presence of immigrants at the moment of startup's foundation and the duration of the startup's life. To understand if this result can be considered statistically consistent it is necessary to look at the *p-value.* The p-value corresponds to the probability to reject the null hypothesis H_0, considered true. In this case, in which the purpose is to test if the results of the linear regression are consistent, the null hypothesis H_0 tested is: *"the two variables are not statistically significant"*, in

other words it is tested if the variable's coefficient is zero, so it is possible to be written: $H_0: \beta_1 = 0$. To carry out an analysis of the results' statistical significance it is necessary to compare the p-value with the *interval of confidence,* that in this case is at 95%. Returning to the specific case of the independent variable *ln_imm_iniziale*, it shows a p-value of about 3.7%, it is lower than 5%, this means that it possible to reject the null hypothesis that the two variables (*durantion* and *ln_imm_iniziali*) are not statistically significant, in other words, it is possible to reject the hypothesis that the variable *ln_imm_iniziali* does not influence the variable *durantion*. The same reasoning can be applied to the control variables. In particular, the two variables, *ln_ricavi_iniziali* and *ln_tot_debiti_iniziali*, show a similar behavior to the independent variable, their coefficient present negative signs, respectively -0.0477657 and -0.070282 variables and the dependent variable. Looking at the statistical significance of these two results, the variable *ln_tot_debiti_iniziali* has a p-value equals to 3.3%, very close but still lower than 5%, so it is possible to reject the null hypothesis that *ln_tot_debiti_iniziali* and *duration* are not statistically consistent. Instead, the *ln_ricavi_iniziali*'s p-value is about 17% (higher than 5%), so the result about this variable is not statistically significant because it is not possible to reject the null hypothesis that the variables *duration* and *ln_ricavi_iniziali* are not statistically significant. On the other hand, the control variables *ln_pop_iniziale* and *ln_dipendenti_iniziali* present coefficients with positive signs, respectively 0.7301652 and 0.3469872, this means that there is a positive relation with the dependent variable. Both the results can be considered statistically consistent from the moment that for both variables the p-value is lower that 5%, respectively 3.2% and 0.6%.

Looking at the dummy variables *sect** it is possible to conclude a general result, the majority of these variables show a positive sign indicating that the duration has a positive relationship with the majority of these industries identified by the variable sect* of reference.

5.2 Variant 2

5.2.1 Model's Fit

Source	SS	df	MS			
				Number of obs	=	801
				F(107, 693)	=	1.07
Model	167.85509	107	1.56873916	Prob > F	=	0.3053
Residual	1014.63929	693	1.46412596	R-squared	=	0.1420
				Adj R-squared	=	0.0095
Total	1182.49438	800	1.47811798	Root MSE	=	1.21

Figure 16: Analysis of model's fit Variant 2

As for the first regression, this framework provides the analysis of the variance (on the left) and the measures of fit of the model (on the right). In this case, R^2 is equal to 0.1420, that the variables inserted in the regression explain the 14.2% of the variance of the dependent variable and the $\overline{R^2}$ is equal to 0.0095, decreasing the portion of dependent variable's variance explained by the regression's variables to 0.95%. even in this case the two measures do not show a high level of fit of the model, and even in this case the reason can be attributed to the fact the duration of the startup is affected by a high number of variables, and many of these are unquantified. Comparing these two values with the first regression's ones, it is possible to see that they are very similar, taking the analysis to the same conclusions. So, this means that the changing of the variable of the absolute value of the immigrants' stock with their rate respect to the entire population, doesn't increase the fitting of the regression, then the new variable doesn't explain a larger portion of the dependent variable's variance compared to what the old did. The same reasoning can be applicable to the *SER*, even in this case its value is very similar to the previous one (1.21 vs 1.206) and shows a low fitting of the model.

5.2.2 Coefficients' Analysis

After the analysis of the model's fit, the study of the regression's coefficients will be carried out. Even in this case it is used a ***linear-log regression***, to consider the variation of the dependent variables in percentage terms.

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_imm_rate_iniziale	-1.796194	3.575744	-0.50	0.616	-8.816786	5.224398
sect1	3.889974	1.764524	2.20	0.028	.4255202	7.354428
sect2	3.443223	1.752199	1.97	0.050	.0029679	6.883478
sect3	2.092428	1.355413	1.54	0.123	-.5687801	4.753636
sect4	3.955026	1.760091	2.25	0.025	.4992749	7.410777
sect5	.6126753	1.755983	0.35	0.727	-2.835009	4.06036
sect6	3.813141	1.750709	2.18	0.030	.3758105	7.250472
sect7	2.88792	1.51977	1.90	0.058	-.0959857	5.871825
sect8	2.453052	1.733209	1.42	0.157	-.9499176	5.856022
sect9	1.396295	1.745705	0.80	0.424	-2.03121	4.8238
sect10	1.831414	1.397494	1.31	0.190	-.9124155	4.575243
sect11	1.958029	1.523145	1.29	0.199	-1.032504	4.948561
sect12	2.873389	1.440887	1.99	0.047	.0443608	5.702417
sect13	3.359272	1.747465	1.92	0.055	-.0716894	6.790233
sect14	3.613926	1.749445	2.07	0.039	.1790775	7.048775
sect15	1.780445	1.749363	1.02	0.309	-1.654243	5.215133
sect16	1.692842	1.749863	0.97	0.334	-1.742826	5.128511
sect17	1.554052	1.391265	1.12	0.264	-1.177549	4.285652
sect18	3.449842	1.733138	1.99	0.047	.0470104	6.852674
sect19	1.999394	1.513953	1.32	0.187	-.9730908	4.971879
sect20	2.132848	1.358818	1.57	0.117	-.535045	4.800741
sect21	.5250401	1.733663	0.30	0.762	-2.878823	3.928903
sect22	4.000399	1.743516	2.29	0.022	.5771923	7.423606
sect23	3.335813	1.750187	1.91	0.057	-.1004917	6.772118
sect24	1.871296	1.522624	1.23	0.219	-1.118213	4.860806
sect25	1.864091	1.738142	1.07	0.284	-1.548565	5.276747
sect26	2.413336	1.396185	1.73	0.084	-.3279235	5.154595
sect27	2.754052	1.366835	2.01	0.044	.070418	5.437687
sect28	3.568822	1.751978	2.04	0.042	.1290004	7.008644
sect29	2.078222	1.439777	1.44	0.149	-.7486271	4.905071
sect30	2.744427	1.382061	1.99	0.047	.0308969	5.457957
sect31	2.871481	1.353832	2.12	0.034	.2133757	5.529586

sect31	2.871481	1.353832	2.12	0.034	.2133757	5.529586
sect32	3.622005	1.752478	2.07	0.039	.1812029	7.062808
sect33	3.504102	1.73784	2.02	0.044	.0920395	6.916164
sect34	2.57924	1.424978	1.81	0.071	-.2185512	5.377031
sect35	2.113477	1.436441	1.47	0.142	-.7068212	4.933775
sect36	2.791911	1.369054	2.04	0.042	.1039198	5.479902
sect37	2.090447	1.373651	1.52	0.129	-.6065705	4.787465
sect38	2.231276	1.727184	1.29	0.197	-1.159865	5.622418
sect39	1.695739	1.754163	0.97	0.334	-1.748373	5.139851
sect40	1.664451	1.444208	1.15	0.250	-1.171097	4.499998
sect41	1.896284	1.521651	1.25	0.213	-1.091316	4.883884
sect42	3.146709	1.757525	1.79	0.074	-.3040038	6.597422
sect43	1.909325	1.750779	1.09	0.276	-1.528143	5.346792
sect44	2.742505	1.443457	1.90	0.058	-.0915689	5.576579
sect45	2.873261	1.397119	2.06	0.040	.1301681	5.616354
sect46	2.978523	1.522957	1.96	0.051	-.0116399	5.968686
sect47	3.644568	1.747371	2.09	0.037	.2137921	7.075343
sect48	.6454982	1.748527	0.37	0.712	-2.787548	4.078544
sect49	1.380573	1.746073	0.79	0.429	-2.047654	4.8088
sect50	3.638795	1.738537	2.09	0.037	.2253633	7.052226
sect51	2.895545	1.325717	2.18	0.029	.2926427	5.498448
sect52	2.45053	1.335271	1.84	0.067	-.1711314	5.072192
sect53	2.948587	1.43682	2.05	0.041	.1275448	5.769628
sect54	3.091963	1.532486	2.02	0.044	.0830898	6.100835
sect55	3.383669	1.745956	1.94	0.053	-.0443284	6.811667
sect56	2.251862	1.440557	1.56	0.118	-.5765174	5.08024
sect57	1.984358	1.279251	1.55	0.121	-.5273141	4.49603
sect58	2.65315	1.51706	1.75	0.081	-.3254343	5.631735
sect59	3.571923	1.726581	2.07	0.039	.1819666	6.961879
sect60	1.934086	1.751343	1.10	0.270	-1.504489	5.372661
sect61	2.243877	1.397905	1.61	0.109	-.5007609	4.988514
sect62	3.396814	1.745702	1.95	0.052	-.0306862	6.824314
sect63	3.724488	1.735317	2.15	0.032	.3173781	7.131598
sect64	3.383669	1.745956	1.94	0.053	-.0443284	6.811667
sect65	2.668796	1.329832	2.01	0.045	.0578136	5.279778

sect65	2.668796	1.329832	2.01	0.045	.0578136	5.279778
sect66	2.454478	1.308186	1.88	0.061	-.114006	5.022962
sect67	2.153723	1.370341	1.57	0.116	-.536794	4.84424
sect68	0	(omitted)				
sect69	2.395968	1.745707	1.37	0.170	-1.031541	5.823477
sect70	2.183669	1.25615	1.74	0.083	-.2826479	4.649986
sect71	2.265944	1.262157	1.80	0.073	-.212167	4.744055
sect72	2.828226	1.440278	1.96	0.050	.0003939	5.656058
sect73	3.480544	1.521795	2.29	0.022	.4926625	6.468425
sect74	3.340688	1.749519	1.91	0.057	-.0943059	6.775681
sect75	3.701776	1.748966	2.12	0.035	.2678695	7.135683
sect76	.3957545	1.521655	0.26	0.795	-2.591853	3.383362
sect77	2.498027	1.275319	1.96	0.051	-.0059246	5.001979
sect78	2.445125	1.317688	1.86	0.064	-.1420149	5.032266
sect79	2.521476	1.515572	1.66	0.097	-.4541872	5.49714
sect80	2.433877	1.263872	1.93	0.055	-.0475998	4.915354
sect81	1.954864	1.336859	1.46	0.144	-.669915	4.579644
sect82	2.453393	1.321964	1.86	0.064	-.1421422	5.048928
sect83	-.6245172	1.751881	-0.36	0.722	-4.064149	2.815114
sect84	1.772232	1.293551	1.37	0.171	-.7675184	4.311981
sect85	3.39383	1.745726	1.94	0.052	-.0337154	6.821376
sect86	2.855095	1.296186	2.20	0.028	.3101723	5.400018
sect87	1.891994	1.521703	1.24	0.214	-1.095707	4.879695
sect88	2.799736	1.398136	2.00	0.046	.0546441	5.544827
sect89	.588844	1.51477	0.39	0.698	-2.385245	3.562933
sect90	3.11894	1.337493	2.33	0.020	.4929157	5.744965
sect91	.3447205	1.749007	0.20	0.844	-3.089268	3.778709
sect92	2.172387	1.311185	1.66	0.098	-.4019855	4.746759
sect93	2.007655	1.329884	1.51	0.132	-.6034301	4.61874
sect94	1.588595	1.523396	1.04	0.297	-1.40243	4.57962
sect95	3.703195	1.750618	2.12	0.035	.2660431	7.140346
sect96	3.375571	1.746307	1.93	0.054	-.0531169	6.804259
sect97	2.549032	1.75636	1.45	0.147	-.8993932	5.997458
sect98	3.476713	1.745806	1.99	0.047	.0490089	6.904416
sect99	3.413376	1.428468	2.39	0.017	.6087314	6.218021
sect100	1.611987	1.733842	0.93	0.353	-1.792227	5.016201
sect101	3.385951	1.521857	2.22	0.026	.3979487	6.373954
sect102	2.001992	1.515276	1.32	0.187	-.9730909	4.977075
sect103	3.447224	1.745681	1.97	0.049	.0197673	6.874681
sect104	1.556897	1.523995	1.02	0.307	-1.435304	4.549097
ln_ricavi_iniziali	-.0527986	.0347739	-1.52	0.129	-.1210735	.0154763
ln_tot_debiti_iniziali	-.072727	.032942	-2.21	0.028	-.1374052	-.0080489
ln_dipendenti_iniziali	.3521091	.127391	2.76	0.006	.1019905	.6022277

Figure 17: Results of linear regression Variant 2

Starting from the independent variable *ln_imm_rate_iniziale*, its coefficient is about -1.79, even in this case as for the prior one, the coefficient exhibits a negative sign, but with a higher value in absolute terms, this means that the dependent variable *duration* has a stronger negative relationship with the variable *ln_imm_rate_iniziale*, respect to the one with the variable *ln_imm_iniziali*. Looking at the statistical significance of this result, it is possible to see that the p-value is very high than (61.6%), of course higher then 5%, so it is not possible to reject the hypothesis that the two variables are not statistically

significant. About the variable *ln_ricavi_iniziali*, its coefficients are very similar to the one of the previous case (-0.052986 vs -0.0477657), so it is confirmed its weak negative relationship with the variable *duration*, and even in this case it is not possible to reject the hypothesis that the result is not statistically consistent, from the moment that the p-value is lower than 5%. The variable *ln_tot_debiti_iniziali* shows a weak negative relationship with the variable duration because the coefficient has a negative sign, and even in this case is very similar to the one of the first regression (-0.072727 vs -0.070282). The p-value is lower than 5% (2.8%) so it is possible to reject the hypothesis that the result is not statistically consistent.

The variable *ln_dipendenti_iniziali* presents a positive sign (about 0.352), so this means that there is a positive relationship between it and the dependent variable, and the result can be considered statistically significant from the moment that the p-value is lower that 5%, a result very similar to the one of the first version regression.

5.3 Variant 3

5.3.1 Model's Fit

Source	SS	df	MS	Number of obs	=	787
				F(107, 679)	=	1.14
Model	176.872228	107	1.65301148	Prob > F	=	0.1754
Residual	985.53692	679	1.45145349	R-squared	=	0.1522
				Adj R-squared	=	0.0186
Total	1162.40915	786	1.47889205	Root MSE	=	1.2048

Figure 18: Analysis of model's fit Variant

As for the two previous linear regressions, the model's fit is commented. In this case, it is possible to see a little improvement of the fit of the model, in particular the R^2 shows a value of 0.1522, then using this model the independent variables explain the 15.22% of the variance of the dependent variable. Even the $\bar{R}^2$ is increased showing a better model's fit, with a value equal to 0.0186. Even if the two values are increased they are still low, this mean that, as for the other cases, there is other variables that explain the variance of the dependent variable that are not including in the model. On the other hand, the *SER* remains almost equal to the previous two.

5.3.2 Coefficients' analysis

Now the coefficients in the output will be analyzed, even in this case it is used a linear-log regression, so it is applied the logarithm function to the independent variables to consider the variation of the variables in percentage terms.

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_imm_SLL_TOT_iniziali	-.1237	.0605094	-2.04	0.041	-.242508	-.0048919
ln_pop_iniziale	.0866046	.0373958	2.32	0.021	.0131794	.1600298
sect1	4.103991	1.757401	2.34	0.020	.6533988	7.554584
sect2	4.082525	1.753274	2.33	0.020	.6400354	7.525014
sect3	2.320239	1.35163	1.72	0.087	-.3336374	4.974116
sect4	4.157678	1.75282	2.37	0.018	.7160798	7.599276
sect5	1.16012	1.752103	0.66	0.508	-2.280071	4.600311
sect6	3.75529	1.748668	2.15	0.032	.3218439	7.188736
sect7	2.915651	1.513767	1.93	0.055	-.0565753	5.887878
sect8	2.505523	1.726122	1.45	0.147	-.8836552	5.894702
sect9	1.467827	1.738689	0.84	0.399	-1.946026	4.88168
sect10	1.939576	1.392182	1.39	0.164	-.7939224	4.673074
sect11	2.123143	1.517352	1.40	0.162	-.8561229	5.102409
sect12	3.009147	1.436628	2.09	0.037	.188379	5.829915
sect13	0	(omitted)				
sect14	3.909467	1.742396	2.24	0.025	.4883362	7.330598
sect15	1.822857	1.743854	1.05	0.296	-1.601137	5.24685
sect16	1.757248	1.743055	1.01	0.314	-1.665177	5.179673
sect17	1.121945	1.425916	0.79	0.432	-1.677789	3.921679
sect18	3.758173	1.729078	2.17	0.030	.3631908	7.153155
sect19	2.051181	1.507865	1.36	0.174	-.9094583	5.01182
sect20	2.245143	1.354347	1.66	0.098	-.4140677	4.904354
sect21	.5594455	1.728019	0.32	0.746	-2.833458	3.952349
sect22	3.931808	1.743207	2.26	0.024	.5090838	7.354533
sect23	3.661166	1.747002	2.10	0.036	.2309902	7.091343
sect24	2.074718	1.517504	1.37	0.172	-.9048469	5.054282
sect25	1.887608	1.731093	1.09	0.276	-1.511331	5.286547
sect26	2.483389	1.391033	1.79	0.075	-.2478537	5.214631
sect27	2.790948	1.361901	2.05	0.041	.1169047	5.464992
sect28	3.998346	1.7465	2.29	0.022	.5691559	7.427535
sect29	2.227715	1.434026	1.55	0.121	-.5879432	5.043373
sect30	2.896331	1.375706	2.11	0.036	.1951817	5.597481

sect31	2.934187	1.349936	2.17	0.030	.2836369	5.584738
sect32	4.040439	1.747672	2.31	0.021	.608947	7.47193
sect33	3.896014	1.731147	2.25	0.025	.4969686	7.295059
sect34	2.627521	1.419227	1.85	0.065	-.1590799	5.414121
sect35	2.218182	1.430506	1.55	0.121	-.5905642	5.026929
sect36	2.892644	1.363554	2.12	0.034	.2153551	5.569932
sect37	2.268197	1.367498	1.66	0.098	-.4168352	4.95323
sect38	2.206373	1.722365	1.28	0.201	-1.175428	5.588175
sect39	1.501849	1.771145	0.85	0.397	-1.975731	4.979429
sect40	1.837106	1.437982	1.28	0.202	-.9863196	4.660532
sect41	1.967513	1.51569	1.30	0.195	-1.00849	4.943517
sect42	3.167145	1.752014	1.81	0.071	-.2728717	6.607161
sect43	1.949922	1.743831	1.12	0.264	-1.474027	5.373871
sect44	2.90065	1.438572	2.02	0.044	.0760669	5.725234
sect45	2.971453	1.391621	2.14	0.033	.2390554	5.703852
sect46	3.033562	1.517343	2.00	0.046	.0543149	6.01281
sect47	3.704063	1.740429	2.13	0.034	.2867932	7.121333
sect48	.712251	1.741667	0.41	0.683	-2.707449	4.131951
sect49	1.39436	1.747519	0.80	0.425	-2.036831	4.82555
sect50	3.644861	1.732174	2.10	0.036	.2437996	7.045922
sect51	2.987403	1.32071	2.26	0.024	.3942373	5.580569
sect52	2.55367	1.330067	1.92	0.055	-.0578684	5.165208
sect53	3.252104	1.430708	2.27	0.023	.44296	6.061248
sect54	4.162215	1.751449	2.38	0.018	.7233083	7.601122
sect55	3.450093	1.738726	1.98	0.048	.0361676	6.864018
sect56	2.32421	1.435053	1.62	0.106	-.4934635	5.141884
sect57	2.086146	1.274507	1.64	0.102	-.416303	4.588596
sect58	2.92196	1.512012	1.93	0.054	-.0468206	5.890741
sect59	3.576096	1.719179	2.08	0.038	.2005498	6.951642
sect60	1.97175	1.744379	1.13	0.259	-1.453275	5.396775
sect61	2.438416	1.39236	1.75	0.080	-.295433	5.172265
sect62	3.466518	1.738694	1.99	0.047	.0526563	6.88038
sect63	3.753888	1.728215	2.17	0.030	.3605995	7.147176
sect64	3.450093	1.738726	1.98	0.048	.0361676	6.864018
sect65	2.748816	1.324778	2.07	0.038	.1476615	5.34997
sect66	2.61748	1.303392	2.01	0.045	.0583167	5.176644
sect67	2.215955	1.36511	1.62	0.105	-.4643901	4.896299
sect68	0	(omitted)				
sect69	2.466905	1.738691	1.42	0.156	-.9469512	5.880762
sect70	2.29484	1.251606	1.83	0.067	-.1626435	4.752323
sect71	2.359181	1.25756	1.88	0.061	-.1099934	4.828355
sect72	3.024424	1.43565	2.11	0.036	.2055773	5.84327
sect73	3.548236	1.515886	2.34	0.020	.5718487	6.524624
sect74	3.347432	1.745229	1.92	0.056	-.0792626	6.774126
sect75	3.745464	1.741784	2.15	0.032	.3255339	7.165395
sect76	.4685086	1.515687	0.31	0.757	-2.507489	3.444506
sect77	2.529822	1.271271	1.99	0.047	.0337271	5.025916
sect78	2.399814	1.320527	1.82	0.070	-.1929931	4.992621
sect79	2.82042	1.511891	1.87	0.063	-.1481239	5.788964
sect80	2.543361	1.259626	2.02	0.044	.0701309	5.016591
sect81	1.770304	1.344313	1.32	0.188	-.8692053	4.409813
sect82	2.501635	1.316987	1.90	0.058	-.084221	5.087492
sect83	-.2578279	1.744533	-0.15	0.883	-3.683156	3.1675
sect84	1.819571	1.288762	1.41	0.158	-.7108665	4.350009
sect85	3.463143	1.738697	1.99	0.047	.0492738	6.877012
sect86	3.032592	1.291647	2.35	0.019	.49649	5.568694
sect87	1.962894	1.515695	1.30	0.196	-1.013119	4.938907
sect88	2.883694	1.393991	2.07	0.039	.1466433	5.620745
sect89	.636275	1.508676	0.42	0.673	-2.325956	3.598506
sect90	3.27636	1.345645	2.43	0.015	.6342351	5.918485
sect91	.6199062	1.743433	0.36	0.722	-2.803262	4.043075
sect92	2.236681	1.306597	1.71	0.087	-.3287752	4.802137

sect93	2.094239	1.32484	1.58	0.114	-.5070367	4.695515
sect94	1.655506	1.517563	1.09	0.276	-1.324174	4.635185
sect95	3.764672	1.743694	2.16	0.031	.3409912	7.188353
sect96	3.439977	1.738764	1.98	0.048	.025977	6.853977
sect97	3.202221	1.755397	1.82	0.069	-.244437	6.648879
sect98	3.544497	1.738835	2.04	0.042	.1303574	6.958637
sect99	3.473864	1.422874	2.44	0.015	.6801021	6.267626
sect100	1.646247	1.72666	0.95	0.341	-1.743988	5.036482
sect101	3.453686	1.51572	2.28	0.023	.477625	6.429747
sect102	2.062123	1.509321	1.37	0.172	-.9013747	5.02562
sect103	3.515717	1.738694	2.02	0.044	.101855	6.92958
sect104	1.592255	1.519203	1.05	0.295	-1.390644	4.575154
ln_ricavi_iniziali	-.0460593	.0348398	-1.32	0.187	-.1144659	.0223473
ln_tot_debiti_iniziali	-.0709794	.0331802	-2.14	0.033	-.1361275	-.0058313
ln_dipendenti_iniziali	.3628399	.1279206	2.84	0.005	.1116724	.6140074

Figure 19: Results of linear regression Variant 3

Even in this case the independent variable taken in consideration, *imm_SLL_TOT_iniziali* exhibit a negative sign, with a value in absolute terms lower than the previous two cases, this means that the negative relationship between *imm_SLL_TOT_iniziali* and *duration* id weaker respect to ne negative relationship present between *imm_rate_iniziale* and *duration* and between *imm_iniziali* and *duration*, in particular the coefficient is equal to -0.1237. Even for the variable *pop_iniziale,* the sign is the same as the one shown in the variant 1 of the linear regression with a lower absolute value, then the relationship between the amount of population present in the municipality and the duration of the startup founded in that municipality is still positive but weaker. For the other three variables *ricavi_iniziali*, *tot_debiti_iniziali* and *dipendenti_iniziali*, nothing is changed, the signs, and then the relationships, are the same of the previous two cases with very similar absolute values. All results can be considered statistically consistent from the moment that for all variables, the output shows a p-value lower than 5%, so, for each independent variables, it is possible to reject the hypothesis that the independent variable and the dependent variable are not statistically dependent for a confidence interval of 95%. The only variable that has a p-value higher than 5% is *ricavi_iniziali*, then for that variable the result is not statistically significant.

6. Discussion and Limitations

The goal of this thesis is to assess the effect of low-skilled immigration in Lombardy's territory on the probability of survival of the female-led innovative startups. To this scope, three linear regressions are carried out doing some changes between one and the other about variable used in the model. Even with some differences in the coefficients' values the results of all three linear regressions show a weak negative relationship between the probability of the female-led innovative startups' survival, represented by the startup life's duration and the presence of immigrants on the Lombardy's territory. The main element that changes between the three regressions is the variable used to represent the level of immigrants. In particular, the first model is used as independent variable, to represent the level of immigrants, the level of immigrant women, between 25 and 65 age, that result residents in the municipality and in the year in which the startup is founded, moreover, other control variables are used linked to the municipalities and to the startups, all considered at the year of startup's foundation. The result can be considered statistically consistent from the moment that the p-value of the variable is higher then 5%, so it is possible to reject the null hypothesis that the two variables are not correlated. In the second scenario the dependent variable representing the level of immigrants is represented by the rate of immigrants present in the municipality and in the year of startup's foundation, constructed as the ratio between the number of immigrants over the number of the whole population, the other variables are unchanged. Even in this case the results show a negative relationship between the startup's duration and the level of immigrants expressed in terms of immigrant's rate, but the difference is that in this case the result is not statistically consistent because the p-value is very high, so it is not possible to reject the hypothesis that the two variables are not statistically dependent. The third linear regression changes the geographical point of view, taking in consideration the Local Labor Systems in the place of the municipality, and taking into consideration the number of immigrants in the LLS in which the startup was founded in the year of startup's creation. The relationship between the two variables is still negative, and the result can be considered statistically significant because the variable's is lower than 5%. These results are not in linear with the initial expectations of this thesis, that were reinforced by the review of the prior studies, in fact the starting idea of this thesis, based on the literature review of the past research, was that the immigrants, in particular low-skilled immigrants that can offer house and children care services, allow women entrepreneurship, resident

in their same geographic area, to dedicate more time to their startup and it will be reflected on the duration of the startup increasing it. But the results show the opposite conclusion. The reason of this unexpected result can be caused by different reasons, some are related to the nature of the problem, others are related to some limitations of the model constructed to represent the problem. They are reported below.

6.1 Problem nature related reasons

The problem that this thesis wants to explore is an articulated and complicated problem by its nature. As already stated in Chapter 2, the startups run their business within a highly uncertain environment, due to the novelty nature of that type of business and the high information asymmetry, for this reason it is not easy to assess. To create the statistical model, they are faced with two main issues related to the problem's nature, reported below.

- The first issue is that, as said in Chapter 2, the performance and the probability of survival of the startups depends on a high number of variables, that often are not easy to assess or to transform them in mathematical variable to insert into a statistical model. For this reason, the created model presents some limitations because it includes only a part of these variables.
- The second issue is linked to the assumption from which the model is based, namely the assumption that increasing the hours that an entrepreneur dedicates to her startup improves its performance and increases its probability of survival, but, as it is said in Chapter 2, this assumption is not universally shared among the studies, and it is not verified in all cases. Then, even if it is true that the presence of low-skilled immigrants in a certain area allows women to dedicate more hours to paid work, it is possible that this does not increase the startup's probability of survival, and that the effect of the immigration on the startups is different.

6.2 Model related reasons

Other than the issues related to the nature of the problem analyzed, there are also issues linked to the difficulty in translating the problem into a mathematical model that would alter the obtained results and deviate them from the expected results. The main issues are reported below.

- The research is made in a limited geographical area, even the Lombardy is chosen as sample because is one of the Italian region with a higher number of innovative startups, and then this leads into a higher number of observation, because of the specificity of the problem, namely the innovative startups with a strong presence of women in the administration, a lot data are eliminated from the database because out of the work's interest, and, in light of this, the data within the geographic area results limited.
- The limitation of the data used is not only in terms of location but also in terms of time. Because of the reasons mentioned in Chapter 4, it is chosen to take into consideration startups born between 2009 and 2019, limiting the period of observation to ten years, and then also the number of observations.
- The choice to take into consideration in the model the value of the variables at the time of the startups' foundation, can generate a model that does not fit perfectly with the representation of the problem, because it does not take into consideration the variation of the variables during each year of startup's life, delating some information.
- The last issue regards the sample of immigrants considered in the linear regression model. In all three variants of regressions the sample of immigrants that are supposed to offer low skilled services is collected in the same way, in particular the sample includes all women between 25 and 65 residents in Italy with a nationality different from the Italian's one. ISTAT's database, from which this information are downloaded, does not provide the information about the level of education of them, then in the variable used to represent the level of low-skilled immigrants are included also possible high-skilled immigrants, that do not provide house and children care services, and influence the duration of the startups in a different manner from the one hypothesized by that work.

7. Conclusions and future improvements

This thesis analyzes a topic very discussed and studied in these current times, namely the women in the labor market, with the particular focus on the innovative startups, including the other hot topic of the immigrants. It tries to put together these three variables to provide a further contribution to the research that faced these three themes. As said in the previous chapter, the model presents different limitations, but, even if the model is not perfect and the results are not aligned with the expected ones, this work can represent a starting point and an inspiration to replicate and develop future research in that direction. To this scope, at the conclusion of the work, some possible improvements are provided.

- Extend the territorial area and the period of observation.
- Create a more complicated and sophisticated model that takes into consideration each variable for each year of startup's observation. A suggestion from which to start is the possibility of changing the dependent variable, then them to use another variable to represent the startup's survival probability, that permits to consider the data available for each year of startup's life and not only the ones at the year of startup's foundation. A possible trial could be creating a startup's survival rate from a year to the next to use as dependent variable.
- Include more control variables in the linear regression. There are different factors that can influence the startups' survival that in this work are not included due to the lack of available information, some of these variable to implement can be: the distinction between the types of startups' investors, such as Business Angel, Venture Capital and banks' loan, the level of education of the immigrants' sample, data not available on the ISTAT's database but that it is useful to select only low-skilled immigrants, and the number of patent owned by each startup, this data is available in the Final Database but only for a low number of startups and only for the year of foundation, this variable can give information about the stage of innovation of the startup.

8. References

- Aidis, R., & Weeks, J. (2016). Mapping the gendered ecosystem: The evolution of measurement tools for comparative high-impact female entrepreneur development. *International Journal of Gender and Entrepreneurship*, *8*(4), 330-352.
- Alperovych, Y., & Hübner, G. (2011). Explaining returns on venture capital backed companies: Evidence from Belgium. *Research in International Business and Finance*, *25*(3), 277-295.
- Alsos, G. A., Isaksen, E. J., & Ljunggren, E. (2006). New venture financing and subsequent business growth in men–and women–led businesses. *Entrepreneurship theory and practice*, *30*(5), 667-686.
- Altman, Sam. (2018). How to Succeed with a Startup. [Video]. YouTube. Available online: https://www.youtube.com/watch?v=0lJKucu6HJc&t=122s
- Amoroso, S., & Audretsch, D. B. (2022). The role of gender in linking external sources of knowledge and R&D intensity. *Economics of Innovation and New Technology*, *31*(1-2), 3-19.
- Amoroso, S., & Link, A. N. (2018). Under the AEGIS of knowledge-intensive entrepreneurship: employment growth and gender of founders among European firms. *Small Business Economics*, *50*, 899-915.
- Arcuri, M. C., Gandolfi, G., & Russo, I. (2023). Exploring the impact of innovation intensity on the growth of female-led entrepreneurial firms. *Journal of Small Business and Enterprise Development*.
- Audretsch, D., Colombelli, A., Grilli, L., Minola, T., & Rasmussen, E. (2020). Innovative start-ups and policy initiatives. *Research Policy*, *49*(10), 104027.
- Bandera, C., & Thomas, E. (2019). To pivot or not to pivot: On the relationship between pivots and revenue among startups.
- Banerjee, A. V., & Duflo, E. (2020). *Una buona economia per tempi difficili*. Gius. Laterza & Figli Spa.
- BarNir, A. (2012). Starting technologically innovative ventures: reasons, human capital, and gender. *Management Decision*, *50*(3), 399-419.
- Bass, B. M. (1995). Theory of transformational leadership redux. *The leadership quarterly*, *6*(4), 463-478.

- Baughn, C. C., Chua, B. L., & Neupert, K. E. (2006). The normative context for women's participation in entrepreneruship: A multicountry study. *Entrepreneurship theory and practice*, *30*(5), 687-708.
- Becker, G. S. (1975). Investment in human capital: effects on earnings. In *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition* (pp. 13-44). NBER.
- Berkus, D. (2006). *Extending the runway*. Boston: Aspatore Books.
- Bhide, A. (2000). *The origin and evolution of new businesses*. Oxford University Press.
- *Borsa Italiana,* Glossario finanziario – Venture Capital, from https://www.borsaitaliana.it/borsa/glossario/venture-capital.html
- Boyer, T., & Blazy, R. (2014). Born to be alive? The survival of innovative and non-innovative French micro-start-ups. *Small Business Economics*, *42*, 669-683.
- Brattström, A. (2019). Working with startups? these are the three things you ought to know about startup teams. *Technology Innovation Management Review*, *9*(11).
- Brüderl, J., Preisendörfer, P., & Ziegler, R. (1992). Survival chances of newly founded business organizations. *American sociological review*, 227-242.
- Buttner, E. H., & Moore, D. P. (1997). Women's organizational exodus to entrepreneurship: self-reported motivations and correlates with success. *Journal of small business management*, *35*, 34-46.
- Caputo, R. K., & Dolinsky, A. (1998). Women's choice to pursue self-employment: The role of financial and human capital of household members. *Journal of Small Business Management*, *36*, 8-17.
- Cardon, M. S., & Kirk, C. P. (2015). Entrepreneurial passion as mediator of the self–efficacy to persistence relationship. *Entrepreneurship theory and practice*, *39*(5), 1027-1050.
- Cardon, M. S., Wincent, J., Singh, J., & Drnovsek, M. (2009). The nature and experience of entrepreneurial passion. *Academy of management Review*, *34*(3), 511-532.
- Carter, N. M., Gartner, W. B., Shaver, K. G., & Gatewood, E. J. (2003). The career reasons of nascent entrepreneurs. *Journal of business venturing*, *18*(1), 13-39.

- Carter, S., Shaw, E., Lam, W., & Wilson, F. (2007). Gender, entrepreneurship, and bank lending: The criteria and processes used by bank loan officers in assessing applications. *Entrepreneurship theory and practice*, *31*(3), 427-444.
- Chaganti, R., DeCarolis, D., & Deeds, D. (1996). Predictors of capital structure in small ventures. *Entrepreneurship Theory and Practice*, *20*(2), 7-18.
- Ciavarella, M. A., Buchholtz, A. K., Riordan, C. M., Gatewood, R. D., & Stokes, G. S. (2004). The Big Five and venture survival: Is there a linkage?. *Journal of business venturing*, *19*(4), 465-483.
- Coleman, S. (2000). Access to capital and terms of credit: A comparison of men- and women-owned small businesses. *Journal of small business management*, *38*(3), 37.
- Colombo, M. G., & Piva, E. (2008). Strengths and weaknesses of academic startups: a conceptual model. *IEEE Transactions on Engineering Management*, *55*(1), 37-49.
- Cooper, A. C., & Bruno, A. V. (1977). Success among high-technology firms. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.
- Cortes, P. (2008). The effect of low-skilled immigration on US prices: evidence from CPI data. *Journal of political Economy*, *116*(3), 381-422.
- Cortes, P. (2022). Immigration, Household Services, and Women's Economic Opportunities.
- Cortes, P., & Pan, J. (2016). When time binds: Returns to working long hours and the gender wage gap among the highly skilled.
- Cortes, P., & Tessada, J. (2011). Low-skilled immigration and the labor supply of highly skilled women. *American Economic Journal: Applied Economics*, *3*(3), 88-123.
- Croce, A., Guerini, M., & Ughetto, E. (2018). Angel financing and the performance of high-tech start-ups. *Journal of Small Business Management*, *56*(2), 208-228.
- DeTienne, D. R., & Chandler, G. N. (2007). The role of gender in opportunity identification. *Entrepreneurship theory and practice*, *31*(3), 365-386.

- Dizionario di economia e finanza, Treccani (2012), from https://www.treccani.it/enciclopedia/startup_%28Dizionario-di-Economia-e-Finanza%29/
- Dolfsma, W., & Seo, D. (2013). Government policy and technological innovation—a suggested typology. *Technovation*, *33*(6-7), 173-179.
- Eagly, A. H., & Carli, L. L. (2004). Women and men as leaders. *The nature of leadership*, 279-301.
- Eagly, A. H., & Carli, L. L. (2004). Women and men as leaders. *The nature of leadership*, 279-301.
- Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological bulletin*, *108*(2), 233.
- England, P. (2006). Toward gender equality: Progress and bottlenecks. *The declining significance of gender*, *63*, 245-265.
- Eskilson, A., & Wiley, M. G. (1999). Solving for the X: Aspirations and expectations of college students. *Journal of Youth and Adolescence*, *28*, 51-70.
- *European Commission,* Benchmarking of business incubators (2002), from https://op.europa.eu/en/publication-detail/-/publication/5f01aafc-ef62-457d-9316-c85e7fc2509e
- *European Commission,* Internal Market, Industry, Entrepreneurship and SMEs, from https://single-market-economy.ec.europa.eu/access-finance/policy-areas/business-angels_en
- Fabowale, L., Orser, B., & Riding, A. (1995). Gender, structural factors, and credit terms between Canadian small businesses and financial institutions. *Entrepreneurship Theory and Practice*, *19*(4), 41-65.
- Fairlie, R. W., & Robb, A. M. (2009). Gender differences in business performance: evidence from the Characteristics of Business Owners survey. *Small Business Economics*, *33*, 375-395.
- Fay, M., & Williams, L. (1993). Gender bias and the availability of business loans. *Journal of Business Venturing*, *8*(4), 363-376.
- Fischer, E. M., Reuber, A. R., & Dyke, L. S. (1993). A theoretical overview and extension of research on sex, gender, and entrepreneurship. *Journal of business venturing*, *8*(2), 151-168.

- Gatewood, E. J., Shaver, K. G., & Gartner, W. B. (1995). A longitudinal study of cognitive factors influencing start-up behaviors and success at venture creation. *Journal of business venturing*, *10*(5), 371-391.
- Gaweł, A., & Mroczek-Dąbrowska, K. (2022). Gender pay gap in explaining female entrepreneurship–industry perspective of selected European countries. *International Journal of Manpower*, *43*(9), 42-59.
- Giarratana, M. S. (2004). The birth of a new industry: Entry by start-ups and the drivers of firm growth: The case of encryption software. *Research Policy*, *33*(5), 787-806.
- Goldin, C., & Katz, L. F. (2011). The cost of workplace flexibility for high-powered professionals. *The Annals of the American Academy of Political and Social Science*, *638*(1), 45-67.
- Goldin, C., Kerr, S. P., & Olivetti, C. (2022). *When the Kids Grow Up: Women's Employment and Earnings across the Family Cycle* (No. w30323). National Bureau of Economic Research.
- Grant, M. (2022). What a Startup Is and What's Involved in Getting One Off the Ground. *Diakses dari https://www. investopedia. com/terms/s/startup. asp*.
- Gross, Bill. (2015). The Single Biggest Reason Why Start-Ups Succeed. [Video]. YouTube. Available online: https://www.youtube.com/watch?v=bNpx7gpSqbY
- Hackett, S. M., & Dilts, D. M. (2004). A real options-driven theory of business incubation. *The journal of technology transfer*, *29*(1), 41-54.
- Hakim, C. (2000). *Work-lifestyle choices in the 21st century: Preference theory*. OUp Oxford.
- Hudson, J., & Khazragui, H. F. (2013). Into the valley of death: research to innovation. *Drug discovery today*, *18*(13-14), 610-613.
- *Investopedia*, The Glass Celling: Definition, History, Effects, and Examples (2022), from https://www.investopedia.com/terms/g/glass-ceiling.asp
- ISTAT (2023), Stranieri e neutralizzati nel mondo del lavoro, from https://www.istat.it/it/files//2023/02/Focus_stranieri-e-naturalizzati-nel-mondo-del-lavoro.pdf
- Jamali, D. (2009). Constraints and opportunities facing women entrepreneurs in developing countries: A relational perspective. *Gender in management: an international journal*, *24*(4), 232-251.

- Jennings, J. E., & McDougald, M. S. (2007). Work-family interface experiences and coping strategies: Implications for entrepreneurship research and practice. *Academy of management review*, *32*(3), 747-760.
- Kasser, T., & Ryan, R. M. (1993). A dark side of the American dream: correlates of financial success as a central life aspiration. *Journal of personality and social psychology*, *65*(2), 410.
- Kerr, W. R., Lerner, J., & Schoar, A. (2014). The consequences of entrepreneurial finance: Evidence from angel financings. *The Review of Financial Studies*, *27*(1), 20-55.
- Kim, J., Kim, H., & Geum, Y. (2023). How to succeed in the market? Predicting startup success using a machine learning approach. *Technological Forecasting and Social Change*, *193*, 122614.
- Klyver, K., Nielsen, S. L., & Evald, M. R. (2013). Women's self-employment: An act of institutional (dis) integration? A multilevel, cross-country study. *Journal of Business Venturing*, *28*(4), 474-488.
- Lukeš, M., Longo, M. C., & Zouhar, J. (2019). Do business incubators really enhance entrepreneurial growth? Evidence from a large sample of innovative Italian start-ups. *Technovation*, *82*, 25-34.
- Malach-Pines, A., & Schwartz, D. (2008). Now you see them, now you don't: gender differences in entrepreneurship. *Journal of Managerial Psychology*, *23*(7), 811-832.
- Mason, C., & Stark, M. (2004). What do investors look for in a business plan? A comparison of the investment criteria of bankers, venture capitalists and business angels. *International small business journal*, *22*(3), 227-248.
- Mayoshe, B., & Nuringsih, K. (2021, August). Understanding Entrepreneurial Passion Among New Entrepreneurs. In *International Conference on Economics, Business, Social, and Humanities (ICEBSH 2021)* (pp. 251-257). Atlantis Press.
- Megginson, W. Weiss. K., 1991. Venture capitalist certification in initial public offerings. *Journal of Finance*, *46*, 29-48.
- Ministero delle Imprese e del Made in Italy (2022), from https://www.mimit.gov.it/it/impresa/competitivita-e-nuove-imprese/start-up-innovative#:~:text=%C3%A8%20un'impresa%20nuova%20o,una%20piattaforma%20multilaterale%20di%20negoziazione

- Minniti, M. (2009). Gender issues in entrepreneurship. *Foundations and Trends® in Entrepreneurship*, *5*(7–8), 497-621.
- Minniti, M., & Naudé, W. (2010). What do we know about the patterns and determinants of female entrepreneurship across countries?. *The European Journal of Development Research*, *22*, 277-293.
- Monitor, G. E. (2022). Global entrepreneurship monitor releases new research highlighting women's entrepreneurship trends across the globe.
- OECD, Innovation in Firms: A Microeconomic Prospective, 2009
- Okrah, J., Nepp, A., & Agbozo, E. (2018). Exploring the factors of startup success and growth. *The business & management review*, *9*(3), 229-237.
- Orhan, M., & Scott, D. (2001). Why women enter into entrepreneurship: an explanatory model. *Women in management review*, *16*(5), 232-247.
- Parasuraman, S., & Simmers, C. A. (2001). Type of employment, work–family conflict and well-being: a comparative study. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *22*(5), 551-568.
- Perry, S. C. (2002). A Comparison of Failed and Non-failed Small Business in the United States: Do Men and Women Use Different Planning and Decision Making Strategies?. *Journal of Developmental entrepreneurship*, *7*(4).
- Rapporto AlmaLaurea 2022, from https://www.almalaurea.it/informa/news/2022/06/15/rapporto-almalaurea-2022
- Salamzadeh, A., & Kawamorita Kesim, H. (2015). Startup companies: Life cycle and challenges. In *4th International conference on employment, education and entrepreneurship (EEE), Belgrade, Serbia*.
- Sarfaraz, L., Faghih, N., & Majd, A. A. (2014). The relationship between women entrepreneurship and gender equality. *Journal of Global Entrepreneurship Research*, *4*, 1-11.
- Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social issues*, *57*(4), 675-688.
- Schumpeter, J. A. (1949). Science and ideology. *American Economic Review*, *39*(2), 345-359.
- Schwartz, M. (2013). A control group study of incubators' impact to promote firm survival. *The Journal of Technology Transfer*, *38*, 302-331.

- Shaver, K. G. (2012). *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer Science & Business Media.
- Skawińska, E., & Zalewski, R. I. (2020). Success factors of startups in the EU—A comparative study. *Sustainability*, *12*(19), 8200.
- Sørensen, M. (2007). How smart is smart money? A two-sided matching model of venture capital. *The Journal of Finance*, *62*(6), 2725-2762.
- Stock, J. H., & Watson, M. W. (2015). Introduction to econometrics (3rd updated edition). *Age (X3)*, *3*(0.22).
- Tasnim, R., Yahya, S., & Zainuddin, M. N. (2014). " I'm Loving It!" What makes the successful entrepreneur affectively committed to entrepreneurial performance?. *Journal of Applied Management and Entrepreneurship*, *19*(2), 27.
- Tverdostup, M., & Paas, T. (2017). Gender-specific human capital: identification and quantifying its wage effects. *International Journal of Manpower*, *38*(6), 854-874.
- Unioncamere - Centro Studi Guglielmo Tagliacarne - Si.Camera, V Rapporto nazionale imprenditoria femminile, 2022
- *United Way*, What is gender equality? Learn the definition with examples (2022), from https://unitedwaynca.org/blog/what-is-gender-equality/
- Verheul, I., Stel, A. V., & Thurik, R. (2006). Explaining female and male entrepreneurship at the country level. *Entrepreneurship and regional development*, *18*(2), 151-183.
- Vossenberg, S. (2013). Women Entrepreneurship Promotion in Developing Countries: What explains the gender gap in entrepreneurship and how to close it. *Maastricht School of Management Working Paper Series*, *8*(1), 1-27.
- Vracheva, V., & Stoyneva, I. (2020). Does gender equality bridge or buffer the entrepreneurship gender gap? A cross-country investigation. *International Journal of Entrepreneurial Behavior & Research*, *26*(8), 1827-1844.
- Watson, J., & Robinson, S. (2003). Adjusting for risk in comparing the performances of male-and female-controlled SMEs. *Journal of business venturing*, *18*(6), 773-788.
- Williams, D. R. (2004). Effects of childcare activities on the duration of self–employment in Europe. *Entrepreneurship theory and practice*, *28*(5), 467-486.

- Wiltbank, R., Read, S., Dew, N., & Sarasvathy, S. D. (2009). Prediction and control under uncertainty: Outcomes in angel investing. *Journal of business venturing*, *24*(2), 116-133.
- Winborg, J., & Landström, H. (2001). Financial bootstrapping in small businesses: Examining small business managers' resource acquisition behaviors. *Journal of business venturing*, *16*(3), 235-254.
- Zaech, S., & Baldegger, U. (2017). Leadership in start-ups. *International Small Business Journal*, *35*(2), 157-177.
- Zilgalvis, P. (2014). The need for an innovation principle in regulatory impact assessment: the case of finance and innovation in Europe. *Policy & Internet*, *6*(4), 377-392.