



**Politecnico  
di Torino**

## **Politecnico di Torino**

Master's degree in Engineering and Management

Academic year 2023/2024

October 2024

### **Supporting Female Entrepreneurs of Northwestern Italy**

The influence of low-skilled immigrated women and childcare services on  
the survival rate of female-led innovative startup

Supervisor: Anna D'Ambrosio

Candidate: Flora Quaranta

# Abstract

The focus of this thesis is to replicate the Patricia Cortes' research, conducted in US, about the positive impact of low-skilled immigrated women on the female participation in the labor market. In particular, this paper wants to verify if there is a positive correlation between the survival rate of female-led innovative startups, the presence of low-skilled immigrated women and the presence of care services such as nurseries and early childhood services, in the North-west of Italy. The basic assumption that is made is that these childcare services provide native entrepreneurial women, who are also mothers, with the opportunity to take care of their children, while low-skilled immigrated women, similarly, give a greater support for household and family care. In this way, entrepreneurial women can devote more time to their professional careers and perform better. The first part of the thesis is dedicated to the background literature which gives an idea of what is meant by an innovative startup, which are the legal requirements in Italy to build a startup, the available sources of funding, and the key factors influencing the survival of startups. The second part focuses on the role of the female entrepreneur, in particular it is given a brief historical overview on how women have managed to establish themselves both socially and professionally over the years, up to the current condition in which women still face socio-cultural obstacles preventing them from taking on entrepreneurial risks; these obstacles are then analyzed and possible solutions to overcome them are suggested. Additionally, this section examines the current situation of immigration, and the impact that low-skilled immigrated women have on female participation in the labor market in north-western Italy. This is followed by the construction of the empirical model based on a linear regression analysis conducted by collecting data on immigration and childcare services from ISTAT (Statistical National Institution) and data on innovative startups and enterprises from AIDA (Computer Analysis of Italian Companies). After data collection a final database has been created using the software STATA and all variables that may affect the duration of a startup have been defined to carry out the empirical analysis.

## Contents

<b>Introduction</b> .....	7
<b>Background Literature</b> .....	9
<b>What is meant by innovative startup?</b> .....	9
<b>What is meant by Innovation?</b> .....	9
<b>Italian context</b> .....	10
<b>Legal requirements for an innovative startup</b> .....	10
<b>Covid-19: startups behavior</b> .....	11
<b>Financing Method for a startup</b> .....	12
<b>Key factors of Entrepreneurship and Firm Survival</b> .....	18
<b>Innovative Startups in Italy</b> .....	24
<b>Women and Professional Career</b> .....	25
<b>Historical framework</b> .....	25
<b>Current framework</b> .....	26
<b>Italian Female Entrepreneurs</b> .....	28
<b>What holds back Italian female entrepreneurs?</b> .....	29
<b>Possible solutions</b> .....	30
<b>The impact of low-skilled immigrants on female labor participation: the case of Northwestern Italy</b> .....	31
<b>Empirical Methods</b> .....	35
<b>Linear regression with a Single Regressor</b> .....	35
<b>Measure of Fit</b> .....	36
<b>Linear Regression with Multiple Regressors</b> .....	39
<b>Measure of Fit</b> .....	40
<b>Empirical Applications</b> .....	42
<b>Description of the Dataset and Variables</b> .....	42
<b>Italian Innovative Startups Database</b> .....	42
<b>International Immigrated Women resident in the North-west of Italy Database</b> .....	45
<b>Care Services Database</b> .....	46
<b>Control Variables database</b> .....	47
<b>Final Database</b> .....	48
<b>Construction of the model: Hypothesis and Research Objectives</b> .....	50
<b>Variables Description</b> .....	53

<b>Descriptive Statistics</b> .....	57
<b>Results</b> .....	67
<b>Version 1</b> .....	67
<b>Multiple linear regression model <i>without</i> control variables:</b> .....	67
<b>Multiple linear regression model <i>with</i> control variables:</b> .....	70
<b>Version 2</b> .....	77
<b>Log-Linear regression model:</b> .....	77
<b>Step-by-step Strategy</b> .....	84
<b>Discussion and Limitations</b> .....	90
<b>Conclusions and future improvements</b> .....	92
<b>References</b> .....	93

# Index of Tables and Figures

Figure 1: Graphic re-elaboration of the trend of the number of innovative startups over the years 2018-2022 .....	24
Figure 2: Share of women in STEM and non-STEM employment over the years 2016-2023 .....	27
Figure 3:anno_osservazione .....	57
Figure 4:provincia .....	58
Figure 5: Summarize donne_immigrate .....	59
Figure 6: Summarize tot_servizi_attivi .....	60
Figure 7: Summarize rate_donne_immigrate .....	60
Figure 8: Summarize tasso_servizi.....	61
Figure 9:duration_startup_guida_femminile when duration max 5 years.....	61
Figure 10:duration when duration max 5 years .....	62
Figure 11:summarize duration_startup_guida_femminile when duration max 5 years .	62
Figure 12: duration_startup_guida_femminile when duration not set max 5 years .....	63
Figure 13: summarize duration_startup_guida_femminile when duration not set max 5 years.....	63
Figure 14: Summarize residenti_laureati.....	63
Figure 15:Summarize unemployment_rate .....	64
Figure 16: Summarize value_added_per_habitant .....	64
Figure 17:Summarize area_km2.....	64
Figure 18: dipendenti.....	65
Figure 19: Summarize tot_debiti .....	65
Figure 20:summarize valore_aggiunto .....	66
Figure 21:Model's Fit without control variables- Version 1-condition a.....	67
Figure 22:Regression coefficients without control variables- Version 1- condition a....	68
Figure 23: Model's Fit without control variables- Version 1-condition b.....	69
Figure 24: Regression coefficients without control variables- Version 1-condition b. ...	69
Figure 25: Model's Fit with control variables- Version 1-condition a. ....	70
Figure 26: Regression coefficients with control variables- Version 1-condition a.....	72

Figure 27: Model's fit and Regression coefficients without the variable i.ateco_2007_2d – Version 1-condition a.....	73
Figure 28: Model's Fit with control variables- Version 1-condition b. ....	74
Figure 29: Regression coefficients with control variables- Version 1-condition b. ....	76
Figure 30: Model's Fit - Version 2-condition a .....	78
Figure 31: Regression coefficients - Version 2-condition a.....	80
Figure 32: Model's Fit - Version 2-condition b. ....	82
Figure 33: Regression coefficients - Version 2-condition b. ....	83
Figure 34: Step-by-step strategy – Version 2- condition a. ....	84
Figure 35: Step-by-step strategy –first step- Version 2- condition a. ....	85
Figure 36: Step-by-step strategy - second step – Version 2- condition a. ....	85
Figure 37: Step-by-step strategy- third step – Version 2- condition a. ....	86
Figure 38: Step-by-step strategy – correlation between variables- Version 2- condition a. ....	87
Figure 39: Step-by-step strategy – Version 2- condition b. ....	87
Figure 40: Step-by-step strategy – first step- Version 2- condition b. ....	88
Figure 41: Step-by-step strategy – second step- Version 2- condition b. ....	88
Figure 42: Step-by-step strategy – third step- Version 2- condition b. ....	89
Table 1: Immigrated women in the Italian North-west's regions - Source: ISTAT .....	33
Table 2: Number of Italian and foreign women employed as domestic workers - Source: INPS .....	34

# Introduction

The gender gap, that is the social and professional difference between men and women, is one of the main topics discussed today. Even though female entrepreneurship in the world has experienced great growth in recent years, with more and more companies founded and run by women, according to many researchers, the main cause of gender gap, in cultures across countries, stands in the vision of the female figure as linked to the domestic and familiar dimension and so, far from a professional career that could lead her to a prominent position in society and in work environment. This gap translates into lack of self-confidence, low expectations of a good balance between personal and professional life as well as a limitation for the economy growth of a Country. This research work identifies the main reasons that discourage women from becoming entrepreneurs and replicates Patricia Cortes' studies conducted in US to the North-west of Italy. Patricia Cortes has found that the presence of low-skilled immigrated women in a certain geographical area has a positive impact on the labor supply of highly skilled native women since they help native women in household and family care. For low-skilled women it is intended all women that don't have a study title. This concept is applied to our specific case but with some differences: the aim is to understand if in addition to low-skilled immigrated women, also the presence of childcare services in a specific geographic area of the North-west of Italy has a positive impact on the survival rate of a female-led innovative startup. To conduct the analysis the research focuses on the provinces of North-west of Italy since it a territory with a consistent number of innovative startups and a high number of immigrated women and childcare services. The availability of a large number of data allows to have a more precise result. The paper is divided as follows: in the second Chapter there is a review of the past literature to better understand the meaning of "innovative startup" and of "innovation", which is the Italian legislation under the foundation of a startup, how the startup founder can finance its business, a brief insights on how startups have managed the Covid-19 emergency, which are the factors that influence the survival rate of a startup and how many startups are active on the Italian territory. In the third Chapter it is framed the role of women in the workforce, particularly in the entrepreneurial environment. In addition, the factors that hold back women from pursuing an entrepreneurial career and the

immigration situation in the North-west of Italy are also analyzed, with the aim of taking the first steps through the application of Cortes' research. Chapter 4 outlines the econometric theory behind the empirical model to be constructed, namely a linear regression analysis with multiple regressors and then, in Chapter 5 all the datasets used to construct the final database are described. The latter is obtained by merging, with the software STATA, 3 main datasets: the one related to immigrated women aged 20 to 65 years old (considered as the working age range), the database containing all the childcare services active on the territory and a database containing a set of useful control variables to better estimate the model. Then it follows the formulation of the hypothesis to be tested, the description of all the relevant variables for the generation of the empirical model and the descriptive analysis which is about summarizing and understanding the basic features of a dataset. Chapter 6 will comment the results obtained while Chapter 7 and Chapter 8 will respectively address the limitations of the model and the conclusions drawn from this study.



# Background Literature

## What is meant by innovative startup?

Before delving deeper into the core of the thesis, it is essential to establish a clear understanding of exactly what the term “innovative startup” means.

According to Iacobucci, D., Iacopini, A., and Micozzi, A. in the book “*Le start-up innovative*”, 2014, “an innovative start-up is a new business that brings significant innovation to the market. This innovation may relate to the products and services offered, or new production methods for existing products and services, or organizational innovations such as a new sales formula”. That’s why governments try to facilitate their establishment since they are considered as a key element to improve the economy and to create new jobs.

The concept of startup extends far beyond and as outlined by leading figures in the field, we observe that a startup:

- “Is a temporary organization looking for a replicable and scalable business model” (Steve Blank, 2012).

This concept of temporariness is linked to the idea that a startup has limited duration in time; in fact, as scalable a startup is able to grow exponentially and in a short time should turn into a large enterprise. In a few words, “it is designed to grow fast” (Paul Graham, 2013).

Instead, for repeatable business model is meant that the logic according to which a startup organizes and offers its own products and services can be replicated in different countries, in different contexts and times, without the need for substantial modifications.

- “A human institution designed to deliver a new product or service under conditions of extreme uncertainty” (Eric Ries, 2011).

## What is meant by Innovation?

More broadly innovation means: “introduce new ways of designing, producing or selling goods or services; it basically means creating a positive change in the existing state of things or, in relation to the literal meaning of the word, alter the order of things established to make new things” (SardegnaImpresa, 2022).

Joseph Schumpeter in "Theory of economic development" (1934), believes that innovation is a critical dimension of economic change and for this reason he defines it as a "creative destruction". According to the author sometimes the innovation process can profoundly influence some economic sectors, and this forces companies to adapt in order not to risk bankruptcy; to quote again Eric Ries, innovation and uncertainty are strictly correlated since innovation is “inherently risky”.

It is possible to say then that determining the true extent of innovation within a new company could be challenging given the uncertainty surrounding the market potential of a new idea during the early stages, moreover nowadays innovation is increasingly based on the application of scientific expertise alongside the innate intuition and creativity of individual entrepreneurs.

Innovation is what drives a startup, and it lacks significance if it not showcases its worth in the marketplace and it is not recognized by potential buyers.

## **Italian context**

Italy has recently introduced a series of measures to foster innovation, with a specific focus on innovative startups. The main object is to increase sustainable growth, technological development, entrepreneurship and employment, attracting talent, innovative enterprise and capital from abroad. Innovative startups are formally acknowledged under legislation (221/2012, conversion law of Decree law 179/2012 called Decreto Crescita 2.0) and are registered in a special section of a register held by the Chamber of Commerce set up by Law 221/2012.

The Italian law considers innovative startups as key points of Italian industrial policy since it is viewed as “young company with a high technological content, with strong growth potential” (Ministero delle Imprese e del Made in Italy).

## **Legal requirements for an innovative startup**

According to law, startups can be a limited liability company, established in Italy, or in another EU country but with a branch in Italy and can access the status of “innovative” only if they respect the following requirements (*Scattoni, Lombardi, Pini, Turi – Innovative startup localization determinants and origin: “A Rome city case study”, 2019*):

- a) have been set up no more than 5 years ago;
- b) have their main business center in Italy;
- c) have a yearly production value not in excess of 5 million euros after their second year of business;
- d) non-profit status
- e) have as an exclusive or prevalent object clause development, production and trade in innovative goods and services of high technological value;
- f) not originate from business split-ups or mergers;
- g) fulfil at least one of the following sub-requirements:
  - g1) R&D expenditures greater than 15% in comparison with the highest amount between the cost and the total production value;
  - g2) one third of the workforce (employees and independent workers) with a PhD or studying for a PhD or two thirds of the workforce with tertiary education;
  - g3) a startup owner with sole rights (inventions, processing software, etc.).

### **Covid-19: startups behavior**

The global health systems have faced an unparalleled strain due to the COVID-19 pandemic and the measures taken to control the spread of the infection have led to an economic downturn, suddenly stopping a significant portion of economic activity; that's why, in response to Covid-19 emergency, Italy has adopted new measures including (Ministero delle Imprese e del Made in Italy): Grants to purchase services for the development of innovative enterprises, support for Venture Capital, tax credit in R&D, extension of the period of residence in the special section of the Commercial Register and others in order to sustain and help startups to survive. Alessandra Luksch, Director of the Observatories Digital Transformation Academy and Startup Intelligence of the Politecnico di Milano - says that in such a dramatic context, the startup ecosystem has shown some ways to face the emergency, by fielding skills, knowledge, patents, products, and new solutions. Interestingly, as reported by Osservatori.net. 2020. "*Effetto startup e COVID: nuovi modelli di business*"<sup>1</sup>, many startups initially have prioritized crafting remedies for the repercussions of the global health crisis, rather than solely

---

<sup>1</sup> Available at: <https://www.osservatori.net/it/ricerche/comunicati-stampa/effetto-startup-covid-nuovi-modelli-business>. Osservatori.net is an official and licensed platform that publishes reports and press releases on various topics, including start-ups and technological innovations. It is run by the Politecnico di Milano and offers content based on official research and studies.

focusing on financial gains. The 63% of high-tech startups in Italy took proactive measures to aid during the emergency, such as organizing fundraising campaigns for charitable contributions, introducing innovative products or services, and offering complementary solutions. Moreover, 30% of high-tech startups adapted their business strategies to face the pandemic, primarily to address emerging market demands. Therefore, as it is clear from the report "Situazione e prospettive delle imprese dopo l'emergenza sanitaria COVID-19", drafted by ISTAT, Covid19 and the consequent crisis have brought difficulties, but also positive incentives to entrepreneurial activity and have not stopped the birth of Italian startups.

## **Financing Method for a startup**

During the startup phase, one of the main challenges is to find the initial capital to start operations, to turn an idea into a concrete reality. The funding methods for startups are different and can be adapted to the specific needs of each business project. They vary depending on the stage in which the startup is in that specific moment, and they play a key role in determining the future path and growth of the company. Exploring available financing options requires a deep understanding of the financial needs, market potential and long-term growth strategies. Now, let's look at the main funding approaches for startups, analyzing pros and cons and their distinctive features.

### ***Bootstrapping***

The first way to finance a business is to have internal equity useful to cover the initial costs; "bootstrapping" in fact, means, self-funding the business. More in detail, according to Investopedia (2017) "when an individual bootstraps, they rely on money other than outside investments. An individual is said to bootstrap when they attempt to establish and build a company from personal finances or the operating revenues of the new company". Start the business using only personal finances can be advantageous as it allows the entrepreneur to have full control over financial decisions and remain independent from investors or lenders who might withdraw their support. However, this financing method can also impose some financial risks on the entrepreneur shoulders and might not ensure sufficient capital for the company to achieve success.

### ***Family, Friends, and Fools (3F)***

Unfortunately, bootstrapping on its own is not enough, that's why entrepreneurs often engage individuals within their close circle of reference which include friends and relatives, to obtain additional funds needed and support the source of capital for the startup. It thus means money borrowed from close people, without interest charges and without a fixed time for repayments. Investors and bankers consider it "patient capital" (Ministero dell'Istruzione, dell'Università e della Ricerca. 2020. "*Scheda Teseo – Strumenti per finanziare una startup*".), which is money that will be repaid later, when the company's profits increase. Thus "3 F" strategy can be beneficial as these investors are personally related with the founders and have confidence in their entrepreneurial abilities, so basically, it's easier to convince them. However, family and friends rarely have high capital to loan, and it is not excluded that who's lending money he's doing it because of a personal interest in demanding for equity participation in the new company. As with most strategies, also this approach carries intrinsic risks since combination of both financial investment and human relation may at times create conflicts of interest. Therefore, it is crucial to establish a formal written document to point out all aspects of involvement in the collaboration, and to prevent misunderstandings whether it involves loans or gifts. The combination of the 3F approach and bootstrapping serves as a valuable funding strategy to navigate the initial phase when the startup is not yet generating enough revenue to cover its operating costs. Despite this, to achieve quicker business growth, it often becomes necessary to secure further financial resources by attracting additional investors.

### ***Incubators and Accelerators***

Incubators are not properly a mean for financing a startup; instead, they are designed to assist entrepreneurs in developing and refining their business models to make them replicable and scalable. Their role is that of support for securing investments, enabling the startup to collect initial data to validate their idea, these data are then shared with potential investors. According to the European Commission, a business incubator is "an organization that accelerates and makes systematic the process of creating new businesses by providing them with a wide range of integrated support services that include the physical spaces of the incubator, business development support services and integration and networking opportunities". In Italy, instead, the D.L 179/2012 (Decreto

Crescita 2.0), in article 25, paragraph 5., introduces the concept of certified incubator for innovative startups, describing it as "a capital company, also in cooperative form [...] that provides services to support the creation and growth of innovative startups". The duration of the incubation period varies based on the specific startup, but it generally does not exceed 36 months. Incubators can be differentiated looking at the different economic requirements and types of companies that gradually emerge in the market. First, it is possible to classify incubators in public and private. Public incubators are such because are established through public sector initiatives and they include entities like the Business Innovation Center (BIC), which offers fundamental services to companies and University Business Incubators (UBI) that focus on transferring scientific and technological knowledge from universities to the business sector. While, among private incubators we find those established by an existing private entity, typically a large corporation, known as Corporate Private Incubators (CPI), and those founded by groups of people, called Independent Private Incubators (IPI). Both types are intended to facilitate the growth of businesses by providing resources during the initial stages of the development of the business model and throughout the subsequent phases of formation and education. Instead, for what concern startup accelerators are legitimate businesses and so, being real companies, they are profit-oriented entities and require payment for their services. Generally, such a payment is given in the form of equity: in exchange for the services provided, the accelerator obtains a share of ownership in the startup typically ranging from 5% to 15%. Startup accelerators also provide the so-called accelerator programs in order to give to startup both financial and technical support during the early stages. These programs offer valuable services and funding opportunities, and they typically have a shorter duration compared to incubator programs, lasting between 1 and 6 months. As mentioned before, accelerators represent a valuable opportunity to raise a startup by granting access to essential resources that might not be obtained otherwise such as office space, mentorship, as well as legal and financial support. Moreover, an important advantage of accelerators is their capacity to accelerate the transition of a product from the idea stage to market within a short timeframe. Nevertheless, there are potential disadvantages to take into account, such as the possibility of losing some control over the business since, given that accelerators obtain equity stakes in the companies they assist, as discussed earlier, it might be

necessary to surrender some degree of authority over critical decisions. Another potential disadvantage of accelerators is the strict deadlines often set to complete the project within a limited period to facilitate a rapid market launch.

### ***Business Angel***

A business angel is a private individual who invests personal capital in an innovative startup and in exchange for this investment, receives a shareholding in the company.

An angel investor offers more than just financial support: they contribute with valuable skills, networks, and experience for these new businesses that have an innovative idea but don't have the adequate resources to accomplish it. Therefore, they differ from venture capital and investment funds precisely because these latter provide capital to already mature and developed companies. Business Angels are able to invest their capital very quickly, no lengthy bureaucratic procedures are required, usually they invest in small projects of which they know and understand the target market, for example Facebook and Uber, were born thanks to the initial investments of these kind of investors. Although angel investors can provide significant benefits to startups, securing such funding is a challenging process that involves certain risks that may impact the business's future. One of the main challenges is about finding the right angel investor; it often requires approaching numerous potential investors before identifying those who are both interested and suitably qualified to invest in the startup. Additionally, accepting angel investment means selling part of the business to investors, which means that part of the capital and control is being given up. Another challenge of angelic financing is to align their objectives and expectations with those of investors who are active partners in the company and therefore participate proactively in strategic decisions.

### ***Bank loans***

Bank loans are one of the most used means through which entrepreneurs finance startups and this is due to several factors. The primary benefit of bank financing is its reputation as a reliable and established funding source. Banks have extensive experience in collaborating with startups and small businesses and they generally provide competitive interest rates compared to other funding sources. Additionally, they offer various financing options, such as loans, lines of credit, and credit cards and the legal supervision from the government, ensure that banks follow set principles and policies.

Meanwhile, the principal drawback of bank financing lies in the stringent requirements to be met for obtaining a loan or credit line since the application process is often time-consuming and complicated. For small and medium-sized enterprises, banks typically operate several evaluations before approving a loan and typically, these assessments include determining the company's potential to generate future income to repay the principal and interest, as well as requiring collateral (such as real estate, deposited securities, or guarantees) to secure the loan in case the company defaults on its repayment obligations. Banks are conservative even regarding new companies since proving their solvency is more difficult compared to experienced companies that have financial statements to disclose, that's why gain loan approval becomes more difficult for new startups. Consequently, it is recommended for a startup to submit a well-developed business plan when approaching a bank, as this can serve as a demonstration of their potential and increase the chance of obtaining the necessary funding.

### ***Business Competitions and Calls***

Another avenue for financing a startup is participating in the so-called business competitions organized by public or private entities at regional or national levels. These competitions are designed for young entrepreneurs, providing startups with a platform to demonstrate their value in a competition with other emerging companies. These contests not only offer an excellent opportunity to "test" the feasibility of the startup's idea, typically submitting a business plan or elevator pitch, but they also serve to improve startup capital, acquire skills, receive advice and services, gain visibility, and establish a network of potential investors. To reduce business risks and minimize the use of personal capital, startups can take advantage of public funding opportunities offered by government entities (such as the European Union, national governments, and regional authorities), as well as foundations and companies through the publication of calls for startups. These initiatives are designed to support specific target groups (including young people, women, and the unemployed) or new business projects and do so through subsidies or non-repayable grants. To obtain these funds, it is necessary to meet the criteria specified in the calls and submit applications within the deadlines. Other types of subsidies include guarantees provided by institutions that assume part of the risk on behalf of companies and vouchers with a specific monetary value that can be used for training or consulting services. It must be said that there are no subsidies that



fully cover the costs of starting a business and, in addition, the period of validity of contributions is often very short and disbursements are made after the expenses have been paid.

### ***Equity Crowdfunding***

Equity crowdfunding is a form of financing that allows to a large number of investors to finance startups by paying an economic contribution of different magnitude, in order to encourage the development of a new project in exchange for shares of the same companies. All this happens through authorized web platforms that allow, even the companies not present on the stock exchange, to be able to buy financial instruments bypassing the long bureaucratic steps that would concern them. It is an approach of investment and capital raising that over the years has taken more and more hold and that's why in Italy has arisen the need to develop a specific regulation. Italy was the first country in Europe to introduce an ad hoc discipline for equity crowdfunding through the issuance of D.L n. 179/2012 (Decreto Crescita bis), which however restricted the field only to innovative startups. Over the years through a new Decree-Law, precisely n.50/2017, the equity crowdfunding was opened to all types of companies and in 2013 was also published a regulation entitled "Consob<sup>2</sup>. 2013. Regulation on the collection of risk capital through online portals" which establishes a series of rules in order to regulate some aspects of the phenomenon and to create a reliable "environment" that can create confidence in investors. One of the main objectives of regulation is also to protect investors as this model is exposed to the risk of fraud, while other potential risks, concern the dilution of the company among a large number of investors, the company may not be successful, and investors may lose all their money and it may also happen that the company is unable to meet its financial obligations to investors.

### ***Venture Capital***

It is possible to understand the role of the Venture Capitalist by making a comparison with the Business Angels investors discussed in the previous paragraphs. This suggestion is proposed by Ministero dell'Istruzione, dell'Università e della Ricerca. 2020. "*Scheda Teseo – Strumenti per finanziare una startup*" which highlights how Venture Capitalists, unlike Business Angels, are not people who invest their capital, but

---

<sup>2</sup> Consob is the acronym of Commissione Nazionale per le Società e la Borsa (National Commission for Companies and the Stock Exchange).

are made up of a group of investors, namely Venture Capital companies. These companies raise money from pension funds, corporations, private investors with high assets and subsidies. Compared to the Business Angels, which offers guidance for startups, the Venture Capitalist is limited to the strategic sharing of the business, often placing one of its members on the Board of Directors of the subsidiary thus becoming a controlling figure and participating in operational decisions. The objective of the Venture Capitalists is to take a shareholding position in the startup and help it realize the project and gain profit from the future sale of the stake. Usually, startups are evaluated on the basis of three elements: a startup needs to have a consolidated team with a precise focus, it wants to commercialize an innovative product and has conducted a careful market analysis. In this way the startup will obtain from the use of Venture Capitalists, international networks, know-how and of course money; at the same time the greater the injection of money into the company by the Venture, the greater the control over the latter will be. Italian Venture Capitalists are mainly small companies and most belong to non-financial services and they differentiate their holdings by investing 29.6% in the software sector, 28.4% in the industrial sectors and 22.0% in research and development.

### **Key factors of Entrepreneurship and Firm Survival**

Before analyzing in detail the key factors of entrepreneurship, it is useful to examine the concept of entrepreneurship through a brief review of the economic literature. This will then be followed by the identification of the determinant factors of the firm survival.

We refer in general to firm because the aim is to analyze this concept from a general point of view but all the discussion that will follow is strictly related also to startups.

Over the years, scholars have developed different theories and measures to describe entrepreneurship since it represents a very broad concept whose definition may vary depending on the context of interest. Joseph Schumpeter is one of the founders of the economic theory of entrepreneurship and in his essay *"The Theory of Economic Development"* published in 1934, defines entrepreneurship as the "ability to introduce innovation into production processes by creating new combinations of resources and new products." This process, like any other economic activity, is subject to a certain level of risk as also stated by Donald F. Kuratko in *Entrepreneurship: Theory, Process,*

*and Practice*" (2016): "Entrepreneurship is the ability to identify opportunities and take risks". It is important to stress that the main driver of entrepreneurship is generate economic value, which means generate profit and growth (Baumol, W. J. 1990. "*Entrepreneurship: Productive, Unproductive, and Destructive*") not only at the level of the individual enterprise but also at the level of the social system. It is, indeed, historically recognized that there is a positive correlation between entrepreneurship and society, since an increase in entrepreneurship also translates into higher productivity, more employment, growth of gross domestic product, increase in innovation and technology processes.

Let's go now deeper into the concept of determinants of entrepreneurship and of firm survival. In both cases these factors are multiple and complex since they include economic, personal, social and institutional aspects.

#### ***Entrepreneurship determinants***

Audretsch, D. B., Thurik, A. R., Verheul, I., & Wennekers, S. (2002), in the book "*Entrepreneurship: Determinants and Policy in a European-US Comparison*", they propose an analysis on the various determinants of entrepreneurship by combining economic, cultural and institutional factors. It is interesting the subdivision made to distinguish such determinants: we have determinants belonging to the demand side of entrepreneurship which focuses on the so-called *product market perspective* so it contains all that variables that allows entrepreneurs to enter a new market such as population characteristics and its related value added, the levels of social and technological development and the determinants belonging to the supply side which instead focuses on the *labor market perspective* and contains variables such as market demand for goods and services. All these variables are intended by the authors as *macro* variables and as such, together with strictly personal characteristics and abilities, they push individuals to take or not the enterprise risk.

#### ***Firm Survival determinants***

In the business lexicon, the term "firm survival" indicates the company's ability to remain active over time, through changes in leadership, shifts in resource providers, and

evolving internal and external conditions of its management<sup>3</sup>. In an increasingly dynamic and competitive global economic environment, survival represents a significant challenge for companies that, over the years, have developed a set of strategies to continue with their operations. Below, we will explore which are the factors that contribute to success and, consequently, to the long-term durability of firms.

Several studies have detected the factors that affect the likelihood of firms surviving in the market and from the past literature we see that they can be classified considering two macro categories: Firm Specific factors and Industry Specific factors which are further divided into several subcategories.

### **Firm Specific Factors**

With the term “Firm Specific factors” are intended all that factors that are difficult to transfer or copy by external parties; among these we found individual and organizational factors.

1. **Individual determinants** are all the aspects strictly related to the personal characteristics of the entrepreneur such as his/ her personality, values, professional skills, motivation for growth of the business and personal background. According to Dobbs and Hamilton (2007), it exists a strong correlation between the survival rate of a company and the personal characteristics of its founder since its past experience and its soft-skills are fundamental for company’s affirmation and consolidation. More in detail, in addition to Dobbs and Hamilton (2007) also the authors Zhou and de Wit (2009) argue that a high level of education and experience in the field are the main drivers of the longevity of a company but Wiklund and Shepherd (2003) have demonstrated that this is true only if education and experience are matched with the motivation to grow which is view as a the driving force behind everything.

2. **Organizational determinants** refer to all the internal characteristics and internal dynamics that takes place inside a company. Among the most important we have:

- **Size and Age:** according to various economists (Evans, 1987; Hall, 1987; Dunne and Hughes, 1994) size and age of the firm have a great impact on the duration of the firm. More in detail, they recognize that, large and mature

---

<sup>3</sup> Source: [https://www.treccani.it/enciclopedia/sopravvivenza\\_\(Dizionario-di-Economia-e-Finanza\)/](https://www.treccani.it/enciclopedia/sopravvivenza_(Dizionario-di-Economia-e-Finanza)/)

enterprises tend to be more stable on the market thanks to economies of scale, to their reputation and capability in facing crisis, but sometimes this kind of firm are also rigid and less likely to innovate with the risk that their products can become obsolete. On the contrary, smaller and younger enterprises are more flexible and are able to better adapt to innovation trends but at the same time they can have a limited productive capacity and limited financial resources. To conclude we have that, according to Elena Cefis, Orietta Marsili. (2006), “the ability to innovate increases the survival probabilities for all firms across most industrial sectors. [...] the innovation premium is highest for small and young firms, which are those at the greatest risk of failure: small, young firms that innovate have a 23% greater chance of surviving than those that do not.”.

- ***Human capital:*** Human capital refers to the capital of each individual worker in terms of personal knowledge and training, attitudes and inclinations, skills and abilities, and even emotions. This capital, in fact, has a concrete and quantifiable value. This is why human capital represents, for businesses, and in particular for startups, an asset as important as financial capital. In fact, good investment and proper human capital management enable companies to achieve competitive advantage in the market and withstand market pressures even in times of crisis, according Birley & Westhead, 1990; Chandler & Hanks, 1994 “human capital of the total workforce plays a more determined role when compared to the entrepreneur alone”.
- ***Organizational structure:*** If the knowledge and preparation of human resources are crucial, equally important is the way in which this type of resource is organized. The careful distribution of tasks among employees, their cohesion and coordination are core for a profitable and durable company. In support of this Anna Brattstrom (2019) proposes a set of characteristics that a team of employees should have: first of all, synergy so work together for the same objective, flexibility in order to be able to

manage changes and homogeneity although the author also recognizes that having an heterogeneous team can equally bring advantages to the firm.

- ***Business strategies:*** in order to construct a successful and profitable firm able to satisfy all stakeholders (customers, suppliers, employees) it is fundamental for the leader to set clear business goals. The choice of an effective strategy is strictly related to the entrepreneur's skills because, as highlighted earlier, a strong growth objective from the founder side is one of the factors that most influences the survival rate. Among various business goals, developing an innovative and competitive product, meaning a product with features not yet seen on the market and with a competitive price is certainly a core objective since such a product acquires high appeal among customers and can lead to an increase in the market share of the company, greater visibility, higher revenues, and thus, greater growth and longevity (Lee, 2010). The second element is the high technological and digital content of the product or of the service which is reachable by direct investments in R&D. In this way, companies, by applying technology to their processes and products, become more efficient than less technological competitors and creates a competitive advantage (Morgan and Hunt, 1999) that is difficult to bridge even in the long term, given the amount of investment and highly specialized skills needed to overcome the gap. To further enhance their competitive advantage, companies can capitalize on research and development investments through the acquisition of patents. The patent is becoming an increasingly central tool as it has taken on growing strategic significance. In fact, owning a substantial portfolio of patents provides the company with greater protection when negotiating the terms of use of a certain technology with other companies and represents an important indicator of how a company is pushing towards innovation. Third element that the entrepreneur should consider in developing business goals is the sustainable approach, which in recent years has become a key factor because of climate change and great environmental problems. Infact, who succeed in developing environmentally friendly solutions, most of the time, gains

credibility and leadership in those sectors which are considered highly strategic.

### **Industry Specific Factors**

3. *Environmental determinants* indicate all the external factors and also the characteristics of the industry in which the company performs such as market dynamics, competition, regulatory environment. Below there will be briefly explained the factors that, according to literature, are the most significant.

- **Entry barriers:** with entry barriers it is meant all the obstacles that make it difficult for new companies to enter a given market<sup>4</sup>. The government in which the company arises has a strong influence on the success of a business and it represents the principal barrier to entry since it can provide financial support to new firms through loans, guarantees and government equity (the so-called “government aid programs”) but it can also hinder its growth through fiscal policies, tariffs or trade restriction policies. Among the other barriers to entry, we found technology challenges, patents, high investments to enter the market and licensing requirements. It is intuitive that, to make sure that the company can enter and survive over time, those barriers need to be low.
- **Intensity of competition and market trends:** Another important aspect that has an effect on survival is the analysis of competitors and potential customers: knowing the products offered by the competition, their characteristics, the price set, allows the company to improve and enhance its service or product, and to better position itself on the market. Instead, knowing the needs of customers, listening to their opinions helps develop a "tailor-made" product that meets their needs and allows the company to achieve the so-called brand loyalty. This is reached by doing a deeper market analysis to understand the latest market trends, how much customers are willing to pay for that specific product, revenues forecasts and so on.

---

<sup>4</sup> Corporate Finance Institute (CFI). n.d. "Barriers to Entry"  
<https://corporatefinanceinstitute.com/resources/economics/barriers-to-entry/#:~:text=Barriers%20to%20entry%20are%20the,or%20education%20and%20licensing%20requirements.>

## Innovative Startups in Italy

Analyzing the Annual Report to the Parliament about the state of implementation of the policies in favor of innovative startups of the year 2023 by Adolfo Urso, Minister of Enterprises and Made in Italy, it is reported that in Italy the number of innovative startups is growing and precisely, at 31 December 2022, the number of innovative startups regularly registered in the Business Register amounted to 14.264, +1.4% compared to 2021. The document, in addition to illustrating the trend of the Italian innovation ecosystem in 2022, also contains some updates to the third quarter 2023 which show, in the first 9 months of 2023, a slight decrease in innovative startups of -3.6% compared to 2022. The greatest loss was recorded in the Lombardia region (-209 units) but, despite this, it is the first Italian region for number of startups. Another significant reduction occurred in the Lazio region, where registrations were reduced by 101 units, but at the same time startups in Liguria increased and, generally in most of the regions of Southern Italy, in particular Campania, Abruzzo, Puglia, Calabria, Sicily, and Molise, have contributed to a growth of +2.7% compared to the end of 2022.

The following graph shows the growing trend of the number of innovative startups during the years 2018-2022 (*Source: Data processing on Infocamere*):

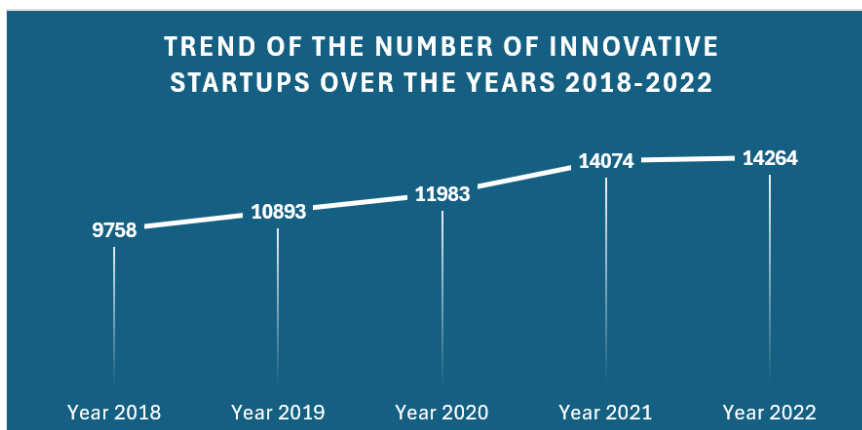


Figure 1: Graphic re-elaboration of the trend of the number of innovative startups over the years 2018-2022

Geographically, almost 35% of innovative startups are located in north-western Italy, with Lombardy dominating with 27.6% of the national total, and Milan continues to remain the most innovative city with the highest number of startups, 2.831 to be exact. For Central Italy instead, Lazio leads the group with almost 13% of the national total and in Rome there are 1.659 companies. In southern Italy, however, Campania is the



leading region with over 1.400 startups operating in the territory with 715 only in the city of Naples. Regarding instead north-east Italy, although the total number of enterprises is of 2.500 this territory has recorded a decrease of the -1,7%. For what concern economic sectors, the distribution of enterprises remains the same compared to 2021, in particular more than half of the startups deals with Information and Communication Services (7,283 enterprises) and among these, 5,695 are involved in software production, IT consultancy and related activities. On the other hand, 23.1% of the total deals with professional, scientific, and technical activities, and of these more than 2,000 startups operate in the field of scientific research and development. The contribution of manufacturing activities such as the manufacture of computers, manufacture of non-electrical or electrical and household equipment and the manufacture of mobile or portable machinery is also significant. The report also describes in detail the situation regarding some specific categories of innovative startups: those with young, female and with international prevalence. The category of innovative young startups namely all those companies whose participation in control and ownership is mainly held by people under the age of 35 years and the category of startups with female prevalence that is, companies mainly owned by women are both increasing, compared to 2021, while the portion of startups with international prevalence remains low.

## **Women and Professional Career**

### **Historical framework**

Historically it is known that in Italy as in the rest of the world there was a clear gender distinction in the world of work and that women's activities had less social recognition. Mainly women were devoted to home and family care but at the same time they were engaged in other activities without receiving an income, on the contrary men were devoted to economically and socially recognized productive tasks. In the recent years, the status of women in the labor market has improved and is closer to that of men. The main social and cultural events that led to the first steps towards a radical socio-cultural transformation were: industrialization during which women also become part of the workforce of industries, especially the textile industry, even if with a lower wage than men, follows the outbreak of the Great War, during which women experienced for the

first time a detachment from males and began to take the first steps towards emancipation and a planning of their rights: they began to search for work to support the family and managed the properties independently. Since the end of the Second World War, Italian women began to fight for their rights and their freedom: On 2 June 1946 they obtained the right to vote and in 1948, with the entry into force of the Italian Constitution, several fundamental principles on equal rights for men and women were sanctioned. Since 1968 the second wave of feminist movement has spread in Italy (the first feminist claims date back to the nineteenth century) in order to achieve equal rights, the respect of the same and the affirmation of its own feminine identity and in 1977 finally comes into force the Law n. 903 on the equal treatment of men and women in matters of work.

### **Current framework**

Although important steps forward have been made during past years, the stereotype that sees the woman best suited to fill roles related to home and children care has not yet been completely overcome. Inequalities still concern labor market participation, equal salary, achieving a balance between work and private life, and presence of women in important corporate and managerial positions. In more detail, gender equality in the world is mainly undermined by some dynamics, explains the European Commission, such as *the glass ceiling effect*, that according to Investopedia can be viewed as those "invisible barriers that marginalized people, such as women and minorities, encounter when they seek career advancements". Indeed, fewer than one in ten CEOs of top companies are women and the profession with the largest disparities in hourly earnings in the EU is that of managers: women earn 23% less than men. This pay difference based on gender is commonly called *gender pay gap*, but women often also face the issue of *gender biases* or *sectoral segregation*. Professions dominated by women are often consistently undervalued and around 24% of the gender pay gap is due to the high concentration of women in relatively low-paying sectors, such as care, health, and education. Gender gap is a real problem of today's society, and it is confirmed by the fact that "*Achieving gender equality and empowering all women and girls*" is the fifth of the 17 Sustainable Development Goals, defined in 2015 with the ONU Agenda 2030. According to the European Commission, the world can be truly sustainable only when all the women of the planet are guaranteed equal access to education, the right to have a

decent work, fair representation at the heads of States and fair participation in political activities, without any gender differences. By consulting data from the European Commission's website, it is learned that the gender pay gap in the EU is still high: it stands at 12.7 % in 2021 and has only changed minimally over the last decade, it means that women earn 13.0 % on average less per hour than men and as highlighted by the Global Gender Gap Report 2023, “despite Europe has the highest gender parity of all regions at 76.3%, it is projected to attain gender parity in 67 years”. The Global Gender Gap Report also takes into account a particular index called *Global Gender Gap Index* which measures the progress towards gender equality on a scale of 0 to 1, where 1 represents full equality and 0 full disparity. The report analyses the current state and evolution of gender equality considering four main sectors: Economic Participation and Opportunities, Educational Attainment, Health and Survival, Political Empowerment. For what concern Italy, it drops by some positions in the world ranking: from 63<sup>rd</sup> in 2022 to 79<sup>th</sup> (out of 146 countries) in 2023 and also, the level of political participation that goes from the 40<sup>th</sup> position to the 64<sup>th</sup>. The good news is that the report shows that the level of female participation in the economy of the nation improves, and Italy is among the European countries in which companies are more active for the promotion of Diversity & Inclusion (D&I) initiatives in order to bridge the gap within the working environment. Another interesting aspect to underline is that despite the increasing number of women who choose scientific studies, fewer women embarking on careers in these areas, in fact, the employment of women in the so-called STEM professions (Science, Technology, Engineering, and Mathematics) is just a bit over 40%, in the NON-STEM professions, the percentage of employed women is more than half with approximately 53% (*Source: LinkedIn Economic Graph*).

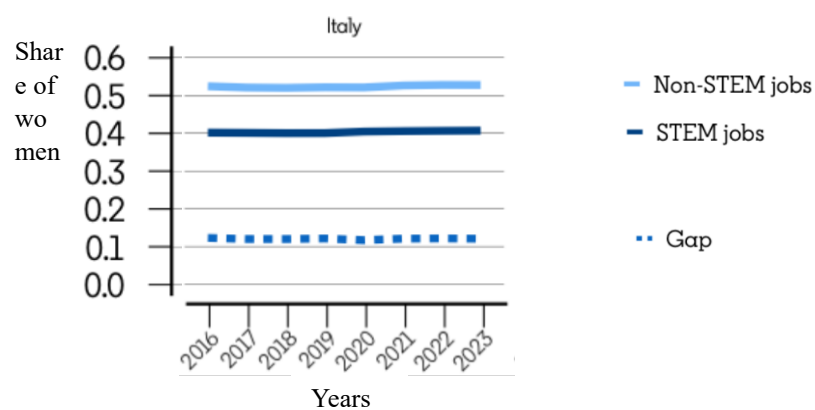


Figure 2: Share of women in STEM and non-STEM employment over the years 2016-2023

## Italian Female Entrepreneurs

In the previous paragraphs it has been repeatedly pointed out that for women to play top roles and enjoy the same opportunities offered to male colleagues, it is a fundamental right and also an important goal for the development of an increasingly fair, democratic and sustainable society. Unfortunately, gender stereotypes that limit the opportunities of women entrepreneurs still persist today, but female entrepreneurship in the world has experienced great growth in recent years, with more and more companies founded and run by women, showing the same skills and competences as male entrepreneurs. According to the report of *Unioncamere* on female entrepreneurship, presented in March 2024, there are 1 million and 325 thousand female companies<sup>5</sup> registered in Italy (22.2% of the national total). Focusing the analysis on innovative startups, at the end of September 2022 there were 2.000 female-led startups recorded in the Register of Innovative Startups, 572 more compared to 2019 (+40% growth). More than 70% of these 2.000 female enterprises operate in business services (1.455), just over 15% in manufacturing (306) and 4.6% in commerce (91), while residual shares are active in other economic sectors. Women's innovation has its beating heart in four regions of Italy that concentrate more than 50% of the total of startups led by women: Lombardy (470), Lazio (263), Campania (204) and Emilia Romagna (143). The considerable increase in innovative women-led startups goes hand in hand with their growing commitment to information and communication services, financial and insurance activities, professional, scientific and technical activities, education, health and care that today represent almost 10% of the female universe that makes enterprise. Another important aspect highlighted by an ISTAT analysis is that female entrepreneurs have a lower average age (49 years) than their male colleagues (52 years), thanks the fact that the majority of women who decide to take the risk of enterprise are young women under 35. The increased presence of young women entrepreneurs helps to mitigate the gender imbalance and more in general lead to improvements in innovation and financial performance. Indeed, the study "*The mix that matters. Innovation Through diversity*", published by BCG (Boston Consulting Group) in 2018, shows that for the companies

---

<sup>5</sup> For those activities mainly led by women it is intended those cooperatives or partnerships with at least 60% of women among the members; corporations with at least two-thirds of shares and governing body members being women; individual companies that have a female owner, and self-employed workers with VAT.

that had a higher percentage of "gender diversity" have obtained 34% of revenues from product and service innovation in the years 2017-2019, compared to 25% of companies with fewer women.

### **What holds back Italian female entrepreneurs?**

Unioncamere shows that women's enterprises have 60% lower productivity than non-female enterprises and have a lower survival rate: three years after birth, 18% of women-led enterprises are closed compared with 14.7% of other enterprises and after 5 years, the probability of survival for a female enterprise is 72.1% against 77% of non-female enterprises. In 2023, women-led enterprises decreased by 11 thousand units (-0.9%), therefore, this paragraph will try to analyze the causes that hinder women entrepreneurs to do business or that lead to the closure of these enterprises. First of all, from the study conducted by the Women in Business Observatory. 2019. "*Rapporto sull'imprenditoria femminile in Italia.*", we learn that "the motivations that push Italian female entrepreneurs to start a business concern 20.5% the desire to do business, 21% the possibility of being creative, 47.5% the search for the work life balance and 44.8% the expansion of its activity". From the same source we also learn that Italian women who seek to pursue a professional career suffer, compared to European colleagues, a greater fear of failure. The issue of manage private life and work is among the most critical aspects since, for 21.9% of female entrepreneurs, it represents an additional load of effort to add to the work. Education and cultural background, in turn, weigh on fear of failure: a perception augmented by the fear of being, for 56% of the respondents, overwhelmed by bureaucracy. All these factors, lead to surrender early than the rest of Europe, among Italian female entrepreneurs, in fact, there is "*a not positive mindset, as well as a clear sense of inadequacy*". Although only 13.5% of female entrepreneurs point to gender stereotypes and prejudices as an obstacle to their own affirmation, a change of awareness in women is necessary for them to have more confidence in themselves: In 35.7% of cases, they admit that they are held back by the fear of failure, and over 4 out of 10 women perceive that they have more difficulty growing a business than men. Probably the reason of such fear and insecurity could also be linked to the contraposition between entrepreneurship and motherhood and indeed there are available data that support this thesis.

According to the *Confartigianato Women's Entrepreneurship Observatory*, despite the fact that Italian women entrepreneurs, many of whom are mothers, are the most enterprising in Europe, Italy is among the European countries with the lowest number of employed women with children. According to the results of the *World Values Survey (WVS)*, a large survey conducted every 5 years in 100 countries, Italy is the only country where 54.1% of respondents agree with the statement that a working mother "harms" preschool children (against an EU average of 30%), in addition Italy is the country with the highest percentage of agreement with the statement that if there is lack of work it is right to give priority to men (25.4%, compared to an EU average of 11.4%). Every year Save the Children publishes the report *"Le (mamme) equilibriste"* to highlight the consequences of parenting both on women and men. In 2023, 7 out of 10 dismissal validations concerned women, mainly mothers and in fact, or 41.7 % of women, the main difficulty lies in reconciling work and childcare due to the lack of care services, while 21.9% pointed out problems related to the organization of work. For men, however, the predominant motivation is professional: 78.9% said that the end of the employment relationship was due to a change of company and only 7.1% reported needs for childcare.

### **Possible solutions**

Identifying the causes of the gender gap in the labour market is important in order to identify policy interventions to bridge the gap. Numerous studies have confirmed that this problem requires concrete and targeted solutions and today, thanks to the research work of the Nobel Prize for Economics 2023 Claudia Goldin, the society has a deeper understanding of the socio-cultural origins of this disparity and the measures to be implemented to address it. In 2014 Claudia Goldin published in the *American Economic Review* the article entitled *"A grand gender convergence: Its last chapter"* where she asks what the last chapter should be to obtain gender equality within the employment sector. It states that the solution does not necessarily involve policy intervention or provide for an improvement in women's negotiating skills but that the way in which jobs are structured and remunerated should be changed. Goldin argues that the gender pay gap would disappear if companies did not have incentives to remunerate disproportionately those who work longer hours or in particular time slots like for example during weekends, holidays or evening hours. Another extremely interesting

article is what Goldin published in 2000 together with Cecilia Rouse entitled "*Orchestrating impartiality: The impact of "blind" auditions on female musicians*" where it is explained that initially the selections to hire professional musicians in American symphony orchestras were carried out in front of a jury, but starting from the 70s, the orchestras introduced blind auditions, in which all candidates performed behind a screen. The results show that, with this procedure, women are more likely to be selected, confirming that previously the female disadvantage was due to a discriminatory attitude. Thanks to previous studies and research it has been possible to identify several strategies that would support female employment and parenthood such as: enhancing the value of young people and their economic independence, adopting more effective welfare incentive policies, implementing legislative actions, and promoting a cultural shift to counter gender stereotypes related to motherhood. Additionally, strengthening care services such as nurseries, parental leave, childcare subsidies, and others would be strongly beneficial. This thesis aims to focus particularly on this last point by analyzing the studies conducted by the economist Patricia Cortes about the impact of the immigration on the labour market with a particular focus on the role of low-skilled immigration in facilitating the labor market participation of highly skilled women.

### **The impact of low-skilled immigrants on female labor participation: the case of Northwestern Italy**

In this section it will be analyzed more in detail the results obtained by Patricia Cortes's research work conducted in United States with the aim of replicate this study on the specific Italian case and it will be given a general perspective of the Italian immigration framework with a special attention on the low-skilled immigrated women and their impact on high-skilled native women labour participation. Let's take into exam her paper entitled "*Low-Skilled Immigration and the Labor Supply of Highly Skilled Women*" written in collaboration with José Tessada, and published in 2011, in which they investigate the effects of low-skilled immigration on the labor force participation of highly skilled women in the United States with a particular interest on how the presence of low-skilled immigrants, who are often employed in household services such as childcare and housekeeping, allows native highly skilled women to increase their labor

supply, to allocate more time to their professional careers, to focus on career advancement and search for jobs that require greater commitment but higher payment. As it is possible to notice, studies focus just on highly skilled women because it was shown that they benefit the most from domestic help with respect with women who carry out ordinary jobs. As a result, it is possible to say that foreign women helping native ones influence the survival rate of female led startups and more in general contribute to higher successful career and higher overall economic output. As highlighted at the beginning, through this thesis work, the aim is to apply Patricia Cortes's research work to investigate whether, even in Northwestern Italy, outsource domestic work to low-skilled immigrants has positive effects on the professional success of high-skilled Italian women, especially for those who are leading an innovative startup or participate in it. Before carrying out the statistical analysis based on Cortes' model, below will be analyzed the current situation of immigration in Italy with special attention to low-skilled women immigrants in north-west Italy.

ISMU ETS Foundation estimates that “on January 1st, 2023, foreigners in Italy are about 5 million and 775 thousand (equal to 8.6% of the total) 55,000 less than the same date of 2022 and the demographic balance shows a significant growth of the foreign population resident in Italy (+110,000 units)”<sup>6</sup>. Very interesting is the recent discovery concerning the *feminization of migration* that is explained by the Osservatorio di Politica Internazionale in the document called “Osservatorio quadrimestrale N.3 - 2023”. The term '*feminization*' of international migrations refers to the growing number of women actively participating in the migration process, but especially to the fact that they do so more independently than in the past, defining new migratory models built around the centrality of the female role. According to ISTAT data, the feminization of migration flows would coincide with the ever-increasing demand for foreign carers and housekeepers by Italian families. Since the '70s, women have emigrated to enter the labor market, predominantly in the service sector and today, women represent the majority of foreign workers employed in the home care sector, 553,085 are the foreign women employed in domestic works which corresponds to the 57,5% of the total. If

---

<sup>6</sup> Source: Agenzia Nova, 2023. “*Migrants: the number of irregular migrants in Italy will decrease in 2023*”. Available at: <https://www.agenzianova.com/en/news/migranti-nel-2023-diminuisce-il-numero-di-irregolari-in-italia/>



until the early 80s it was mainly the upper middle-class families who could afford it economically, in recent years, the need of women to enter the labour market has increasingly led high-skilled women to hire migrant women to do for them the domestic work. Immigrant workers have thus entered a labour sector that is not very attractive for Italian women, given the low pay, the impossibility of professional growth, the burden and heaviness of the tasks, in addition to the low social consideration and over the years they have replaced Italian women generally from rural areas of Northern Italy, the South and the islands, that used to offer these care services to local high skilled women. Just to give some numbers, according to Patricia Cortes (2022), the 17,2% of all female migrants are employed as housekeeper, babysitter or caregiver in US; in Italy, around 19% of immigrants offer same care services and on the contrary just 1% of Italian citizens are employed in this kind of, low-skilled works (ISTAT). Now turning to the specific case study, historically, the North-west of Italy<sup>7</sup> is the area in which more foreign immigrants and, consequently, more immigrants involved in domestic works are concentrated. According to ISTAT, there are 913.744 immigrated women in this area (below it is provided a table highlighting the immigrated women for each North-west's regions), and the most of them are employed as colf and as domestic carers.

<b>Regions</b>	<b>Immigrated women</b>
<b>Piedmont</b>	219.366
<b>Valle d'Aosta</b>	4.594
<b>Liguria</b>	77.131
<b>Lombardy</b>	612 653

*Table 1: Immigrated women in the Italian North-west's regions - Source: ISTAT*

---

<sup>7</sup> For North-west of Italy is intended the set of the following regions: Piedmont, Lombardy, Liguria, Valle d'Aosta.

Since it has not been possible to derive with precision the number of foreign women operating in the domestic sector in the northwestern Italy, below it is provided a table that shows the number of both Italian and foreign women employed in domestic works in the interested regions.

<b>Regions</b>	<b>Women employed as domestic workers</b>
<b>Piemonte</b>	58.657
<b>Valle d'Aosta</b>	1.593
<b>Liguria</b>	25.697
<b>Lombardia</b>	140.961

*Table 2: Number of Italian and foreign women employed as domestic workers - Source: INPS*

Not only Northwestern Italy is the area with the highest number of low-skilled immigrant women, but it is also the area in which innovative female companies are mainly located. More in detail the 30.2% of which 24.6% in Lombardy, Piedmont (4.2%), Liguria (1.4%) and Valle d'Aosta (0.2%). From the researches it is known that the turnover of the innovative female-led startups is in strong increase and this increment is probably due, in addition to a greater awareness and propensity of women to take risks in the business environment, also to the presence of these women who compensate for the lack of public services and help them in household chores, in childcare and elderly care. Therefore, it is possible to consider that Patricia Corte's thesis is, theoretically, valid also for the Italian case but, to assert this with certainty, it is necessary to verify it empirically and this is what will be done in the next chapter.

# Empirical Methods

In this chapter it will be provided an introduction of the econometrics' theory which is behind the statistical model that will be constructed, using the linear regression method, to verify if there is a positive correlation between the low-skilled immigrated women in the north-west of Italy, the presence of childhood services in the same area and the survival rate of startups headed by local high-skilled women. The aim is to give empirical content to theories treated in the previous paragraphs in order to obtain real world data and do accurate forecasts. The main source that will be used to elaborate this chapter will be Stock, J. H., & Watson, M. W. 2011. “*Introduction to econometrics*” (3rd ed.).

## Linear regression with a Single Regressor

Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. The simplest form of linear regression involves only two variables: a dependent variable and an independent variable, it is called *simple linear regression* and is described mathematically by a first-degree polynomial in one independent variable  $X$ :

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

Where:

- $Y_i$  is the dependent variable or “regressor”
- $X_i$  is the independent variable or “regressand”
- $\beta_0$  is the intercept (the value of  $Y$  when  $X$  is zero)<sup>8</sup>
- $\beta_1$  is the slope of the line (how much  $Y$  changes for one unit change in  $X$ )
- $u_i$  is the regression error (representing unobserved variables, factors other than  $X$  that affect  $Y$ . It also includes errors in the measurement of  $Y$ )

The intercept  $\beta_0$  and the slope  $\beta_1$  are the coefficients of the population regression line, also known as the parameters of the population regression line and since they are unknown, they need to be estimated starting from the data available from the samples.

---

<sup>8</sup> In some econometric applications, the intercept has an economic interpretation, in other applications, has no real-world meaning.

To estimate these coefficients, since it is a widely used tool in various fields such as economics, finance, social science, we use the Ordinary Least Squares Estimator (OLS) which selects the regression coefficients so that the estimated regression line is as close as possible to the observed data, where closeness is measured by the sum of the squared mistakes made in predicting  $Y$  given  $X$ . It means that OLS wants to find the best-fitting line through the observed data that minimizes the average squared difference between the actual values of  $Y$  and the prediction based on the predicted value.

The OLS estimators of  $\beta_0$  and  $\beta_1$  are:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{s_{XY}}{s_x^2}$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}$$

Where:

- $\bar{X}$  and  $\bar{Y}$  are the means of the two variables in the sample
- And knowing,  $\widehat{\beta}_0$  and  $\widehat{\beta}_1$  it is possible to obtain:
- The OLS predicted values  $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i, i = 1, \dots, n$
- The OLS residuals  $\widehat{u}_i = Y_i - \widehat{Y}_i, i = 1, \dots, n$

### Measure of Fit

After having obtained the OLS estimators and consequently the OLS regression line, also called the sample regression line or sample regression function, there's the need to check the acuity of the estimate, that is how well the statistical model approximates and describes the actual data.

The most common statistics to verify the goodness of fitting are:

#### ***The $R^2$***

$R^2$  is called coefficient of determination and it measures the fraction of variance of  $Y_i$  explained by  $X_i$ .

$R^2$  ranges from 0 to 1 and in detail:

- An  $R^2$  of 0 indicates a model whose predictive variables do not explain at all the variability of the  $Y_i$  around its average, in other words, there is no fit;
- An  $R^2$  of 1 indicates a model whose independent variables fully explain the variability of  $Y_i$  around its mean. That is, knowing the values of the independent variables it is possible to predict exactly what the value of  $Y_i$  will be, there is a perfect fit.

Usually, the greater the value of the  $R^2$ , better the ability of explanatory variables to predict the values of the dependent variable and so the more the model has a high predictive power. The  $R^2$  index evaluates how far individual observations deviate from the regression line and in general, if two regression models are constructed on the same data set, the model with the highest coefficient of determination will be the one that will have the least discrepancies between the observed and expected y values. However, this interpretation in some situations can be misleading. A model with a high R-squared value can be wrong.

While, especially in some fields of study, such as behavioral sciences, it is normal to observe  $R^2$  values below 50%, this does not mean that the regression model is not good but just that, by its nature, the dependent variable that has been analyzed depends on so many different factors, many of which have not been measured. The limit of having a low value of this index stands in obtaining not precise forecasts. On the other hand, a high  $R^2$  is a necessary but not sufficient condition to be able to make precise forecasts.

Now, the dependent variable  $Y_i$  can be written as the sum of the predicted value  $\hat{Y}_i$ , plus the residual  $\hat{u}_i$ .

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

The  $R^2$  can be written as the ratio of the explained sum of the squares (ESS), that is the sum of squared standard deviations of the predicted values of  $Y_i$  from their average to the total sum of squares (TSS) which is the sum of squared deviation of  $Y_i$  from its average):

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

Or more simply,  $R^2$  can be expressed 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:

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

Being  $TSS = ESS + SSR$ , the  $R^2$  also can be expressed as:

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

### ***The Standard Error of the Regression (SER)***

It quantifies the average distance that the observed values fall from the regression line, measured in the units of the dependent variable.

The SER is measured as follow:

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

Let's focus on the  $n - 2$  denominator. It is divided by  $n - 2$  to compensate for a slight downward distortion caused by the estimation of the two regression coefficients  $\beta_0$  and  $\beta_1$  which have caused the loss of two degrees of freedom of the data, so to correct for this lost the divisor is the factor  $n - 2$ .

## Linear Regression with Multiple Regressors

The simple linear regression model becomes a multiple linear regression model if the considered independent variables are two or more, called  $X_{1i}, X_{2i}, \dots, X_{ni}$ .

This model permits estimating the effect on  $Y_i$  of changing one variable  $X_{1i}$  while holding the other regressors ( $X_{2i}, \dots, X_{ni}$ ) constant.

In this way the use of multiple variables as regressors allows to improve predictions and the model can be expressed by:

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

Where:

- $Y_i$  is the dependent variable;
- $X_{1i}$  and  $X_{2i}$  are the two regressors;
- $\beta_0$  is the unknown population intercept;
- $\beta_1$  is the effect on  $Y$  of a unit change in  $X_{1i}$ , controlling for  $X_{2i}$  (holding constant  $X_{2i}$ );
- $\beta_2$  is the effect on  $Y$  of a unit change in  $X_{2i}$ , controlling for  $X_{1i}$ ;
- $u_i$  is the model's random error (residual) term or regression error

As said before, the regression error  $u_i$  represents, among other things, the influence of omitted variables on the dependent variable. The term **omitted variable** refers to any important variable not included as an independent variable in the regression model that might influence the dependent variable, in other words: the bias consists in attributing the effect of the missing variables to those that were included.

The coefficients of the multiple regression model can be estimated from data using OLS, but if there are omitted variables it may lead to biased and inconsistent estimates.

In order to obtain more accurate estimates, in the multiple regression model, additional variables are included, the so-called **control variables**.

As before, we need to estimate the unknown population coefficients  $\beta_0, \dots, \beta_k$  using a sample of data, so that, the OLS estimators ( $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ ) are used and the expression of the straight line becomes:

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

Being the residual value equal to:

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

## Measure of Fit

In case of multiple regression there are three main statistics used to measure the accuracy and reliability of the regression line.

### *The Standard Error of the Regression (SER)*

The SER is a measure of the spread of the distribution of Y around the regression line and in case of multiple regression it is calculated as follows:

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

The difference with the linear regression with a single regressor stands just in the denominator that, in this case, is  $n - k - 1$  instead of  $n - 2$ . It adjusts for the downward bias introduced by estimating, in this case,  $k + 1$  coefficients: the  $k$  slope coefficients plus the intercept.

### *The $R^2$*

The regression  $R^2$  is the fraction of the sample variance of  $Y_i$  predicted by the regressors.

Being:

- $R^2 = \frac{ESS}{TSS}$
- Explained sum of squares:  $ESS = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$
- Total sum of squares:  $TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2$



The result is that  $R^2$  is 1 minus the fraction of the variance of  $Y_i$  not explained by the regressors:

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

However,  $R^2$  can give a misleading estimate of how well the regression fits the data because when more variables are added to a model,  $R^2$  always increases, even if the new variables do not improve the fit of the model.

To address this issue, the adjusted  $R^2$  is used.

***The Adjusted  $R^2$  ( $\bar{R}^2$ )***

It provides a more accurate measure of fit for multiple regression model since it does not necessarily increase when a new variable is added.

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

Where;

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

Thus, it is always true that  $\bar{R}^2 < R^2$  since  $\frac{n-1}{n-k-1} > 1$  always.

# Empirical Applications

## Description of the Dataset and Variables

In this section it will be provided a detailed description of the datasets used to conduct the analysis and the variables included within it with the aim of constructing the *Final Database*. In particular the Final Database will be a combination of four datasets: the one related to the Italian Innovative Startups, the one related to International Immigrated Women resident in the North-west of Italy, the one related on Care Services and, finally, the database containing a set of control variables. All the observations will be carried out considering a time span that goes from 2013 to 2022 because of the availability of data.

### Italian Innovative Startups Database

As highlighted in the previous paragraph *Legal requirements for an innovative startup*, innovative startups are no more considered as such after 5 years of activity, if they survive, they have the opportunity to transform into innovative SMEs (Small Medium Enterprises) without losing the available benefits. Since, in our analysis we will consider both startups and enterprises that are no more in the startup status we will sum two databases: the one downloaded from the Innovative Startups register provided by the Ministero delle Imprese e del Made in Italy and AIDA which is the database containing the financial statements, personal data and products of all companies with share capital operating in Italy (excluding banks, insurance companies and public bodies).

The dataset is in panel format, therefore each observation, which basically are the enterprises and the startups, is represented by an id associated with each year of observation of the startup/enterprise. More in detail, database rows are the observations and database columns are the variables. In total we have 28 variables which will be briefly described below:

- *id* it is the company identification code that allows to uniquely identify it;
- *anno\_osservazione* is the year in which the variables are observed;
- *anno\_costituzione* shows the year of foundation of the startup;

- ***comune*** is the local area in which the startup is based;
- ***codice\_istat\_comune*** it is the code provided by ISTAT to uniquely identify each municipality in Italy;
- ***provincia*** it is the name of the province in which the startup operates;
- ***codice\_istat\_provincia*** it is the code provided by ISTAT to uniquely identify each province in Italy;
- ***regione*** it is the name of the region in which the startup conducts its business, in our specific analysis we will consider four main regions: Piedmont, Valle d'Aosta, Liguria and Lombardia;
- ***stato\_giuridico*** this variable corresponds to the legal condition of the startup determined by how it was established and registered with the relevant authorities. A startup can be in the legal status of “*attiva*” which means that the startup is doing business, is active, and in the legal status of “*cessata*” or “*in liquidazione*” when the startup is winding up all business affairs and, as a consequence, its operations ends;
- ***ateco\_2007*** it is a set of alphanumeric codes introduced by ISTAT used to carry out a classification of the economic activities based on the sector on which the startup or the company operates;
- ***dipendenti*** stands for the number of employees of each startup;
- ***tot\_immob\_immateriali*** the term “*immateriali*” refers to intangible fixed assets owned by the startup/enterprise and that represent a capitalizable costs for it they are indicated in thousand of euros;
- ***tot\_immob\_materiali*** are durable tangible assets that are part of a company's permanent organization, whose economic utility extends beyond the limits of a single financial year. They are expressed in thousand of euros;
- ***tot\_attivo*** is the total amount of assets owned by the startup/enterprise expressed in thousands of euros;
- ***tot\_patrimonio netto*** it represents the size of the equity owned by the company's shareholders in thousands of euros;
- ***tot\_debiti*** stands for the total amount of debts accumulated by the startup/enterprise in thousands of euros;

- ***capitale\_sociale*** is the capital invested in the startup/enterprise by its members that is the total amount of equity owned by the startup/enterprise expressed in thousands of euros;
- ***ricavi*** stands for revenues in thousands of euros, so the amount of money actually earned at the end of the year from selling a particular asset or service;
- ***risultato\_operativo*** is the financial result of a company's main operating activities, usually expressed as the difference between income and operating expenses. It thus indicates the ability of the firm to generate profits from its core business and it is expressed in thousands of euros;
- ***valore\_aggiunto*** it represents the value that the inputs used by the enterprise, capital and labour, have 'added' to the inputs purchased from outside, so as to obtain a given production. It is calculated as the difference between the value of goods and services produced and the value of intermediate products and it is expressed in thousands of euros;
- ***utile*** indicates the income the startup/enterprise has produced, in thousand of euros, at the end of the year. It is the difference between revenues and expenses of an enterprise and if this difference is positive it is commonly called profit or surplus, otherwise it is called loss;
- ***EBITDA*** it is the acronym of Earnings Before Interest, Taxes, Depreciation, and Amortization. It focuses on the company's core profitability by excluding expenses related to financing, accounting decisions, and tax environments and it is often used to compare the profitability of companies within the same industry, as it removes the effects of capital structure, taxation, and non-cash accounting items. It indicated in thousands of euros;
- ***diritti\_brevetto\_industriale*** refers to the legal protections owned by a startup, namely the number of intellectual properties conferred on an industrial invention of the startup/ enterprise;
- ***data\_iscrizione\_sezione\_startup*** is the date on which a startup is officially registered as an innovative startup in the Startup Register and thus can benefit from specific legal advantages for startups;

- ***data\_uscita\_sezione\_startup*** it indicates the date in which the startup ceases to be considered an innovative startup and so it's no more registered in the Startup Register;
- ***data\_cessazione*** it is the date in which the company no longer exists as a legal entity and is deleted from the Companies Register;
- ***prevalenza\_femminile\_complessiva*** is the percentage of women involved in the company, considering both ownership (shareholders) and management (managers). It is a *time-variant variable* which means that its value varies with the years.

The prevalence of women is usually classified in different levels, such as:

- NO: if  $(\% \text{ shareholders} + \% \text{ managers})/2 \leq 50\%$ ;
  - Maggioritaria: if  $(\% \text{ shareholders} + \% \text{ managers})/2 > 50\%$ ;
  - Forte: if  $(\% \text{ shareholders} + \% \text{ managers})/2 > 66\%$
  - Esclusiva: if  $(\% \text{ shareholders} + \% \text{ managers})/2 = 100\%$ .
- ***prevalenza\_femminile\_amm*** is the percentage of women involved in the company considering just the management (administrators) and also in this case it is a *time-variant variable* and can assume four values:
    - NO: if  $(\% \text{ administrators}) \leq 50\%$ ;
    - Maggioritaria: if  $(\% \text{ administrators}) > 50\%$ ;
    - Forte: if  $(\% \text{ administrators}) > 66\%$ ;
    - Esclusiva: if  $(\% \text{ administrators}) = 100\%$ .

## **International Immigrated Women resident in the North-west of Italy Database**

This database has been downloaded from ISTAT (Istituto Nazionale di Statistica) which is the Italian Statistical Institute and main producer of official statistics to support citizens and public decision-makers. Data have been downloaded and filtered following some steps. First of all, it was necessary to select the section “Popolazione e famiglie” then “Stranieri e Immigrati” and filter for “Sesso” by selecting “Femmina”. In order to obtain a set of data containing all the international immigrated women in each province of the North-west of Italy and in the working age, it has been selected also the age range

between 20 and 65 years. Data are available from 2009 to 2024 but since 2024 data are approximate because the year is still in progress, the selected range of years is 2009-2023. Then, another database has been downloaded and merged with the previous one with the aim of consider also the entire population, meaning consider both international and Italian citizens in each province of the Italian territory under analysis, so in this case the procedure is the same, but it is necessary to select for both the labels “Cittadinanza” and “Sesso” the filter “Totale” without specifying the working age.

It is important to underline that to construct this database it has been made an assumption: All the immigrated women are low-skilled and they help native women with household tasks and childcare. It seems to be a strong assumption but the reason behind this choice is that, from empirical studies it has emerged that the majority of immigrated women are low-skilled and because of this they are often employed as colf (housekeepers) or as babysitters. In addition to this, women employed in these types of jobs are almost never covered by a formal employment contract but work off the books. As a result, there are very few data available on how many low-skilled women are effectively employed in these tasks.

The dataset is structured as follow: the rows indicate the observations which are the provinces and the year of observation, and the 4 columns indicate the following variables:

- *provincia* which is the name of the province
- *anno\_osservazione* is the year of observation
- *tot\_popolazione* contains the number of the total population in the specific province and for that specific year of observation
- *donne\_immigrate* is the total number of international immigrated women between the ages of 20 and 65 in the specific province for that specific year of observation

## Care Services Database

The source for downloading this database is again ISTAT which provides a dataset showing all the existing socio-educational services, both public and private ones, available across the Italian territory from 2013 to 2022.

Among the available services there are:

- **traditional and workplace nurseries** offered to children aged 3 months to 3 years;
- **Spring sections** which are services open to children aged 24 to 36 months;
- **Supplementary early childhood services** like play areas, home-based services and parent-child centers

More in detail, the database has been downloaded by selecting the category “Assistenza e Previdenza” then the sections “Servizi Sociali”, “Servizi socio educativi per la prima infanzia” and “Servizi sul territorio provincia e comune” in order to display data for each province of Northwestern Italy. At the end, to obtain a unique number containing the total number of the public and private socio-educational services for childcare, all the available services have been summed, for each province and for each year.

In particular, we have 3 columns representing the 3 variables:

- *provincia* which is the name of the province in which the services are located
- *anno\_osservazione* is the year of observation
- *tot\_servizi\_attivi* represents the total number of all the care services, both public and private, for each province and for each year of observation

## **Control Variables database**

In order to obtain a final dataset that was as comprehensive as possible, it was deemed necessary to consider also a dataset containing some important control variables. As mentioned in the section “*Linear Regression with Multiple Regressors*”, control variables are all those variables that can be included in the research study to determine if they have an influence on the outcome.

The original control variable database was made up of a lot of information not useful for our discussion, so it has been cleaned up by all the redundant and unnecessary data; the result is a dataset composed of ten observations, hence 10 variables:

- *provincia* which is the name of the province
- *anno\_osservazione* is the year of observation

- ***working\_population*** refers to the total number of people able or available to work<sup>9</sup>.
- ***unemployment\_rate*** it is the percentage of people in the working-age that have no job. It is a crucial measure of an economy. It tells how much the economy is using resources (inputs) to produce GDP (output). It is given by the total number of unemployed people over the total labor force which contains both employed and unemployed people.

$$\text{Unemployment rate} = \left( \frac{\text{Unemployed people}}{\text{Total labour force}} \right) \times 100$$

- ***value\_added\_per\_habitant*** it is an economic indicator that measures the quantity of economic value produced by each individual in a specific geographic area and so it is a measure of the productivity and economic wellness of a population. It is obtained by dividing the value added which is the value created by an industry net of costs of intermediate goods and services (which are all the goods used for the production) and the total population of the specific geographic area.

$$\text{Value added per habitant} = \left( \frac{\text{Total value added}}{\text{Total population}} \right)$$

- ***number\_firm*** is the total number of firms in the specific province and in that specific year of observation
- ***area\_km2*** it is the area of the reference province measured in  $km^2$  for each year of observation
- ***resident\_laureati*** is the number of people with a bachelor's degree resident in the specific province and observed in the specific year of observation.

## Final Database

The final database is the result of a merger of the four datasets described above made with STATA which is a widely used statistical software for data manipulation, data visualization, statistics and automated reporting and it is extremely useful to handle large datasets and to conduct extensive survey. Beyond academia it is frequently used in

---

<sup>9</sup> Investopedia. n.d. “*Working-Age Population: Definition, Importance, and Example*” Available at: <https://www.investopedia.com/terms/w/working-age-population.asp#:~:text=The%20working%2Dage%20population%20is,%2D64%20or%2015%2D64.>



private industry, business forecasting, public health research and in several fields like economics, biomedicine and political science. The aim in constructing this merged database is to have an optimized set of data for the realization of our linear regression model and to do so the so-called “*merge keys*” have been defined. These latter are those values that must be present in all datasets to be merged and must have unique values that allow the data to be combined correctly, ensuring that the observations are properly aligned between the various files, those values are the variables “*provincia*” and “*anno\_osservazione*”.

The database obtained is characterized by 118 variables visible on the columns which will be explained below and 2.454 observations visible on the rows. In addition to all the variables explained before other additional variables have been added:

- “*data\_fine*” indicates the date in which the startup is no more active. This variable has been created following 3 steps:
  - a) “*data\_fine*” is set equal to “*data\_cessazione*”<sup>10</sup> if it is an available information for a certain enterprise;
  - b) If the enterprise does not have as available information “*data\_cessazione*”, “*data\_fine*” will coincide with the last year in which the startup has produced revenues +1 year in order to assume that it will exit from the Enterprises Register the next year;
  - c) If no revenues data are available, “*data\_fine*” will coincide with “*data\_uscita\_sezione\_startup*”, hence the date of exit from the startup section.
- “*startup\_prev\_femminile*” is a dummy variable created to identify all the female led innovative startups. By definition, a dummy variable is a binary variable that assumes 1 or 0 as values; in our case it assumes value 1 if the percentage of women involved in the company, considering both ownership and management, is MAGGIORITARIA, FORTE or ESCLUSIVA and 0 otherwise.
- “*duration*” is created by subtracting to “*data\_fine*” the year of creation of the startup, hence, “*anno\_costituzione*”. This variable is used in two different ways: once it is set equal to a maximum value of 5 years with the aim of analyzing the

---

<sup>10</sup> *data\_cessazione* is the date in which a company closes and it is a date provided by AIDA, the database containing all information about registered enterprises.

duration of innovative startups, that as said before, can be considered as such only for 5 years, after that, they exit from the startup register and become enterprises. Then, the duration variable is not set equal to a maximum value of five years since it is interesting also to evaluate the total duration of those startups that become enterprise. This variable contains the information about the duration of all startups/enterprises, namely those led by men and those led by women.

- “*duration\_startup\_guida\_femminile*” is the effective duration we want to estimate since it represents the duration of female-led innovative startup. In fact, this variable is equal to the variable “*duration*” if the dummy variable “*startup\_prevalenza\_femminile*” is equal to 1.
- “*ateco\_2007\_2d*” is added to consider just the first two digits of the ATECO code; it is constructed as *ateco\_2007\_* variable divided by 10000.

## **Construction of the model: Hypothesis and Research Objectives**

The core of the research, as the background literature has highlighted in the previous chapters, is to demonstrate that the presence of low-skilled immigrated women and the presence of care services such as nurseries and early childhood services, in the specific geographic area, can help native entrepreneurs women, creators of innovative startups, to dedicate more time to their professional career, improve in terms of work performance and as a consequence increase the survival rate of the startup since these services can replace them in their historical role as home and child carers. The analysis consists in identifying if there is a proportional relationship between the survival rate of female-led innovative startup, the number of low-skilled international immigrated women and the number of care services. The evaluation is carried out creating a multiple linear regression model considering as geographical area the North-west of Italy that includes 4 main regions: Piedmont, Lombardia, Liguria and Valle d’Aosta and on a time span that goes from 2013 to 2022. The survival rate of the startup (our dependent variable) is measured through the variable *duration\_startup\_guida\_femminile* intended, once as the period in which the female-led startup has been active on the market under the status of startup and then, after 5 years as an enterprise. Below it will be explained

more in detail the hypothesis to be tested and the independent variables that will be taken into account to carry out our tests.

The hypothesis we are going to test is the following:

**Hp1:** “The survival rate of a female-led innovative startup established in the North-west of Italy increases by an increase in the number of low-skilled international immigrated women and by an increase in the care services present in North-west Italy provinces”

The result is a multiple linear regression model:

$$duration\_startup\_guida\_femminile_i = \beta_0 + \beta_1 donne\_immigrate_{jt} + \beta_2 tot\_servizi\_attivi_{jt} + u_{ijt}$$

Where:

- $duration\_startup\_guida\_femminile_i$  is the dependent variable indicating the life span of the  $i$ -nth female-led startup. It is measured in years;
- $\beta_0$  is the intercept;
- $donne\_immigrate_{jt}$  is the first independent variable representing the total amount of international immigrated women resident in the province  $j$  at the year  $t$ ;
- $tot\_servizi\_attivi_{jt}$  is the second independent variable indicating the total number of the childcare services present in the province  $j$  at the year  $t$ ;
- $\beta_1, \beta_2$  are the coefficient of the independent variables;
- $u_{ijt}$  is the regression error representing all the omitted factors linked to the  $i$ -nth startup, the  $j$  province and the year  $t$ ;

In order to obtain a more accurate estimation of the model it is useful to introduce also the control variables.

The linear regression model becomes the following:

$$\begin{aligned} duration\_startup\_guida\_femminile_i \\ = \beta_0 + \beta_1 donne\_immigrate_{jt} + \beta_2 tot\_servizi\_attivi_{jt} \\ + \beta_3 control\ variable_{ijt} + u_{ijt} \end{aligned}$$

The hypothesis to test and the linear regression model remain unchanged, the only difference is that we are going to consider a vector called *control variable*<sub>ijt</sub> indicating the set of control variables related to the *i*-nth startup to the province *j* and the year *t*, and  $\beta_3$  is the coefficient of the control variables' vector;

Then, another version of the same empirical model is created by changing the two control variables:

- *donne\_immigrate* becomes *rate\_donne\_immigrate* calculated as:

$$rate\_donne\_immigrate = \frac{donne\_immigrate}{tot\_popolazione}$$

- *tot\_servizi\_attivi* becomes *tasso\_servizi* calculated as:

$$tasso\_servizi = \frac{tot\_servizi\_attivi}{areakm2}$$

The linear multiple regression model becomes:

$$\begin{aligned} &duration\_startup\_guida\_femminile_i \\ &= \beta_0 + \beta_1 rate\_donne\_immigrate_{jt} + \beta_2 tasso\_servizi_{jt} \\ &+ \beta_3 control\_variable_{ijt} + u_{ijt} \end{aligned}$$

These two variants will be analyzed considering two conditions:

1. The duration is set to a maximum value of 5 years to evaluate the duration of female-led innovative startups since, by law, a startup "is a new business or an established company for not more than 5 years".
2. The duration is not set to a maximum value of 5 years, hence there will be considered also startups that exit from the startup register because they become enterprises with the aim of evaluate their total duration.

The next paragraph will provide a more detailed explanation of the variables introduced above to better understand the rationale behind the selection of these variables for creating the model.

## **Variables Description**

The variables that have been implemented to develop the research study can be classified in: dependent variable, independent variables and control variables.

### **The Dependent variable**

#### *duration\_startup\_guida\_femminile*

The *dependent* variable represents the variable to be predicted and intuitively, it is so-called because its value depends on the values assigned to the other variables.

The variable under analysis is once the duration of a female-led startup, meaning the time span in which the startup is active. According to the Italian regulations, the maximum duration of a startup is 5 years<sup>11</sup>; after this period, a startup becomes an enterprise and therefore exists the startup register and enters into the enterprise register. Since not all startups manage to survive for 5 years and become enterprise it is not possible to have the end date of these startups so in these cases to indicate the termination date it is considered the date when the last financial statements are available + 1 year. The reason behind this choice stands in the fact that if a company has recorded revenues in a given year, it means that it was still active and then died the following year. This variable was thus created by making this assumption and by subtracting to the startup's termination date the date in which the startup has been founded. Then this variable will be also evaluated taking into consideration all that startups that have scaled up, meaning all that startups that have passed the initial stage (5 years) and are becoming consolidated enterprises.

### **The Independent variables**

With independent variable we refer to the variable which is controlled and manipulated by the researcher.

The independent variables used in our regression models are:

---

<sup>11</sup> The standard "lifetime" of an Innovative Startup is five years from the date of incorporation. However, the "Decreto Rilancio" (Relaunch Decree) has provided an exception to this rule due to the needs imposed by the COVID emergency. As a result, the lifetime of Innovative Startups registered in the Special Section of the Business Register as of May 19, 2020, is extended by one year, increasing from five years to six. Source: [https://www.studioassociatomsc.com/la\\_mossa\\_giusta/prorogata-di-12-mesi-iscrizione-startup-e-pmi-innovative-659/](https://www.studioassociatomsc.com/la_mossa_giusta/prorogata-di-12-mesi-iscrizione-startup-e-pmi-innovative-659/)

### ***donne\_immigrate***

This variable indicates the number of international immigrated women between 20-65 years old resident in a certain province of the North-West of Italy in a certain year of observation. The age range is so defined since it is considered the most appropriate to address women that chooses not to continue with the academic career and are able to be employed as housekeepers or as babysitters, for example. These women are assumed to be the ones that substitute native women entrepreneurs in the house and children care allowing them to be more profitable and bring forward their professional escalation. The basic assumption is that there is a positive correlation between this variable and the duration of a female-led startup born in the same geographic area already mentioned, hence, when the number of international immigrated women increase by one unit, the duration will increase too.

### ***tot\_servizi\_attivi***

The variable “*tot\_servizi\_attivi*” is the second most important independent variable of which we want to test the effects on the duration of a female-led innovative startup. It includes all the childcare services, both private and public, available in a certain province of the North-west of Italy and in a specific year, for all female entrepreneurs of that geographic area. These data are available on ISTAT database from 2013 to 2022 and, as already pointed out in the description section of the database, this variable is a total number containing the sum of all the following childcare services’ categories: traditional and workplace nurseries (3 months to 3 years old), spring sections (24 to 36 months) and supplementary early childhood services. Much attention is given to the effect this variable may have on female-led startup duration as it is believed that entrepreneurial mothers, who can access these services, are better able to balance work and family, so it is expected a positive correlation between childcare services and startup duration.

### ***rate\_donne\_immigrate***

The variable *rate\_donne\_immigrate* is the ratio between the number of international immigrated women and the total population of a certain province of the north-west of Italy observed in a certain year. The absolute number of immigrant women may not fully reflect the impact on the life of startups, but the rate may better represent the

integration of these women into the social and labor environment of the province. An area with a high rate of immigrant women with respect to the total population may have a more favorable environment for female entrepreneurship.

### ***tasso\_servizi***

This variable represents the number of childcare services per square kilometers. To construct this rate, it has been considered as denominator no more the total population but the extension in  $km^2$  of the geographical area. The aim is to consider the spatial distribution of childcare services, not just their quantity, since a high number of services might not be helpful if they are distributed over a large area; for example, it may be difficult for mother entrepreneurs living far from the service centers to access them. A higher rate could imply that childcare services are closer and therefore potentially more accessible.

To conclude, the choice behind the construction of these two rates stands in the idea of “normalize” their values with respect to a total. In other words, we want to minimize the variability between geographical areas of different size and hence, the distortion of the final results.

### **The Control variables**

Below there will be explained the control variables that have been chosen because considered the most important ones:

#### **valore\_aggiunto**

With term “valore\_aggiunto” it is meant value added, so it is an indicator of the dimension and economic capability of a company. A higher value added suggests that the startup is able to generate profits from its core activity which is crucial for its long-term sustainability.

#### **value\_added\_per\_habitant**

The value added per habitant is a measure of the wealth of the population, so a high value added per habitant means a greater willingness to consume which again means increase in the demand for goods and services and potential increase of the startup duration.

**tot\_debiti**

The total number of debts it's a measure of the financial stability of a company. A higher value of total debts may indicate a higher financial risk for the startup since it means that they are forced to spend a significant portion of their revenues on paying debts, thus reducing financial resources for investments and growth. So, it is important to consider the effects of this variable on the startup life.

**dipendenti**

Include the number of employees as a control variable is equally significant since it is considered an indicator of the firm dimension and also a measure of the operational efficiency. A consistent number of employees could mean that the startup is able to manage higher volumes of projects and activities and that it is a large and potentially long-lived startup. At the same time, a higher number of employees translate also in higher fixed costs such as wages and contributions.

**i.ateco\_2007\_2d**

The industry sector in which the startup operates may affect its survival rate, as some sectors may be more stable or profitable than others. For our discussion it has been considered useful to take just the two digits of the ATECO code, indicating the specific industrial sector, in order to reduce the variability and consider less observations since this variable is treated as a dummy variable by placing "i." before the variable name. In the practice, it is more correct transform categorical variables like the ATECO code in binary variable (0 or 1) in order that each sector is represented separately in the empirical model, and it is possible to evaluate the specific effect on the dependent variable of the single sector.

**unemployment\_rate**

The unemployment rate is a measure of the economy status of a certain geographic area. A lower unemployment rate could indicate a stronger local economy, which might better support new startups, while a higher unemployment rate could reduce the level of entrepreneurship e the trust in starting new businesses.



## **area\_km2**

A larger area can offer more opportunities to reach a consistent number of customers, more space for offices, laboratories and production facilities and in some cases, also better infrastructures and better public services. All these aspects are favorable to the growth and longevity of the startup.

## **residenti\_laureati**

The number of residents with a master's degree can be viewed as potential human resources that might be employed in a local startups since one of the fundamental requirements to be part of an innovative startup is that: "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;", so for a startup, having access to this pool of educated people could help developing and enhancing its products and services and be more durable.

# **Descriptive Statistics**

In this chapter it will be provided a concise overview of the main characteristics of the most important data collected to construct the linear multiple regression model with the aim of facilitating understanding and interpretation.

First of all, it is interesting to start with the description of "*anno\_costituzione*" and "*provincia*" since they can give us an idea of the entrepreneurial scene.

## **1. anno\_costituzione**

anno_costit uzione	Freq.	Percent	Cum.
2013	100	4.07	4.07
2014	138	5.62	9.70
2015	163	6.64	16.34
2016	196	7.99	24.33
2017	233	9.49	33.82
2018	289	11.78	45.60
2019	303	12.35	57.95
2020	349	14.22	72.17
2021	403	16.42	88.59
2022	280	11.41	100.00
Total	2,454	100.00	

Figure 3:anno\_osservazione

This variable indicates the year of foundation of the female-led startup. Looking at the column “Frequenze” it is possible to understand how much female-led startups have been founded in the north-west of Italy each year of observation and it is clear that from 2013 to 2022 the number of startups born on that geographical area is almost tripled.

## 2. provincia

provincia	Freq.	Percent	Cum.
Alessandria	18	0.73	0.73
Asti	7	0.29	1.02
Bergamo	126	5.13	6.15
Biella	4	0.16	6.32
Brescia	117	4.77	11.08
Como	44	1.79	12.88
Cremona	22	0.90	13.77
Cuneo	56	2.28	16.06
Genova	86	3.50	19.56
Imperia	5	0.20	19.76
La Spezia	10	0.41	20.17
Lecco	17	0.69	20.86
Lodi	14	0.57	21.43
Mantova	18	0.73	22.17
Milano	1,440	58.68	80.85
Monza e della Brianza	66	2.69	83.54
Novara	28	1.14	84.68
Pavia	38	1.55	86.23
Savona	9	0.37	86.59
Sondrio	1	0.04	86.63
Torino	259	10.55	97.19
Valle d'Aosta/Vallée d'Aoste	12	0.49	97.68
Varese	55	2.24	99.92
Verbano-Cusio-Ossola	1	0.04	99.96
Vercelli	1	0.04	100.00
Total	2,454	100.00	

Figure 4:provincia

In the north-west of Italy, as said in the previous chapters, there are 4 main regions:

- Lombardia which counts 12 provinces:
  - Milano
  - Bergamo
  - Brescia
  - Como
  - Cremona
  - Lecco
  - Lodi
  - Mantova

- Monza e Brianza
- Pavia
- Sondrio
- Varese
- Piemonte which is characterized by 8 provinces:
  - Torino
  - Alessandria
  - Asti
  - Biella
  - Cuneo
  - Novara
  - Verbano-Cusio-Ossola
  - Vercelli
- Liguria with 4 provinces:
  - Genova
  - Imperia
  - La Spezia
  - Savona
- Valle d’Aosta with just one province:
  - Aosta

Looking again at the “Frequency” column, the area with the higher number of female-led innovative startups, as highlighted also in the background literature, is the Lombardia region with 1.440 only in the province of Milano considered as the most innovative city.

### 3. **donne\_immigrate**

Variable	Obs	Mean	Std. Dev.	Min	Max
donne_immig~e	2,454	117188	64347.57	3326	182271

*Figure 5: Summarize donne\_immigrate*

The relation between the mean (117,188) and the standard deviation (64,347.57) suggests a distribution of data with strong variability. The standard deviation is quite

large compared to the average, indicating that the values of the number of immigrant women resident in each province of the north-west of Italy differ significantly from the average. This may be due to territorial or economic differences between the areas surveyed, which can affect the presence of immigrant women in different ways.

#### 4. tot\_servizi\_attivi

Variable	Obs	Mean	Std. Dev.	Min	Max
tot_servizi_attivi	2,454	262.967	121.583	29	572

Figure 6: Summarize tot\_servizi\_attivi

The table indicates that on average we have almost 263 childcare services active on the territory under exam and also in this case, as the case of the variable *donne\_immigrate*, we have a significant variability among the observations indicated by a standard deviation of 121.583 which suggests that some provinces could have more availability of childcare services with respect to others.

#### 5. rate\_donne\_immigrate

Variable	Obs	Mean	Std. Dev.	Min	Max
rate_donne_immigrate	2,454	.0461007	.0081207	.0206601	.056225

Figure 7: Summarize rate\_donne\_immigrate

This variable has been constructed to decrease the strong correlation between immigrated women and the total population, hence it has been created a rate in order to consider both variables in just one value. Being a rate, it produces percentual results: we have a mean of 0.00461 which indicates that the 4,61% of the total population is composed by immigrated women aged 20 to 65 years and the standard deviation is a quite low value (0.0081) that could indicate that the immigrated women with respect to the total population are distributed quite homogeneous with just little local variations.

## 6. tasso\_servizi

Variable	Obs	Mean	Std. Dev.	Min	Max
tasso_servizi	2,454	.1358874	.0585822	.0128268	.2639357

Figure 8: Summarize tasso\_servizi

The variable represents the childcare services active on the geographical area, the table indicates that on average there are 0.136 childcare services per  $km^2$  available in considered area and the standard deviation (0.0586) indicates a large dispersion of the data with respect to the mean. The mean number of childcare services seems a bit low and it could probably indicate that the north-west area is too large with respect to the services accessible to families.

## 7. duration\_startup\_guida\_femminile

As said in the previous paragraphs, the aim of this paper is to evaluate the duration considering both: startups and consolidated enterprises. To do so we need to impose two conditions that will give different values of frequency.

- a. The duration is set to a maximum value of 5 years to evaluate the duration of female-led innovative startups since, by law, a startup "is a new business or an established company for not more than 5 years"

This condition is imposed by replacing the value of the duration equal to 5 years for that startups that have a duration higher than 5 years. Therefore, we will have a major frequency of female-led startup with duration equal to 5 years as per below picture (895) and of course the value of the mean will be higher than expected.

duration_startup_guida_femminile	Freq.	Percent	Cum.
1	380	15.48	15.48
2	479	19.51	34.99
3	392	15.97	50.96
4	309	12.59	63.54
5	895	36.46	100.00
Total	2,455	100.00	

Figure 9: duration\_startup\_guida\_femminile when duration max 5 years

Below it is shown also the frequency table of the variable “*duration*” to show the big difference between the number of startups led by men and the ones led by women. By subtracting to the total number of startups (the one present in below figure) the number of female-led startups, more or less the difference, for each year of duration, is of 1000 startups more for men.

<i>duration</i>	Freq.	Percent	Cum.
1	1,281	14.37	14.37
2	1,769	19.85	34.22
3	1,466	16.45	50.66
4	1,162	13.04	63.70
5	3,236	36.30	100.00
Total	8,914	100.00	

Figure 10: *duration* when *duration* max 5 years

Let’s look at the most important synthesis measures of the variable *duration\_startup\_guida\_femminile*:

Variable	Obs	Mean	Std. Dev.	Min	Max
<i>duration_sve</i>	2,455	3.350305	1.508835	1	5

Figure 11: summarize *duration\_startup\_guida\_femminile* when *duration* max 5 years

It shows a mean of 3,35 which means that female-led startup in the north-west of Italy tend to survive for a period of about 3-4 years, with a moderate dispersion of 1,51 and it is a quite consistent result for what has been pointed out at the beginning.

- b. The duration is not set to a maximum value of 5 years to evaluate the duration of startups that scaled up and become enterprises.

In this case, no condition is imposed, so we will visualize also the startups that have exit the startup register and have entered the enterprise register. In this case the frequency of startups with a life of 5 years is a much smaller number (304) with respect to the case a. since the frequency is distributed on more years (from 1 to 10) and it seems that the majority of startups have a lifespan of 2 years.

duration_startup_guida_femminile	Freq.	Percent	Cum.
1	380	15.48	15.48
2	479	19.51	34.99
3	392	15.97	50.96
4	309	12.59	63.54
5	304	12.38	75.93
6	191	7.78	83.71
7	147	5.99	89.69
8	94	3.83	93.52
9	98	3.99	97.52
10	61	2.48	100.00
Total	2,455	100.00	

Figure 12: duration\_startup\_guida\_femminile when duration not set max 5 years

Let's also analyze for this case the synthesis measures:

Variable	Obs	Mean	Std. Dev.	Min	Max
duration_sve	2,455	3.94664	2.410755	1	10

Figure 13: summarize duration\_startup\_guida\_femminile when duration not set max 5 years

The variable shows a Mean of 3,95 indicating that the average duration of the female-led startups that have also scaled up is almost of 4 years. The value of the standard deviation, equal to 2,41 suggests also a quite high variability. This result can be justified by the fact that there are less startups that last over 5 years and so the average stays below 5 years.

## 8. residenti\_laureati

Variable	Obs	Mean	Std. Dev.	Min	Max
residenti_~i	2,454	11869.56	5813.145	534	18013

Figure 14: Summarize residenti\_laureati

There are on average 11,870 residents with a bachelor's or a master's degree, the value of the standard deviation but also the gap between min e max suggests that there are significant differences in the number of graduated residents between the different provinces.

## 9. unemployment\_rate

Variable	Obs	Mean	Std. Dev.	Min	Max
unemployment_rate	2,454	6.815991	1.654669	2.815521	13.56989

Figure 15: Summarize unemployment\_rate

Also in this case, by looking simply at the value of min and max, we have a consistent gap between the two values indicating that there are provinces with a higher unemployment rate and provinces with a lower unemployment rate. The mean value seems to be quite high and it could be linked to multiple factors, the most plausible could be the recent economic crisis caused by the pandemic of Covid-19.

## 10. value\_added\_per\_habitant

Variable	Obs	Mean	Std. Dev.	Min	Max
value_added_per_habitant	2,174	74487.12	7842.861	55507.57	85957.15

Figure 16: Summarize value\_added\_per\_habitant

The table shows that on average the economic contribution of each habitant to the local economy is almost of 74,487 euros. The variability is not so high but there's a quite large difference between min and max maybe due to the differences between the provinces considered.

## 11. area\_km2

Variable	Obs	Mean	Std. Dev.	Min	Max
area_km2	2,454	2463.856	1861.534	405.4018	6894.823

Figure 17: Summarize area\_km2

As we have seen also for the other variables that are related to the territory there is always a certain variability, also in this case the results of standard deviation, the gap between min and max and especially the mean suggest that the areas are large and that they are different from each other in terms of dimensions.



## 12. dipendenti

dipendenti	Freq.	Percent	Cum.
0	1,292	77.55	77.55
1	183	10.98	88.54
2	78	4.68	93.22
3	35	2.10	95.32
4	17	1.02	96.34
5	19	1.14	97.48
6	5	0.30	97.78
7	7	0.42	98.20
8	3	0.18	98.38
9	8	0.48	98.86
10	1	0.06	98.92
11	1	0.06	98.98
13	3	0.18	99.16
14	1	0.06	99.22
15	1	0.06	99.28
16	1	0.06	99.34
17	3	0.18	99.52
18	1	0.06	99.58
19	1	0.06	99.64
20	1	0.06	99.70
21	1	0.06	99.76
24	1	0.06	99.82
25	2	0.12	99.94
30	1	0.06	100.00
Total	1,666	100.00	

Figure 18: dipendenti

For the variable indicating the number of employees it is interesting to look at the frequency column in which we see that most of the startups have 0 employees since often startups start as individual projects with a single entrepreneurs that manage all the activities.

## 13. tot\_debiti

Variable	Obs	Mean	Std. Dev.	Min	Max
tot_debiti	1,723	85.9942	233.3031	0	3672

Figure 19: Summarize tot\_debiti

Despite the mean value of total debts is not so high there's a large variability linked to the fact that many startups start their activities by financing them with their equity capital so they don't ask for loans while other ask for large fundings.

#### 14. valore\_aggiunto

Variable	Obs	Mean	Std. Dev.	Min	Max
valore_agg~o	1,723	.1567034	70.67115	-1044	897

*Figure 20: summarize valore\_aggiunto*

The value that immediately catches the eye is the negative value of the minimum of the value added produced by a company, meaning that there are some enterprises in serious financial difficulty with losses that exceed the value added generated. It can be considered a significant result since, as we have mentioned before, there are very few startups that are able to survive over 2 years of activity.

# Results

In this chapter it will be presented and discussed the final results of the linear regression analysis conducted to estimate the effect of the number of low-skilled immigrated women and of childcare services on the female-led innovative startups in a certain province of North-west of Italy.

There will be analyzed two versions of the linear regression and for both variants the results will be observed taking in consideration the two conditions already mentioned:

- a. The duration is set to a maximum value of 5 years to evaluate the duration of female-led innovative startups
- b. The duration is not set to a maximum value of 5 years to evaluate the duration of startups that scaled up and become enterprises

## Version 1

The hypothesis Hp1 (see the chapter “*Construction of the model: Hypothesis and Research Objectives*”) is tested once, considering the model without control variables in order to see the “absolute” effect of the independent variables on the duration of female-led startups and then, to obtain a more accurate result, there will be introduced also the control variables. For both, there will be applied the two conditions a. and b.

### Multiple linear regression model *without* control variables:

$$\begin{aligned}
 &duration\_startup\_guida\_femminile_i \\
 &= \beta_0 + \beta_1 donne\_immigrate_{jt} + \beta_2 tot\_servizi\_attivi_{jt} + u_{ijt}
 \end{aligned}$$

#### Condition a.

#### Model's Fit

Source	SS	df	MS	Number of obs	=	2,454
Model	20.8947706	2	10.4473853	F(2, 2451)	=	4.61
Residual	5560.31631	2,451	2.26859091	Prob > F	=	0.0101
				R-squared	=	0.0037
				Adj R-squared	=	0.0029
Total	5581.21108	2,453	2.27525931	Root MSE	=	1.5062

Figure 21: Model's Fit without control variables- Version 1-condition a.

The statistics of model's fit, when we're estimating the duration of female-led startups considering just the effect of independent variables, show not good results. The principal statistics we look at are:

- **The p-value associated to test F:** Assuming that we have chosen as significance level 0,05, our p-value associated to test F, is equal to 0,0101 which is a quite good result since it is lower than 0,05. It means that we can reject the null hypothesis that the coefficients of the independent variables are equal to 0 ( $\beta_1=\beta_2=0$ ) that in other words means that the results are significant and that the independent variables have an effect on the dependent variable.
- **The adjusted  $R^2$ :** we look directly at this value since it indicates a more precise measure of the goodness of the model with respect to the standard  $R^2$ . It is equal to 0.0037 meaning that the only the 0,37% of the variability of the dependent variable is explained by the regression model.
- **Root MSE (Root Mean Squared Error):** is the measure of the standard deviation of the residuals (difference between the observed values and the predicted values) so it represents, on average, how far the model's predictions diverge from real data. It is a value that needs to be near 0 and it is expressed in the same unit of measurement of the dependent variable. In our case the RMSE is a bit high: the average error of the previsions is of 1.5062 years.

### Analysis of Regression Coefficients

duration_startup~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tot_servizi_attivi	2.65e-06	.0002688	0.01	0.992	-.0005244	.0005297
donne_immigrate	-1.44e-06	5.08e-07	-2.83	0.005	-2.43e-06	-4.40e-07
_cons	3.518862	.079894	44.04	0.000	3.362196	3.675529

Figure 22:Regression coefficients without control variables- Version 1- condition a.

The coefficients obtained do not respect our prevision since the one related to the childcare services is very low and the one related to the immigrated women has a negative effect on the duration of female-led startups. To understand if the values of coefficients are significant, we have to look, again at the p-values of each coefficient indicated in the column  $P > |t|$ . Luckily, the coefficient of the childcare services is not significant because  $<0,05$  but the other results significant. As we know from the

literature there are several factors that have an influence on the survival rate of a startup, hence, to obtain better results we should introduce the control variables.

**Condition b.**

**Model's Fit**

Source	SS	df	MS	Number of obs	=	2,454
Model	67.1456515	2	33.5728257	F(2, 2451)	=	5.80
Residual	14186.1779	2,451	5.78791428	Prob > F	=	0.0031
				R-squared	=	0.0047
				Adj R-squared	=	0.0039
Total	14253.3236	2,453	5.8105681	Root MSE	=	2.4058

Figure 23: Model's Fit without control variables- Version 1-condition b.

Now, let's look at the Model's fit when no condition on the duration is set.

- **The p-value associated to test F:** it is equal to 0,0031 (<0,05), it is much more near 0 than the previous case, so the model results statistically significant as before.
- **The adjusted R<sup>2</sup>:** it has slightly improved but still very low (0,39%), so it still does not exist a significant relation between independent and dependent variables.
- **Root MSE:** this value is getting worse, so the accuracy of the model is lower than before: previsions diverge from the observed data of 2.4058 years.

**Analysis of Regression Coefficients**

duration_startup~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
tot_servizi_attivi	.0005773	.0004293	1.34	0.179	-.0002645 .0014191
donne_immigrate	-2.76e-06	8.11e-07	-3.40	0.001	-4.35e-06 -1.17e-06
_cons	4.119631	.1276136	32.28	0.000	3.86939 4.369873

Figure 24: Regression coefficients without control variables- Version 1-condition b.

For what concern the coefficients they continue having same characteristics as the case a.: the coefficient of the childcare services is very small and not statistically significant, the coefficient of immigrated women is significant but negative meaning that if the number of immigrated women increases, the duration of scaled up startups decrease of very small value, but the issue is that it decreases, while an increase in the number of childcare services active in a certain province leads to a small increase in the duration of

scaled up startups. Also in this case, those results are not as expected because control variables are not included.

**Multiple linear regression model *with* control variables:**

$$\begin{aligned}
 &duration\_startup\_guida\_femminile_i \\
 &= \beta_0 + \beta_1 donne\_immigrate_{jt} + \beta_2 tot\_servizi\_attivi_{jt} \\
 &+ \beta_3 control\ variable_{ijt} + u_{ijt}
 \end{aligned}$$

**Condition a.**

**Model's Fit**

Source	SS	df	MS	Number of obs	=	1,478
Model	902.516418	73	12.3632386	F(73, 1404)	=	10.38
Residual	1672.38209	1,404	1.19115534	Prob > F	=	0.0000
				R-squared	=	0.3505
				Adj R-squared	=	0.3167
Total	2574.89851	1,477	1.74333007	Root MSE	=	1.0914

Figure 25: Model's Fit with control variables- Version 1-condition a.

By introducing all that variables that may have an impact on the duration of female-led innovative startups, the fit increases:

- **The p-value associated to test F:** it is equal to 0 that of course is <0,05) so we can strongly reject the null hypothesis and the model is statistically significant.
- **The adjusted R<sup>2</sup>:** it is a quite good value because we have that the 35,05% of the variability in the dependent variable is explained by the independent variables in our model.
- **Root MSE:** unfortunately, this value is still not near 0, but it equal to 1.0914.

## Analysis of Regression Coefficients

duration_startup_guida-ve	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tot_servizi_attivi	.0051516	.000551	9.35	0.000	.0040707	.0062326
donne_immigrate	.0000681	3.26e-06	20.86	0.000	.0000617	.0000745
unemployment_rate	.0704745	.0207968	3.39	0.001	.0296783	.1112706
value_added_per_habitant	-.0000973	.0000138	-7.08	0.000	-.0001243	-.0000704
residenti_laureati	-.0007268	.000041	-17.72	0.000	-.0008072	-.0006463
area_km2	-.0001921	.0000274	-7.02	0.000	-.0002457	-.0001384
dipendenti	.0223947	.0139585	1.60	0.109	-.0049871	.0497765
tot_debiti	.0003058	.0001437	2.13	0.034	.0000239	.0005877
valore_aggiunto	-.0005145	.0004752	-1.08	0.279	-.0014467	.0004177
ateco_2007_2d						
3	.3877137	.7073989	0.55	0.584	-.9999589	1.775386
10	.4783924	.4995231	0.96	0.338	-.5014996	1.458284
11	.8021973	1.142801	0.70	0.483	-1.439584	3.043979
13	.3500762	.6322546	0.55	0.580	-.8901892	1.590342
14	.3191858	.5846578	0.55	0.585	-.827711	1.466083
15	.2165749	.6325977	0.34	0.732	-1.024364	1.457513
16	-.5083518	1.148074	-0.44	0.658	-2.760476	1.743773
18	1.483045	.7066178	2.10	0.036	.0969046	2.869185
20	.178873	.41372	0.43	0.666	-.6327029	.9904489
21	.6400469	1.138382	0.56	0.574	-1.593067	2.87316
22	.1638936	.4824122	0.34	0.734	-.7824327	1.11022
24	.4875275	.7117624	0.68	0.493	-.9087048	1.88376
25	.4541351	.469776	0.97	0.334	-.4674033	1.375674
26	.3256119	.3827156	0.85	0.395	-.425144	1.076368
27	-.1980535	.4315858	-0.46	0.646	-1.044676	.648569
28	-.1033617	.3826093	-0.27	0.787	-.8539091	.6471858
29	.2177334	.5221535	0.42	0.677	-.8065516	1.242018
30	.0521255	.6348298	0.08	0.935	-1.193192	1.297443
31	.6472714	.7063432	0.92	0.360	-.7383303	2.032873
32	-.5072121	.4301659	-1.18	0.239	-1.351049	.3366251
35	.6710446	.4688852	1.43	0.153	-.2487464	1.590836
36	.3765921	1.141862	0.33	0.742	-1.863347	2.616531
38	.6628529	.6331475	1.05	0.295	-.5791642	1.90487
41	.5936052	.5833024	1.02	0.309	-.550633	1.737843
43	.9218789	.7074898	1.30	0.193	-.465972	2.30973
45	.3335936	.8360994	0.40	0.690	-1.306545	1.973732
46	.4150865	.3904531	1.06	0.288	-.3508479	1.181021
47	.0409926	.3588677	0.11	0.909	-.6629821	.7449674
49	-.2095834	1.139759	-0.18	0.854	-2.445398	2.026231
52	.2373455	.582028	0.41	0.683	-.9043926	1.379084
53	.4877875	.8368272	0.58	0.560	-1.153779	2.129354
55	.9414851	.632643	1.49	0.137	-.2995423	2.182512
56	.0854794	.4510171	0.19	0.850	-.7992605	.9702193
58	-.0717818	.3720547	-0.19	0.847	-.8016248	.6580613
59	.3580027	.4842206	0.74	0.460	-.591871	1.307877
61	.837604	.8384832	1.00	0.318	-.8072109	2.482419

62	.0854702	.321971	0.27	0.791	-.5461257	.7170662
63	-.0029342	.3296369	-0.01	0.993	-.6495681	.6436998
64	.3180364	.5498264	0.58	0.563	-.7605333	1.396606
66	.1143503	.8361648	0.14	0.891	-1.525917	1.754617
68	-.1162039	.630791	-0.18	0.854	-1.353598	1.12119
69	.4927556	.8359696	0.59	0.556	-1.147129	2.13264
70	.3797443	.35323	1.08	0.283	-.3131712	1.07266
71	.0214756	.3838905	0.06	0.955	-.7315852	.7745364
72	.0796074	.3272186	0.24	0.808	-.5622826	.7214975
73	.174067	.3833992	0.45	0.650	-.5780299	.926164
74	.1474793	.3634153	0.41	0.685	-.5654163	.8603748
75	-.2721319	1.139635	-0.24	0.811	-2.507703	1.963439
77	.6816608	.5198796	1.31	0.190	-.3381636	1.701485
78	-1.167099	1.138839	-1.02	0.306	-3.401108	1.06691
79	-.3407813	.4136352	-0.82	0.410	-1.152191	.4706283
80	-2.362389	1.138268	-2.08	0.038	-4.595278	-.129501
81	.0103333	1.145167	0.01	0.993	-2.23609	2.256757
82	.3843111	.397592	0.97	0.334	-.3956273	1.16425
85	.3533946	.3976831	0.89	0.374	-.4267225	1.133512
86	.6693861	.5197683	1.29	0.198	-.3502201	1.688992
87	.9890407	.8392591	1.18	0.239	-.6572961	2.635377
88	.0341248	.6335398	0.05	0.957	-1.208662	1.276911
90	.7887454	.7074698	1.11	0.265	-.5990663	2.176557
91	-1.276365	.7076498	-1.80	0.071	-2.66453	.1118002
93	-.2526583	.5840938	-0.43	0.665	-1.398449	.8931323
94	.3039177	1.138958	0.27	0.790	-1.930326	2.538161
95	-.2390154	1.138881	-0.21	0.834	-2.473107	1.995076
96	.3828567	.5227684	0.73	0.464	-.6426345	1.408348
_cons	10.0009	.9952639	10.05	0.000	8.048537	11.95327

Figure 26: Regression coefficients with control variables- Version 1-condition a.

Even if the fit of the model seemed to be quite good the coefficients of the two independent variables show a positive but very weak relationship with the duration of the startup. Another controversial aspect is linked to the following control variables:

- Unemployment rate should have a negative impact on the duration, but its coefficient is positive and statistically significant. This is unusual because it is as if we're saying that, all other things being equal, an increase in the unemployment rate is associated with an increase in the lifespan of female-led startups which is something not realistically realizable especially if high unemployment is expected to create a difficult economic environment for enterprises.
- The value added created by the startup and the value added per habitant have negative coefficients while instead they should be positive because both are a measure, respectively, of the wealth of the startups and of the potential customers. In addition, value added is not statistically significant.



- The number of graduated residents has a negative coefficient meaning that as the number of graduated residents increases, the duration of startups tends to decrease which is again something that is not true since they represent potential new skills and professionalism. Same for the number of total debts: this variable as a positive and statistically significant coefficient and we don't expect that if debts increase also the life of a startup increases.
- To consider the industrial sector in which the startups operate we consider the variable `i.ateco_2007_2d` that on STATA is perceived as a dummy variable and so each sector is represented separately. Looking at the coefficients and at the p-values of each industrial sector we see that most of them are negatives and are all not statistically significant, so the industrial sector does not have a relevant impact on the duration of the startup at a statistical level. It could be considered to remove the industrial sectors from the set of control variables, but the result remains almost the same as shown in the figure below:

Source	SS	df	MS	Number of obs	=	1,478
Model	830.262613	9	92.2514014	F(9, 1468)	=	77.62
Residual	1744.6359	1,468	1.18844407	Prob > F	=	0.0000
				R-squared	=	0.3224
				Adj R-squared	=	0.3183
Total	2574.89851	1,477	1.74333007	Root MSE	=	1.0902

duration_startup_guida~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tot_servizi_attivi	.0051589	.0005402	9.55	0.000	.0040993	.0062185
donne_immigrate	.0000676	3.18e-06	21.26	0.000	.0000614	.0000739
unemployment_rate	.0716922	.0201343	3.56	0.000	.0321971	.1111874
value_added_per_habitant	-.0000959	.0000134	-7.15	0.000	-.0001222	-.0000696
residenti_laureati	-.0007242	.00004	-18.11	0.000	-.0008026	-.0006458
area_km2	-.0001865	.0000267	-7.00	0.000	-.0002388	-.0001342
dipendenti	.021346	.0135343	1.58	0.115	-.0052027	.0478947
tot_debiti	.0003243	.0001364	2.38	0.018	.0000568	.0005919
valore_aggiunto	-.0003961	.0004602	-0.86	0.390	-.0012989	.0005068
_cons	10.00329	.9235118	10.83	0.000	8.191749	11.81484

Figure 27: Model's fit and Regression coefficients without the variable `i.ateco_2007_2d` – Version 1-condition a.

**Condition b.**

**Model's Fit**

Source	SS	df	MS	Number of obs	=	1,478
Model	3350.71027	73	45.9001407	F(73, 1404)	=	14.62
Residual	4407.3628	1,404	3.13914729	Prob > F	=	0.0000
				R-squared	=	0.4319
				Adj R-squared	=	0.4024
Total	7758.07307	1,477	5.2525884	Root MSE	=	1.7718

Figure 28: Model's Fit with control variables- Version 1-condition b.

Also, for the case in which we evaluate the duration of consolidated startups we have that by introducing the control variables the fit of the model increases:

- **The p-value associated to test F:** it is equal to 0 that of course is  $<0,05$ ) so we can strongly reject the null hypothesis, and the model is statistically significant.
- **The adjusted  $R^2$ :** it is a quite good value because we have that the 43,19% of the variability in the dependent variable is explained by the independent variables in our model.
- **Root MSE:** still not near 0, but it equal to 1.7718 a bit higher than case a.

## Analysis of Regression Coefficients

duration_startup_guida~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tot_servizi_attivi	.0074824	.0008945	8.36	0.000	.0057276	.0092372
donne_immigrate	.000117	5.30e-06	22.10	0.000	.0001067	.0001274
unemployment_rate	.3429022	.0337612	10.16	0.000	.2766744	.4091301
value_added_per_habitant	-.0001561	.0000223	-6.99	0.000	-.0001999	-.0001123
residenti_laureati	-.001236	.0000666	-18.57	0.000	-.0013666	-.0011054
area_km2	-.0002969	.0000444	-6.68	0.000	-.0003841	-.0002097
dipendenti	.0513092	.0226601	2.26	0.024	.006858	.0957604
tot_debiti	.00037	.0002333	1.59	0.113	-.0000876	.0008276
valore_aggiunto	-.0011495	.0007714	-1.49	0.136	-.0026628	.0003637
ateco_2007_2d						
3	.06197	1.148381	0.05	0.957	-2.190758	2.314698
10	.4157681	.8109185	0.51	0.608	-1.174974	2.00651
11	.4246635	1.855207	0.23	0.819	-3.214612	4.063939
13	.2727734	1.026393	0.27	0.790	-1.740655	2.286202
14	-.3827837	.9491249	-0.40	0.687	-2.244639	1.479072
15	-.108971	1.02695	-0.11	0.916	-2.123492	1.90555
16	-1.336009	1.863766	-0.72	0.474	-4.992075	2.320057
18	2.218005	1.147113	1.93	0.053	-.0322351	4.468245
20	-.2677378	.671627	-0.40	0.690	-1.585238	1.049763
21	.8163642	1.848033	0.44	0.659	-2.80884	4.441568
22	.0084477	.783141	0.01	0.991	-1.527805	1.5447
24	1.287734	1.155465	1.11	0.265	-.9788889	3.554357
25	.0330111	.7626275	0.04	0.965	-1.463001	1.529023
26	.1932398	.6212949	0.31	0.756	-1.025526	1.412006
27	-.5157857	.7006301	-0.74	0.462	-1.89018	.8586089
28	-.4376334	.6211223	-0.70	0.481	-1.656061	.7807944
29	-.4924388	.8476564	-0.58	0.561	-2.155248	1.170371
30	.4717644	1.030573	0.46	0.647	-1.549865	2.493394
31	-.1734147	1.146667	-0.15	0.880	-2.42278	2.075951
32	-1.210407	.6983251	-1.73	0.083	-2.58028	.1594664
35	.4996644	.7611814	0.66	0.512	-.9935109	1.99284
36	-.2693477	1.853682	-0.15	0.884	-3.905632	3.366937
38	-.136033	1.027842	-0.13	0.895	-2.152305	1.880239
41	.4963925	.9469247	0.52	0.600	-1.361147	2.353932
43	2.301701	1.148529	2.00	0.045	.0486845	4.554718
45	.1372259	1.357312	0.10	0.919	-2.525351	2.799803
46	-.2040345	.633856	-0.32	0.748	-1.447441	1.039372
47	-.4553607	.5825807	-0.78	0.435	-1.598183	.6874616
49	-.8710416	1.850269	-0.47	0.638	-4.500631	2.758548
52	-.7689237	.9448557	-0.81	0.416	-2.622405	1.084557
53	1.995705	1.358493	1.47	0.142	-.6691903	4.660599
55	.7220063	1.027023	0.70	0.482	-1.29266	2.736672
56	-.0991645	.7321745	-0.14	0.892	-1.535438	1.337109
58	-.3440779	.6039882	-0.57	0.569	-1.528894	.8407387
59	-.2222502	.7860766	-0.28	0.777	-1.764261	1.319761
61	3.05493	1.361181	2.24	0.025	.3847617	5.725099

62	-.4307044	.522683	-0.82	0.410	-1.456028	.5946193
63	-.5171877	.5351278	-0.97	0.334	-1.566924	.5325485
64	.3387658	.8925802	0.38	0.704	-1.412169	2.0897
66	-.6370566	1.357418	-0.47	0.639	-3.299842	2.025729
68	-1.032307	1.024017	-1.01	0.314	-3.041075	.9764615
69	1.990078	1.357101	1.47	0.143	-.6720856	4.652242
70	.1548524	.5734285	0.27	0.787	-.9700166	1.279721
71	-.21047	.6232023	-0.34	0.736	-1.432978	1.012038
72	-.2101005	.5312019	-0.40	0.693	-1.252135	.8319345
73	-.0603907	.6224046	-0.10	0.923	-1.281334	1.160553
74	-.3034993	.5899632	-0.51	0.607	-1.460804	.8538049
75	-.1080618	1.850067	-0.06	0.953	-3.737255	3.521131
77	.874934	.8439649	1.04	0.300	-.7806341	2.530502
78	-1.78389	1.848774	-0.96	0.335	-5.410547	1.842768
79	-1.337094	.6714894	-1.99	0.047	-2.654324	-.0198632
80	-4.180291	1.847847	-2.26	0.024	-7.805129	-.5554525
81	.1506925	1.859048	0.08	0.935	-3.496119	3.797504
82	-.0904269	.6454451	-0.14	0.889	-1.356568	1.175714
85	-.0775893	.645593	-0.12	0.904	-1.34402	1.188842
86	1.691861	.8437843	2.01	0.045	.0366473	3.347075
87	.4796289	1.362441	0.35	0.725	-2.19301	3.152268
88	-1.073124	1.028479	-1.04	0.297	-3.090646	.9443972
90	1.19793	1.148496	1.04	0.297	-1.055023	3.450883
91	-1.807909	1.148788	-1.57	0.116	-4.061436	.4456174
93	.0301979	.9482094	0.03	0.975	-1.829862	1.890258
94	1.148201	1.848968	0.62	0.535	-2.478837	4.775239
95	-.9098496	1.848842	-0.49	0.623	-4.536641	2.716941
96	-.0268722	.8486546	-0.03	0.975	-1.69164	1.637895
_cons	13.42341	1.615697	8.31	0.000	10.25397	16.59285

Figure 29: Regression coefficients with control variables- Version 1-condition b.

The considerations made for case a. are valid also for case b. since the number of coefficients is different but the signs and the significance of all the variables explained before remain the same. The only difference is that now, the total number of debts has still a positive coefficient, but it is no more statistically significant, so we cannot conclude saying that there's a positive correlation between the total number of debts and the duration of a consolidated startup, which is a reasonable outcome.

## Version 2

The second version is the one in which the hypothesis to be tested remains the same (Hp1) but the independent variables are changed and are transformed into two ratios because of the strong correlation between variables<sup>12</sup>.

To develop this regression, it has been chosen a particular type of regression: the **log-linear regression model**. It consists in transforming the dependent variable through the natural logarithm while the independent variables are kept in their linear form; our specific version will also contain some control variables in the log form too. The reason behind the choice of using this type of regression is linked to the fact that we want to examine the *percentual variation* of the duration of a female-led innovative startups (when duration is set maximum to 5 years) and of female-led consolidated enterprises (when duration is not set to a maximum of 5 years), in addition, the model it's able to manage potentially non-linear relations between variables.

### Log-Linear regression model:

$$\begin{aligned} \ln\_duration\_startup\_donne_i \\ &= \beta_0 + \beta_1 rate\_donne\_immigrate_{jt} + \beta_2 tasso\_servizi_{jt} \\ &+ \beta_3 control\ variable_{ijt} + u_{ijt} \end{aligned}$$

The vector *control variable*<sub>ijt</sub> contains a set of control variables related to the startup *i* to the province *j* and the year *t*.

- Some have been transformed in the log form:
  - *ln\_laureati* is the log form of the variable *resident\_laureati*, hence the number of degreed people resident in a certain province of the North-west of Italy
  - *ln\_value\_added\_inhabitant* is the log form of the variable *value\_added\_per\_habitant* which indicated the wellness of the population

The coefficients of these logarithmic control variables will indicate a percentual change in the dependent variable for a percentual change of each control variable.

---

<sup>12</sup> For more details refer to paragraph "Variable Description".

For example, if the coefficient of the variable *ln\_laureati* is 0,10, an increase of 1% in *residenti\_laureati* means an increase of 0,10% in the *duration\_startup\_guida\_femminile*.

- Some are kept in a linear form:
  - *area\_km2*
  - *dipendenti*
  - *valore\_aggiunto*
  - *unemployment\_rate*
  - *tot\_debiti*

The coefficients of these linear control variables will indicate a percentual change in the dependent variable for each unit increase in these control variables.

In other words, if the coefficient of the variable *area\_km2* is -0,02, an increase of one unit in *area\_km2* means a decrease of 2% in the *duration\_startup\_guida\_femminile*.

### Condition a.

#### Model's Fit

Source	SS	df	MS	Number of obs	=	1,478
Model	45.4519685	73	.622629705	F(73, 1404)	=	3.57
Residual	244.747771	1,404	.174321775	Prob > F	=	0.0000
				R-squared	=	0.1566
				Adj R-squared	=	0.1128
Total	290.19974	1,477	.196479174	Root MSE	=	.41752

Figure 30: Model's Fit - Version 2-condition a

Let's look at the Model's Fit when we're estimating the variation of the female-led innovative startup.

As before we look at the three most important measures of goodness of the model:

- **The p-value associated to test F:** it is equal to 0 (<0,05), we can strongly reject the null hypothesis, and the model is statistically significant.
- **The adjusted R<sup>2</sup>:** the value is not so high, it is equal to 0,1128, meaning that just the 11,20% of the variability in the dependent variable is explained by the independent variables in our model.

- **Root MSE:** a very low value of RMSE (near 0) suggests that the model has a higher predictive accuracy. Our value is quite good because it indicates that the average error of the previsions is of 0,41752 years.

## Analysis of Regression Coefficients

ln_duration_startup_donne	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
rate_donne_immigrate	7.579706	3.123356	2.43	0.015	1.452759 13.70665
tasso_servizi	1.326344	.4207253	3.15	0.002	.5010257 2.151661
ln_laureati	.1594601	.0343677	4.64	0.000	.0920424 .2268777
ln_value_added_inhabitant	-3.148818	.3333659	-9.45	0.000	-3.802767 -2.494869
area_km2	-.0000633	.0000121	-5.25	0.000	-.000087 -0.0000397
dipendenti	.0126251	.0053281	2.37	0.018	.0021733 .0230769
valore_aggiunto	-.0003131	.0001817	-1.72	0.085	-.0006695 .0000433
unemployment_rate	.0401866	.0074001	5.43	0.000	.0256701 .0547031
tot_debiti	.000106	.000055	1.93	0.054	-1.93e-06 .0002139
ateco_2007_2d					
3	-.1553227	.2706036	-0.57	0.566	-.6861536 .3755083
10	-.0256443	.1908695	-0.13	0.893	-.4000644 .3487759
11	.5316102	.4368188	1.22	0.224	-.3252776 1.388498
13	-.0045955	.2416631	-0.02	0.985	-.4786552 .4694643
14	.0861028	.2230414	0.39	0.700	-.3514276 .5236331
15	.1284475	.2417622	0.53	0.595	-.3458065 .6027015
16	.0394787	.4391272	0.09	0.928	-.8219373 .9008947
18	.3082215	.2706043	1.14	0.255	-.2226108 .8390539
20	.1190618	.1583635	0.75	0.452	-.1915927 .4297163
21	.4249988	.4352911	0.98	0.329	-.4288922 1.27889
22	.1320691	.1845793	0.72	0.474	-.2300118 .49415
24	.2613686	.2721367	0.96	0.337	-.2724697 .7952069
25	.025335	.1799696	0.14	0.888	-.3277032 .3783732
26	.0632243	.1464521	0.43	0.666	-.2240641 .3505127
27	-.0335157	.1650962	-0.20	0.839	-.3573776 .2903462
28	-.085189	.1460634	-0.58	0.560	-.371715 .201337
29	.0575518	.1997457	0.29	0.773	-.3342802 .4493839
30	-.0574703	.242712	-0.24	0.813	-.5335876 .418647
31	.2041652	.2701254	0.76	0.450	-.3257276 .734058
32	-.1182126	.1644841	-0.72	0.472	-.4408737 .2044484
35	.277383	.1790539	1.55	0.122	-.073859 .6286249
36	.2241383	.436878	0.51	0.608	-.6328655 1.081142
38	.3451422	.2422095	1.42	0.154	-.1299893 .8202738
41	.3546613	.2230384	1.59	0.112	-.0828631 .7921856
43	.0573249	.2704961	0.21	0.832	-.4732951 .587945
45	.0720994	.3197349	0.23	0.822	-.5551103 .699309
46	.1550711	.1492763	1.04	0.299	-.1377575 .4478997
47	-.0003973	.1372282	-0.00	0.998	-.2695916 .268797
49	-.1135006	.4354965	-0.26	0.794	-.9677944 .7407933
52	.1501163	.2228729	0.67	0.501	-.2870835 .5873161
53	.1353266	.3201597	0.42	0.673	-.4927164 .7633695
55	.3964883	.241822	1.64	0.101	-.0778832 .8708597
56	.0314859	.1724877	0.18	0.855	-.3068754 .3698473
58	-.0190951	.1420906	-0.13	0.893	-.2978278 .2596376
59	.2123159	.1853365	1.15	0.252	-.1512504 .5758822
61	.1908567	.3212549	0.59	0.553	-.4393346 .821048
62	.0486593	.123156	0.40	0.693	-.1929303 .2902489
63	.030515	.1260138	0.24	0.809	-.2166806 .2777106
64	.2356618	.210095	1.12	0.262	-.176472 .6477956
66	.0817233	.3197239	0.26	0.798	-.5454646 .7089113
68	.0231306	.241356	0.10	0.924	-.4503267 .4965878
69	.1320758	.3198196	0.41	0.680	-.4953 .7594516
70	.159406	.1350744	1.18	0.238	-.1055634 .4243753
71	-.0053505	.1468406	-0.04	0.971	-.2934011 .2827001
72	.0475732	.1251053	0.38	0.704	-.1978402 .2929866
73	.028295	.1465547	0.19	0.847	-.2591947 .3157848
74	.113302	.1390096	0.82	0.415	-.1593869 .3859908
75	-.0781324	.4360447	-0.18	0.858	-.9335017 .7772369
77	.2748904	.1987343	1.38	0.167	-.1149577 .6647385
78	-.4990876	.4351595	-1.15	0.252	-.1.35272 .3545452
79	-.0214852	.1583844	-0.14	0.892	-.3321807 .2892104
80	-.4918557	.4352447	-1.13	0.259	-.1.345656 .3619443
81	.0736212	.4381209	0.17	0.867	-.7858208 .9330632
82	.1047288	.1519837	0.69	0.491	-.1934109 .4028685
85	.0608804	.1520719	0.40	0.689	-.2374322 .359193
86	.2729158	.1990347	1.37	0.171	-.1175216 .6633532
87	.3936584	.3211204	1.23	0.220	-.2362691 1.023586
88	.1842235	.2421882	0.76	0.447	-.2908662 .6593133
90	.1248137	.2705669	0.46	0.645	-.4059452 .6555726
91	-.5605526	.2704976	-2.07	0.038	-.1.091176 -.0299295
93	-.0896895	.2233371	-0.40	0.688	-.5277998 .3484209
94	.3938568	.4353546	0.90	0.366	-.4601587 1.247872
95	-.125846	.4351658	-0.29	0.772	-.9794913 .7277992
96	.0938237	.1995616	0.47	0.638	-.2976474 .4852948
cons	34.35293	3.530791	9.73	0.000	27.42674 41.27913

Figure 31: Regression coefficients - Version 2-condition a.



Let's start with the two independent variables of interests:

- ***rate\_donne\_immigrate***: it given by the immigrated low-skilled women aged 20-65 years resident in a certain province of the northwestern Italy and the total population of that area. Its coefficient is positive (7,579706) and statistically significant which is exactly the expected result. This suggests that an increase of one unit in the number of immigrated women is associated with an increase of 757,9726% in the duration of a female-led innovative startup.
- ***tasso\_servizi***: this rate was constructed as the total number of childcare services active on a certain province over the area in  $Km^2$  of the geographical territory. Also in this case we got the expected result since we have that an increase of one unit in the childcare services leads to a an increase in the duration of a female-led innovative startup of 132,6%.

For what concern the other variables we have that:

- ***ln\_laureati***: the number of graduated people, as predicted, has a positive correlation with the dependent variable with a positive and significant coefficient of 1.594621. It can be argued that a more educated environment is positive associated with the duration of a startup lead by women (an increase of 1% of graduated residents brings to an increase of 1,59% for the duration).
- ***dipendenti***: the coefficient linked to the number of employees is again positive and significant but not too high suggesting that an increase of one unit in the human resources increases the duration of the startup of 1,26%.

Now, let's analyze the issues of the model:

- ***tot\_debiti***: this time, the number of total debts owned by the startup has positive coefficient but it is not statistically significant so there's not a robust evidence of the effects of this variable on the dependent variable.
- ***ln\_value\_added\_inhabitant***: the value added per habitant has a negative coefficient (-3,148813) and it is highly significant, which is not a good news: when the value added per habitant increases of 1 %, the durations decreases of 3,15%.

- **valore\_aggiunto**: same as the previous case happens for the value added created by the startups.
- **unemployment\_rate**: an increase of one unit in the unemployment rate seems to be favorable for the increase of the duration of a startup (an increase of 4,0186%) which again not realistic.
- **i.ateco\_2007\_2d**: As in the first version a lot of the coefficient relative to the ateco codes are not significant indicating that the specific sector does not have a determinant impact on the startup's duration.

**Condition b.**

**Model's Fit**

Source	SS	df	MS	Number of obs	=	1,478
Model	105.280219	73	1.44219479	F(73, 1404)	=	5.57
Residual	363.45743	1,404	.258872813	Prob > F	=	0.0000
				R-squared	=	0.2246
				Adj R-squared	=	0.1843
Total	468.737649	1,477	.317357921	Root MSE	=	.5088

Figure 32: Model's Fit - Version 2-condition b.

The fit of the model is quite the same when the aim is to evaluate the duration of all that startups that have scaled up. The p-value is still 0 (<0,05), the adjusted  $R^2$  increases a bit (from 0,1128 to 0,1843 ) and the RMSE gets a little worse.

## Analysis of Regression Coefficient

ln_duration_startup_donne	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
rate_donne_immigrate	16.76629	3.806177	4.41	0.000	9.299885 24.2327
tasso_servizi	2.139941	.5127032	4.17	0.000	1.134194 3.145687
ln_laureati	.1618516	.0418811	3.86	0.000	.0796952 .2440079
ln_value_added_inhabitant	-4.312389	.4062456	-10.62	0.000	-5.109303 -3.515476
area_km2	-.0000812	.0000147	-5.52	0.000	-.00011 -0.0000523
dipendenti	.0195174	.0064929	3.01	0.003	.0067806 .0322542
valore_aggiunto	-.0004608	.0002214	-2.08	0.038	-.0008951 -.0000265
unemployment_rate	.0885519	.0090179	9.82	0.000	.0708618 .1062419
tot_debiti	.0001138	.000067	1.70	0.090	-.0000177 .0002452
ateco_2007_2d					
3	-.317087	.3297623	-0.96	0.336	-.9639668 .3297929
10	-.0753655	.2325969	-0.32	0.746	-.5316404 .3809094
11	.5472719	.532315	1.03	0.304	-.4969466 1.59149
13	-.0607429	.2944949	-0.21	0.837	-.6384404 .5169546
14	-.0416878	.2718022	-0.15	0.878	-.5748699 .4914944
15	.0951232	.2946156	0.32	0.747	-.482811 .6730574
16	-.0531141	.535128	-0.10	0.921	-1.102851 .9966225
18	.373766	.3297632	1.13	0.257	-.2731156 1.020648
20	.0523259	.1929845	0.27	0.786	-.3262431 .4308949
21	.5542697	.5304533	1.04	0.296	-.4862968 1.594836
22	.127202	.2249315	0.57	0.572	-.3140361 .5684401
24	.4073358	.3316305	1.23	0.220	-.2432089 1.057881
25	-.0310275	.219314	-0.14	0.888	-.461246 .399191
26	.0403021	.178469	0.23	0.821	-.3097926 .3903968
27	-.0897748	.7011897	-0.44	0.657	-.4838885 .3054389
28	-.1441765	.1779954	-0.81	0.418	-.4933421 .2049891
29	-.0725942	.2434135	-0.30	0.766	-.5500876 .4048992
30	-.0284895	.2957731	-0.10	0.923	-.6086943 .5517154
31	.0561823	.3291795	0.17	0.865	-.5895543 .701919
32	-.2134443	.2004432	-1.06	0.287	-.6066447 .1797561
35	.2648083	.2181982	1.21	0.225	-.1632212 .6928379
36	.1512747	.5323871	0.28	0.776	-.8930852 1.195635
38	.2656327	.2951608	0.90	0.368	-.313371 .8446363
41	.3858776	.2717985	1.42	0.156	-.1472972 .9190524
43	.1951053	.3296313	0.59	0.554	-.4515176 .8417282
45	.0230874	.3896346	0.06	0.953	-.7412412 .7874161
46	.0781902	.1819107	0.43	0.667	-.2786559 .4350362
47	-.0867421	.1672286	-0.52	0.604	-.414787 .2413027
49	-.2794094	.5307036	-0.53	0.599	-1.320467 .761648
52	.01908	.2715968	0.07	0.944	-.5136993 .5518593
53	.2949412	.3901522	0.76	0.450	-.4704029 1.060285
55	.3868972	.2946886	1.31	0.189	-.1911801 .9649745
56	.0155366	.2101965	0.07	0.941	-.3967964 .4278697
58	-.0767908	.173154	-0.44	0.657	-.4164593 .2628777
59	.1369619	.2258543	0.61	0.544	-.3060863 .5800102
61	.4794839	.3914868	1.22	0.221	-.2884782 1.247446
62	-.0299859	.1500801	-0.20	0.842	-.3243912 .2644194
63	-.0418114	.1535626	-0.27	0.785	-.3430482 .2594255
64	.2716207	.2560254	1.06	0.289	-.2306128 .7738542
66	-.0630262	.3896211	-0.16	0.872	-.8273284 .701276
68	-.1249709	.2941207	-0.42	0.671	-.7019343 .4519924
69	.2890561	.3897378	0.74	0.458	-.475475 1.053587
70	.1377774	.164604	0.84	0.403	-.1851189 .4606737
71	-.0634053	.1789425	-0.35	0.723	-.4144288 .2876182
72	.0059269	.1524555	0.04	0.969	-.2931382 .3049919
73	-.0078845	.1785941	-0.04	0.965	-.3582246 .3424555
74	.0517124	.1693995	0.31	0.760	-.280591 .3840159
75	-.0769076	.5313717	-0.14	0.885	-1.119276 .9654604
77	.3304277	.242181	1.36	0.173	-.1446479 .8055034
78	-.654367	.530293	-1.23	0.217	-1.694619 .3858849
79	-.145178	.19301	-0.75	0.452	-.523797 .233441
80	-.6973724	.5303968	-1.31	0.189	-1.737828 .3430832
81	.107345	.5339017	0.20	0.841	-.9399861 1.154676
82	.025418	.18521	0.14	0.891	-.3379002 .3887362
85	-.0044146	.1853175	-0.02	0.981	-.3679436 .3591144
86	.4350523	.2425471	1.79	0.073	-.0407415 .9108462
87	.3722533	.391323	0.95	0.342	-.3953874 1.139894
88	.0696752	.2951348	0.24	0.813	-.5092775 .6486279
90	.1406935	.3297175	0.43	0.670	-.5060985 .7874855
91	-.639028	.3296331	-1.94	0.053	-1.285655 .0075984
93	-.0332951	.2721625	-0.12	0.903	-.567184 .5005938
94	.6283864	.5305307	1.18	0.236	-.4123318 1.669105
95	-.2931808	.5303007	-0.55	0.580	-1.333448 .7470862
96	.0186198	.2431893	0.08	0.939	-.4584337 .4956734
cons	46.71411	4.302684	10.86	0.000	38.27373 55.15449

Figure 33: Regression coefficients - Version 2-condition b.

The results obtained considering the model under the condition b. are the same as case a., just a little change in the values of the coefficients that have increased for *rate\_donne\_immigrate*, *tasso\_servizi*, and also for the other variables including the negative coefficients. So, the issues remain the same as before.

### Step-by-step Strategy

In this regard, it could be interesting introducing a “step-by-step” strategy which consists in taking the basic log-linear regression model, remove the three variables that does not show realistic results, hence, *unemployment\_rate*, *ln\_value\_added\_inhabitant* and *valore\_aggiunto*, and add them, one by one to the regression equation to see alone their effects on the dependent variable. For convenience the *i.ateco\_2007\_2d* variable is excluded because it does not have an impact on the final results.

Also in this case we will consider the two conditions:

#### Condition a.

Before implementing the step-by step strategy we need to look at the results of the model obtained by eliminating the variables mentioned before:

Source	SS	df	MS	Number of obs	=	1,666
Model	12.9301392	6	2.1550232	F(6, 1659)	=	6.77
Residual	528.380771	1,659	.318493533	Prob > F	=	0.0000
				R-squared	=	0.0239
				Adj R-squared	=	0.0204
Total	541.31091	1,665	.325111658	Root MSE	=	.56435

ln_duration_startu~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
rate_donne_immigrate	-11.43359	3.104755	-3.68	0.000	-17.52324 -5.343935
tasso_servizi	-1.603629	.5092598	-3.15	0.002	-2.602489 -.6047694
ln_laureati	.0893246	.0386363	2.31	0.021	.0135435 .1651057
area_km2	-.0000464	.000013	-3.58	0.000	-.0000718 -.000021
dipendenti	.0165816	.0063182	2.62	0.009	.004189 .0289741
tot_debiti	.0000272	.0000616	0.44	0.659	-.0000937 .0001481
_cons	1.115501	.2101089	5.31	0.000	.703394 1.527607

Figure 34: Step-by-step strategy – Version 2- condition a.

The results obtained are definitely, out of expectation suggesting that the three eliminated variables could be important to better test our hypothesis and to better explain the variation of the duration.

## First step

The first step consists of adding the variable *ln\_value\_added\_inhabitant*:

Source	SS	df	MS	Number of obs	=	1,478
Model	30.4218202	7	4.34597432	F(7, 1470)	=	24.59
Residual	259.77792	1,470	.176719673	Prob > F	=	0.0000
				R-squared	=	0.1048
				Adj R-squared	=	0.1006
Total	290.19974	1,477	.196479174	Root MSE	=	.42038

ln_duration_startup_donne	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rate_donne_immigrate	7.666743	3.07658	2.49	0.013	1.631787	13.7017
tasso_servizi	1.600202	.4144585	3.86	0.000	.7872088	2.413195
ln_laureati	.1719207	.0335925	5.12	0.000	.1060264	.2378151
area_km2	-.0000563	.0000119	-4.74	0.000	-.0000796	-.000033
dipendenti	.0106744	.0049457	2.16	0.031	.000973	.0203758
tot_debiti	.0001086	.0000525	2.07	0.039	5.56e-06	.0002116
ln_value_added_inhabitant	-3.445097	.3220461	-10.70	0.000	-4.076815	-2.813378
_cons	37.84077	3.403132	11.12	0.000	31.16525	44.51628

Figure 35: Step-by-step strategy –first step- Version 2- condition a.

By considering this variable as a control variable for our model the results increase a lot. Unfortunately, the issue related to this variable persists, the coefficient is again highly negative and statistically significant.

## Second step

In the second step, the previous variable is eliminated and on its place is added the variable *valore\_aggiunto*:

Source	SS	df	MS	Number of obs	=	1,666
Model	13.5754702	7	1.93935289	F(7, 1658)	=	6.09
Residual	527.73544	1,658	.318296405	Prob > F	=	0.0000
				R-squared	=	0.0251
				Adj R-squared	=	0.0210
Total	541.31091	1,665	.325111658	Root MSE	=	.56418

ln_duration_startu~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rate_donne_immigrate	-11.44071	3.103798	-3.69	0.000	-17.52849	-5.352938
tasso_servizi	-1.607597	.5091098	-3.16	0.002	-2.606163	-.6090313
ln_laureati	.0891405	.0386246	2.31	0.021	.0133824	.1648986
area_km2	-.0000461	.000013	-3.56	0.000	-.0000715	-.0000207
dipendenti	.019521	.0066451	2.94	0.003	.0064873	.0325546
tot_debiti	.0000232	.0000617	0.38	0.707	-.0000978	.0001442
valore_aggiunto	-.0002894	.0002033	-1.42	0.155	-.0006881	.0001093
_cons	1.115727	.210044	5.31	0.000	.703748	1.527707

Figure 36: Step-by-step strategy - second step – Version 2- condition a.

The inclusion of the value added created by the business as a control variable makes the results worse again and its coefficient is still negative and not significant. Probably this result might be due to the fact that the information about the value added created by the business is not available for all the observations, there are a lot of missing values that limit the ability of the model to capture the real effect of the variable because there is a smaller number of observations.

### Third step

In the third step, the previous variable is eliminated, and it is integrated the variable *unemployment\_rate*:

Source	SS	df	MS	Number of obs	=	1,666
Model	75.3836818	7	10.7690974	F(7, 1658)	=	38.32
Residual	465.927228	1,658	.281017629	Prob > F	=	0.0000
				R-squared	=	0.1393
				Adj R-squared	=	0.1356
Total	541.31091	1,665	.325111658	Root MSE	=	.53011

ln_duration_startu~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rate_donne_immigrate	-3.968431	2.959057	-1.34	0.180	-9.772313	1.835451
tasso_servizi	-1.564249	.4783684	-3.27	0.001	-2.502519	-.625979
ln_laureati	.0488924	.0363933	1.34	0.179	-.0224893	.1202741
area_km2	-.0000668	.0000123	-5.45	0.000	-.0000909	-.0000428
dipendenti	.012698	.0059406	2.14	0.033	.0010462	.0243499
tot_debiti	.0000551	.0000579	0.95	0.342	-.0000586	.0001687
unemployment_rate	.1237307	.0082998	14.91	0.000	.1074516	.1400098
_cons	.3405407	.2040922	1.67	0.095	-.0597649	.7408462

Figure 37: Step-by-step strategy- third step – Version 2- condition a.

The introduction of *unemployment\_rate* has worsened the coefficients of the two principal independent variables and its coefficient is still positive and strongly significant. Probably the inclusion of this variable may have introduced collinearity with the others control variables creating changes in their coefficients and in their significance.

Indeed, by verifying the correlation between the variables contained in the last regression model, we see that *unemployment\_rate* is negative correlated with *rate\_donne\_immigrate* and *ln\_laureati* which indeed, have lost their significance.

	ln_dur~e	rate_d~e	tasso~i	ln_lau~i	area_km2	dipend~i	tot_de~i	unempl~e
ln_duratio~e	1.0000							
rate_donne~e	-0.1060	1.0000						
tasso_serv~i	-0.0806	0.6503	1.0000					
ln_laureati	-0.0865	0.7710	0.6799	1.0000				
area_km2	0.0017	-0.4111	-0.6570	-0.1411	1.0000			
dipendenti	0.0623	0.0536	0.0632	0.0504	-0.0360	1.0000		
tot_debiti	0.0272	0.0391	0.0288	0.0375	-0.0248	0.2805	1.0000	
unemploye~e	0.3394	-0.2546	-0.2158	-0.1007	0.2927	0.0189	-0.0301	1.0000

Figure 38: Step-by-step strategy – correlation between variables- Version 2- condition a.

### Condition b.

When evaluating the variation of the duration of consolidated enterprises when the variables *ln\_value\_added\_inhabitant*, *valore\_aggiunto* and *unemployment\_rate* are eliminated, what we obtain is the same situations as before: more or less the values of the coefficients are the same with the only difference that this time the variable *tasso\_servizi* is not statistically significant since the beginning, same for *ln\_laureati*.

Source	SS	df	MS	Number of obs	=	1,666
Model	17.1946231	6	2.86577052	F(6, 1659)	=	6.35
Residual	748.624135	1,659	.451250232	Prob > F	=	0.0000
				R-squared	=	0.0225
				Adj R-squared	=	0.0189
Total	765.818758	1,665	.459951206	Root MSE	=	.67175

ln_duration_startu~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rate_donne_immigrate	-12.30528	3.695607	-3.33	0.001	-19.55383	-5.05674
tasso_servizi	-1.10196	.6061747	-1.82	0.069	-2.290908	.0869877
ln_laureati	.0571003	.045989	1.24	0.215	-.0331024	.1473029
area_km2	-.000037	.0000154	-2.40	0.017	-.0000673	-6.72e-06
dipendenti	.0235059	.0075206	3.13	0.002	.008755	.0382568
tot_debiti	.0000327	.0000734	0.45	0.655	-.0001112	.0001767
_cons	1.449451	.2500938	5.80	0.000	.9589179	1.939983

Figure 39: Step-by-step strategy – Version 2- condition b.

### First step

By introducing the variable *ln\_value\_added\_inhabitant*, again the fit of the model increases, the coefficients have a realistic shape but the added variable still exhibits the issues already pointed out: negative and significant coefficient.

Source	SS	df	MS	Number of obs	=	1,478
Model	62.1838328	7	8.88340468	F(7, 1470)	=	32.12
Residual	406.553816	1,470	.276567222	Prob > F	=	0.0000
				R-squared	=	0.1327
				Adj R-squared	=	0.1285
Total	468.737649	1,477	.317357921	Root MSE	=	.5259

ln_duration_startup_donne	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
rate_donne_immigrate	16.27933	3.848807	4.23	0.000	8.729587 23.82907
tasso_servizi	2.629757	.5184883	5.07	0.000	1.612701 3.646813
ln_laureati	.2047405	.0420243	4.87	0.000	.1223065 .2871745
area_km2	-.0000718	.0000149	-4.83	0.000	-.0001009 -.0000426
dipendenti	.0178974	.0061871	2.89	0.004	.005761 .0300338
tot_debiti	.0001196	.0000657	1.82	0.069	-.9.24e-06 .0002485
ln_value_added_inhabitant	-5.094512	.4028802	-12.65	0.000	-5.884794 -4.304231
_cons	55.64124	4.257324	13.07	0.000	47.29017 63.99232

Figure 40: Step-by-step strategy – first step- Version 2- condition b.

## Second step

As before, the introduction of *valore\_aggiunto* does not have a positive impact on the model, the results are quite the same as before:

Source	SS	df	MS	Number of obs	=	1,666
Model	18.7244959	7	2.67492798	F(7, 1658)	=	5.94
Residual	747.094262	1,658	.450599676	Prob > F	=	0.0000
				R-squared	=	0.0245
				Adj R-squared	=	0.0203
Total	765.818758	1,665	.459951206	Root MSE	=	.67127

ln_duration_startuave	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
rate_donne_immigrate	-12.31626	3.692947	-3.34	0.001	-19.55959 -5.07293
tasso_servizi	-1.10807	.6057467	-1.83	0.068	-2.296179 .0800388
ln_laureati	.0568169	.0459561	1.24	0.217	-.0333212 .1469551
area_km2	-.0000365	.0000154	-2.37	0.018	-.0000668 -6.25e-06
dipendenti	.0280317	.0079064	3.55	0.000	.0125241 .0435393
tot_debiti	.0000266	.0000734	0.36	0.717	-.0001174 .0001705
valore_aggiunto	-.0004456	.0002418	-1.84	0.066	-.0009199 .0000287
_cons	1.4498	.2499135	5.80	0.000	.9596203 1.939979

Figure 41: Step-by-step strategy – second step- Version 2- condition b.

## Third step

With the implementation of *unemployment\_rate*, this time, the most important variables: *rate\_donne\_immigrate*, *tasso\_servizi* and *ln\_laureati* become negative and not significant which is not properly the bad news because this means that there is no statistical evidence of a substantial effect of that variable on the dependent variable.



Source	SS	df	MS	Number of obs	=	1,666
Model	145.58503	7	20.7978615	F(7, 1658)	=	55.60
Residual	620.233728	1,658	.374085481	Prob > F	=	0.0000
				R-squared	=	0.1901
				Adj R-squared	=	0.1867
Total	765.818758	1,665	.459951206	Root MSE	=	.61163

In_duration_startu~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rate_donne_immigrate	-1.601766	3.414066	-0.47	0.639	-8.298102	5.09457
tasso_servizi	-1.045497	.5519263	-1.89	0.058	-2.128043	.0370489
ln_laureati	-.0008712	.0419895	-0.02	0.983	-.0832292	.0814867
area_km2	-.0000663	.0000141	-4.69	0.000	-.000094	-.0000386
dipendenti	.0179377	.0068541	2.62	0.009	.0044941	.0313813
tot_debiti	.0000727	.0000668	1.09	0.277	-.0000584	.0002038
unemployment_rate	.1774047	.009576	18.53	0.000	.1586224	.1961871
_cons	.3383151	.2354751	1.44	0.151	-.1235448	.800175

Figure 42: Step-by-step strategy – third step- Version 2- condition b.

By the analysis conducted it seems that the issues related to the 3 variables persists also considering separately, their effects on the dependent variable whether considering the duration of innovative startups or the duration of startups that have scaled up.

In the next chapter we will discuss the possible reasons behind these problems, and we will point out the limitations found in the model.

## Discussion and Limitations

The main objective of this paper is to estimate the effects of low-skilled immigrated women and the childcare services present in the provinces of North-west of Italy on the duration of an innovative female-led startups and on the lifespan of all that startups that have lived for more than 5 years. These startups exit from the startup register and, if they have all the legal requirements, can be registered in the enterprises' register and be legally considered consolidated enterprises. To do so, two regression models have been created making some adjustments on the variables implemented and considering the two conditions discussed earlier. Both models exhibit more or less the same characteristics: there is a positive correlation between the number of immigrated women, the number of childcare services active on the territory and the survival rate of a female-led business measured in terms of duration in years. In other terms, it means that an increase in the number of immigrated women and in the number of childcare services promotes the long-term duration of a startup founded by a woman entrepreneur in the north-west of Italy, so our hypothesis can be considered satisfied. However, the model shows some non-realistic results linked especially to 3 variables: the unemployment rate which exhibits a positive and statistically significant coefficient, the value added created by the company has a negative coefficient and it's not significant and the value added per capita has a negative coefficient and it is significant. It is believed that the behavior showed by these variables stands in the concept of "**Necessity Entrepreneurship**". According to Robert W. Fairlie, Fank M. Fossen. 2018. "*Opportunity versus Necessity Entrepreneurship: Two Components of Business Creation*" there are two different motivations for starting a business: "*opportunity*" entrepreneurship and "*necessity*" entrepreneurship. With the term "*necessity*" entrepreneurship is intended a condition in which people start businesses not to catch market opportunities but for economic necessity like it may happen because of lack of work or because of the need of searching adequate alternative income. This kind of entrepreneurship is often motivated by adverse economic conditions like for example high unemployment rates. Therefore, we can justify the positive and significant coefficient associated with the variable *unemployment\_rate* by saying that the lack of job opportunities in a certain territory could lead entrepreneurs to keep their businesses alive longer since they represent the

only source of income, so a necessity more than a strategic choice. This assertion is consistent with our central topic which are female-led innovative startups since we have remarked many times in the literature that women entrepreneurs, who decide to found a business, are driven by a sense of “ransom” of their professional role in the labor market since it is known that women are treated differently in terms of wages and responsibilities in many types of jobs. For what concerns the variable *ln\_value\_added\_inhabitant*, hence the value added per capita, we could interpret its negative and significant coefficient by saying that in certain geographical areas in which there’s a high economic wealth there is also a higher competition and so there are less incentives to start a business or it is harder to maintain it alive. While the variable *valore\_aggiunto*, so the value added created by the company has a negative coefficient but it is not statistically significant, this might be due to the fact that there are a lot of missing values in our database that limit the ability of the model to capture the real effect of the variable because there is a smaller number of observations, as highlighted before in analyzing the step-by-step strategy for condition a. This latter point represents one of the limits of our research: the availability of data. During the data research phase to construct the four databases, one of the issues encountered was being able to find datasets that covered the entire period of interest (2013-2022), for example the database about innovative startups in the North-west of Italy contains data from 2010 to 2022, the database about the immigrated women aged 25-56 years in the same area contains data from 2009 to 2023, the childcare services active on the territory are available from 2013 to 2022 and the control variables’ database contains data from 2012-2022. We understand well that by merging the four databases in a unique one, several pieces of information are lost for all the years that are missing. Another limit of the model stands in the fact that there are multiple factors that can influence the startup survival rate and many of these are not available pieces of information. For example, among the control variables, it could have been interesting to consider the age of the startup founder or whether she has children or not, in order to understand if the level of professional experience and the presence of children could have an effect on the entrepreneurship risk-taking, but unfortunately this information is private and is not available. We also thought to consider as an additional dataset the one about Residential Care Facilities that includes social care facilities for non-self-sufficient people such as the elderly or people with disabilities. This information would have

further enriched the care services database and probably contributed to a more precise estimation; however, these data were available on ISTAT just at the regional level.

## **Conclusions and future improvements**

In this paper, we address to a highly discussed topic: the gender gap. Even today, women in the workforce does not enjoy the same opportunities provided to men. This highlights the necessity for some women to create their own business. However, the entrepreneurial path remains quite complex for women who are often seen as those responsible for household and childcare. It is exactly on this last point that our thesis focuses: taking as area of interest the North-west of Italy (as it has the highest number of immigrants and startups), it is possible that low-skilled immigrant women and childcare services can replace native women, founders of innovative startups, in household and family care roles and contribute to the longevity of the firm? The answer to the question is yes; from the results of our empirical model, it emerges that there is a positive and significant correlation between these two aspects and the duration of female-led startups and so our hypothesis is confirmed. However, some issues have emerged linked to some important variables such as the value added created by the startup which exhibits a negative coefficient, and it is not statistically significant maybe due to the fact that for this variable there are a lot of missing values in the database. For this reason, we suggest working on a final database that contains values for each observation. Moreover, it could be interesting having more information about the availability of capital, both from private and public sources, the presence of support network in the territory such as incubator or accelerator and also, considering an extended period of time to conduct the research will help having a more robust and sophisticated model.

## References

- Scattoni, Lombardi, Pini, Turi. 2019. “*Innovative startup localization determinants and origin: A Rome city case study*”
- Iacobucci, D., Iacopini, A. & Micozzi, A., 2014. “*Le start-up innovative*”
- Aburrà, L. and Grandi, A., 2003. “*Il ruolo degli incubatori nella creazione di nuove imprese*”. Torino: Rosemberg & Sellier.
- Bartolomei, G. 2015. “*Come finanziare una start-up innovativa: Guida pratica per l'accesso ad agevolazioni e contributi europei, nazionali e regionali per le start-up innovative, incubatori certificati e PMI innovative*”.
- Cortes, Patricia, and Jessica Pan. 2015. “*When Time Binds: Returns to Working Long Hours and the Gender Wage Gap among the Highly Skilled.*”
- Cortes, Patricia, and Jessica Pan. 2017. “*Occupation and Gender.*”
- Cortes, Patricia, and Josè Tessada. 2011. “*Low-Skilled Immigration and the Labor Supply of Highly Skilled Women.*”
- Cortes, Patricia. 2022. “*Immigration, Household Services, and Women's Economic Opportunities*”. Boston University, United States.
- Goldin, C., & Olivetti, C. 2013. “*Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply. The American Economic Review*”, 103(3), pp. 257-262.
- Goldin, Claudia. 2014. “*A Grand Gender Convergence: Its Last Chapter.*”

- Ministero degli Affari Esteri e della Cooperazione Internazionale. 2021. *“Le startup italiane alla prova: dalla crisi pandemica COVID-19 alle opportunità del Next Generation EU – Rapporto ICE 2020-2021.”*
- Andreas Kuckertz, Leif Brändle, Anja Gaudig, Sebastian Hinderer, Carlos Arturo Morales Reyes, Alicia Prochotta, Kathrin M. Steinbrink, Elisabeth S.C. Berger – Journal of Business Venturing Insights, Volume 13. 2020. *“Startups in times of crisis – A rapid response to the COVID-19 pandemic”*
- Steve Blank. 2012. *“The Startup Owner's Manual: The Step-by-Step Guide for Building a Great Company”*.
- Paul Graham. 2013. *“Startups in 13 Sentences”*.
- Eric Ries. 2011. *“The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses”*.
- Donald F. Kuratko. 2016. *“Entrepreneurship: Theory, Process, and Practice”*.
- Baumol, W. J. 1990. *“Entrepreneurship: Productive, Unproductive, and Destructive”*.
- Joseph Schumpeter. 1934. *“The Theory of Economic Development”*.
- Audretsch, D. B., Thurik, A. R., Verheul, I., & Wennekers, S. 2002. *“Entrepreneurship: Determinants and Policy in a European-US Comparison”*.
- Elena Cefis, Orietta Marsili. 2006. *“Survivor: The role of innovation in firms' survival”*.
- Robert Dobbs, Robert Hamilton. 2007. *“Enterprise Survival: Insights from the 2007 Small Business Economy Report”*.
- Haibo Zhou, Gerrit de Wit. 2009. *“The Determinants of Firm Survival: Evidence from China”*.
- Stock, J. H., & Watson, M. W. 2011. *“Introduction to econometrics”* (3rd ed.).

- Robert W. Fairlie, Fank M. Fossen. 2018. “Opportunity versus Necessity Entrepreneurship: Two Components of Business Creation”
- Startupbusiness, Donatella Cambosu. 2023. “Startup innovative: cosa sono, come si fondano e quali sono i vantaggi fiscali”.  
Available at: <https://www.startupbusiness.it/startup-innovative-cosa-sono-come-costituirle-i-vantaggi/97912/>
- Fatture in Cloud. n.d. “Cos’è una startup e in cosa differisce dall’azienda tradizionale”  
Available at: <https://www.fattureincloud.it/come-aprire-un-attivita/startup-e-azienda-tradizionale/>
- SardegnaImpresa. 2022. “Cosa significa innovare e perché è importante farlo?” Available at: <https://www.sardegnaimpresa.eu/it/approfondimenti/ricerca-e-innovazione/cosa-significa-innovare-e-perche-e-importante-farlo#:~:text=Innovare%20significa%20quindi%20aderire%20alle,con%20i%20bisogni%20dei%20clienti>
- Ries, E. 2010. “What is a startup?”, Startup Lessons Learned.  
Available at: <https://www.startuplessonslearned.com/2010/06/what-is-startup.html>
- Blank, E. 2010. “What is a Startup?”  
Available at: <https://steveblank.com/2010/01/25/whats-a-startup-first-principles/#:~:text=Steve%20Blank%20defines%20it%20as,startup%2Dfirst%2Dprinciples%2F%E2%80%A6>
- Inside Marketing. n.d.. “Innovazione”.  
Available at: <https://www.insidemarketing.it/glossario/definizione/innovazione/>
- Startup Bakery, Danilo De Felice. 2022. “Startup innovative: quali sono i requisiti e vantaggi?”.  
Available at: <https://startup-bakery.com/startup-innovative-requisiti>

- Ministero delle Imprese e del Made in Italy (MIMIT). n.d.. “Startup innovative”. Available at: <https://www.mimit.gov.it/it/impresa/competitivita-e-nuove-imprese/startup-innovative>
- StartupItalia, Chiara Buratti. 2024. “Mamme hi-tech e controcorrente”. Available at: <https://startupitalia.eu/unstoppable-women/mamme-vs-gender-gap-promama/>
- Osservatori.net Digital Innovation. n.d.. “Effetto startup: con il Covid19 nuovi modelli di business e iniziative per l'emergenza”. Available at: <https://www.osservatori.net/it/ricerche/comunicati-stampa/effetto-startup-covid-nuovi-modelli-business>
- ISTAT. 2022. “Situazione e prospettive delle imprese dopo l'emergenza sanitaria COVID-19”. Available at: [https://www.istat.it/it/files/2022/02/REPORT-COVID-IMPRESA\\_2022.pdf](https://www.istat.it/it/files/2022/02/REPORT-COVID-IMPRESA_2022.pdf)
- Investopedia.n.d..“*Bootstrapping Definition, Strategies, and Pros/Cons*”. Available at: <https://www.investopedia.com/terms/b/bootstrapping.asp>
- Crowdfund Me. 2022. “Finanziare una start-up” Available at: <https://www.crowdfundme.it/blog/equity-crowdfunding/finanziare-una-start-up/>
- Brof Business Stuff. n.d. “*Business angels: Pro e contro.*” Available at: <https://brofbusiness.com/business-angels-pro-e-contro/>
- FasterCapital. n.d.. “*I pro e i contro dei finanziamenti bancari per le startup.*” Available at: <https://fastercapital.com/it/contenuto/I-pro-e-i-contro-dei-finanziamenti-bancari-per-le-startup.html>
- Assolombarda. 2023. “*Osservatorio Open Innovation e Corporate Venture Capital 2023*”.



Available at: <https://www.assolombarda.it/centro-studi/osservatorio-open-innovation-e-corporate-venture-capital-2023>

- European Commission. n.d. “*The Gender Pay Gap Situation in the EU*”. Available at: [:https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu\\_it](https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_it)
- World Economic Forum. 2023. “*Global Gender Gap Report 2023*”. Available at: <https://www.weforum.org/publications/global-gender-gap-report-2023>
- Lara, S. 2023. “*Women's Journey in STEM: From Graduation to Leadership*”. Available at: <https://economicgraph.linkedin.com/blog/Womens-journey-in-stem-from-graduation-to-leadership>
- Fondazione Gi Group, Valore D, D’Avenia, A., Rosina, A., Seghezzi, F., 2023. “*Donne, lavoro e sfide demografiche.*” Available at: <https://www.valored.it/ricerche/donne-lavoro-e-sfide-demografiche/>
- Startup Italia 2023. “*Quegli ostacoli che limitano le imprese al femminile e segnano il gender gap*” Available at: <https://startupitalia.eu/unstoppable-women/quegli-ostacoli-che-limitano-le-imprese-al-femminile-e-segnano-il-gender-gap/>
- Grant Thornton. 2024. “*Women in business 2024*” Available at: <https://www.bgt-grantthornton.it/insights/articoli/2024/women-in-business-2024/>.
- Forbes Italia. 2024. “*Occupazione femminile: SumpUp racconta le storie di sei imprenditrici*” Available at: <https://forbes.it/2024/03/07/occupazione-femminile-sumpup-racconta-le-storie-di-sei-imprenditrici/>.
- AGI. 2024. “*In Italia una donna su 5 perde il lavoro dopo la maternità*” Available at: <https://www.agi.it/cronaca/news/2024-05-09/maternita-in-italia-una-donna-su-cinque-perde-lavoro-26317116/>.

- Osservatorio Domina. 2023. “*Rapporto annuale sul lavoro domestico*”  
Available at: <https://www.osservatoriolavorodomestico.it/rapporto-annuale-lavoro-domestico-2023>.
- ISTAT. 2024. “Donne impretrici, più giovani e più istruite”. Available at: [https://www.istat.it/it/files//2024/03/STATISTICATODAY\\_IMPREDITRICI.pdf](https://www.istat.it/it/files//2024/03/STATISTICATODAY_IMPREDITRICI.pdf).
- Save the Children. 2023. “*Le equilibriste: la maternità in Italia nel 2023*”.  
Available at: <https://www.savethechildren.it/cosa-facciamo/pubblicazioni/le-equilibriste-la-maternita-in-italia-2023>.
- Nobel Prize. 2023. “*Popular science prize 2023*”.  
Available at: <https://www.nobelprize.org/uploads/2023/10/popular-economicsciencesprize2023.pdf>.
- Startup Italia, Chiara Buratti. 2024. “*Startup innovative: la mappa dei distretti più vantaggiosi secondo Cribis*”.  
Available at: <https://startupitalia.eu/startup/startup-innovative-la-mappa-dei-distretti-piu-vantaggiosi-secondo-cribis/>.
- Agenzia Nova, 2023. “*Migrants: the number of irregular migrants in Italy will decrease in 2023*”  
Available at: <https://www.agenzianova.com/en/news/migranti-nel-2023-diminuisce-il-numero-di-irregolari-in-italia/>
- Kenneth Benoit. 2011. “*Linear Regression Models with Logarithmic Transformations*”  
Available at: <https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf>