Politecnico di Torino

Master's degree in Engineering and Management
A.a. 2023/2024
Sessione di laurea 04/2024

Female entrepreneurship and immigration in the Piedmont region

Is it possible that the low-skilled immigration of women influences the survival rate of female-led entrepreneurship?

Relatrice: Anna D'Ambrosio

Candidata: Federica Quarone

Abstract

This master's degree thesis aims to analyze if there is a correlation between the survival rate of female-led entrepreneurship and the number of low-skilled immigrant women, who provide home and children care, in the Piedmont region.

After the introduction, a study of the background literature is made. Here it is explained what is a start-up and the causes that lead it to failure. Then, the situation of female entrepreneurship is analyzed. After an analysis of the immigrant situation in the Piedmont region and Italy, a focus on Patricia Cortes' research (the study that inspired this thesis) is made. In the end, it is possible to find a little section about Piedmont's start-ups.

In the second chapter, a theoretical explanation of the statistical method that is used to evaluate the relationship between female entrepreneurship and low-skilled women is provided.

To create the models, it was necessary to collect data from two big databases: ISTAT (the Statistical National Institution), regarding immigrant women with age between 25 and 65 years old, and AIDA (Computer Analysis of Italian Companies) regarding data about innovative start-ups and enterprises. Using the statistical software STATA, a final database was created and, after defining the different variables that can influence the duration of life of a start-up (the selected dependent variable), the different linear regression models were implemented.

The results obtained suggest that considering the immigrant stock for a bigger geographical area (for example by Local Labor System or by Province) is more accurate than considering them for each municipality. However, also considering it in a large geographical area, in this analysis the results achieved are all statistically insignificant.

Contents

Index of Figures

Introduction

In this thesis a case study conducted by the economist Patricia Cortes in the United States of America is replicated in an Italian region, the Piedmont. The study focus is about the possible relationship between the low-skilled immigration of women and the labor success of native women. The term low-skilled women include all the women that don't have a study title, including a degree. In America, the relationship between these two variables results positive. Also in Italy, the expectation was a positive correlation but, after the implementation of different linear regressions, a positive correlation is emerged, but with a high p-value indicating that the variable used were statistically insignificant.

To achieve this result, research about the women work situation has been made. In the background literature it is emerged that the main cause of the gender gap is principally due to the less time that women can dedicate to their careers, because the female gender is the one who dedicate more time to take care of children and home. It is also emerged that the types of works carried out by low-skilled immigrants' women are related to home services and children care. For this reason, it was expected that in the Piedmont municipalities in which more immigrants' women lived, the innovative start-ups founded by native women have more possibility to survive during time.

Of course, this is not the only cause of failure of the female start-ups, but there are some reasons that unites all the start-ups. The principal causes are analyzed in the background literature.

In chapter 2 is available a theoretical explanation of the statistical method utilized to study if effectively there is a positive correlation between these two factors.

The data about the Piedmont region immigration and the female-led women start-ups are taken from ISTAT (population data), AIDA (start-ups and enterprises) and a database of the principal Italian start-up. Then, a unique database explained in Chapter 3 is created using the statistic software STATA.

To create the linear regressions, it was necessary to define the dependent variable (the *duration* of life of the start-up analyzed), the independent variable (four different to evaluate different cases) and the control variables, other variables besides the independent one that can influence the dependent variable.

In the following chapter, the different variants of linear regressions are explained and the results obtained are obtained. The four models with the different independent variables turned out to be not statistically significant. This means that the variables chosen don't

describe properly the change of the independent variable. However, it is emerged that increasing the geographical area considered, the p-values (that indicates the statistical significance) have better values than the ones obtained in the regressions in which the independent variable was referred to municipalities.

For this reason, the several problems encountered during this analysis, including the lack of data and the different approximations made, are described in the last chapter.

Background literature

What is a startup?

"A startup is a temporary organization designed to search for a repeatable and scalable business model" according to Steve Blank. A startup is temporary because, as we will see, only organizations that have less than 5 years are considered start-ups, after those, if they survive, they become enterprises.

Repeatable refers to time; it refers to the start-up's capacity to repeat more times the same operations and obtain the same results, for example, be able to obtain a profit thanks to a service that is produced internally and that is part of the core business.

Scalable because the goal of a start-up is to increase revenues faster than the increase in costs.

Moreover, the business model must be new and innovative. For this reason, another definition of a startup says that it is "a company working to solve a problem where the solution is not obvious, and success is not guaranteed". In fact, to solve a problem that is never resolved before (innovation), the probability of failure is higher. Indeed 90% of startups fail after the first year of life, for many different reasons.

Countries realized that every state that promotes innovative enterprises would be a Country with a good probability of increasing economics.

In 2012, Italy has started to promote the development of innovative organizations, the start-ups, to increase the Italian economy, decrease younger unemployment, sustain innovation and social mobility, and attract funds and talent from foreign countries.

In Italy, the normative DL 179/2012, art.25, comma 2 (*Start-Up Act*) defines the following requirements for an innovative enterprise:

- newly established or have been incorporated for less than 5 years;
- headquarters in Italy, or in another EU/EEA Member State (provided a production facility or a branch in Italy);
- annual turnover lower than 5 million;
- do not distribute their profits;
- mission statement: the development, production and commercialization of innovative product or services with clear technological component;
- do not result of a company merger or split-up, or of a business or branch transfer.

To consider the start-up as innovative, at least one of three criteria must be met:

- R&D expenditure corresponds to at least 15% of the higher value between turnover and annual costs;
- The total workforce includes at least 1/3 of PhDs, PhD students or researchers, or at least 2/3 of the team hold a master's degree;
- 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.

In Italy, during the third semester of 2022, the innovative startups were 14.708, the highest number ever. There is also an increase in rate of the medium production, with a total production of 2 billion euros, +700 million of the prior year.

Otherwise, statistics show that 90% of startups fail during the first year. In the following chapter, we analyze what makes a startup successful and why the rate of failure is so high.

Startup phases and causes of unsuccess.

The startup life is divided into different stages:

- *Early stage*: this stage refers to the development of the idea of a product or service, with a description of specific characteristics. In this stage to develop the idea, the startup needs some funds: the investors in this stage take a lot of risk because they don't know yet if the project will have success or not. So, during this phase, it doesn't matter only the idea, but also the team has a crucial role because it is necessary to be able to convince investors that the startup's product is valuable and the market fit. After selling the product to early customers, the team must show the sales dynamics and the prospect of growth.
- *Venture-funded stage (Growth):* This phase starts when the startup has received a Series A fund and it is ready to expand its customer base on a large scale. Now more funds are necessary to support growth operations and establish sales function. The goal of this stage is the growth in the market and the capture of part of it.
- *Late stage*: This stage is all about performance. Investors are venture capitalists, equity firms, growth firms, corporate venture capitalists, and family offices. In this

phase, typically startups are looking for expansion opportunities or are considering an exit.

Watching these phases, it is easy to understand why only 1 startup to 10 survive.

From past literature, several reasons have emerged that can lead to the failure of the startup, both internal and external, because having only a good idea isn't enough.

First, the *entrepreneur* plays a fundamental role. From the investors' point of view, it is better a startup founded by a team than by a single member, because in a team it is possible to compare ideas with each other and it is easier to find errors if present. Especially in innovative startups, the importance of the team is in evidence since the probability that a single person has all the digital competencies necessary to develop the project is very low.

A team is more efficient if the study backgrounds of members that compose the group are different. In this way, there are competencies necessary to do all the activities related to the management and implementation of the project, without asking for external suggestions because there are already enough suitable people's points of view in the team. But there aren't only positive consequences in having a large team; in fact, one of the causes is the disputes that may arise within the group, due to different expectations about the project or different levels of interest in achieving the objectives. So, it is important to find an equilibrium within the group and to communicate in the clearest possible way, in this way, it is possible to find a fast solution when there are discordances.

Another problem that can emerge, is that, even if all the members of the team have a degree or a master's degree, or even Ph.D., they usually underestimate the *budgeting problem*, until the time in which they realize that they don't have enough money, especially in the beginning stage. It can happen not because they receive little money, but because, probably, they don't consider in the right way the equilibrium between cost and revenues, which initially are not higher for a startup.

That's why startups created by enrepreneurs who have already worked in other circumstances, and not by newly graduated students, have a higher probability to survive and have success because this kind of problem is considered from the beginning and the right portion of attention is dedicated to them.

So, the construction of the perfect team is essential to the success of the startup.

Another problem which we have already talked about is the lack of money. During the different stages, the startup is financed by different kinds of investors.

In the beginning, when there is only the idea and the volunteer to raise a startup, the typical form of financing is given by founders, friends, and relatives. So, the team needs to have a personal fund or must be able to convince relatives and friends to believe in their project.

If the funds are not enough, during the early stage the investors who invest in the startup and take most of the risk are Business Angels. Business Angels are high net worth individuals who invest directly in startups usually in exchange for a minority ownership share. They have industry expertise and invest professionally, trying to have a return on the investment. To minimize their risk in investing in the early stage of the startup, they have a diversified portfolio and sometimes they don't have only monetary motivation, but they enjoy helping young businesses succeed.

During the other stages of the startup, venture capitalists and professional investors invest in the project to help the growth. Of course, they take less risk with respect to business angels, because the business has already started and it is clear if it can have success or not.

The are lots of differences between BAs and VCs, in addition to the stage of investing, including the fact that the investment done by professional investors is done to disinvest and they have the expectation to multiply their investment by 10, so they have a monetary motivation. For this reason, they are not interested in the future of the firm, the only important thing is the return of the investment. BAs have a hands-on approach, they give help and suggestions to the team, while pure investors don't participate in decisions taken by the organization.

In Europe, there are some policy interventions in the angel market, as they take a lot of risks in investing in the early stage. In Italy, for example, with "*Decreto Crescita*", Business Angels who invest in Innovative Startups can take advantage of a tax benefit equal to 30% of the amount invested as capital, whether it is a natural person or an investment company. During the COVID-19 pandemic, with "*Decreto Rilancio*" the tax benefit was 50% with a maximum amount of 100 thousand euros for innovative startups. The investor has to maintain the investment for at least 3 years, otherwise, he will lose the incentive.

Policy creates more business opportunities through Business Angel Networks (BAN), to connect BAs with entrepreneurs, through an educational program, to also increase the participation of women in BAs, because the angel market is predominantly comprised of male investors. In 2015 women angels represented only 25.3% of the

angel market, due to the gender gap, risk aversion, low knowledge of what BAs do, and little time (later we will see why women have typically less time than men). But during the last years, the percentage of women business angels has increased reaching 27% in 2022, the highest historic value.

Every country has its own Business Angel Network, in Italy IBAN (Italian Business Angel Network) is the principal group of members of angel investing, it gives the list of local B.A.N. of each region. During 2022, investment done by business angels increased by 77% with respect to 2021, especially on ICT (Information and communication technology) startups.

However, if there are so many BAs, why it is not easy to find funders for startup developers?

As we said before, the idea is not the only thing that matters. The team of the startup must convince investors that their product or service can have success and can give gain to them. Investors prefer to invest in startups in which, in addition to the team, there are external entrepreneurs, who have advanced experience in entrepreneurship and can help in the management of the new organization. This is why investors, during the growth stage, prefer to invest in startups in which BAs have already invested and control the minority of it, giving their contribution.

The thing that matters the most is that the entrepreneur and the team must convince the investors that their product has a competitive advantage with respect to all the other similar products in the market. In fact, another important thing that a startup must consider is the timing market: the product or service must be realized in the perfect market time. It means that people start to need this product to solve a kind of problem that, in the moment of development, cannot be resolved in any other way. It is rare to create a product or a service that doesn't exist yet, but the goal of a startup is to create a product that can have additional functions and is better than the previous one so much so that it can take a part of that market.

After obtaining the money to start producing the product, a lot of startups fail to manage that money. During the business plan planning, a plan that shows the next business choices of the organization is essential to make understand investors what the organization will do. Usually, entrepreneurs fail in predicting the costs and don't use their resources in the best possible way. For this reason, external expertise are needed. If there are more than one, they solve the budgeting problem, but they can create

another kind of problem: different discussions and ideas that lead to a slowdown of operations. So, it is important to find the right equilibrium.

Another thing that entrepreneurs underestimate is the opinion of customers. Typically, the startups make the error of not considering the suggestions of the first customer, which instead is so important because changes and improvements to the product are less costly if done at the beginning phase.

So, it is important to conduct a survey to first customers to understand the things that can be changed and, in this way, capture a major market when the product is finished.

So, things that really matter to not fail from the internal organization of a startup are a good team able to find funders to realize the idea, external expertise to manage the costs and the plan, good product with characteristics better than the other already existing in the market and a good time market and, in the end, take care of the customers opinion.

What about external factors that influence the success or unsuccess of a startup?

It is important to also consider the external factors because the problems aren't only in the internal organization, but sometimes, even if a startup has all the internal factors to the best of its possibilities, it can fail.

The first thing to consider is the Government of the place in which the startup is created.

This is relevant because according to the state in which the startup was founded, it will have different chances of success. In Italy, there are different benefits to innovative startups, for example:

- digital and free constitution: on February 27th, 2016, it was introduced the possibility for innovative startup, born in the form of s.r.l., of drawing up a memorandum and articles of association through a standard model typed, validating the content with digital signature;
- Incentive on the investment in innovative startups and innovative PMI (mentioned before, the 30% of tax deduction from IRPEF);
- Free access to guarantee funds for PMI: easier access to credit for a maximum of 2.5 million euro (DL 179/2012 art.30);
- Smart & Start Italia: it is a zero-rate financing for entrepreneurial projects with costs between 100 thousand and 1.5 million euros. The financing covers 80%

of the eligible expenditures. If the startup is founded by a woman or by younger than 35 years old guys, the coverage can rise to 90%;

- ICE agency services: give assistance from a financial, normative, or social point of view. The startups have a discount of 30% on these services. ICE agency is the promoter of the Global Start-up Program, which provides internationalization training and the possibility to receive a financial contribution from foreign incubators (DL 179/2012);
- Innovative start-up can transform into an innovative PMI entering into Enterprises register having the access to benefits of innovative PMI, including the exemption from the payment of stamp duty and secretarial fees due for the obligations relating to entries in the Commercial Register, and from the payment of the annual fee due in favor of chambers of commerce;
- Possibility to receive funding through a campaign of equity crowdfunding (DL 179/2012 art.26).

In addition to the benefits that the Government can offer, the influence given by the environment that surrounds the startup is relevant. In fact, from past literature, it emerges that the influence that other enterprises that are in the same area can increase the performance of the start-up. So, it is important to evaluate all the possible competitors that are in the same area of the market. The network effect generated by the presence of other enterprises inside the same territorial area leads to an increase in the economic benefits to all the start-ups in the same area, as much as the increase in the magnitude of the ecosystem. In this case, the innovative product or service must have something that gives it an advantage over the others.

The dimension of the organization matters because small organizations have more difficulties in receiving demands and this leads to less revenue and consequent failure. The market has a higher influence on the success of the startup: when a new innovative product is invented from zero and put on the market, doesn't always find customers available to buy it. People tend to be habitual and because of this they are often averse to buying new products, they prefer using the ones that they already have. For this reason, has a higher probability of success something that already exists on the market, but that has something plus concerning the precedent product. For example, the phone: when smartphones started to have an incorporated camera, people changed their old phones with the new ones, but before the introduction of the

camera, they refused to change their phones with something that had the same function as what they already hold.

For this reason, it is important to identify and study the competitors of the startup, understanding which are their strategies, which are their scopes and how to use in the better way their own resources. Often entrepreneurs don't consider this phase of the study of competitors important and, due to this, they fail because this loss can be a higher obstacle for the growth of the startup.

Obviously, market trends and inflation also influence the success of a startup.

In a period of high inflation, costs are higher, and the startup needs more financing and it is more difficult to find investors disposed to invest larger amounts of money. But this is not the only problem, having higher costs also means a higher price of the product in the market and less demand from customers, and, consequently, less return to the startup. For this reason, it is also important to study the prices of similar products sold by competitors, searching to have a price competitive advantage; if two products are similar, but have different prices, the customers will choose the product that costs less.

To have an advantage over its competitors, the startup also must invest money in marketing and adv. This is one of the causes of failure because they don't consider fundamental making advertisements for the startup, but this is the only way to make themselves known by customers. This is the crucial point for a startup: as it is new, it has to find more channels of communication to make itself as more known as possible. Without it, remains in a narrow circle of knowledge (family, friends, and colleagues) and will never succeed in the market.

In addition to all these lacks that a startup can have, the Journal of Entrepreneurship, Management and Innovation made a study similar to what will be shown below. In fact, they made a linear regression for a dependent variable which is the survival rate of a startup. The independent variable and the control variables that can influence the survival rate, in their opinion, are:

- how a startup is financed (that can be through angels, equity companies, venture capital, debt, government funding, and FFF sources);
- intellectual property (the number of patents that the startup owns);
- collaboration with other entities, such as university or company partnerships, human or social capital;

- a binary variable for the founder demographics (to differentiate the origin of the founders);
- the industry;
- the interactions.

As previously articulated, enterprises funded by government entities exhibit a higher propensity to failure in contrast to those backed by angel investors or creditors. The establishment of a competitive advantage assumes great importance, particularly in the context of university partnerships and commercial collaborations. While the number of patents, copyrights, and trademarks in isolation may not exert a discernible influence, when factored into the equation, a higher number of patents and copyrights contributes to a diminished risk of failure.

The educational backgrounds and previous experiences of the founders within corporate entities have a statistically significant impact. Having more experience before starting a business makes it more likely that the business will succeed for a longer time. Also, the level of education of the founders has a positive effect on how long the business survives.

This study explores "founder identities," finding that founders' traits like where they're from and how old they are can affect how likely their business is to succeed. Older entrepreneurs tend to be more resilient than younger ones. Different industries also have different success rates.

Also, looking at founders' backgrounds, who they work with, and the investors they have shown big differences in survival rates. For example, entrepreneurs' experience matters most when they have angel investors.

Overall, this study shows that many things affect whether a startup survives or not, like education, experience, where the founders come from, and who invests in them.

Women and the labor market

In this section, it is analyzed the literature about women and the labor market. On October 9^{th,} 2023, the Economic Nobel Prize was won by the economist Claudia Goldin for her numerous studies about the gender gap. She discovered the key factors of the gender gap in the labor market, and she provided a detailed explanation of what women earn and of their participation in the labor market over the centuries. Her

research shows the causes of change and the most important causes of the still-existing gender gap.

The women's story in the labor force began at the end of the Second World War. The role in women's society was only enclosed within the walls of the house: they had to cook, take care of children, clean the house and clothes, and so on.

During WWII men were forced to leave their families and cities to go to fight for the State, so women found themselves alone and started to do the kind of work that, since that moment, was done only by men. For this reason, at the end of the War, women started to fight for their rights and their freedom, by receiving their first acknowledgments, for example in Italy on June 2nd, 1946 they received the right to vote, and in 1948, with the entered into force of the Italian Constitution, some fundamental principles are protected in the field of equal rights for men and women.

In 1968, there was the fight of the feminist movement which asked for equal rights and the affirmation of one's own feminine identity. To protect these rights and respect for the principles laid down in the Constitution, Law No. 903 on equal treatment for men and women in matters of employment entered into force on 9 December 1977.

Despite this, there is still present a divergence between men and women in the labor market.

In the paper "A Grand Gender Convergence: Its Last Chapter" by Claudia Goldin it is demonstrated that in the last century, the success of converging roles of men and women is becoming a reality. In the economist's opinion, what needs to change to obtain equality in the labor market is the change in the structure of the work. It is necessary to have more time flexibility and have a different way to remunerate for the working hours. Right now, the wage per hour is higher for people who work long hours or work in particular hours (for example during the night) than the wage earned by people who work during the normal timetable.

Until 1970, most unemployed women weren't in the workforce since they were married or having their first child.

A way to evaluate the gender gap is by considering the earnings per unit time wage.

Over time, the gender wage gap decreased as human capital investments between females and males converged. The differences in the medium salary that still exist are because men are disposed to work long hours or during particular hours of the day or night. After all, from surveys it emerges that, in case of children or parents to look for, women take care of them, sometimes also renouncing some work hours. So, the

different numbers of work per day have a large effect on the time-adjusted earnings in some occupations. This is why men and women start working with equal earnings and then, when they grow and maybe they have children, the earnings change. Of course, the gender wage gap differs by occupation. For example, for what concerns technology occupations, it is found that it has greater gender equality in earnings. Why? Because this kind of work permits to have more time flexibility and men and women who have the same educational background earn the same wage.

To prove that these statements also apply to Italy, as Data Istat of 2016 shows, 97% of women take care of their children, against 73% of men. The gap is higher if we refer to the house cure and cooking, only 13% of men take care of this housework, while 81% of the female population usually do this work.

Italy and other states surveyed the "Use of time", to understand what people do during their time, choosing between leisure, sports, or taking care of the house and children. It is emerged that, among European states, Italy is the 5th state for the time dedicated to work not remunerated; people between 20 and 74 years old dedicate to them around 3h30' per day. But the most important data is that Italian women are the ones who dedicate more time to those work (5h02'), while Italian men are in the last position, as they dedicate less than 2 hours per day. As we can see, there is a big gender gap in Italy, even if in the last few years is decreasing. Regarding the participation in the cure of sons younger than 18 years old (the age at which one became an adult in Italy), data shows 73% of mothers against 46.8% of fathers. But the gap is even bigger if we look at the time spent on what researchers called "repetitive activities", such as cooking and housekeeping: 3h30' for women, which is 76.5% of their time in domestic work, and only 43' for men, which correspond to 48.3% of their time in domestic work.

Through these data is easy to understand that this time that women dedicate to family and house, could be one of the problems that prevent women from having a successful career and increasing their employment status.

In Italy, only 7.1% of families declared to ask for the help of a housekeeper or cleaning lady (maid) in domestic work.

The last data INPS available (2020) shows that in Italy domestic workers are more than 920 thousand, of which 87.6% are women. Fewer than 70% of them are foreign people, 48.5% from extra-UE states, and 20.3% from UE states, Italians are only

31.2%. This data cannot be precise due to the illegal work, often people try to avoid taxes by paying in cash their workers.
This is why we decided to consider only women in this research (and not men) because they are the ones that most deal with these jobs.

Women Entrepreneurs

Female-led business in Italy refers to enterprises founded by women, societies in which there are at least 60% female employees, or joint-stock companies whose shareholdings are no less than two-thirds women and whose management bodies are made up of at least two-thirds women. We can make a distinction by saying that if the participation of women is higher than 50% we have a "majority presence", if it is higher than 60% we have a "strong presence" and if all the members and the holder are women, we have an "exclusive presence" (Figure 1). The 90% of the female enterprises are of the last type described.

Grado di imprenditorialità femminile	Società di capitali	Società di persone e cooperative	Ditte individuali	Altre forme giuridiche
MAGGIORITARIO	*% di cariche + % di quote >50%*	*> 50% "soci"*	*-*	*> 50% "Amministratori"*
FORTE	*% di cariche + % di quote >4/3*	*> 60% "soci"*	*-*	*> 60% "Amministratori"*
ESCLUSIVO	*100% di cariche +100% di quote*	*100% "soci"*	*Titolare*	*100% Amministratori"*

Figure 1 Female entrepreneurship grades

In Italy, at the end of 2022, female enterprises were more than 1 million and 346 thousand, which corresponds only to 22.8% of Italian entrepreneurship. The distribution has a geographic variety, it depends on each region.
In this study, it is considered only the female innovative startups/enterprises, but some authors claim that there are differences between the women entrepreneurs and, probably, these differences influence the survival rate of the startup.
To make an example, a paper affirms that there are 4 different kinds of entrepreneurs: the innovators (the ones that we consider), the traditional, the housewife and the radicals.

The *traditional* are the women who work in positions in which there is no chance to advance in the position of the career and their only motivation is that they want to contribute to the family income, without any wishing to affirm themselves and be realized in the work.

The *innovators*, instead, decide to start their business to solve the problem of a low chance of making a career in dependent work. Their goal is to have success and the work represents an important aspect of their life. So, in their opinion, starting their own business is the only way to achieve their career goals.

The *housewife*, instead, is highly connected with traditional work, but for them, entrepreneurship is a way to feel grateful, even if there is no monetary return and it permits them to be independent by some domestic work. Most of the time, their entrepreneurship activity is a second work, as a pastime they care very much.

In the end, the *radicals*, are feminist entrepreneurs who start their businesses with the sole aim of improving the living conditions of women in the future. So, they do it due to their beliefs and they are not interested in making money or having success in their working careers.

Other authors describe female entrepreneurs differently, but it is interesting to notice that the *innovators* category is present in all the subdivisions seen. The description of this category is similar to the one written before, in fact another author describes these women as women who don't have children and, for this reason, have dedicated more time to work and have managed to succeed. To overcome the obstacles they found, they decided to open their own business to be fully satisfied and avoid all the discrimination that can be found in working places. Their desire for success and inventiveness is very similar to male entrepreneurs.

After explaining these distinctions between female entrepreneurs, it is possible to analyze the different types of female startups in Italy. A press communication by Unioncamere admits that the female innovative startups listed at the end of 2022 are 2 thousand, 572 more than the previous year. Above 70% of these enterprises are in the service sector, 15% are in manufacturing activities and 4.6% are in the commercial sectors. The Italian regions in which there are more startups are Lombardia, Lazio, Campania, and Emilia Romagna (Figure 2). After, it is analyzed the case of Piemonte. If we consider all the female enterprises, and not only the innovative ones, it is interesting to see that they are placed more in the Italian Center and in the

"Mezzogiorno" (South Italy) area, especially in Molise, Basilicata, Abruzzo, Sicilia e Umbria.

Startup femminili per regione

Dati al 30 settembre 2022

	30 settembre 2022									
	Agricoltura	*Attività manifatturiere*	*Costruzioni*	*Commercio*	*Trasporti*	*Turismo*	*Assicurazioni e credito*	*Servizi alle imprese*	*Altri settori*	***TOTALE***
ABRUZZO	1	14	3	1	0	0	0	29	0	**48**
BASILICATA	2	0	0	2	1	0	0	31	1	**37**
CALABRIA	0	4	0	2	0	0	0	31	4	**41**
CAMPANIA	1	26	2	6	1	4	0	150	14	**204**
EMILIA-ROMAGNA	0	35	0	7	0	1	0	96	4	**143**
FRIULI-VENEZIA GIULIA	0	1	0	1	0	0	0	20	3	**25**
LAZIO	1	24	1	9	3	2	0	208	15	**263**
LIGURIA	0	5	0	4	0	0	0	24	3	**36**
LOMBARDIA	5	47	2	31	0	4	1	359	21	**470**
MARCHE	0	15	1	1	0	0	0	46	6	**69**
MOLISE	0	3	0	0	0	0	0	15	4	**22**
PIEMONTE	0	15	1	4	1	0	0	63	6	**90**
PUGLIA	0	16	0	5	0	1	0	77	4	**103**
SARDEGNA	0	5	0	1	1	0	0	23	2	**32**
SICILIA	2	23	3	7	2	2	0	84	1	**124**
TOSCANA	2	17	0	8	0	0	0	76	1	**104**
TRENTINO-ALTO ADIGE	0	7	0	1	0	0	0	18	0	**26**
UMBRIA	2	11	0	0	0	0	0	19	0	**32**
VALLE D'AOSTA	0	0	0	0	0	0	0	3	0	**3**
VENETO	2	38	1	1	0	0	0	83	3	**128**
ITALIA	**18**	**306**	**14**	**91**	**9**	**14**	**1**	**1.455**	**92**	**2.000**

Fonte: InfoCamere - startup.registroimprese.it

Figure 2 Report UnionCamere Start-up distribution

It is interesting to notice that the highest number of female startups is in Lombardia, followed by Lazio and Campania, while the region that have less startups are Valle D'Aosta and Molise, probably also due to the fact that they are smaller than other regions. Anyway, if we consider macro-area (such as North, Center and Sud Italy), they are almost equally distributed.

In addition to geographical reasons, there is another thing that differs for female startups: the difference between women entrepreneurs and men ones. In fact, after the beginning phase, the organization of the following steps is different between females and males. There are two different approaches called the "*strategic approach*" and the "*evolutive approach*". The strategic approach consists of scheduling activities and setting periodical goals to be achieved. The final goal is the growth of the enterprise and the result is based only on economic parameters. A high amount of money is invested in the activity with a credit and managerial abilities are necessary for the success and development. This kind of approach is typical of the men entrepreneurs, their idea to found a startup arises from a project in which development is clearly

planned. The money is taken from credit banks and, in addition to partners, from the beginning there are employees.

The evolutive approach, more typical for women, to realize the product or service used only personal resources and the achievement of the goal is measured not from an economic point of view but depends on the satisfaction of the customer. Women based their ideas more on their instincts than on planned work. This female approach, today, is considered a positive thing because, in case of uncertainty, it is better to concentrate on communication and informality to permit the enterprise to be more flexible and to survive in a dynamic environment.

Motivations that bring women to create their own business

One of the main reasons women start their own businesses is because they want to be independent and feel fulfilled in what they do. Nevertheless, the notion of independence carries diverse connotations for women depending on the stage of life they are in.

For young people, starting their own businesses is often seen as a great way to avoid the prevalent problem of not having a job. It gives them a chance to get into the job market and use their skills. Conversely, for more mature individuals, independence takes on a weightier significance, signifying economic self-sufficiency and the opportunity to realize their professional aspirations.

Another aspect of wanting independence is shown by people who have spent many years taking care of their children and now want to go back to work. For these women, independence signifies a chance to regain their economic autonomy and reclaim their professional identity.

For some of them, independence translates into the liberty to engage in activities they are passionate about, accompanied by the unrestricted control of their own time. It is a manifestation of the aspiration to prioritize personal interests and desires, directing their lives according to their individual preferences.

For this reason, there exist four distinct avenues through which individuals venture into the realm of entrepreneurship:

1. Transition from Dependence to Independence: The most prevalent impetus involves the shift from a conventional employment status to one that is self-sufficient and autonomous. This change often takes place within the same industry as the previous employment, leveraging familiarity and expertise.

2. The Path of "Entrance": This path is common for young women who choose not to follow the usual route of gradually advancing in a job where they depend on others. Instead, they want to be independent and find fulfillment in their work. Many of these women start businesses in fields like serving others. They are very talented and put a lot of time and effort into their work. They aim to succeed and grow professionally.
3. The "Return" Category: This group includes women who, after taking time off from work, want to start working again. Once their children are grown up, they feel like starting their own businesses. These women often gravitate toward sectors traditionally associated with femininity, such as the handicraft industry.
4. Embracing Tradition: this last option involves a dedication to maintaining family traditions and histories. Women in this category assume prominent roles in family enterprises, perpetuating the businesses initially founded by relatives. This shift in roles often stimulates innovations within the company. At times, the entry of women into these enterprises becomes a strategic necessity to support new ventures, especially in cases where the familiar enterprise is a secondary occupation for the original founder.

To sum up, women start their own businesses mainly because they want more flexibility with their time and they want to establish themselves professionally while working towards their goals.

Patricia Cortes, during her research, found that the successful of native American women depends also on the low-skilled immigrated women in the same city. She demonstrates that, with high level of low- skilled immigrated women, who often do the housekeeping or babysitting work, the successful in the career of high-killed women is higher. In fact, as it is said before, women, in addition to work, in most cases have also to do the housework, such as taking care of children, cooking, clean. For this reason, they can dedicate less hours to work with respect to men. But, with the help that an households or babysitter can give, the high-skilled women can dedicate more hours to work and, in this way, can aspire to have the same success and salary as men.

In her papers she considered only high-skilled native women because it was shown that the help of someone who does household services gives benefits only to high-skilled women and not to women that do normal works.

The intent of this thesis is to try to see if even here in Italy, in particular in the Piedmont region, female immigration affects the success of Italian women, especially the ones that decided to create or participate to an innovative startup.

Immigration

In order to reconstruct the study conducted in America, it is also important to analyze the immigration situation in Italy.

The Organization for Economic and Co-Operation and Development (OECD) site, shows an international immigration database, in which for each year from 2000 to 2021 there is the number of total migrants in each Country, including Italy. Filtering the database by gender, the number of women migrated to Italy during the year is represented in Figure 3.

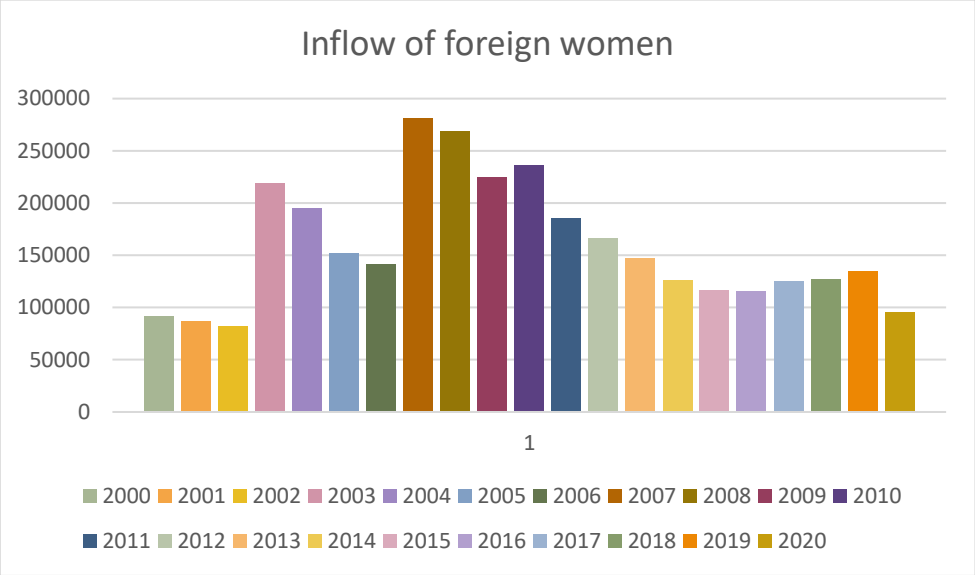

Figure 3 Inflow of foreign women in Italy

The higher inflow is in 2007 and the lower is during 2000, probably due to the Covid-19 pandemic.

Below (Figure 4) is shown the total number of immigrants in Italy (not only the female one):

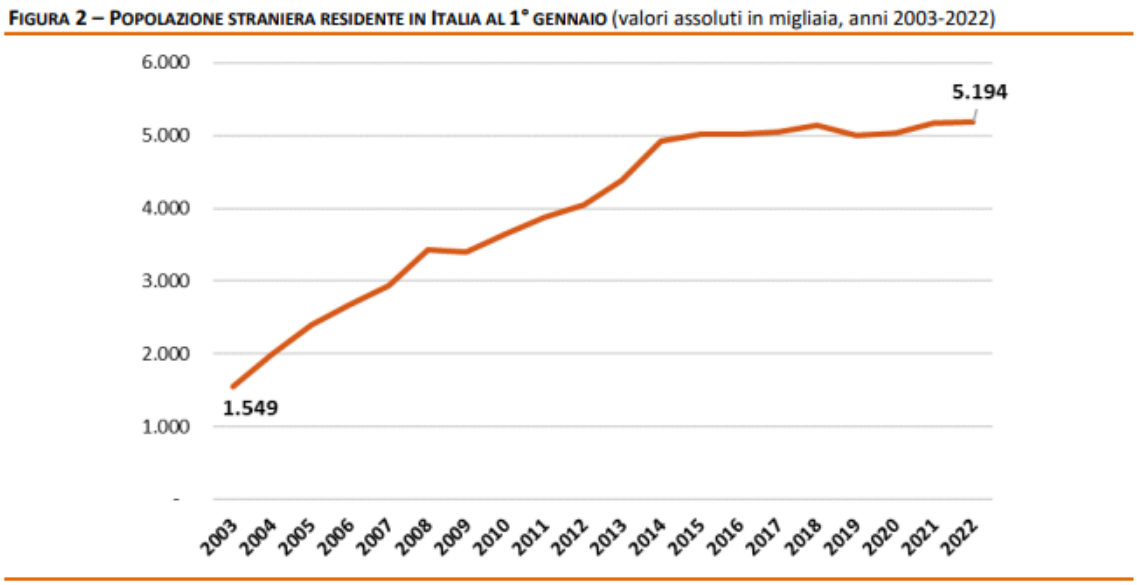

Figure 4 Total number of immigrants resident in Italy

Instead in the graph above (figure 5), it is possible to see the country of origin for migrants of any gender:

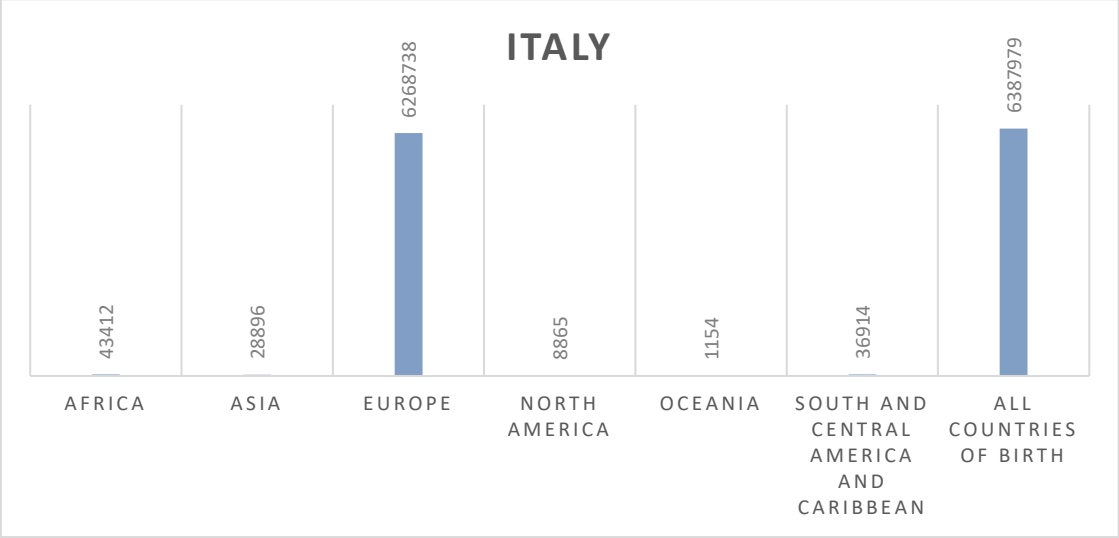

Figure 5 Country of origin

As it can be seen, most foreigners come from other member states of the European Union.

OECD also offers a database in which is shows the female employment rate by place of birth and educational attainment for women in the range of 25 years old to 65 years old.

To make an example, in Figures 6 and 7 are shown the rates of 2019:

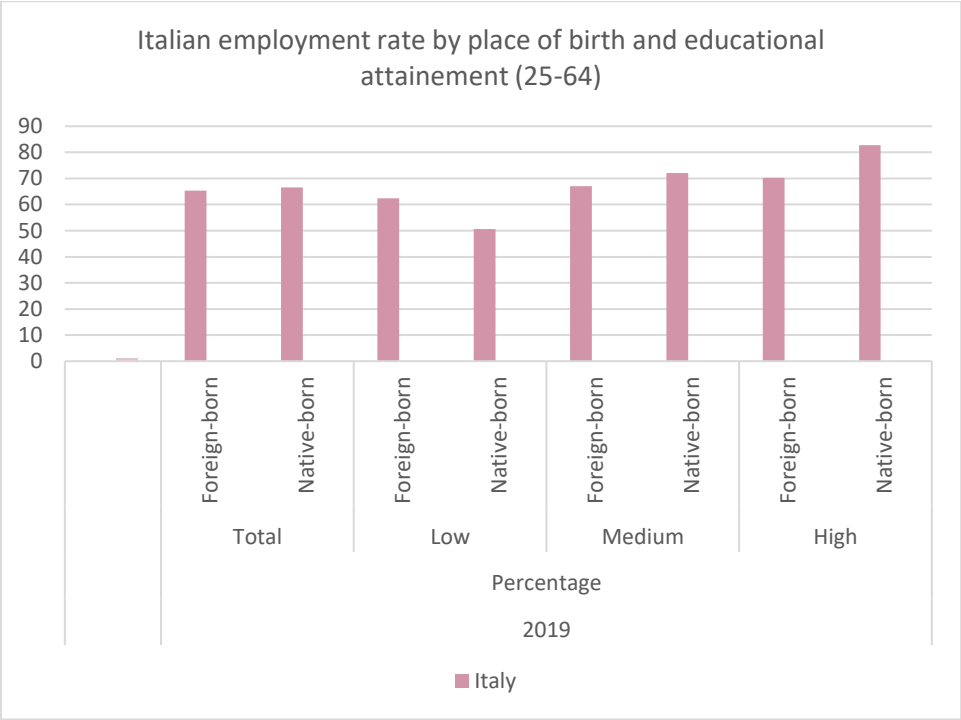

Figure 6 Italian employment rate by place of birth and educational attainment

Year	2019							
Unit	Percentage							
Educational attainment	Total		Low		Medium		High	
Place of birth	Foreign-born	Native-born	Foreign-born	Native-born	Foreign-born	Native-born	Foreign-born	Native-born
Italy	65,3	66,5	62,4	50,6	67,1	72,1	70,2	82,7

Figure 7 Percentage of employment by place of birth and educational attainment

It is interesting to notice that the rate of foreign women is higher than the native ones for what regards the low educational attainment, so it indicates that all types of work in which no special specializations or qualifications are required are mostly performed by foreign women. These works include those mentioned above, including maid, caregiver, babysitting, and cleaning lady.

For what concerns medium and high educational attainment, native-born women have a higher percentage of employment and the delta between foreign- and native-born employment rates is even higher for the high-skilled women. This is probably due to the fact that most of the women who move to Italy are low-skilled. In fact, a statistics studio revealed that in 2019 (to have the same data time as the database before) the percentage of foreign with a university degree was only 10.7%, the ones with a high school diploma were 34.4% and the low-skilled, with only a middle school diploma, were 54.9%. These percentages are similar to all the other years and each year the highest one is about low-skilled women.

"Ministero del lavoro e delle Politiche sociali" reveals that among the sectors in which foreigners are most employed are collective and personal services (around 35% as we said before), restaurants and hotels, agriculture, and constructions. This confirms what has been said before, because for collective and especially personal services any specialization is required.

In Patricia Cortes researches the incidence of low-skilled women influencing the survival rate of the female startup: this means that the foreign women help the native ones in works out of the working hours, permitting them to have more time to dedicate to work.

As it is shown in the previous chapter, the women take care of work relating to housing, childcare, and parental care, so, as the working hours of high-skilled women increase, it means that the foreigners help them in this kind of work.

The annual report published from National Observatory Domina 2023 demonstrate that the domestic work in Italy is increasing during the last years. With Eurostat data it is possible to compare the Italian situation with European one. In Italy the percentage distribution of services work (native and not) is 5.7% and the ones which have a family as an employer are the 35.2%, the highest respect to the other European states. In fact, in other countries the service employee is usually hired by association/firms instead of the family itself.

In Europe 2 million of domestic workers have family as employer. There are a lot of irregular workers, for this reason probably this data is underestimated.

In Italy, INPS data shows that during 2018 the domestic workers hired by Italian families were 859.233. This data varies so much over the years and the normative "*c.d. sanatoria*" of 2012, increased the number of this kind of employee because relation that existed before became regular.

There is a decrease of domestic workers during the years, as this graph shows (Figure 8):

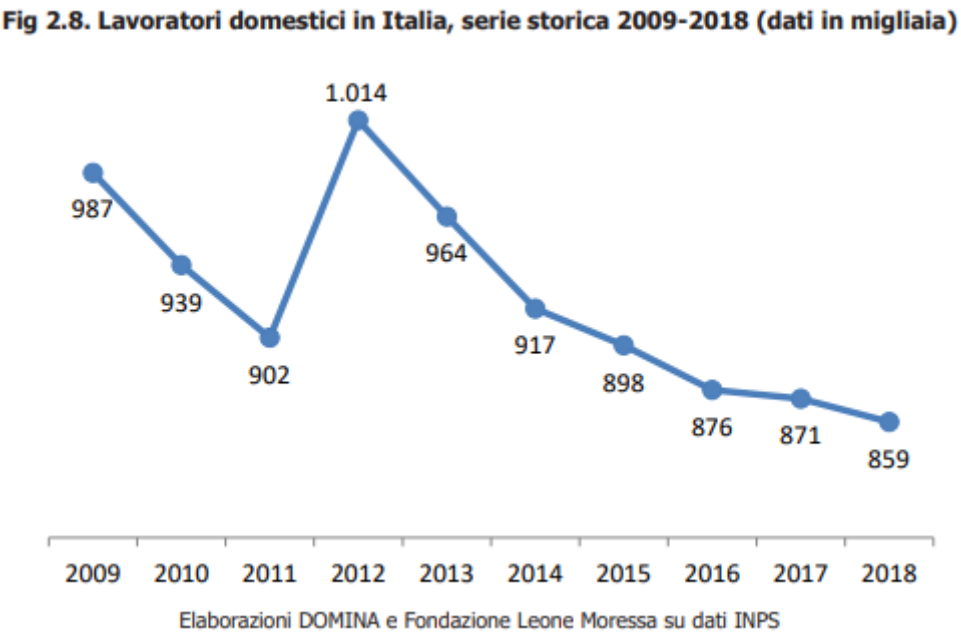

Figure 8 Domestic work in Italy from 2008 to 2018

This data include colf and caregivers without distinction of gender.

During 2018 the composition of domestic workers was:

- 40% came from East Europe;
- 28.6% were native people;
- 8% from Philippine;
- 6.8% from Sud America;
- 5.4% from East Asia.

From what concerns the age of these workers, those under 30 are only 6.1% of maid and 5.4% of caregivers. Instead, over 50 years old are 43.6% among maid and 56.6% among, caregivers.

The female component is 88.4% of all the domestic workers, this is why in our study we consider only foreign women.

As said before, it is needed to consider both irregular immigrants (without residence permits) and irregular contracts of work (clandestine work). For example, in 2018 the irregular rate was 58.3% of 2 million domestic workers.

As my research consider only the region of Piedmont, in Piedmont during 2019 there were 68.666 domestic workers, in Turin 22.109; 92.3% were women and 48.1% came from East Europe.

Interestingly, the middle age of employers was 66 years old and it is predominant (55%). Forecasts say that the number of caregivers is destined to increase due to the increase of over 80 and the decrease in children (0-14 years old).

If we consider data of 2023 about Piedmont, Domina annual report shows that the number of total regular domestic workers is equal to 68234, of which 51.8% colf and 48.2% caregiver. The middle age is 49.8 years old and the female gender is predominant (90.7%). The origins of these workers are:

- 40.9% from Est Europa;
- 31.7% from Italy;
- 6% from Asia;
- 12.3% from America;
- 8.7% from Africa;
- 0.4% from West Europa.

In Figure 9, it is possible to see the number of colf each 1000 residents (in fourth column) and the number of caregivers for each 100 residents with 79 or more years old ("badanti", in the last column).

Province	COLF	Distr. %	Colf ogni 1.000 abitanti	BADANTI	Distr. %	Badanti ogni 100 anziani + 79 anni
Torino	21.911	62,0%	10,0	18.334	55,8%	9,4
Alessandria	3.011	8,5%	7,4	2.646	8,0%	6,7
Asti	1.594	4,5%	7,7	1.477	4,5%	8,0
Biella	1.125	3,2%	6,7	1.382	4,2%	7,9
Cuneo	3.612	10,2%	6,2	4.332	13,2%	9,2
Novara	2.388	6,8%	6,6	2.526	7,7%	8,7
Verbano-Cusio-Ossola	707	2,0%	4,6	1.130	3,4%	7,8
Vercelli	1.010	2,9%	6,1	1.049	3,2%	6,7
PIEMONTE	**35.358**	**100,0%**	**8,3**	**32.876**	**100,0%**	**8,7**

Figure 9 Colf and Caregivers by Piedmont region 2023

As it possible to notice, Turin is the province in which more domestic workers operate.

Patricia cortes studies

To better understand the goal of this research, it is necessary to analyze Patricia Cortes' work, as our research endeavors to replicate a study she conducted in the United States but within the context of Italy.

As previously outlined at the outset of our analysis, the primary objective of this research is to investigate whether, akin to the situation in the United States, the immigration of low-skilled women can have a positive impact on the career trajectories of native women in Italy.

In her paper titled "Low-Skilled Immigration and the Labor Supply of Highly Skilled Women," Cortes posits a compelling thesis. She asserts that immigrants lacking formal degrees or educational qualifications predominantly find employment in the service sector, especially in roles such as private household occupations. Furthermore, her research covers a significant correlation: in regions where there is a significant influx of immigrants, the price for these service-based occupations tends to decrease. Consequently, native women are allowed to substitute a portion of the time they would otherwise invest in household activities by dedicating more hours to their work.

Crucially, Cortes' study focused exclusively on highly skilled women in the United States. This decision was guided by her observation that women earning higher wages tend to benefit more of the help from the labor contributed by immigrant women, given the resultant reduction in service prices. Moreover, highly skilled women can leverage this dynamic to request additional work hours, thereby advancing in their careers. For example, they may use flexible domestic support to manage responsibilities such as childcare or eldercare.

To execute her research, Cortes used a statistical linear regression. The main point of her analysis was about how bringing in low-skilled immigrants could lower prices.

Their analysis using linear regression focused on understanding how women's labor choices (referred to as "n") changed over time ("t") and across different cities ("c"). They looked at several factors including how many hours women worked per week, their likelihood of working 50 or 60 hours a week, and wages ("w") in different regions (“j”). They also considered whether more women immigrating to a city was linked to cities that already had a lot of immigrants from the same places (which was a theory supported by previous research). To make their analysis stronger, they included other factors like using men as a comparison group.

The outcomes of this regression analysis revealed fascinating results. Notably, low-skilled immigration exerted an impact on the labor supply decisions of women earning higher wages, thereby underscoring its significance. Moreover, it emerged that low-skilled immigration significantly enhanced the likelihood of native women dedicating more hours to their work, providing valuable insights into the intricate dynamics of labor supply and immigrant influence on the workforce.

In another research conducted by Patricia Cortes, titled "When Time Binds: Returns to Working Long Hours and the Gender Wage Gap among the Highly Skilled," an important point emerges. This study underscores a critical factor underlying the persistent gender wage gap: the division of labor within households. It becomes evident that women, who typically shoulder a disproportionate weight of household and childcare responsibilities, often find themselves constrained in terms of the time they can allocate to their professional work compared to their male counterparts. In essence, the problem is not the wage per hour, but the amount of time that individuals can dedicate to their work.

This temporal constraint is particularly significant in professions where success is directly correlated with the quantity of time invested. Therefore, it is fundamental the ability to balance professional commitments with familiar responsibilities. In this context, the presence of a support system, such as low-skilled immigrant workers capable of assisting with childcare and household duties, permits to high-skilled women to achieve greater professional success.

Furthermore, Cortes' research considers also the different work choices made by men and women, revealing another contribute to the gender wage gap. These disparities can be attributed to different factors. One such factor is the differential propensity for risk-taking between genders. Additionally, excluding the progress made in recent years, there remains some gender discrimination within certain segments of the job market. These and many other factors influence the distinct career paths chosen by men and women, amplifying the wage disparity.

In summary, Patricia Cortes' thorough study highlights how the difference in pay between men and women is complex. It's not just about different wages but also about how women spend their time, the jobs they choose, and their roles at home. The research also shows how low-skilled immigrant workers could help reduce these disparities, making it easier for highly skilled women to succeed in their careers. Overall, these findings give us important information about the many factors that

create gender differences at work and give us a good starting point for looking at similar issues in Italy.

Piedmont female entrepreneurship

Given the focus of this research on female entrepreneurship in the Piedmont region, it is interesting to look closely at a study by Bertolini and Goglio about female entrepreneurship in Piedmont, especially in the province of Cuneo. Their choice of Cuneo as a focal point for this research is due to several reasons. First, the province of Cuneo was selected due to the high degree of enterprise fragmentation within its boundaries. Furthermore, at the time of this study, Cuneo had the highest density of enterprises compared to other provinces in Piedmont.

The study reveals many reasons why women start their own businesses. Mostly, they want to be financially independent and have a successful career. Often overlooked in existing literature is the crucial role played by male counterparts, such as husband or boyfriend, in providing both psychological and material support crucial for the realization of women's entrepreneurial aspirations.

Some women, instead, affirm that they found the courage to become entrepreneurs because during their work experiences they found adverse working conditions. Also, some people start their own businesses to continue important family customs and traditions, passing them down through the generations.

Gender disparities in entrepreneurship emerge particularly within specific sectors. Industries such as machinery production or architecture, traditionally regarded as male-dominated domains, exhibit more pronounced gender differences. Older entrepreneurs, usually in their 50s and 60s, often say that these ideas about gender are old-fashioned. In today's world, people's personal qualities are important in showing that these ideas are wrong and making things fairer. This means women are seen as equals to men in business.

When it comes to managing time between work and family, we see different trends. The research shows that businesses without many employees find it harder to balance work and family. But when there are two or more women leading the business, they can manage work and family life better. This shows how sharing experiences and support can help overcome these challenges.

As a result, a big difference appears between people who have help from their family and friends and those who don't have this advantage.

Furthermore, the presence of public institutions that offer childcare support, such as kindergartens, assumes principal importance. However, these resources are not uniformly accessible and distributed. This makes it hard for businesses to grow and be known, especially on a big scale. The substantial time commitment required to manage both professional and familiar responsibilities can impose constraints that are often more oppressive for women.

Another notable dimension to consider is the correlation between the prosperity of startups and their investment in social capital. Enterprises that invest more in social networks and relationships tend to demonstrate resilience even in case of economic changes, maintaining a stable revenue stream.

Friendships and connections are really important for sharing information and knowledge, which helps businesses grow.

Other than how much time they spend, there are other reasons why businesses led by women may not do as well. High price competition emerges as a challenge in the business landscape. The necessity for increased investments in collaborations, which often demand a more substantial level of engagement from entrepreneurs, can be a significant hurdle. This high level of investment, in terms of both time and resources, can cause pressure on women entrepreneurs.

In summation, it becomes evident that the challenges and complexities encountered by women-led enterprises in Piedmont is like those experienced by their counterparts in other regions. The many different challenges women entrepreneurs face, like managing time, getting family support, dealing with competition, and investing money, show how complicated the world of female entrepreneurship is.

Empirical methods

In the following chapter, a brief explanation of the statistic method used to verify if there is a correlation between the survival rate of the start-ups guided by women and the number of women migrated to Italy. All the definitions that will be given in the following chapters are taken textually from the book "Introduction to Econometrics" by Stock-Watson.

Linear Regression with a Single Regressor

The method utilized to verify if the correlation between the two variables explained before (the duration of life of a female start-up and the number of foreign women resident in Piedmont) is the linear regression. It represents a method of estimation of the linear relationship between two variables. The general equation of linear regression is:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

This is a linear regression model with a single regressor. Y is the dependent variable and X is the independent variable, also called regressor. The right side of the equation is the population regression function or population regression line. β_0 is the intercept and β_1 is the slope of the line; they are the coefficients of the population regression line and they are called parameters of the population regression line.

β_1 (the slope) is the change in Y due to a unit change in X, while β_0 is the value of Y when the independent variable is equal to zero.

The term u_i is the error term and represents all the other variables that can affect the dependent variable but that are not considered in this model.

Estimate coefficients

To estimate β_0 and β_1 we use the Ordinary Least Squares (OLS) estimator, which estimates the value in the closest possible way to the observed data. It minimizes the average squared difference between the actual values of Y and the predicted based on the estimated line, the so-called predicted value.

The OLS estimators of the slope β_1 and the intercept β_0 are:

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})\text{^}2} = \frac{s_{XY}}{s_x^2}$$

$$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X}$$

Whereas the OLS predicted values and residuals are respectively:

$$\widehat{Y_i} = \widehat{\beta_0} + \widehat{\beta_1}X_i, i = 1, \ldots . n$$

$$\widehat{u_i} = Y_i - \widehat{Y_i}, i = 1, \ldots . n$$

There are situations in which the value of the intercept is different from zero, but it can be meaningful. In this case the intercept has only a geometrical interpretation.
We use the OLS estimator because it is the most used by economists and statistics.

Measure of Fit

After having obtained these values, it is necessary to check the goodness of the fit, it means how well the regression line describes the data. To measure it, the regression R^2 and the standard error of the regression are used.

The R^2

The regression R^2 is the fraction of the sample variance of Y_i explained by X_i. It is unitless and it ranges between 0 (that means that there is not fit) and 1 (that means that there is a perfect fit).
While the standard error of the regression (SER) is an estimator of the standard deviation of the regression error u_i, it measures the dispersion of its distribution.
We can rewrite the value of the dependent variable as a sum of the predicted value $\widehat{Y_i}$ and the residual $\widehat{u_i}$:

$$Y_i = \widehat{Y_i} + \widehat{u_i}$$

Mathematically the R^2 can be written as the ratio of the explained sum of the squares (ESS) to the total sum of squares (TSS):

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\widehat{Y_i} - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

The ESS is the sum of squared standard deviations of the predicted values of Y_i from their average, while the TSS is the sum of squared deviation of Y_i from its average.
Another way to express the R^2 is to write it in terms of the fraction of the variance of Y_i not explained by X_i. The sum of squared residuals, or SRR, is the sum of the squared OLS residuals:

$$SRR = \sum_{i=1}^{n} \hat{u}_i^2$$

So, the R^2 can be written as:

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

The R^2 of the regression of Y on the single regressor X is the square of the correlation between Y and X.
If R^2 is near 1 means that the regressor is good at predicting Yi, while an R^2 near 0 indicates that the regressor is not a good at predicting Y_i.
When the X doesn't explain the variance of Y, the coefficient of determination is equal to zero (ESS=0). Otherwise, when X explains all the variance of Y, the coefficient of determination is equal to one (ESS=TSS).

The Standard Error of the Regression (SER)

The SER is a measure of the spread of the observations around the regression line, measured in the units of the dependent variable.
Then, the formula of the standard error of regression is:

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

The division by $n-2$ to correct for a slight downward bias is introduced because the regression coefficients are estimated, so "two degrees of freedom" of the data are lost.
Another way to check for the goodness of the fit is that of using the root mean squared error (RMSE), which differs from SER only for the divisor n instead of $n-2$.

There are cases in which a regressor is binary, which means that the only value it can have is equal to zero or equal to one. This kind of regressor is usually called “dummy”. When it is present, the value of $\beta_1$doesn’t represent the slope anymore, because if the dummy is equal to zero, the expected value of Yi is β_0, while if is equal to 1, the expected value is equal to $\beta_0 + \beta_1$.

Linear Regression with Multiple Regressors

In real cases, there are more than one factor that can have an influence on the dependent variables. For this reason, in this analysis, the linear regression with multiple regressors is used and, in this model, in addition to the independent variable, there are also the control variables.

In the following paragraphs, the differences between the previous model are exposed.

The omitted variables bias

Omitted variables bias, also known as omitted variable bias or confounding, is a common issue in regression analysis. It occurs when a regression model does not include one or more important independent variables (predictors) that are correlated with both the dependent variable and the included independent variables. As a result, the effect of the omitted variables on the dependent variable is incorrectly attributed to the included variables, leading to biased and potentially inaccurate coefficient estimates.

The multiple regression Model and the control variables

To avoid the problem of variable bias, in addition to the linear regression with one regressor, we will use the linear regression with multiple regressor. This model permits estimating the effect on Y_i of changing variable while holding the other regressors constant. It is useful to do residual analysis, which means to analyze if there are other variables that can influence the value of the dependent variable Y.

The general expression is:

$$Y_{i=}\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, with\ i = 1, \ldots . n$$

Where:

- Y_i is the dependent variable;
- X_{1i} and X_{2i} are the two regressors;

- β_0 is the unknown population intercept;
- β_1 is the effect on Y of a unit change in X_{1i}, controlling for X_{2i} (holding constant X_{2i});
- β_2 is the effect on Y of a unit change in X_{2i}, controlling for X_{1i};
- u_i is the regression error (as before, it represents the omitted factors that can influence the value of Y). This value is homoscedastic if the variance of the conditional distribution of u_i is constant for each i and thus doesn't depend on the values of the different X.

The variable X_{0i} (the one of β_0) is called constant regressor because its value is always equal to one and the intercept β_0 is called constant term in the regression.

OLS estimators

As in the case of linear regression with a single regressor, also in this model it is necessary to estimate the value of the coefficients that are unknown (β_0, β_1,…, β_k). Also, in this case the OLS estimators are used and the population regression line that predict the value of Y_i is

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

With a residual value equal to:

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

So, the population multiple regression model with OLS estimators is:

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

Measure of Fit in Multiple Regression

In multiple regression there are two statistics measures of fit.

The SER

The standard error of the regression SER, which estimate the standard deviation of the error term:

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

The divisor $n-k-1$ adjusts for the downward bias introduced by estimating $k+1$ coefficients (the k slope coefficients plus the intercept) and this is the only difference that there is between SER for multiple regression and the one of single regression. In fact, in formula x the divisor $n-2$ adjusts for the downward bias introduced by estimating two coefficients (the slope and interception of the regression line).

The regressor R^2

The regressor R^2, as explained before, is the fraction of the sample variance of Y_i explained by the regressors. So, the R^2 is 1 minus the fraction of the variance of Y_i not explained by the regressors.

The mathematical definition is the same one of the single regressor:

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

In this multiple regression model, the regressor increases whenever a regressor is added, unless the estimated coefficient is equal to zero. For this reason, it isn't right to use this regressor to estimate the fit, because not necessarily the addition of a regressor means that the fit is better.

Adjusted R^2

To better measure the fit, the adjusted regressor R^2 is used. It is needed to deflate or reduce the regressor R^2 by some factor and this is what the adjusted R does. The formula is:

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

In this formula the ratio of the sum of squared residuals to the total sum of squares is multiplied by the factor $(n-1)/(n-k-1)$. This factor is always greater than 1, so the adjusted R^2 is always less than R^2. Adding a regressor has two opposite effects on the

adjusted R^2. The SSR falls while the factor increases. Whether the R^2 adjusted increases or decreases depends in which of these two effects is stronger. In the end, it is interesting the fact that the adjusted R^2 can be negative.

Non linear Functions of a single independent variable

In this study it is used one of the three logarithmic regression model: the lin-log model. In this case, the variable X is in logarithm while the dependent variable Y is not.
There can be three different combinations using the log:

- X is in logarithm and Y is not;
- Y is in logarithm and X is not;
- Both X and Y are logarithms.

In our model a 1% change in X is associated with a change in Y of $0.01\beta_1$.
The regression model is:

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

It is easier to analyze the results in percentage. Also in this case, to estimate the fit of the results, it is used the adjusted R^2 (as explained before).

Empirical Applications

Data Construction

Data collection for a linear regression analysis is a critical step in the statistical modeling process.

The principal data that are necessary for this study are: number of all the women immigrated in the Piedmont region, in Italy, and the number of innovative female-lead startups, including the ones that survive and became enterprises over the years.

The timespan that we considered for our research is from 2009 to 2019. It is decided to consider only this interval of time for two reasons: the 2009 is the first year in which the number of observations were significant for the scope of the analysis, while the 2019 is chosen to mitigate the *sensoring problem*, that is avoid considering the start-up born in a recent year as dead while they are still alive.

We collect data about startups and enterprises from AIDA in these years.

While, for what concern immigration data, we collect data from ISTAT and we decide to collect the foreign women immigrated in Italy from 25 to 65 years old. We decided to consider this range of age because it is the period when a person usually devotes his time to work.

We collect these data because the goal of our study is to see if the number of low-skilled women immigrated in Piedmont can influence the survival rate of the Italian startup led by native women.

After collecting them separately, we create a unique database to connect the number of startups and the number of the women immigrants in the same municipality and in the same year.

In the following paragraphs these databases are better explained.

Database startup and AIDA

The first database is made up of 28 variables on columns and 8247 observations on rows and it is in the panel format.

It is the sum of two databases: the one is the startup register and the second one is the AIDA database, a database in which it is possible to find all the enterprises registered. It was fundamental to take both because a startup, after 5 years, cannot be considered as a startup anymore and for this reason is delated by the startup register and is moved to the

AIDA database. So, for the startups that survive for more than 5 years it was necessary to take data from both databases.

It shows for each year, the characteristics of the startup/enterprise in a municipality of the Piedmont region in Italy.

The variables present in this database are:

- "*id* " to identify univocally each startup/enterprise;
- *"anno_osservazione"* which indicates the year of observation for the data of that startup;
- *"anno_costituzione"* which is the year in which the start-up is subscribe to the register of startups;
- *"comune"* which is the municipality in which the startup operates;
- *"codice_istat_comune"* is the referral code to find the city in the ISTAT database. Each municipality in Italy has its own code to be identified;
- *"provincia"* in which we find the province of the city where the startup/enterprise is sited;
- *"codice_istat_provincia"*, as for each municipality, we find the ISTAT code for each province of the Piedmont, so this column refers to the column "*provincia*";
- *"regione"* indicates the region in which the startup/enterprise is sited, in this case there is only Piedmont because in this study only the Piedmont region is analyzed;
- *"stato_giuridico"* can assume different values: "cessata (in liquidazione)" means that the enterprise is closed and it is in liquidation; "in liquidazione", the enterprise is in a situation of liquidation; "attiva" means that the company is currently active in the market;
- *"data_cessazione"* contains the date in which the enterprise failed and closed;
- *"ateco_2007"* contains the code that indicates the sector in which the startup/enterprise operates. It is a code given by ISTAT to facilitate the statistics studies about economics activities;
- *"tot_immobili_immateriali"* indicates the value in thousands of intangible properties owned by each startup/enterprise;
- *"tot_immobili_materiali"* indicates the value in thousands of tangible properties owned by each startup/enterprise;
- *"tot_attivo"* indicates the value in thousands of total assets owned by the startup/enterprise;
- *"tot_patrimonio_netto"* indicates the value in thousands of total owners' equity;

- *"tot_debiti"* indicates the value in thousands of total debts owned by the startup/enterprise;
- "*capitale sociale"* indicates the value in thousands of the authorized capital owned by the startup/enterprise. Authorized capital refers to the maximum amount of capital that a company is legally allowed to issue in the form of shares;
- *"ricavi"* indicates the value in thousands of revenues collected in that year by the startup/enterprise. Revenues represent the total income generated by a business through its primary activities;
- *"risultato_operativo"* indicates the value in thousands of operating income of the startup/enterprise. Operating income represents the profit a company generates from its core business operations. It is calculated by subtracting the operating expenses of a business from its gross profit. Operating income excludes interest and taxes, focusing solely on the costs directly associated with producing and selling goods or services;
- *"valore_aggiunto"* indicates the value in thousands of value added owned by the startup/enterprise. Value added refers to the increase in the value of a product or service at each stage of the production or distribution process. It is calculated by taking the difference between the total value of a firm's output and the total value of intermediate goods used in the production process;
- *"utile"* indicates the value in thousands of income of the startup/enterprise. Income refers to the money earned;
- *"EBITDA"* indicates the value in thousands of EBITDA of the startup/enterprise. EBITDA stands for Earnings Before Interest, Taxes, Depreciation, and Amortization. It is calculated by starting with a company's operating and adding back depreciation and amortization expenses. It is commonly used to assess a company's profitability and operational efficiency;
- *"dipendenti"* indicates the number of employees that work in each startup/enterprise,
- *"diritti_brevetto_industriale"* indicates the number of intellectual properties owned by the startup/enterprise;
- *"data_iscrizione_sezione_startup"* indicates the date in which the startup is registered in the startup register;
- *"data_uscita_sezione_startup"* indicates the date in which the startup exit from the startup database;

- *"prevalenza_femminile_complessiva"* can assume 4 values:
    - NO: if (% authorized shares+ % managers)/2 <= 50%;
    - Maggioritaria: if (% authorized shares+ % managers)/2 > 50%;
    - Forte: if (% authorized shares+ % managers)/2 > 66%
    - Esclusiva: if (% authorized shares+ % managers)/2 =100%.
        - It indicates if we are talking about a female enterprises or not;
- *"prevalenza_femminile_amm"* can assume 4 different values:
    - NO: if (%administrators) <= 50 %;
    - Maggioritaria: if (%administrators) >50%;
    - Forte: if (%administrators) > 66%;
    - Esclusiva: if (%administrators) =100%.
        - It indicates if the enterprise is a female- led business or not, because it considers the number of women inside the administration.

As said at the beginning, this database is in a panel format, so in each row it is possible to find the id and the year considered. For this reason, the same id can be present several times, indicating the same company in different years of observation.

Database ISTAT

This database is the one useful to see how many foreign women live in each municipality in Piedmont from 2001 to 2019. The database is in the panel format (as the previous described before) and the data concerns women with age that varies from 25 to 65 years old.

The database is composed of 3 variables on columns and 22439 observations on rows.

The variables are:

- "*Comune"* which indicate the municipality taken into consideration;
- "*anno_osservazione"* which indicates the year that is considering;
- "*imm_"* indicates the total number of foreign women with an age between 25-65 years old resident in that municipality during that year;
- *"pop_*" indicates the total number of people resident in that municipality during the year considered.

This database wasn't ready, but it was necessary to create it. In fact, on ISTAT site are available data of immigrant resident in each Italian municipality for each range of age.

So, after having filtered the gender equal to female, the database for each age was downloaded and at the end a new database was created by summing all the number of women resident in the municipality considered with the age 25-65 years old. In this way, in this database it is possible to find the entire female foreign population resident in that municipality having an age between 25 and 65 years old.
As this database is in the panel format, it is possible to find the same municipality in different rows due to the change of year considered.

Final database

To make the linear regression was necessary to have a unique database with all data useful for the regression. For this reason, using STATA, the two databases have been merged.
Stata is a statistical software package used for data management, statistical analysis, and graphics. It provides a suite of applications for data manipulation and statistical analysis that is used in various fields, including economics, sociology, political science, epidemiology, and other disciplines that require advanced statistical techniques.
Through a merge using the variables "*comune*" and "*year*", the final database is created. It is composed by 37 variables on the columns (without dummies) and 178 observations on the rows.
The variables in the final database are:

- "*id*" to identify univocally each startup/enterprise;
- *"anno_osservazione"* which indicates the year of observation for the data of that startup;
- *"anno_costituzione"* which is the year in which the stratup is subscribe to the register of startups;
- *"comune"* which is the municipality in which the startup operates;
- *"codice_istat_comune"* is the referral code to find the city in the ISTAT database; in fact, each municipality in Italy has its own code to be identified;
- *"provincia"* in which we find the province of the city where the startup/enterprise is sited;
- *"codice_istat_provincia"*, as for each municipality, we find the ISTAT code for each province of the Piedmont, so this column refers to the column "*provincia*";
- *"regione"* indicates the region in which the startup/enterprise is sited, in this case there is only Piedmont because in this study only the Piedmont region is analyzed;

- *"stato_giuridico"* can assume different values: "cessata (in liquidazione)" means that the enterprise is closed and it is in liquidation; "in liquidazione", the enterprise is in a situation of liquidation; "attiva" means that the company is currently active in the market;
- *"data_cessazione"* contains the date in which the enterprise failed and closed;
- *"ateco_2007"* contains the code that indicates the sector in which the startup/enterprise operates. It is a code given by ISTAT to facilitate the statistics studies about economics activities;
- *"tot_immobili_immateriali"* indicates the value in thousands of intangible properties owned by each startup/enterprise;
- *"tot_immobili_materiali"* indicates the value in thousands of tangible properties owned by each startup/enterprise;
- *"tot_attivo"* indicates the value in thousands of total assets owned by the startup/enterprise;
- *"tot_patrimonio_netto"* indicates the value in thousands of total owners' equity;
- *"tot_debiti"* indicates the value in thousands of total debts owned by the startup/enterprise;
- "*capitale sociale"* indicates the value in thousands of the authorized capital owned by the startup/enterprise. Authorized capital refers to the maximum amount of capital that a company is legally allowed to issue in the form of shares;
- *"ricavi"* indicates the value in thousands of revenues collected in that year by the startup/enterprise. Revenues represent the total income generated by a business through its primary activities;
- *"risultato_operativo"* indicates the value in thousands of operating income of the startup/enterprise. Operating income represents the profit a company generates from its core business operations. It is calculated by subtracting the operating expenses of a business from its gross profit. Operating income excludes interest and taxes, focusing solely on the costs directly associated with producing and selling goods or services;
- *"valore_aggiunto"* indicates the value in thousands of value added owned by the startup/enterprise. Value added refers to the increase in the value of a product or service at each stage of the production or distribution process. It is calculated by taking the difference between the total value of a firm's output and the total value of intermediate goods used in the production process;

- *"utile"* indicates the value in thousands of income of the startup/enterprise. Income refers to the money earned;
- *"EBITDA"* indicates the value in thousands of EBITDA of the startup/enterprise. EBITDA stands for Earnings Before Interest, Taxes, Depreciation, and Amortization. It is calculated by starting with a company's operating and adding back depreciation and amortization expenses. It is commonly used to assess a company's profitability and operational efficiency;
- *"dipendenti"* indicates the number of employees that work in each startup/enterprise,
- *"diritti_brevetto_industriale"* indicates the number of intellectual properties owned by the startup/enterprise;
- *"data_iscrizione_sezione_startup"* indicates the date in which the startup is registered in the startup register;
- *"data_uscita_sezione_startup"* indicates the date in which the startup exit from the startup database;
- *"prevalenza_femminile_complessiva"* can assume 4 values:
    - NO: if (% authorized shares+ % managers)/2 <= 50%;
    - Maggioritaria: if (% authorized shares+ % managers)/2 > 50%;
    - Forte: if (% authorized shares+ % managers)/2 > 66%
    - Esclusiva: if (% authorized shares+ % managers)/2 =100%.
        - It indicates if we are talking about a female enterprises or not.
- *"prevalenza_femminile_amm"* can assume 4 different values:
    - NO: if (%administrators) <= 50 %;
    - Maggioritaria: if (%administrators) >50%;
    - Forte: if (%administrators) > 66%;
    - Esclusiva: if (%administrators) =100%.
        - It indicates if the enterprise is a female- led business or not, because it consider the number of women inside the administration.
- *"data fine"* indicates the date in which the start-up/enterprise stops to exist: it is calculated as equal to "*data_cessazione*" if this variable is different from zero, otherwise is equal to "*data_uscita_sezione_startup*". If also this variable is not available for the start-up considered, it is decided to consider as "*data_fine*" the last year in which "*ricavi*" data are available +1.

    - This variable can assume 2021 as maximum value, for this reason if the "*data_fine"* is higher than 2021, it was substitute with 2021. This is done because we don't know if a start-up born in recent years survive or not more than 5 years;
- *"duration"* is equal to "*data_fine"* less *"anno_costituzione"* and it indicates the number of years in which the start-up exists. The maximum value that this variable can assume is 5 years, for this reason even if duration is higher than 5, it is substituted with 5. It was decided 5 because a start-up after 5 years is not a start-up anymore and, for this reason, if the duration>5 it means that this start-up became an enterprise;
- *"imm_corrente"* is a renaming of "*imm*", indicates the total number of immigrant women with an age between 25-65 years old resident in that municipality during the year considered;
- *"pop_corrente"* is a renaming of "*pop*", indicates the total number of people resident in that municipality during the year considered;
- *"imm_iniziale"* indicates the total number of foreign women with an age between 25-65 years old resident in that municipality during the year considered;
- *"pop_iniziale"* indicates the total number of people resident in that municipality during the year in which the start-up is created;
- *"imm_rate_iniziale"* is a rate equal to "*imm_iniziale*"/"*pop_iniziale*" and indicates how many women immigrates are present in that municipality in proportion to the total population;
- *"ateco_2007_3d"* is equal to the variable "*ateco_2007"*/1000 because only the firsts three numbers are significant to indicate the sector;
- *"sect1" to "sect59"* these are variables dummy created for the regression. They indicates the sectors in which the start-up/enterprise operates: if the variable is equal to '1' it means that the sector in which that enterprise operates is that one, otherwise the value is equal to '0';
- *"ricavi_iniziali*" is equal to the variable "*ricavi*" in the year of "*anno_costituzione"*;
- "*tot_debiti_iniziali*" is equal to the variable "*debiti*" in the year of "*anno_costituzione"*;
- "*dipendenti_iniziali*" it refers to the variable "*dipendenti"* in the year of "*anno_costituzione"*;

- "*ln_imm_iniziale"* is the natural logarithm transformation of "*imm_iniziale";*
- *"ln_pop_iniziale"* is the natural logarithm transformation of "*pop_iniziale";*
- *"ln_ricavi_iniziale"* is the natural logarithm transformation of "*ricavi_iniziali";*
- *"ln_tot_debiti_iniziali"* is the natural logarithm transformation of "*tot_debiti_iniziali" ;*
- *"ln_dipendenti_iniziali"* is the natural logarithm transformation of "*dipendenti_iniziali";*

These last 5 variables are equal to the variable with the same name but transformed into a log scale, this is due to the fact that our regression is in a linear-log form and for this reason the independent variables are on a natural logarithm scale.

Later, to create the different variants of the linear regression model, it was necessary to add to the databases new variables, merging the database of Local Labor System to the one created before. This database was downloaded by the ISTAT site, because it is a subdivision of the region created by ISTAT to consider the geographical area in which a person can work, near the municipality of residence.

The new columns present in the new database are:

- "*sll*" indicates the Local Labor System to which the municipality considered belongs;
- "*immi_sll*" is the sum of the *imm_iniziale* of the Local Labor System considered. It is the total number of immigrant women resident in the municipalities under the Local Labor System of the start-up during the year in which the start-up is created;
- "*immi_prov"* is the sum of the *imm_iniziale* of the municipalities under the same province in which the start-up is sited during the year in which it is created.

Also in this case, the logarithmic form is created for the variables used as independent variables:

- "*ln_immi_sll"* is the natural logarithm transformation of "*immi_sll";*
- *"ln_immi_prov"* is the natural logarithm transformation of "*immi_prov".*

Hypothesis formulation and model construction

As it is emerged from past literature, women are the ones that more take care of family and house, reducing their time availability to work and have success in their career. A way to help these women to have more success in their work, can be given by the low-skilled

women migrated to Italy. In fact, as it is emerged from statistical studies (as it is possible to see in the past literature above), most of these women are employed in the sector of personal and service care. In this way, they can help native women to reduce their time dedicated in family and personal work, increasing their possibility to grow in their career. For this reason, in addition to the variables related to the survival of a start-up, it is interesting to also consider a variable which indicates the number of low-skilled foreign women resident in the same municipality in which the start-up is founded.

In this study, the female-led startups sited in the municipalities of Piedmont region are considered to see if effectively there is a relationship between the stock of immigration in this area and the survival rate of these start-ups, over a period of more than 10 years.

As explained before, the years considered are from 2001 to 2019.

This analysis is made by creating different linear regressions with different independent variables but maintaining the same dependent variable.

The principal hypothesis of this analysis is:

***Hp1*:**" The survival rate of female-led innovative start-up sited in Piedmont region increases with the increase of the immigrants' stock resident in Piedmont' municipalities, measured in terms of start-up duration".

The first linear regression is a single linear regression:

$$duration_i = \beta_0 + \beta_1 ImmStock_{jt} + \varepsilon_{ijt}$$

In this linear regression, the dependent variable $duration_i$, which indicates the duration in year of the start-up *i,* depends only on the independent variable $ImmStock_{jt}$, which represents the immigrants' stock resident in the city *j* at year *t.*

To evaluate if the hypothesis is consistent or not, also a multiple linear regression model is created:

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

The variable utilized are:

- $duration_i$ is the dependent variable indicating duration in year of the start-up *i*;
- $ImmStock_{jt}$ is the independent variable indicating the number of immigrants' stock resident in the municipality j at the year t;

- X'_{ijt} is the vector containing all the control variables (better explained below) related to the start-up *i,* at time *t*, in the municipality *j;*
- β_1 is the coefficient of the independent variable;
- β_2 is the coefficient of the vector of the control variables;
- ε_{ijt} is the residual error referred to the start-up *I,* in the municipality *j,* in the year *t.*

Subsequently, another multiple linear regression is created to test the same hypothesis but with a different independent variable. In fact, in this case, the independent variable is the so called "immigration rate":

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

The independent variable Imm_Rate_{jt} is created by dividing the number of immigrants resident in the municipality *j* in the year *t* by the total population resident in the same municipality in the same year.

In the end, another aspect to test the hypothesis is seen.
In the previous linear regressions, was taken for granted the fact that the service provided by women was in the same municipality in which the start-up is sited, but this can cause a lack of accuracy of the models. In fact, people do not always work in the same municipality in which they live, because they can move to another city, of course near the one in which they live.
For this reason, in the last linear regression, the geographical area considered is larger.
To do so, it has been used the so called "Local Labor System". The "Local Labor System" are an Italian geographical division made by ISTAT, the National Institution of Statistic, an Italian public institution of research. The definition given by ISTAT is the following: "Local Labor Systems" (SLL) represent a spatial grid whose boundaries, regardless of the administrative structure of the territory, are defined using daily home/work (commuting) flows recorded during the General Census of population and housing".
So, in this case the area selected by the variable also consider women who move from their municipality for work.

To add this variable, it was necessary to make another merge with the database; in this way, each municipality is assigned to one SLL and the new analysis can be made with a new hypothesis:

***Hp2*:** "The survival rate of female-led innovative start-up sited in Piedmont region increases with the increase of the immigrants' stock resident in Piedmont' Local Labor System, measured in terms of start-up duration".

The linear regression model, in this case, is:

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

The new independent variable is the number of immigrants in the Local Labor System *j* during the year *t.*

To evaluate the differences between the Local Labor System and the Piedmont' provinces, in the last variant of regression the independent variable represented the female immigrant resident in the same province of the start-up. In this way, a bigger geographical area respect to the Local Labor System division is considered.

The initial hypothesis is close to the previous one, but in this case is the following:

***Hp3*:** "The survival rate of female-led innovative start-up sited in Piedmont region increases with the increase of the immigrant' stock resident in Piedmont' provinces, measured in terms of start-up duration".

Consequently, the linear regression model in this case is:

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

The independent variable is the total number of immigrants women resident in the selected year in the same province in which the start-up is sited.

Variables descriptions

In this section will be explained the variables utilized in the different linear regressions of the analysis.

First, the dependent variable is described and it is explained why this was the chosen one. Then, the dependent and the control variables are presented and described.

The dependent variable: duration

As the goal of this study is to see if the survival of start-up is influenced by other factors, the dependent variable chosen for this regression is the duration. As explained above, the duration indicates the year in which the start-up survives. The variable duration is created subtracting the year of born from the year of "*data_cessazione*" (data in which the start-up closed) found in the database AIDA. As not all the start-up survived for more than 5 years and became enterprises, it is not possible to have the value "*data_cessazione*" for everyone, so it is made a proxy using as end year the last year in which is available the economic account plus one. The maximum value that this variable can assume is 5 years, because it is the maximum number of years in which a start-up can exist before becoming an enterprise and be moved from the start-up database to the enterprise database.

Through the different regressions it is possible to see if this variable really depends on the other selected or not.

Independent variables

$ImmStock_{jt}$

The objective of this study is to investigate whether the migration of low-skilled women from another country has an impact on the survival rate of innovative start-ups created by native women. To explore this relationship, we have selected the dependent variable, "*ln_imm_iniziale,"* representing the natural logarithm of the count of foreign women resident in the same municipality where the start-up originates. Logarithmic transformation was employed to facilitate the interpretation of results in terms of percentage changes.

This variable considers the number of foreign women with an age between 25-65 years old resident in the Piedmont' municipalities in the years between 2001 and 2021.

Anticipating insights from Patricia Cortes' research, our hypothesis posits a positive correlation between the chosen independent variable (number of low-skilled migrant women) and the dependent variable (survival rate of native women's start-ups).

Imm_Rate_{jt}

In the second linear regression the independent variable is the Immigration rate. As explained above, this rate is constructed dividing the number of women immigrants (*imm_iniziale*) in the municipality in which there is the headquarter of the start-up in the year considered by the number of the total population resident in the same municipality (*pop iniziale*) (including Italian e non-Italian people). This variable is used to effectively see if the immigrants move to the cities where there are many people from the same country as them.

Imm_LLS_{jt}

In the third linear regression, instead, the independent variable is the number of immigrants women resident in the same Local Labor System in which there is the headquarter of the start-up considered.

As explained above, to create this variable it was necessary to take from ISTAT the code of the Local Labor System in Piedmont and assign each municipality to the correspondent one.

Using this variable, it is possible to also consider the women that move from their city municipality to another one to work.

Imm_Prov_{jt}

In the last linear regression, the independent variable is the total number of immigrants women resident in the same province in which is sited the start-up analyzed.

In the same database created by commune, there were also the different provinces and through the statistic software STATA it was possible to calculate the total number of immigrants resident in the different municipalities under the same province. In this way, increasing the geographical area considered, more cases of women who don’t work in the same municipality in which they live are included.

Control variables

As explained in chapter 3, the control variables are the variables that can influence the dependent variable but that are not included in the dependent variable. For this reason is important to insert also this one in the regression models to reduce the possibility of omitted variable bias and to decrease the value of errors, which is higher when there is a higher number of omitted variables.

As control variables it is decided to use *"ln_pop_iniziale", "sect1-sect59", "ln_ricavi_iniziali", ln_tot_debiti_inizali"* and *"ln_dipendenti_iniziali".*

These variables are considered the ones that can influence the dependent variables, seeing the background literature.

Pop_iniziale

This variable represent the total population resident in the same municipality in which the start-up is sited. This variable is useful to evaluate if the duration of the start-up depends on the number of the population, because the higher the population variable, the higher is the demand of the product.

Moreover, it is also useful to evaluate if effectively the number of immigrants depends on the number of the population already resident in that geographical area.

sect1-sect59

This is a variable dummy that indicates the industrial sector in which the start-up considered operates. The dummy variable assume the value 1 if the start-up operates in the sector considered and 0 otherwise.

The dependent variable duration can be influenced by the sector: some sectors have more probability to survive due to a higher demand in the market, while in others sectors is more complicated to have a success.

ln_ricavi_iniziali

Another control variable is the turnover of the start-up declared in the first year of life of the start-up. It is expected that the ones with a higher turnover have also a higher possibility to survive over time.

ln_tot_debiti_inizali

The variable "*tot_debiti_iniziali"* shows the level of debts of the start-up during the first year of life. This variable is important to see if the level of initial debt influence in a positive or negative way the dependent variable.

In the background literature it is emerged the fact that the survival rate of a start-up depends also by the type of investors that decide to invest in it. It is not possible to use this variable because in the databases there is not the information about the type of investor (if Business Angels or Venture Capitals).

ln_dipendenti_iniziali

The last control variable considered is "*dipendenti_iniziali"*, which represents the number of employees of the start-up during the first year of life.

From background literature it is emerged that, most of the time the start-up that are founded by more than one persons with different backgrounds are the ones that have higher possibility to survive. Despite this, there can be also the possibility of having a conflict of interest between the different team members. So, through the use of this variable, it is possible to see if effectively the duration of the start-up depends on the number of employees or not.

Descriptive statistics

In this chapter, the characteristics of the databases utilized for the analysis are analyzed more specifically. Each column of the database is detailed presented.

anno_costituzione

First, in the table below are shown the years considered as beginning year in the database. As it is possible to observe, the year in which more start-up are born is 2016, while the year with less start-up is 2009.

anno_costituzione	Freq.	Percent	Cum.
2009	3	1.69	1.69
2010	5	2.81	4.49
2011	7	3.93	8.43
2012	10	5.62	14.04
2013	13	7.30	21.35
2014	19	10.67	32.02
2015	18	10.11	42.13
2016	29	16.29	58.43
2017	27	15.17	73.60
2018	18	10.11	83.71
2019	29	16.29	100.00
Total	178	100.00	

Figure 10 anno_costituzione

provincia

Another data that are possible to observe are the Piedmont provinces to see which is the city in which the start-up are sited.

provincia	Freq.	Percent	Cum.
Alessandria	7	3.93	3.93
Asti	4	2.25	6.18
Biella	3	1.69	7.87
Cuneo	22	12.36	20.22
Novara	15	8.43	28.65
Torino	124	69.66	98.31
Vercelli	3	1.69	100.00
Total	178	100.00	

Figure 11 provincia

The Piedmont provinces are 8: Torino, Cuneo, Asti, Alessandria, Vercelli, Novara, Biella e Verbano-Cusio-Ossola. As it is possible to notice, the province Verbano-Cusio-Ossola is not present in the database as no start-up present in our database is born in this province. As it was expected, the province in which there are more start-ups is Turin, the region capital, followed by Cuneo and Novara. There is a big gap between the number of start-up between this provinces.

Comune

In thie column “comune” it is possible to observe the municipals in which the start-ups are rose.

comune	Freq.	Percent	Cum.
Alba	5	2.81	2.81
Alessandria	2	1.12	3.93
Arquata Scrivia	1	0.56	4.49
Asti	2	1.12	5.62
Biella	2	1.12	6.74
Bra	1	0.56	7.30
Bruino	1	0.56	7.87
Bussoleno	1	0.56	8.43
Carbonara Scrivia	1	0.56	8.99
Carmagnola	1	0.56	9.55
Casale Monferrato	2	1.12	10.67
Castel Rocchero	1	0.56	11.24
Cherasco	1	0.56	11.80
Colleretto Giacosa	1	0.56	12.36
Cossato	1	0.56	12.92
Cuneo	6	3.37	16.29
Dronero	1	0.56	16.85
Galliate	1	0.56	17.42
Grignasco	1	0.56	17.98
Grugliasco	2	1.12	19.10
Lenta	1	0.56	19.66
Manta	1	0.56	20.22
Massino Visconti	1	0.56	20.79
Monasterolo di Savigliano	1	0.56	21.35
Mondovì	1	0.56	21.91
Novara	12	6.74	28.65
Orbassano	1	0.56	29.21
Pianezza	1	0.56	29.78
Pinerolo	1	0.56	30.34
Pino Torinese	1	0.56	30.90
Piovà Massaia	1	0.56	31.46
Prarolo	1	0.56	32.02
Saluzzo	1	0.56	32.58
San Mauro Torinese	1	0.56	33.15
Santo Stefano Belbo	1	0.56	33.71
Savigliano	2	1.12	34.83
Sommariva del Bosco	1	0.56	35.39
Torino	110	61.80	97.19
Tortona	1	0.56	97.75
Venaria Reale	2	1.12	98.88
Vercelli	1	0.56	99.44
Vinovo	1	0.56	100.00
Total	178	100.00	

Figure 12 comune

The total number of municipalties in Piedomnt region is 1180, while in our final database we have only 42 cities.

As for the provinces, also for municipals Turin is the first one for what regards the number of start-up developed in, followed by Novara, Cuneo and Alba.

Turin is the most industrialized city in Piedmont and it also has the University center in Piedmont, this is probably the reason why the start-up are established in this city.

Stato_giuridico

For the start-up considered, there are only two legal status:

stato_giuridico	Freq.	Percent	Cum.
Attiva	100	56.18	56.18
In liquidazione	78	43.82	100.00
Total	178	100.00	

Figure 13 stato_giuridico

The majority are active, but 43.82% of them are in a state of liquidation, so they don't have long life.

Data_fine

In figure 14 it is possible to see the year in which the start-ups stop to live.

data_fine	Freq.	Percent	Cum.
2015	1	0.56	0.56
2017	2	1.12	1.69
2018	7	3.93	5.62
2019	3	1.69	7.30
2020	1	0.56	7.87
2021	164	92.13	100.00
Total	178	100.00	

Figure 14 data_fine

2021 is the year in which the majority of start-up stop to exist due to our assumptions in which we consider 2021 as *data_fine* also for the enterprises that are still alive, because we don't know if them are still alive or not, in this way we have as maximum life year 5 years.

Diritti_brevetto_industriale

It is interesting to notice that the number of patents, that can influence the success of a start-up, is the following one:

diritti_brevetto_industriale	Freq.	Percent	Cum.
0	38	90.48	90.48
1	1	2.38	92.86
2	1	2.38	95.24
3	1	2.38	97.62
4	1	2.38	100.00
Total	42	100.00	

Figure 15 diritti_brevetto_industriale

This data is not available for all the start-up inside the database, but for the ones in which the data is present, the majority doesn't have any patent and only 3 start-ups have respectively 1, 2, 3 and 4 patents.

As data is available only for 42 startups, this variable is not used as control variable in the regression model, even if background literature sustained that it is an important factor that influence the duration of the start-up.

Prevalenza_femminile_complessiva

In the following table it is possible to see the number of start-up that have a total female presence in the start-up.

prevalenza_femminile_complessiva	Freq.	Percent	Cum.
ESCLUSIVA	16	8.99	8.99
FORTE	144	80.90	89.89
MAGGIORITARIA	12	6.74	96.63
NO	6	3.37	100.00
Total	178	100.00	

Figure 16 prevalenza_complessiva_femminile

The meaning of *ESCLUSIVA, FORTE, MAGGIORITARIA* and *NO* are:

- *Esclusiva* if the female presence is total, corresponding to 100%;
- *Forte* if the female presence correspond to a percentage higher than 60%;
- *Maggioritaria* if the percentage is higher than 50%;
- *No* if there is not a prevalence of women inside the start-up.

Only 16 start-ups have an exclusive female presence, but the majority have a "*forte"* presence (144 start-ups).

Imm_corrente

For what regards women immigration data, a summarize table is shown:

Variable	Obs	Mean	Std. Dev.	Min	Max
imm_corrente	178	29977.52	22402.05	5	49551

Figure 17 imm_corrente

This indicate that the medium number of women immigrated in each municipality in which there is a start-up is around 29977.52. The standard deviation is high because the number of foreign women depends a lot from one municipality to the other. The minimum number of this data is 5 (it means that in the municipality considerated there are only 5 women migrated from another country and stabilized there). Instead, the higher number founded is 49551 and it corresponds to the city of Turin.

It is interesting to notice that these data change if we look to the initial database for what regards immigration in all the region municipalities and not only the ones in which there are start-ups.

```
. summarize imm_
```

Variable	Obs	Mean	Std. Dev.	Min	Max
imm_	21,258	97.50875	1139.251	0	49551

*Figure 18 imm_*

In fact, in the observations through different years and different municiplaties, it is emerged that the medium is lower with respect to the previous one. This can be correlated to what we expect from this analysis: female start-ups are founded in the city in which more women migrants live.

Another different data is that there is at least a municipality in Piedmont in which no foreign woman is resident in.

These differences between the two database, as said before, is due to the fact that in the final database only 42 municipalites are considerated (the ones in which there is at least one start-up).

dipendenti

The number of employees that emerged is in Figure 19.

dipendenti	Freq.	Percent	Cum.
0	85	78.70	78.70
1	13	12.04	90.74
2	4	3.70	94.44
5	1	0.93	95.37
6	1	0.93	96.30
8	1	0.93	97.22
11	1	0.93	98.15
16	1	0.93	99.07
17	1	0.93	100.00
Total	108	100.00	

Figure 19 dipendenti

The majority have 0 employees because, at the beginning, the ones that work for the start-ups are the entrepreneurs itselfs. In fact only 6 start-ups through the others count more than 2 employees.

Ricavi

The reason just discussed can be valid also for what reguards revenues. In fact, as the tables below show, the majority have revenues equal to zero, because at the beginning it is difficult having a high level of revenues.

```
. summarize ricavi
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ricavi	112	86.30357	495.3456	0	4960

```
. tab ricavi
```

ricavi	Freq.	Percent	Cum.
0	53	47.32	47.32
1	3	2.68	50.00
2	7	6.25	56.25
3	2	1.79	58.04
4	2	1.79	59.82
5	2	1.79	61.61
6	2	1.79	63.39
7	1	0.89	64.29
9	1	0.89	65.18
12	1	0.89	66.07
14	1	0.89	66.96
15	2	1.79	68.75
16	3	2.68	71.43
17	1	0.89	72.32
18	2	1.79	74.11
21	1	0.89	75.00
22	1	0.89	75.89
23	1	0.89	76.79
27	1	0.89	77.68
29	1	0.89	78.57
31	1	0.89	79.46
32	1	0.89	80.36
35	2	1.79	82.14
36	1	0.89	83.04
40	1	0.89	83.93
43	1	0.89	84.82
52	1	0.89	85.71
58	1	0.89	86.61
59	1	0.89	87.50
60	1	0.89	88.39
68	1	0.89	89.29
71	1	0.89	90.18
90	1	0.89	91.07
95	1	0.89	91.96
97	1	0.89	92.86
102	1	0.89	93.75
108	1	0.89	94.64
132	1	0.89	95.54
365	1	0.89	96.43
551	1	0.89	97.32
593	1	0.89	98.21
1605	1	0.89	99.11
4960	1	0.89	100.00
Total	112	100.00	

Figure 20 ricavi (in thousands)

In fact only two start-ups exeed 1 million in revenues and the medium value is around 86 thousands.

Duration

In the figure 21 showed below it is possible to see the frequency for each duration.

duration	Freq.	Percent	Cum.
1	4	2.25	2.25
2	30	16.85	19.10
3	21	11.80	30.90
4	27	15.17	46.07
5	96	53.93	100.00
Total	178	100.00	

Figure 21 duration

As explained above, the dependent variable can assume value between 1 and 5 years, because we made a proxy and we consider that a start-up can live for a maximum of 5 years, to avoid the so called "*sensoring problem*". For this reason, the start-ups that survive for more than 5 years, are included in the last row of the table, and they are the ones with higher frequency (96 start-ups).

The second duration with more frequency is 2 years, with 30 start-ups followed by 4 years with 27 start-ups.

It is interesting to see also the summary of the variable duration (figure 22).

Variable	Obs	Mean	Std. Dev.	Min	Max
duration	178	4.016854	1.241808	1	5

Figure 22 duration summarized

The medium value of years of duration variable is 4.01. This means that in Piedmont, according to data collection, the female-led innovative start-ups have a medium life of 4.01 years. The standard deviation is 1.24, relatively low due to the proxy of a maximum of 5 years done.

Regression variables

Now results of both independent and control variables are shown.

As explained above, the independent variable is the number of women immigrated in the comune considerated in the year in which the start-up is founded (*imm_iniziale*), while the control variables are:

- The total number of people that live in the comune considerated (both italian and foreign) in the year in which the start-up is founded→ *pop_iniziale;*
- The dummy variable that indicates the sector in which the start-up operates → *sect1-sect59;*
- The value of revenues at the beginning year of the start-up → *ricavi_iniziali;*
- The value of debts at the beginning year of the start-up→ *tot_debiti;*
- The number of employees during the first year of live→ *dipendenti_iniziali.*

Imm_iniziale

```
. summarize imm_iniziale
```

Variable	Obs	Mean	Std. Dev.	Min	Max
imm_iniziale	178	29841.11	22346.67	5	49551

Figure 23 imm_iniziale summarized

In Figure 23 the independent variable imm_iniziale summarized shows that the medium number of immigrants women resident in each municipality in Piedmont during the first years of life of the start-up is equale to 29841. The higher standard deviation (equal to 22346.67) is due to the fact that there is a lot of difference in the number of women resident in some municipalities resoect to others.

Pop_iniziale

The control variable *pop_iniziale* summerized shows that the medium number of people resident in the each municipality has a mean of 559797 people. Also in this case, as the depoenden variable, the standard deviation has an high value because of the different number of people resident in different municipalities.

Variable	Obs	Mean	Std. Dev.	Min	Max
pop_iniziale	178	559797.1	409694.3	395	894314

Figure 24 pop_inziale summarized

Ricavi_iniziali

Variable	Obs	Mean	Std. Dev.	Min	Max
ricavi_ini~i	178	54.30337	394.4894	0	4960

Figure 25 ricavi_inziali summarized

In Figure 25 is shown the summary of the control variable *ricavi_iniziali* that has a mean value of 54.30k euro and a higher standard deviation due to the high difference in the turnover between different start-ups (as it is possible to see in the figure 20, most of the start-up has a turnover equal to 0 while others in thousands).

Tot_debiti_iniziali

The variable *tot_debiti_inziali* indicates the value of debts owned by the start-ups during their first year of life. The medium assumed by this variable, in the Piedmont case, is equal to 64.47k. The standard deviation is equal to 454.74k, the high value due to the high difference between the start-ups. In fact the minimum value of debts found in the database is equal to zero, while the maximum value is equal to 5.993 million.

Variable	Obs	Mean	Std. Dev.	Min	Max
tot_debiti~i	178	64.47753	454.7406	0	5993

Figure 26 tot_debiti_iniziali summarized

Dipendenti_iniziali

As already explained before, the variable *dipendenti_iniziali* indicates the number of employees of the start-up during the first year of life. The mean value of this control variable summarized is 0.47. In fact, the most of start-up has a value of employee equal to zero due to the fact that during the first year of life only entrepreneurs work inside the start-up.

Variable	Obs	Mean	Std. Dev.	Min	Max
dipendenti~i	178	.4719101	2.099717	0	17

Figure 27 dipendenti_iniziali summarized

In fact, if we observe the frequency of the number of employees during the first year (figure 28), only 23 start-ups have more than one employee.

dipendenti_ iniziali	Freq.	Percent	Cum.
0	155	87.08	87.08
1	13	7.30	94.38
2	4	2.25	96.63
5	1	0.56	97.19
6	1	0.56	97.75
8	1	0.56	98.31
11	1	0.56	98.88
16	1	0.56	99.44
17	1	0.56	100.00
Total	178	100.00	

Figure 28 dipendenti_iniziali

Results

Single linear regression

```
. regress duration ln_imm_iniziale
```

Source	SS	df	MS	Number of obs	=	178
				F(1, 176)	=	2.99
Model	4.55753923	1	4.55753923	Prob > F	=	0.0856
Residual	268.391899	176	1.52495397	R-squared	=	0.0167
				Adj R-squared	=	0.0111
Total	272.949438	177	1.54208722	Root MSE	=	1.2349

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_imm_iniziale	.0695482	.0402299	1.73	0.086	-.0098469	.1489433
_cons	3.380826	.3793724	8.91	0.000	2.632122	4.129531

Figure 29 Single linear regression results

In the figure 29 is shown the result of the lin-log regression using only the dependent and the independent variable. In this case, the dependent variable duration is in a linear form, while the variable *imm_iniziale* is transformed in a logarithmic scale. The figure shown the correlation between the year of life of a start-up and the number of the migrated women resident in the same municipality of the start-up headquarters in the year in which it is created.

There are two tables, the first one is the Anova table, in which it is possible to observe the different sources of variance derived by Model, the variance that derives by the independent variables and the control variables, and the Residual, the ones that it is not explained by the independent variables. It is possible to observe also the Total one, the sum of the previous explained.

SS are the Sum of Squares, associated with the three sources of variance: Total, Model and Residual. These can be computed in many ways. $\frac{SS\ Model}{SS\ Total}$ is equal to 0.0167, the value of R-Square R^2. This is because R^2 is the proportion of the variance explained by the independent variables, hence can be computed by $\frac{SS\ Model}{SS\ Total}$.

The F-statistic assesses the overall significance of the regression model. It's calculated as the ratio of the variance explained by the model (Mean Square due to the model) to the variance not explained by the model (Mean Square due to error).

The column **df** contains the degrees of freedom associated with the sources of variance. The total variance has N-1 degrees of freedom. In this case, there were N=178 observations, so the **df** for total is 177. The model degrees of freedom correspond to the number of predictors minus 1 (K-1). In this case there are only two predictors (*ln_imm_iniziale* and the intercept), for this reason the model has 1 degree of freedom.
MS are the Mean Squares, the Sum of Squares divided by their respective DF. These are useful to compute the F ratio, to test the significance of the predictors in the model.
The values on the right side of the table already explained:

- Number of observations is the number of observations used in this regression analysis;
- The F-value is the Mean Square Model divided by the Mean Square Residual, yielding F=2.99. The p-value associated with this F value is 0.0856. This value is used to test the null hypothesis that all coefficients of the model (except the intercept) are equal to zero. In this context, higher F-values indicate that the model is explaining a significant amount of variance in the dependent variable compared to the variability within the model. If the p-value were greater than 0.05 (the alpha value), the group of independent variables does not show a statistically significant relationship with the dependent variable, or that the group of independent variables does not reliably predict the dependent variable. So, in this case, there is not a statistically significant relationship between the two variables. (This is the opposite of what we expected).
- R-squared is the proportion of variance in the dependent variable (*duration*) which can be predicted from the independent variables (in this case *ln_imm_iniziali*). This value indicates that 1.67% of the variance in duration can be predicted from the variable *ln_imm_iniziali.*
- The adjusted R-square attempts to yield a more honest value to estimate the R-squared for the population, when more variables are added to the model. In this case, the value is equal to 0.0111.
- Root MSE is the standard deviation of the error term and is the square root of the Mean Square Residual (or Error).

Now we can analyze the last table, the Parameter Estimates.
The column in which there is the duration shows the dependent variable at the top (*duration*) with the predictor variables below it (in this case only *ln_imm_iniziali* and

_cons). The *_cons* variable represents the constant, called Y intercept, the predicted value of duration when all other variables are 0.

The column **Coeff** contains the values for the regression equation for predicting the dependent variable from the independent variable. The regression equation model is:

$$duration_i = \beta_0 + \beta_1 ImmStock_{jt} + \varepsilon_{ijt}$$

The column of estimates (coefficients or parameter estimates, from here on labeled coefficients) provides the values for β_0 and β_1. In this case, we have only two values because we consider only the independent variable, without the control variables.

So, in this case:

$$duration_i = 3.38 + 0.069 ln\_ImmStock_{jt} + \varepsilon_{ijt}$$

These estimates show the relationship between the independent variable and the dependent variable. If it is positive, it indicates a positive relationship between the two variables, while if it is negative, it indicates an inverse relationship.

As it is a lin_log model, it means that a change of 1% in the *imm_iniziali* variable correspond to 0.0695482 units change in the *duration* variable.

To test if the coefficients are significant or not, it is necessary to see the column of t-value and p-value (explained after).

Std. Err. are the standard errors associated with the coefficients. The standard error is used for testing whether the parameter is significantly different from 0 by dividing the parameter estimate by the standard error to obtain a t-value (see the column with t-values and p-values). The standard errors can also be used to form a confidence interval for the parameter, as shown in the last two columns of this table.

The t-value and 2-tailed p-value are used in testing the null hypothesis that the coefficient (parameter) is 0. Coefficients having p-values less than alpha are statistically significant. In this case, alpha is 0.05, coefficients having a p-value of 0.05 or less would be statistically significant (i.e., you can reject the null hypothesis and say that the coefficient is significantly different from 0).

In this case the value is higher than 0.05 meaning that this coefficient is not statistically significant.

The column [95% Conf. Interval] shows a 5% confidence interval for the coefficient. The confidence intervals are related to the p-values such that the coefficient will not be statistically significant if the confidence interval includes 0.
In this case, this interval confirm the fact that the variable is not statistically significant. While the other variable, *_cons,* is statistically significant because the p-value is less than 0.05 and there isn't the 0 inside the confidence interval, indicating that there are others variable that can influence the duration value.

Multiple Linear Regression Version 1

In the tables below are shown the results of the principal regression model. In fact, here, In addition to the independent variable (*ln_imm_iniziali*) there are also the control variables:

- *ln_pop_iniziale;*
- *sect1-sect59;*
- *ln_ricavi_iniziali;*
- *ln_tot_debiti_inizali;*
- *ln_dipendenti_iniziali;*
- *_cons.*

The model is:

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

In the Anova table below (figure 30) , the sum of squares of the Model has an higher value that corresponds to 101.57, it means that there is a lot of dispersion of data points in this model.
Also the total Sum of squares is high, indicating high dispersion for all the regression.
Because in this regression more variables are used, the degrees of freedom are higher with respect to the previous one. In fact, in this case we have 63 degrees of freedom for the model and 114 for the Residual.
The F-statistic value in this value is 1.07 and the p-value associated with it is 0.3682. This value is greater than 0.05 (the alpha value), so unfortunately the group of independent variables does not show a statistically significant relationship with the dependent variable.

The Adjusted R-squared value is equal to 0.0251, indicating that 2.51% of the variance in duration can be predicted from the variables used in this model. It is a lower value, and it indicates that these additional input variables are not adding value to the model.
The root MSE is equal to 1.2261 and the lower the value of the Root Mean Squared Error, the better the model is. A perfect model (a hypothetical model that would always predict the exact expected value) would have a Root Mean Squared Error value of 0. So, in this case this value confirms that the model is not good.

Source	SS	df	MS	Number of obs	=	178
				F(63, 114)	=	1.07
Model	101.571092	63	1.61223956	Prob > F	=	0.3682
Residual	171.378346	114	1.50331882	R-squared	=	0.3721
				Adj R-squared	=	0.0251
Total	272.949438	177	1.54208722	Root MSE	=	1.2261

Figure 30 Anova Table version 1

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_imm_iniziale	-.5203324	.4969194	-1.05	0.297	-1.504726	.4640611
ln_pop_iniziale	.6551972	.5550171	1.18	0.240	-.4442874	1.754682
sect1	2.965495	2.082349	1.42	0.157	-1.159623	7.090613
sect2	.2806664	1.882664	0.15	0.882	-3.448877	4.01021
sect3	2.564612	2.058936	1.25	0.215	-1.514125	6.643349
sect4	1.806126	2.06523	0.87	0.384	-2.285078	5.89733
sect5	1.890474	1.798388	1.05	0.295	-1.672119	5.453067
sect6	2.564612	2.058936	1.25	0.215	-1.514125	6.643349
sect7	2.608553	2.059444	1.27	0.208	-1.47119	6.688296
sect8	.5582818	2.058267	0.27	0.787	-3.519129	4.635692
sect9	3.688226	1.991038	1.85	0.067	-.2560046	7.632456
sect10	3.24374	2.090879	1.55	0.124	-.8982754	7.385755
sect11	1.196924	2.091522	0.57	0.568	-2.946365	5.340213
sect12	2.250167	1.782577	1.26	0.209	-1.281105	5.781439
sect13	2.479045	2.05372	1.21	0.230	-1.589359	6.547449
sect14	1.783506	1.909201	0.93	0.352	-1.998607	5.565619
sect15	3.369283	1.772872	1.90	0.060	-.1427634	6.881328
sect16	1.570294	2.05927	0.76	0.447	-2.509104	5.649693
sect17	3.223043	1.895712	1.70	0.092	-.5323482	6.978434
sect18	2.20974	1.817713	1.22	0.227	-1.391135	5.810616
sect19	2.771059	2.070855	1.34	0.184	-1.331289	6.873407
sect20	.180125	1.84215	0.10	0.922	-3.46916	3.82941
sect21	3.223187	1.948315	1.65	0.101	-.6364111	7.082785
sect22	1.809855	1.791747	1.01	0.315	-1.739581	5.359291
sect23	3.346305	2.097766	1.60	0.113	-.8093534	7.501963
sect24	2.935976	2.08321	1.41	0.161	-1.190846	7.062798
sect25	2.653582	2.057592	1.29	0.200	-1.422493	6.729656
sect26	3.197727	1.90196	1.68	0.095	-.5700422	6.965495
sect27	3.216676	2.110808	1.52	0.130	-.9648177	7.39817
sect28	3.43261	2.115289	1.62	0.107	-.7577606	7.62298
sect29	2.584578	2.060515	1.25	0.212	-1.497286	6.666442
sect30	3.273315	2.117614	1.55	0.125	-.9216631	7.468293
sect31	.6149869	2.139803	0.29	0.774	-3.623946	4.85392
sect32	3.499966	2.146917	1.63	0.106	-.753061	7.752993
sect33	1.310723	1.813667	0.72	0.471	-2.282136	4.903583
sect34	1.416208	2.119933	0.67	0.505	-2.783362	5.615778
sect35	2.574635	2.059685	1.25	0.214	-1.505586	6.654855
sect36	2.393873	2.109056	1.14	0.259	-1.78415	6.571897
sect37	-.0921747	2.081818	-0.04	0.965	-4.216241	4.031891
sect38	2.539917	1.891254	1.34	0.182	-1.206643	6.286477
sect39	3.32276	2.099112	1.58	0.116	-.8355656	7.481085
sect40	3.04443	2.079436	1.46	0.146	-1.074918	7.163777
sect41	1.485894	1.636277	0.91	0.366	-1.755557	4.727346
sect42	1.83754	1.66769	1.10	0.273	-1.46614	5.141221
sect43	2.61622	1.729199	1.51	0.133	-.8093096	6.04175
sect44	2.486259	1.86271	1.33	0.185	-1.203756	6.176274
sect45	0	(omitted)				
sect46	2.235626	1.660532	1.35	0.181	-1.053876	5.525127
sect47	1.239383	1.878047	0.66	0.511	-2.481013	4.95978
sect48	3.192442	2.108418	1.51	0.133	-.9843179	7.369202
sect49	.4345443	2.113288	0.21	0.837	-3.751862	4.620951
sect50	1.153023	2.142748	0.54	0.592	-3.091743	5.39779
sect51	3.253612	2.115055	1.54	0.127	-.9362959	7.443521
sect52	2.58966	1.76965	1.46	0.146	-.9160032	6.095322
sect53	2.009878	1.881165	1.07	0.288	-1.716695	5.736451
sect54	3.493002	2.126641	1.64	0.103	-.719857	7.705862
sect55	1.113124	1.786934	0.62	0.535	-2.426778	4.653027
sect56	2.584578	2.060515	1.25	0.212	-1.497286	6.666442
sect57	3.60055	2.144189	1.68	0.096	-.6470716	7.848172
sect58	2.069766	1.743501	1.19	0.238	-1.384096	5.523627
sect59	1.892623	2.070198	0.91	0.363	-2.208423	5.993669
ln_ricavi_iniziali	-.034589	.091032	-0.38	0.705	-.2149228	.1457448
ln_tot_debiti_iniziali	-.1737685	.0833927	-2.08	0.039	-.3389688	-.0085682
ln_dipendenti_iniziali	.1841367	.4304106	0.43	0.670	-.6685033	1.036777
_cons	-.9322498	2.988948	-0.31	0.756	-6.853332	4.988833

Figure 31 Linear regression results version 1

Also by the estimators table it is possible to see that the model is not good.
What was expected by this study was a positive correlation between all these variables (or, if not all, by most of them) and the dependent variables, because the initial idea was that all these factors can help a natives women to create their own start-up and keep growing it during the years. Instead, as it can be seen by the column "coefficient", except for *ln_pop_iniziale* (which indicates the total number of residents in that municipality) and the variables *sect1-sect59* (which indicates the sector in which the start-up operates), all coefficients are negative. This means that, instead of improving the survival of the startup, they have a negative effect and therefore worse the dependent variable.
In addition to this, even if the variables *ln_pop_iniziale* and the dummy variables referred to the sector in which the start-ups operate have a positive correlation with the independent variable, they are not statistically significant. In fact, the values of the p-value of all variables are higher than 0.05 (alpha value) and in all the confidence intervals there is the 0 inside it, meaning that the model is not appropriate for our variable. Also, the *_cons* variable has a p-value higher than 0.05, so there is not a significant intercept.
The value of the independent variable changes from the regression done before. In fact, in the single regression done before, the coefficient was positive by little, instead, here his value corresponds to -0.52033, meaning that a variation of 1% in the *ln_imm_iniziali* variable corresponds to -0.52033 units change in the *duration* variable, but in the opposite way. In fact, if the variable *ln_imm_iniziali* increases, the duration variable decreases by the coefficient value units. Otherwise, if the *ln_imm_iniziali* decreases, the dependent variable increases.
This is completely the opposite of what we expected. Our theory affirmed that the migration of low-skilled women from other countries should improve the native women careers and, consequently, their own business, but all these data show the opposite. In fact, the negative sign of the coefficient indicates that, instead of improving the survival rate of the start-up, the higher the number of women migrants, the lower the possibility to have a higher value of survival rate of the start-up (in our case of the *duration*).
The dummies variables, that are not in logarithmic form, must be interpreted in a different way. In fact, they must be compared to each other. For example, a change in *sect1* improves the duration by 2.96 more with respect to all the other variables.
To conclude, this linear regression is also not statistically significant to represent the causes of change of the dependent variable.

The coefficient of the variable *ln_ricavi_iniziali* is equal to -0.034589, indicating that the higher the value of initial revenues, the less the probability that the start-up survives. In this case, it was expected to have a positive coefficient between this variable and the dependent one because having more revenues means also having more money to invest in the activity. Despite this, what has emerged from past literature is that the management of money is one of the principal causes of failure for start-ups, so also this negative coefficient can be reasonable. In fact, the start-ups that have less money, manage them in a better way with respect to the ones that instead have more, because when you have less money, you pay much more attention to where to invest the money and don't waste it on unnecessary expenses for the company's business.

The same reasoning can be done for the variable *ln_tot_debiti_iniziali*, in fact also the coefficient of this variable is negative, probably due to the fact that having more debt, means also having more money to be managed, and if the founders don't have right financial education, it is easy to waste them in things that are not useful for the survive of the start-up. Otherwise, this is the unique variable of the model that is statistically significant, in fact its p-value is equal to 0.039 and its confidence interval doesn't contain the 0, meaning that it is a good variable to predict the change in the duration.

Instead, the number of employees, represented by the variable *ln_dipendenti_iniziali*, has a positive correlation with the dependent variable, indicating that a 1% increase in the variable *ln_dipendenti_iniziali* corresponds to an increase of 0.1841367 units of the *duration* variable.

So, the more the number of the employees inside the business, the higher the years of life of the start-up.

The constant variable *_cons,* different from the other regression, is not significant and has a p-value higher than 0.05.

Most of the p-values of this regression model are higher than 0.05 and most of the confidence interval contains the value zero, indicating that none of these variables are significant to explain the dependent variable.

Multiple Linear Regression Version 2

In this second variation of regression, it was decided to put as independent variable the initial immigration rate.

This rate is caluculated as:

$$\frac{Imm\_iniziale}{Pop\_iniziale}$$

It indicates the rate of women immigrants in the total population resident in the same municipality and in the same year in which the start-up is founded.

As the principal regression model, also in this case it is a lin-log model, in fact, the rate "*imm_rate_iniziale"* is transformed in a log scale.

The model of this multiple lin-log regression is:

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

In the following Anova table (figure 32), as the previous ragressions, it is possible to observe some data that indicates if the model is better or worst than the previous ones.

In this model, the Sum of Squares of the model is equal to 97.103, less than the principal regression model, indicating that in this regression the dispersion of data is decresed respect to the other.

The degrees of freedom are quite the same as before because the number of variables used in both model differs only by one.

The F-statistic value in this model corresponds to 1.02 and the p-value associated with it is 0.4484. Also in this model, the value is even greater than 0.05 (the alpha value), so also in this case the independent variable and the control variables does not show a statistically significant relationship with the dependent variable.

The Adjusted R-squared value is equal to 0.3558, indicating that 35.58% of the variance in duration can be predicted by the variables used in this model. It is a better value respect the previous one, indicating that these additional input variables are adding value to the model.

The root MSE is equal to 1.2366, higher than before, indicating that the model is even worse than the previous one. The better is the model, the less is the value of root MSE, also in this case this value confirms that the model is not good to explain the variation of the variable duration.

Source	SS	df	MS	Number of obs	=	178
				F(62, 115)	=	1.02
Model	97.1030422	62	1.5661781	Prob > F	=	0.4484
Residual	175.846396	115	1.5290991	R-squared	=	0.3558
				Adj R-squared	=	0.0084
Total	272.949438	177	1.54208722	Root MSE	=	1.2366

Figure 32 Anova table version 2

Even if the previous index indicates that the model is not good, the coefficient related to the variable *ln_imm_rate_iniziale* is positive, indicating that the higher the rate, the higher is the duration of the life of the start-up.

In fact, in this case, the caoefficient value is equal to 0.1026677, indicating that if the variable icrease by 1%, the duration will increase by 0.102667 units. The difference with respect to the previous one is the direct correlation, in fact if the independent variable increase, the dependent variable increase too, while in the previous regression model, if the independent variable increase, the dependent variable (the *duration*) decreases.

The coefficient related to the dummies variables indicating the sectors in which the start-ups operate are similar to the ones of the previous model. In fact, also in this regression there are some positive coefficient, someone more higher than others (for example *sect21* has a coefficent equal to 3.036451 while *sect4* 1.401474), and some with negative coefficent (for example the *sect20* equal to -0.06718). This result has to be seen as: the duration of the start-ups that work in the sector indicated by the variable *sect21* is 3.036451 higher respect to the others that work in the other sectors.

The variables *ln_ricavi_inziali* and *ln_debiti iniziali,* that indicates the revenues and the level of debt of the first year of start-up, as the model showed before, have a negative coefficients, indicating that the higher is this value, the less is the duration of life of the start-up.

The level of initial debt indicated by the variable *ln_debiti_iniziali* is statistically sognificant, because the value of the p-value associated to it is equal to 0.048.

Looking to background literature, this is due to the difficult in managing money in the beginning phase, investing them in things that are not necessary to the success of the start-up.

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_imm_rate_iniziale	.1026677	.3492226	0.29	0.769	-.5890752	.7944105
sect1	2.367712	2.071283	1.14	0.255	-1.735101	6.470525
sect2	.1509634	1.897243	0.08	0.937	-3.607111	3.909038
sect3	2.357588	2.073086	1.14	0.258	-1.748798	6.463973
sect4	1.401474	2.069403	0.68	0.500	-2.697616	5.500564
sect5	1.692832	1.810165	0.94	0.352	-1.892756	5.278419
sect6	2.357588	2.073086	1.14	0.258	-1.748798	6.463973
sect7	2.381272	2.072895	1.15	0.253	-1.724734	6.487278
sect8	.3596046	2.072682	0.17	0.863	-3.745979	4.465188
sect9	3.513823	2.005456	1.75	0.082	-.4585994	7.486244
sect10	3.0097	2.104367	1.43	0.155	-1.158646	7.178046
sect11	1.10604	2.108516	0.52	0.601	-3.070525	5.282605
sect12	1.980656	1.790981	1.11	0.271	-1.566932	5.528244
sect13	2.374597	2.070383	1.15	0.254	-1.726433	6.475627
sect14	1.650664	1.923883	0.86	0.393	-2.160178	5.461505
sect15	3.292213	1.786997	1.84	0.068	-.2474854	6.831911
sect16	1.356984	2.073212	0.65	0.514	-2.74965	5.463618
sect17	2.724945	1.88982	1.44	0.152	-1.018425	6.468315
sect18	1.6995	1.808812	0.94	0.349	-1.883409	5.282408
sect19	2.539236	2.084243	1.22	0.226	-1.589248	6.66772
sect20	-.06718	1.85173	-0.04	0.971	-3.7351	3.60074
sect21	3.036451	1.961969	1.55	0.124	-.8498315	6.922734
sect22	1.737559	1.806532	0.96	0.338	-1.840834	5.315952
sect23	2.910101	2.100461	1.39	0.169	-1.250509	7.070711
sect24	2.705417	2.096761	1.29	0.200	-1.447864	6.858697
sect25	2.414224	2.070351	1.17	0.246	-1.686743	6.51519
sect26	2.615168	1.887676	1.39	0.169	-1.123955	6.354291
sect27	2.965445	2.123847	1.40	0.165	-1.241487	7.172377
sect28	2.924456	2.11271	1.38	0.169	-1.260416	7.109329
sect29	2.354028	2.073856	1.14	0.259	-1.753882	6.461938
sect30	3.02506	2.13084	1.42	0.158	-1.195725	7.245844
sect31	.1854755	2.143314	0.09	0.931	-4.060017	4.430968
sect32	3.239467	2.159961	1.50	0.136	-1.039001	7.517934
sect33	1.199622	1.827998	0.66	0.513	-2.421289	4.820533
sect34	1.195887	2.134203	0.56	0.576	-3.031559	5.423332
sect35	2.355897	2.073443	1.14	0.258	-1.751194	6.462989
sect36	1.850482	2.103155	0.88	0.381	-2.315463	6.016427
sect37	-.4475613	2.089437	-0.21	0.831	-4.586333	3.69121
sect38	2.134996	1.89273	1.13	0.262	-1.614138	5.88413
sect39	2.948757	2.105789	1.40	0.164	-1.222405	7.119919
sect40	2.828658	2.093472	1.35	0.179	-1.318106	6.975422
sect41	1.249716	1.644562	0.76	0.449	-2.007844	4.507276
sect42	1.543416	1.673227	0.92	0.358	-1.770924	4.857756
sect43	2.394203	1.739255	1.38	0.171	-1.050926	5.839332
sect44	2.424358	1.878192	1.29	0.199	-1.295978	6.144694
sect45	0	(omitted)				
sect46	1.889517	1.662513	1.14	0.258	-1.403601	5.182636
sect47	1.020537	1.889864	0.54	0.590	-2.72292	4.763994
sect48	2.953492	2.121907	1.39	0.167	-1.249597	7.156581
sect49	.2139395	2.127356	0.10	0.920	-3.999943	4.427822
sect50	1.184756	2.175829	0.54	0.587	-3.125142	5.494654
sect51	3.001173	2.12809	1.41	0.161	-1.214164	7.216511
sect52	2.510523	1.784184	1.41	0.162	-1.023601	6.044648
sect53	1.561468	1.879134	0.83	0.408	-2.160736	5.283671
sect54	3.045575	2.128965	1.43	0.155	-1.171495	7.262645
sect55	.6991187	1.786114	0.39	0.696	-2.838829	4.237067
sect56	2.354028	2.073856	1.14	0.259	-1.753882	6.461938
sect57	3.249036	2.152663	1.51	0.134	-1.014976	7.513048
sect58	2.025891	1.758202	1.15	0.252	-1.45677	5.508551
sect59	1.506897	2.075748	0.73	0.469	-2.604761	5.618554
ln_ricavi_iniziali	-.0393052	.0917536	-0.43	0.669	-.2210515	.1424411
ln_tot_debiti_iniziali	-.1680833	.0840428	-2.00	0.048	-.3345559	-.0016106
ln_dipendenti_iniziali	.1142572	.4321754	0.26	0.792	-.741799	.9703134
_cons	2.942225	2.0162	1.46	0.147	-1.051479	6.935928

Figure 33 Linear regression results version 2

Multiple Linear Regression Version 3

As explained above, in this last variant multiple linear regression, the independent variable is composed by the Local Labor System variable.

The Piedmont region has 25 Local Labor Systems: Acqui Terme, Alba, Alessandria, Asti, Biella, Borgomanero, Borgosesia, Bra, Canelli, Casale Monferrato, Chieri, Cossato, Cuneo, Ivrea, Mondovì, Novara, Novi Ligure, Pinerolo, Saluzzo, Savigliano, Susa, Torino, Tortona and Vercelli.

Each Local Labor System includes the municipalities in a ray of tot kilometers, in this way also people who work in another municipality different from the one in which they live are considered.

A new database has been created and, in this case, the independent variable refers to the Local Labor System, not to the municipality as the previous linear regression.

The model, as explained above is:

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

In this case, the independent variable is the total number of immigrants women resident in the same Local Labor System of the start-up.

Also here, to facilitate the lecture of the results, a logarithmic form of the independent variable is created.

The results obtained are shown in the following tables.

Source	SS	df	MS	Number of obs	=	178
				F(62, 115)	=	1.09
Model	100.881453	62	1.62712022	Prob > F	=	0.3449
Residual	172.067985	115	1.49624335	R-squared	=	0.3696
				Adj R-squared	=	0.0297
Total	272.949438	177	1.54208722	Root MSE	=	1.2232

Figure 34 Anova Table version 3

The data dispersion of the Model is higher respect to the version 2, its value in this case is equal to 100. 881, as Figure 34 shows.

The F-statistic value in this model corresponds to 1.09 and the p-value associated with it is 0.3449. Also in this model, the value is even greater than 0.05 (the alpha value), so also in this case the independent variable and the control variables does not show a statistically

significant relationship with the dependent variable, otherwise in this case we have a 34.49% while in the version 2 it was 44.84% and in the version 1 36.82%. So, it is better of the other 2 cases.

The Adjusted R-squared value is equal to 0.0297, indicating that 2.97% of the variance in duration can be predicted by the variables used in this model. It is a worse value respect the previous one, indicating that these additional input variables don't add value to the model.

The root MSE is equal to 1.2232, similar to the previous model. As it is a value higher than 1, also in this case it confirms that the model is not good to explain the variation of the variable duration.

In the Figure 35, is it possible to see the coefficients of the linear regression.

The coefficient of the independent variable, in this case the logaritm of the number of immigrants women resident in the same Local Labor Sytstem where the start-up considered is sited in the year in which it is founded is positive, equal to 0.0586888. It is interesting to notice that, even if in the version 2 the coefficient was positive. The p-value associated to the independent variable was around 70%, while in this case it corresponds to 10.9%, indicating that this variable influence more the variations of the dependent variable. Even if it is decresed, it idicates that also in this case it is not statistically significant because it is higher than alpha vale, 0.05.

Also in this regression, some coefficients of the dummy variables indicating the sector in which the start-up operates are positive, while other are negative. However, all these variables have a p-value higher than 0.05, meaning that they are not statistically significant, so the sector in which the start-up operates is not a good variable to predict the changes of the dependent variable.

The coefficient of the variable *ln_ricavi_iniziali* is negative, probably for the reasons explained in the other models, and also in this case it is not statistically significant.

Also the coefficient of the variable *ln_tot_debiti_iniziali* is negative, meaning that the higher the level of debt during the first year of life of the start-up, the less is the duration of life of it. However, this variable is statistically significant, having a p-value equal to 0.035 and a 95% coefficient interval that doesn't contain the value 0. For this reason, the variable can influence in a statistical way the dependent variable duration.

As the previous models, the coefficient of the variable *ln_dipendenti_iniziali* indicating the number of employees during the first year of life of the start-up is positive, meaning

that the higher the number of employees, the higher the duration. Otherwise, even in this case, the variable is not statistically significant.

It is interesting to notice that in this case the constant variable *_cons* is statistically significant, differently from the other cases, with a p-value equal to 0.001. This indicate that in this case the intercept term has an effect on the dependent variable.

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_immi_sll	.0586888	.0363025	1.62	0.109	-.0132194	.1305971
sect1	-.9108965	2.054436	-0.44	0.658	-4.980339	3.158546
sect2	-3.496569	1.900848	-1.84	0.068	-7.261783	.2686451
sect3	-1.242244	2.070778	-0.60	0.550	-5.344057	2.859569
sect4	-1.917549	2.054564	-0.93	0.353	-5.987245	2.152147
sect5	-1.915683	1.81463	-1.06	0.293	-5.510115	1.678749
sect6	-1.242244	2.070778	-0.60	0.550	-5.344057	2.859569
sect7	-1.265428	2.079611	-0.61	0.544	-5.384738	2.853882
sect8	-3.245713	2.071056	-1.57	0.120	-7.348076	.8566501
sect9	-.1333887	2.02461	-0.07	0.948	-4.143751	3.876973
sect10	-.6255631	2.10502	-0.30	0.767	-4.795202	3.544076
sect11	-2.595772	2.105699	-1.23	0.220	-6.766756	1.575212
sect12	-1.529584	1.7811	-0.86	0.392	-5.0576	1.998431
sect13	-1.184771	2.066496	-0.57	0.568	-5.278103	2.90856
sect14	-2.028343	1.912781	-1.06	0.291	-5.817194	1.760507
sect15	0	(omitted)				
sect16	-2.258393	2.072089	-1.09	0.278	-6.362803	1.846018
sect17	-.5585106	1.87573	-0.30	0.766	-4.27397	3.156949
sect18	-1.68686	1.798934	-0.94	0.350	-5.250202	1.876482
sect19	-1.082675	2.075006	-0.52	0.603	-5.192863	3.027514
sect20	-3.464279	1.776368	-1.95	0.054	-6.982921	.0543632
sect21	-.6081972	1.964373	-0.31	0.757	-4.499241	3.282847
sect22	-1.876771	1.81119	-1.04	0.302	-5.464389	1.710847
sect23	-.4556852	2.072857	-0.22	0.826	-4.561615	3.650245
sect24	-.8951107	2.078471	-0.43	0.668	-5.012161	3.221939
sect25	-.9174463	2.054562	-0.45	0.656	-4.987138	3.152246
sect26	-.7333592	1.870562	-0.39	0.696	-4.438583	2.971864
sect27	-.6417738	2.094075	-0.31	0.760	-4.789734	3.506186
sect28	-.3361597	2.072108	-0.16	0.871	-4.440607	3.768287
sect29	-1.287351	2.074559	-0.62	0.536	-5.396654	2.821952
sect30	-.5697085	2.097508	-0.27	0.786	-4.724469	3.585052
sect31	-3.045239	2.090418	-1.46	0.148	-7.185955	1.095477
sect32	-.3524428	2.116849	-0.17	0.868	-4.545512	3.840627
sect33	-2.454885	1.83521	-1.34	0.184	-6.090084	1.180313
sect34	-2.373554	2.102836	-1.13	0.261	-6.538867	1.79176
sect35	-1.264158	2.072569	-0.61	0.543	-5.369518	2.841203
sect36	-1.4527	2.06655	-0.70	0.484	-5.546138	2.640738
sect37	-3.842647	2.110576	-1.82	0.071	-8.023292	.3379973
sect38	-1.25595	1.870655	-0.67	0.503	-4.961358	2.449458
sect39	-.3656498	2.072199	-0.18	0.860	-4.470278	3.738979
sect40	-.7959389	2.10208	-0.38	0.706	-4.959754	3.367877
sect41	-2.311042	1.644745	-1.41	0.163	-5.568965	.9468799
sect42	-1.961524	1.666092	-1.18	0.241	-5.261732	1.338684
sect43	-1.172075	1.74046	-0.67	0.502	-4.619592	2.275442
sect44	-1.184771	1.876775	-0.63	0.529	-4.902302	2.532759
sect45	-3.689649	1.768869	-2.09	0.039	-7.193439	-.18586
sect46	-1.543356	1.652571	-0.93	0.352	-4.816781	1.730068
sect47	-2.605968	1.885847	-1.38	0.170	-6.341467	1.129532
sect48	-.6331754	2.09135	-0.30	0.763	-4.775737	3.509386
sect49	-2.978451	2.083185	-1.43	0.155	-7.10484	1.147938
sect50	-1.973819	2.096085	-0.94	0.348	-6.12576	2.178122
sect51	-.6040497	2.096561	-0.29	0.774	-4.756934	3.548834
sect52	-1.146935	1.798189	-0.64	0.525	-4.708802	2.414932
sect53	-1.812971	1.881609	-0.96	0.337	-5.540077	1.914134
sect54	-.3017967	2.083714	-0.14	0.885	-4.429233	3.82564
sect55	-2.739102	1.780655	-1.54	0.127	-6.266237	.7880337
sect56	-1.287351	2.074559	-0.62	0.536	-5.396654	2.821952
sect57	-.0298128	2.097607	-0.01	0.989	-4.184768	4.125143
sect58	-1.622525	1.76134	-0.92	0.359	-5.1114	1.86635
sect59	-1.808974	2.055722	-0.88	0.381	-5.880963	2.263015
ln_ricavi_iniziali	-.0316288	.0905058	-0.35	0.727	-.2109032	.1476457
ln_tot_debiti_iniziali	-.1774741	.0831917	-2.13	0.035	-.3422609	-.0126874
ln_dipendenti_iniziali	.1305626	.4243666	0.31	0.759	-.710026	.9711512
_cons	5.480415	1.662034	3.30	0.001	2.188245	8.772585

Figure 35 Linear regression results version 3

Multiple Linear Regression Version 4

In this last regression model, to increase even more than the previous version the geographical area considered, the independent variable considers the total number of immigrants women resident in the same province of the start-up during the first year of life.

The linear regression model is:

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

As the previous ones, the variable are transformed into a logarithmic scale.

Source	SS	df	MS	Number of obs	=	178
				F(62, 115)	=	1.06
Model	99.3232266	62	1.60198753	Prob > F	=	0.3866
Residual	173.626212	115	1.50979314	R-squared	=	0.3639
				Adj R-squared	=	0.0209
Total	272.949438	177	1.54208722	Root MSE	=	1.2287

Figure 36 Anova Table version 4

Also in this case, the p-value of the model indicates that it is not statistically significant. In fact, the p-value is equal to 0.3866, while it should be under the alpha value that is equal to 0.05.

The Adjusted R-squared value is equal to 0.0209, indicating that 2.09% of the variance in duration can be predicted by the variables used in this model. It is even lower than the version 3, meaning that the variables used in this model are less suitable to predict the changes in the duration variable.

Also the Root MSE equal to 1.2287 shows that the model is not statistically significant.

A difference from the previous models can be seen by the coefficients obtained by the regression regarding the sectors in which the start-ups operate (Figure 37). In fact, in this case all the coefficients are positive, except one, indicating that all the sectors, except the *sect37* have a positive correlation with the dependent variable. The duration of the start-ups that work in the sector indicated by the variable *sect37* is -0.1562 lower respect to the others that work in the other sectors.

However, also in this table the p-value shows that all the sectors are statistically insignificant.

The initial revenues represented by *ln_ricavi_iniziali* have a negative correlation with the duration and also in this case are not statistically significant.
The total initial debt, instead, is statistically significant, having a p-value equal to 0.038 and has a negative coefficient, meaning that a variation of 1% in the *ln_tot_debiti_iniziali* variable corresponds to -0.175 units change in the *duration* variable.
To conclude, also here the initial number of employees has a positive influence on the *duration* varibale, otherwise it is statistically significant as it has a p-value equal to 73.3%.
So, also by increasing the geographical area esaminated, the variables chosen don't have the results of the hypotesis expected and are not sufficient to describe the changes in the dependent variable.

duration	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_immi_prov	.0531955	.042617	1.25	0.214	-.0312206	.1376116
sect1	2.783676	2.079663	1.34	0.183	-1.335736	6.903087
sect2	.2158894	1.884105	0.11	0.909	-3.51616	3.947939
sect3	2.483343	2.057395	1.21	0.230	-1.591961	6.558646
sect4	1.468575	2.05702	0.71	0.477	-2.605986	5.543135
sect5	1.810538	1.796015	1.01	0.316	-1.747022	5.368098
sect6	2.483343	2.057395	1.21	0.230	-1.591961	6.558646
sect7	2.465866	2.055038	1.20	0.233	-1.604768	6.5365
sect8	.4801987	2.057309	0.23	0.816	-3.594935	4.555333
sect9	3.589514	1.988491	1.81	0.074	-.3493045	7.528332
sect10	3.104569	2.086328	1.49	0.139	-1.028046	7.237183
sect11	1.13153	2.086604	0.54	0.589	-3.001631	5.264691
sect12	2.081049	1.779693	1.17	0.245	-1.44418	5.606279
sect13	2.535436	2.05926	1.23	0.221	-1.543562	6.614434
sect14	1.697334	1.910631	0.89	0.376	-2.087258	5.481926
sect15	3.640545	1.785569	2.04	0.044	.1036762	7.177413
sect16	1.468575	2.05702	0.71	0.477	-2.605986	5.543135
sect17	3.130111	1.903666	1.64	0.103	-.6406856	6.900907
sect18	2.017943	1.809774	1.12	0.267	-1.566872	5.602757
sect19	2.644048	2.067764	1.28	0.204	-1.451794	6.739891
sect20	.1933365	1.844506	0.10	0.917	-3.460275	3.846948
sect21	3.111112	1.946687	1.60	0.113	-.7448995	6.967124
sect22	1.845808	1.797279	1.03	0.307	-1.714255	5.405871
sect23	3.236819	2.100455	1.54	0.126	-.9237792	7.397417
sect24	2.828314	2.081357	1.36	0.177	-1.294454	6.951081
sect25	2.747696	2.075537	1.32	0.188	-1.363544	6.858936
sect26	2.96616	1.894115	1.57	0.120	-.7857173	6.718038
sect27	3.080824	2.10832	1.46	0.147	-1.095352	7.257001
sect28	3.247879	2.114685	1.54	0.127	-.9409054	7.436663
sect29	2.442458	2.056523	1.19	0.237	-1.631119	6.516034
sect30	3.151285	2.115812	1.49	0.139	-1.03973	7.342301
sect31	.5170946	2.147212	0.24	0.810	-3.736118	4.770307
sect32	3.366924	2.145163	1.57	0.119	-.8822318	7.616079
sect33	1.272207	1.817399	0.70	0.485	-2.32771	4.872124
sect34	1.345358	2.120567	0.63	0.527	-2.855076	5.545792
sect35	2.463389	2.056904	1.20	0.234	-1.610943	6.537721
sect36	2.172563	2.106636	1.03	0.305	-2.000278	6.345405
sect37	-.1562792	2.087438	-0.07	0.940	-4.291092	3.978533
sect38	2.367531	1.889557	1.25	0.213	-1.375317	6.11038
sect39	3.327917	2.113984	1.57	0.118	-.859478	7.515312
sect40	2.934204	2.07559	1.41	0.160	-1.17714	7.045547
sect41	1.395451	1.635989	0.85	0.395	-1.845128	4.636031
sect42	1.729354	1.665761	1.04	0.301	-1.570198	5.028905
sect43	2.549608	1.727991	1.48	0.143	-.8732095	5.972425
sect44	2.535436	1.866991	1.36	0.177	-1.162715	6.233586
sect45	0	(omitted)				
sect46	2.13348	1.662293	1.28	0.202	-1.159202	5.426163
sect47	1.121098	1.874723	0.60	0.551	-2.592368	4.834563
sect48	3.087324	2.10726	1.47	0.146	-1.086754	7.261401
sect49	.4700258	2.112916	0.22	0.824	-3.715255	4.655306
sect50	1.692914	2.153945	0.79	0.434	-2.573636	5.959464
sect51	3.118127	2.112662	1.48	0.143	-1.066649	7.302904
sect52	2.553383	1.77247	1.44	0.152	-.9575392	6.064305
sect53	1.857432	1.88195	0.99	0.326	-1.870349	5.585212
sect54	3.386247	2.130888	1.59	0.115	-.834633	7.607126
sect55	.9194563	1.77934	0.52	0.606	-2.605075	4.443987
sect56	2.442458	2.056523	1.19	0.237	-1.631119	6.516034
sect57	3.653295	2.16465	1.69	0.094	-.6344605	7.941051
sect58	2.067178	1.747278	1.18	0.239	-1.393843	5.528199
sect59	1.831072	2.078579	0.88	0.380	-2.286193	5.948337
ln_ricavi_iniziali	-.0337713	.0909267	-0.37	0.711	-.2138796	.146337
ln_tot_debiti_iniziali	-.1754918	.0836049	-2.10	0.038	-.3410971	-.0098866
ln_dipendenti_iniziali	.1458976	.4264826	0.34	0.733	-.6988823	.9906775
_cons	1.826136	1.7629	1.04	0.302	-1.665829	5.318102

Figure 37 Linear regression results version 4

Discussion and limitations

The goal of this thesis is to evaluate if there is a positive correlation between the number of low-skilled women immigrated in Italy and the success of native women innovative start-up.
To evaluate this, 4 different linear regressions have been implemented, with different independent variables but with the same dependent one: the duration of the start-up considered.

The first one has as dependent variable the number of foreign women resident in the same municipality in which the start-up is sited. The immigrant women considered in this study are the ones with an age between 25 and 64 years old.
The results obtained in this first regression show a negative correlation between these two variables, going against the initial hypothesis. Furthermore, the p-value is higher than 0.05, indicating that the regression is not statistically significant.

In the second linear regression, the independent variable is the immigration rate, created dividing the number of women immigrants by the number of the total population resident in that municipality.
In this case the coefficient obtained by the regression is positive (in accordance with the initial hypothesis), but also in this case the p-value has a higher value, indicating that this variable is not good to explain the variation of the dependent variable.

The third variant utilized as independent variable the number of foreign women resident in the same Local Labor System in which the start-up considered is sited. In this regression, the geographical area considered is larger than the previous one. For this reason, it is possible to evaluate also the numerous cases in which the women work in a municipality different from the one in which they live.
The coefficient obtained between the variables is positive, indicating that the highest the number of immigrant women, the higher the duration of the start-up in the Local Labor System. However, also in this case the p-value indicates that the variable is not statistically significant, but in this case is lower than the previous one, meaning that considering a larger geographical area it is possible to obtain better results.

In the end, the last linear regression uses as independent variable the number of immigrant women resident in the same province of the start-up considered. The coefficient obtained is positive, but it has a worst p-value than the previous model.
These results are not consistent with the initial expectations of this thesis, supported also by the literature analysis considering previous studies and research. In fact, the initial idea was that the number of immigrant women residing in the same municipality where a start-up founded by a native woman is located could positively influence its success.
The opposite results obtained may have different causes.

Limitations

During this research, several limitations have been encountered.
Firstly, through the literature analysis, it was challenging to identify the factors that can influence the success of start-ups, and certainly, besides those listed, there are many others. Among these, only some were used in the various regressions as control variables because information for start-ups was not available. For example, it was not possible to determine how the various start-ups were financed, whether by Business Angels or Venture Capital, and this, according to the literature, is one of the factors that most influences the success of a start-up. Also, the background of the founders could not be analyzed because the data were not collected.
Certainly, the decisions made in this research also influenced the results.

- The period considered for the analysis, from 2001 to 2021. By choosing a larger range of years considered, it has been possible to collect more data and, consequently, obtain more precise results from the different regressions.

- The lack of knowledge about the educational background of the analyzed foreign women. In fact, in this study, all the foreign women from 24 to 65 years old are considered as low-skilled women, meaning that they don’t have any degree or diploma. This is inconsistent with the reality, because not all the immigrant women have a lower degree of education and they do work different from the childcare and home services.

- The decision to concentrate the study on the Piedmont region. This is not one of the regions in which there is a higher density of innovative start-ups with a female control. Expanding the area of consideration, more accurate data can be collected and analyzed.

In the following chapter, some suggestions for future improvements regarding this study, to obtain more accurate data, are exposed.

Conclusions and future improvements

This thesis addresses highly debated topics nowadays: innovative start-ups, women's work, and immigration in Italy.

In fact, especially the last topic has been a subject of discussion in recent years, particularly in Italian politics, where different viewpoints have emerged: some argue that immigration could be a positive factor for the Italian economy, while others argue the opposite.

This thesis seeks to correlate these three themes with the initial hypothesis that the influx of immigrant women into Italy could help Italian women succeed more in their innovative start-ups. Although the results obtained from various regressions were not statistically significant and therefore do not confirm the initial hypothesis, this could be a starting point for future studies. Indeed, once the limitations explained in the previous chapter are overcome, significant results can be obtained, especially considering the following considerations:

- It has been observed that results improve when considering a larger geographical area, so for future research, it is advisable to use variables that consider larger geographical areas;
- Expand the temporal space: collecting data over a longer period of time will certainly yield more precise results and a bigger number of data;
- To assess the assistance that can be provided to Italian women to succeed in their start-ups and dedicate more time to work, it is important to evaluate the presence of nursery school in the city of residence, as it has emerged that the female employment rate in Italy is higher in cities where childcare services, such as nursery schools, are more available;
- Add other control variables that were not used in this study due to lack of data, such as the way the start-up is funded, whether through business angels or venture capital, or other factors described in the literature analysis that are not utilized in this research.

References

1. Istat. 2019. *"I tempi della vita quotidiana. Lavoro, conciliazione, parità di genere e benessere soggettivo."* Available at: www.istat.it.

2. "*Startup innovative, record in Italia: sono 14.708*." 28 Dicembre 2022. il Sole 24 Ore.

3. De Luca, Massimo, Chiara Tronchin, and Enrico Di Pasquale. September 2019. *Primo rapporto annuale sul lavoro domestico. Analisi, statistiche, trend nazionali e locali*. National Observatory Domina.

4. Cortes, Patricia, and Jessica Pan. March 2017. *Occupation and Gender*.

5. Cortes, Patricia. December 11, 2022. *Immigration, Household Services, and Women's Economic Opportunities.* Boston University, United States.

6. Cortes, Patricia, and Jessica Pan. October 2015. *When Time Binds: Returns to Working Long Hours and the Gender Wage Gap among the Highly Skilled.*

7. Cortes, Patricia, and Josè Tessada. July 2011. *Low-Skilled Immigration and the Labor Supply of Highly Skilled Women.*

8. Lechler, Thomas. June 2011. *Social Interaction: A Determinant of Entrepreneurial Team Venture Success.*

9. Manzo, Cecilia, and Ivana Pais. *I fondatori di startup in Italia tra agency e struttura.*

10. Ballista, Luca. *I driver del successo delle digital Start-Up italiane.*

11. Paoloni, Paola, and Giuseppe Modaffari. October 29, 2018. *Female-Owned Innovative Startups in Italy: Status Quo and Implications.*

12. Matricano, Diego. 2022. *Economic and social development generated by innovative startups: does heterogeneity persist across Italian macro-regions?*

13. Bertolini, Sonia, and Valentina Gogli. *L'imprenditoria femminile come strumento di innovazione per lo sviluppo locale.*

14. Unioncamere, Comunicato stampa. January 5, 2023. *Imprese: cresce l'innovazione al femminile +572 start up innovative in 2 anni.* Roma.

15. Goldin, Claudia. April 2014. *A Grand Gender Convergence: Its Last Chapter.*

16. Goldin, Claudia, and Tanya Devani. Spring 2017. *Narrowing the wage gap: an interview with Claudia Goldin.* Harvard International Review, Vol. 38, No. 2, pp. 68-70.

17. Goldin, Claudia, and Joshua Mitchell. Winter 2017. *The New Life Cycle of Women's Employment: Disappearing Humps, Sagging Middles, Expanding Tops*. The Journal of Economic Perspectives, Vol. 31, No. 1, pp. 161-182.

18. Goldin, Claudia, and Claudia Olivetti. May 2013. *Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply.* The American Economic Review, Vol. 103, No. 3, pp. 257-262.

19. Ministro del lavoro e delle politiche sociali. March 2023. *Rapporti di lavoro nel IV trimestre 2022.*

20. Mason, Colin, and Richard T. Harrison. *Barriers to Investment in the Informal Venture Capital Sector.*

21. Franchi, Maura. 1992. *Donne imprenditrici-le regole del gioco.* Franco Angeli, Milano.

22. Favretto, Giuseppe, and Riccardo Sartori. 2007. *Le età dell'impresa: giovani imprenditori e lavoratori esperti.* Franco Angeli, Milano.

23. Ujwary-Gil, Anna, Anna Florek-Paszkowska, and Bianka Godlewska-Dzioboń. JEMI, *Entrepreneurship and innovation in the age of digital transformation*.

24. Krzyżanowska, Magdalena, and Jolanta Tkaczyk. April 2013. *Identifying Competitors: Challenges for Start-Up Firms.*

25. OECD, Innovation in Firms: A Microeconomic Prospective, 2009

26. Adhanan, E.-T. *What are the three stages of a startup?*. Silicon Valley Bank

27. Watkins, D., & Watkins, J. 1984. *The female entrepreneur: Background and determinants of business choice—Some British data*.

28. Patrizia, & Vicarelli, G. 1994. *Donne nelle professioni degli uomini.* Franco Angeli, Milano.

29. Stock, J. H., & Watson, M. W. (2018). *Introduction to Econometrics.* Pearson, London.

30. Istat. (n.d.). Donne e Uomini - ISTAT. Available at: https://www.istat.it/donne-uomini/bloc-3d.html?langit

31. Integrazione Migranti. (n.d.). Cresce il lavoro domestico in Italia, oltre 920mila i lavoratori regolari. Available at: https://integrazionemigranti.gov.it/it-it/Ricerca-news/Dettaglio-news/id/2309/Cresce-il-lavoro-domestico-in-Italia-oltre-920mila-i-lavoratori-regolari

32. Unioncamere. (n.d.). Imprese: cresce l'innovazione al femminile +572 start up innovative in 2 anni. Available at: https://www.unioncamere.gov.it/comunicazione/comunicati-stampa/imprese-cresce-linnovazione-al-femminile-572-start-innovative-2-anni

33. Italia in Dati. (n.d.). Immigrazione in Italia - Dati e Statistiche. Available at: https://italiaindati.com/immigrazione

34. Statista. (n.d.). Foreign population composition in Italy by educational level 2021. Available at: https://www.statista.com/statistics/1248244/foreign-population-composition-in-italy-by-educational-level/

35. Imprenditoria Femminile. (n.d.). Home - Imprenditoria Femminile. Available at: https://www.imprenditoriafemminile.camcom.it/

1. Ministero dello Sviluppo Economico. (n.d.). Fondo a sostegno dell'impresa femminile. Available at: https://www.mimit.gov.it/index.php/it/incentivi/fondo-a-sostegno-dell-impresa-femminile

2. EconomyUp. (n.d.). Imprenditoria femminile: cos'è e quante la fanno davvero. Available at: https://www.economyup.it/innovazione/imprenditoria-femminile-cose-e-quante-la-fanno-davvero/

3. Istat. (n.d.). Istituto Nazionale di Statistica. Available at: https://noi-italia.istat.it/pagina.php?id=3&categoria=4&action=show&L=0#:~:text=Nel%202022%2C%20il%20tasso%20di,autoctoni%20(64%2C9%25)

36.

37. Informazione Fiscale. (n.d.). Business Angel: definizione, cosa fanno, chi sono. Available at: https://www.informazionefiscale.it/business-angel-definizione-cosa-fanno-chi-sono

38. UCLA Office of Academic Research and Collaboration. (n.d.). Regression Analysis. Available at: https://stats.oarc.ucla.edu/stata/output/regression-analysis/

39. IBAN. (2023, July 25). Survey IBAN 2022. Available at: https://www.iban.it/2023/07/25/survey-iban-2022/

40. Ministero dell'Innovazione e delle Tecnologie. (n.d.). Art. 25 - Decreto Legislativo 179/2012. Available at: https://www.mimit.gov.it/images/stories/Art25-dl179-2012.pdf

41. Ministero dell'Innovazione e delle Tecnologie. (n.d.). Sostegno alle startup innovative: Smart Start Italia. Available at: https://www.mimit.gov.it/it/incentivi/sostegno-alle-startup-innovative-smart-start-italia

42. ICE - Agenzia per la promozione all'estero e l'internazionalizzazione delle imprese italiane. (n.d.). Global Start-Up Program. Available at: https://www.ice.it/it/settori/startup-innovazione/global-start-up-program

43. Istituto Nazionale di Statistica (ISTAT). (n.d.). Archivio. Retrieved from https://www.istat.it/it/archivio/230110

44. Osservatorio Domina sul Lavoro Domestico. (2023). *Focus regionale 2023: Analisi, statistiche e trend del lavoro domestico Piemonte. Il Rapporto, 3° Rapporto Annuale sul Lavoro Domestico.*