POLITECNICO

MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

An analysis of the potential role of Universities on the relocation choices of European start-ups.

TESI DI LAUREA MAGISTRALE IN MANAGEMENT ENGINEERING

INGEGNERIA GESTIONALE

Author: Federico Ridarelli

Student ID: 10807017
Advisor: Massimo Gaetano Colombo
Co-advisor: Anita Quas
Academic Year: 2021-22

Abstract

This study analyses whether the availability of human resources, intended as availability of graduated students in a NUTS3, affect the relocation tendency of start-ups. This relation is analysed considering two aspects: if the availability of students reduces the relocation probability and if the regions characterised by a higher availability of students are targeted as destinations of these migrations. Parallelly, it is analysed how being backed by VC affects the relation availability of graduates-relocation of start-ups. The countries considered in the analyses are six: Czech Republic, Denmark, France, Germany, Netherland, and Portugal. In order to test the hypotheses regarding these connections, data have been retrieved form three different databases: ETER, used to get information about the graduates, VICO 5.0 to get information about the companies, and EUROSTAT to have some significant characteristics that define the regions considered in the analysis. The tests have been done through the exploitation of the software STATA. The results of this thesis underline that increasing the availability of graduates negatively affect the relocation tendency, thus verifying the first hypothesis, but that availability does not grow after relocation, confuting the second hypothesis. However, going more in depth by considering the areas of study of graduates does not provide significant results, meaning that, apparently, there is not a specific connection between the field of business of firms and the area of study of graduates. Conversely, dividing graduates from reputable to less reputable, based on the QS's rankings of their alma mater, allows to conclude that prestigious universities deter more than the others and that their availability grows after firms' relocation, thus start-ups seem to move towards them. To conclude, being backed by a VC increases the inclination to relocate and the availability of graduates just partially dampened this effect.

Key-words: Venture Capital, Universities, Start-ups, graduated students, relocation.

Abstract in italiano

Questo studio analizza come la disponibilità di capitale umano, da intendere come disponibilità di laureati in uno specifico NUTS3, influenzi la tendenza a rilocare da parte delle start-up. La relazione è analizzata considerando due aspetti: se la disponibilità di studenti riduca la probabilità di rilocazione e se le regioni caratterizzate da una maggiore disponibilità di laureati siano considerate come destinazioni dalle aziende che rilocano. Parallelamente, è stato analizzato come essere supportati da VC influisca sulla relazione disponibilità laureati-rilocazione delle start-up. I paesi oggetto di analisi sono stati sei: Danimarca, Francia, Germania, Paesi Bassi, Portogallo e Repubblica Ceca. Per testare le ipotesi riguardanti questi legami sono stati presi dati da tre diversi database: ETER è stato usato per i dati concernenti i laureati, VICO 5.0 è servito per ottenere le informazioni circa le aziende ed EUROSTAT è stato impiegato per avere alcune caratteristiche significative delle regioni considerate nell'analisi. I test sono stati realizzati tramite l'utilizzo del software STATA. I risultati di questa tesi sottolineano come ad un aumento della disponibilità dei laureati si ha una diminuzione della tendenza a rilocare, confermando, perciò, la prima ipotesi, ma questa disponibilità non aumenta in seguito alla rilocazione, confutando la seconda ipotesi. Andando più in profondità con l'analisi, considerando le area di studio dei laureati, i risultati sembrano, apparentemente, dire che non vi è una specifica connessione tra le aree di lavoro delle aziende e le aree di studio dei laureati. Risultati diversi si ottengono considerando una divisione tra laureati provenienti da università prestigiose e laureati provenienti da università meno prestigiose, ottenuta considerando il ranking definito da QS delle loro alma mater; infatti è possibile affermare che le università prestigiose riducano maggiormente la tendenza a rilocare rispetto alle altre e che la disponibilità di laureati dalle università prestigiose aumenti in seguito alla rilocazione, dunque suggerendo che le start-ups sembrino muoversi verso queste università prestigiose. In conclusione, essere supportati da VC aumenta l'inclinazione a rilocare, sia in generale che verso aree dove la disponibilità di laureati è maggiore.

Parole chiave: Venture Capital, Università, start-ups, laureati, rilocazione

Contents

Introduction

This thesis is aimed at unveiling the possible relationship between the relocation of start-ups and the availability of social resources, intended as availability of universities. The reasons why this topic is analysed are the demographical and economic consequences of the relocations over the regions involved, (van Dijk and Pellenbarg, 2000). Despite of this scenario, the as is situation regarding the academic literature, is not satisfactory, indeed this theme has not been analysed deeply by the scholars, that have focused just on a few potential relocation determinants such as the size, the age and lately the VCs' availability. It implies that the stream undertaken by this thesis is undiscovered. The choice to focus on a new potential determinant comes from the interpretation of the results of the previous studies, that show quite ambiguous, and sometimes contradictory, conclusions. Indeed, in the last few years, the debate concerning the migration topic is characterized by the analysis of the impact of VCs' availability on the relocation of start-ups. Among the others, there are two remarkable studies: the first is the one conducted by De Prijcker and colleagues, (De Prijcker et al., 2019) where they concluded that a lower availability of VCs leads to a higher tendency to relocate, and these migrations have, as a destination, areas where the availability of VCs' is higher (Silicon Valley and Technology Square). The second important work is the one conducted by Colombo and colleagues, (Colombo et al., 2019), and they found out that start-ups located in areas characterised by a lower availability of VCs do not present a higher inclination to relocate compared to start-ups closer to the VCs' hubs, therefore the topic is still vexed. Moreover, there is an intrinsic problem related to the analysis of the VCs that is the highly heterogeneity in their distribution across the countries, both the European ones and the U.S. This is not a trivial problem because the results might depend on the specific characteristics of the regions rather than on the (higher) availability of VCs. Conversely, universities, especially considering the European situation, present a more homogenous distribution across the countries, and, hopefully, they will provide significant contributions to the topic.

This thesis is structured on 7 chapters, where the first is the introduction. The following chapters are:

2- Literature Review, it introduces the most important theories regarding the relocation phenomenon, the most used terminologies, and, in the second part, will

be presented those elements that, at the moment, the literature considers as relocation determinants.

3- Hypotheses, where are presented the hypotheses to be tested and the related literature on which they are underpinned.

4- Materials & Descriptive Analyses, here there is a description of the databases used, the related data retrieved, and descriptive analyses done on the sample obtained.

5- Methodologies & Variables, it is aimed at describing how data have been treated, the econometric analyses performed to test the hypotheses previously introduced and the variables used in these analyses.

6- Results, here the results and the related explanations are shown. Per each hypothesis tested, it is indicated the analysis carried out.

7- Conclusion, it represents the end of this thesis where it is described the contribution provided by the work, which are the limitations of the results achieved and the possible future research.

1   Literature Review

This chapter is aimed at showing to the readers the state of art of the academic literature in order to understand why the topic of this thesis is chosen and the decisions made subsequently. As mentioned in the introduction, there is no literature about the relation between relocation and availability of social resources. These two topics are analysed separately therefore the first step was to collect the most significant articles about them and try to discover possible connections. Since the begging, the focus was on quantitative data due to the fact the second part of this thesis is an econometric analysis aimed at studying, impartially, that connections.

First of all, it will be presented the main consequences of the migration firm phenomenon, such as the spreading of the regional economic gap, underlining why it is an important topic that should be deepened even more by the academic literature. Then, there will be presented the main theories related to location (neo-classical, behavioral, institutional, and evolutionary theory), and discussed, using the conclusions of previous works, if these theories may be applied also to the relocation decisions. Some examples will be presented for seek of clearness. At this point, it is the moment to introduce some concepts very connected to the one of relocation, that are: entrepreneurial ecosystems, agglomeration economies and innovation, knowledge spillover and tacit knowledge, underlighting why they are fundamental to better understand the firm relocation process. Here, there will be shown a broad presentation of the main effects of these concepts on relocation. Then, there will be the description of the decision-making process that is presented not to be a sequence of steps that all the managers must follow but is presented as an instrument to help them by playing the role of a benchmark. In the last part of this review takes place the discussion about the determinants of the relocation process, where the topic of university and research centres will have a deepening because that thesis is aimed at assessing the relation between knowledge institutions and start-ups, especially in spatial terms.

In the first chapter the topics are discussed separately because the literature, as far, has considered the topics, basically, independent. The few exceptions are mainly concentrated in the very last paragraph of this chapter.

1.1. Firms' relocation: why is important

Firm relocation (or migration) is a "particular form of location adjustment that can be defined as a firm's change of address from Location A to Location B", (De Bok and Sanders, 2005). It is important to stress immediately, that this phenomenon was studied mostly for established companies rather than start-ups, (van Dijk and Pellenbarg, 2000). This clarification is not eligible because the same event could be caused by different factors. Indeed, it is possible to affirm that "some start-ups are actually relocations" and that relocation decisions is somehow path dependent, since the previous location is a fundamental parameter for the learning process of a firm (for instance the quality of competitors), and therefore affects the level of knowledge of the firm itself. (Manjón-Antolín and Arauzo-Carod, 2011).

Having said that, it is worth of mention the reason why relocation is an important topic that should be investigated even more, indeed the mobility of firms is greater than is assumed, (van Dijk and Pellenbarg, 2000). Usually, firms that migrate are quite small and young due to several reasons that will be discussed deeply later in the chapter. According to the literature, the moment in which firms decide or think to perform their first relocation is after the very first part of their life that is the most critical one in which there is the highest probability of failing. To give a number for clarification, about 180.000 jobs were involved by firm migration in 1995, in The Netherlands, with an average size of the migrant firm of 2.8 employee. More in detail, the inter-provincial migrations, in the same year in Netherlands, were just 6300 involving 17.000 employees, (van Dijk and Pellenbarg, 2000). This means that most relocations consist in a short distance migration, in agreement with the theories of Williamson, (Williamson, 2016), that affirm that the cost to move a firm increases as the relocation covers a longer distance because start-ups face more problems in monitoring the opportunistic behaviour of their new partners. However, here the key point is underlining that these small firms are growing firms that would contribute to the future economic wealth of the region in case they remain there. So this exodus of firms entails the loss of a future growth and the shift of the richness within a country, (van Dijk and Pellenbarg, 2000), therefore firm relocation phenomenon can be identified as a possible reason of the persisting regional economic gap in a country, impacting on the regional planning policy. In addition, across the last decades the number of firms relocated increased significantly and it is reasonable saying that nowadays numbers are much bigger. Other problems that may arise after firms' migration are traffic congestion & environmental pollution, (van Dijk and Pellenbarg, 2000). For many public policies is central understanding how to improve the local economy by attracting new firms and on the other hand, prevent the existing ones to go away. Having a better understanding of the drivers behind firm relocation would ensure that the policies

proposed are effective. For instance, transport investments is seen as a policy that can spur a firm to relocate in case it presents a high car dependency, (Nilsen et al., 2020).

The topic of optimal relocation choices is one of the most difficult among the strategic decisions themes. This is due to several factors: first of all, relocation is not reversible in the short term and usually very costly in the long term. Secondly, once relocated in the chosen location, start-ups become part of that context by building strong connections with different local actors (customers, suppliers, etc.), (Lee, 2022). This means that a wrong decision would imply not only the necessity of a further relocation (sustaining the related costs) but also undertaking huge opportunity costs due to collaborations with local players that are not correctly chosen. The reason why, in many cases, start-ups must consider, at least, relocation is because they face limited internal resources and the only solution to overcome this issue is relying on external resources providers, such as Venture Capitalists, Universities, governmental institutions etc., (Lee, 2022). That limitation of internal resources become problematic when start-ups grow enlarging its size, number of employees, geographic market coverage, clients etc., and this is the moment when a start-up starts to look for other locations. However, as it is better explained later in the chapter, relocation is not always driven by internal needs, indeed when a start-up faces threating changes of the environment, for instance new restrictive regulations, new competitors etc, it has to look to new possible locations in order to survive, (Lee, 2022). In the same work, the author underlines that is difficult for start-ups undertake optimal location choices because of the constraints mentioned before that is the main reason why they relocate more than once in a period of time that is not so large.

1.2. Firm's relocation theory

To better understand this firm migration theme is important to introduce the most significant theories about location and control if that theories can be applied also for relocation and eventually assess their accuracy. In particular, Hayter (Hayter, 1997) identified three main approaches: neo-classical, behavioural & institutional. All of these approaches share that they consider the firm as an "active decision-making agent", (De Bok and Sanders, 2005).

The neoclassical approach is aimed at maximizing the profit, following the standard economic theories, (De Bok and Sanders, 2005). The decision of the optimal location is done considering a set of rational factors through that are defined a group of alternatives. Here, the main determinants are those that affect the economic benefits such as the transportation costs, labour costs, rent expenses, external economies,

and market size, (Manjón-Antolín and Arauzo-Carod, 2011). In an equilibrium situation, relocation does not occur because the optimal location is always the same, however, in the real world, companies face both external changes (new policies, for instance) and internal one (enlargement of the employees), leading to a variation of the firm's profit and so to relocation, (De Bok and Sanders, 2005). Moreover, firms in order to maximize their profits pursue economies of scale but in many cases can be realised only in particular context such as cities where the market is larger, (De Bok and Sanders, 2005), and so relocation becomes necessary. To give an example, this approach was used by Frenkel to demonstrate why metropolitan areas attract high-tech industries which employ high technologies and are involved in the process of innovation, (Frenkel, 2001).

Behavioural approach starts from the same considerations of the previous theory, but it introduces the concept of bounded rationality. It means that people undertake decisions without a perfect knowledge therefore the main drivers for relocation are not external but internal such as size, age, previous experience, (Manjón-Antolín and Arauzo-Carod, 2011). Thus, maximization of the profit is not, anymore, the main goal and the optimal location is substituted by the satisfactory location, (De Bok and Sanders, 2005). The neo-classical theory can be seen as a sort of specific case of the behavioural theory indeed when firms have high level information and the possibility to use it then the decision maker is very similar to the "homo economicus" characterizing the neo-classical approach, (De Bok and Sanders, 2005). However, that case is not a real one, but it can be seen as a sort of benchmark. In this theory what counts is the perception of the reality rather than the reality itself underlining the importance of the spatial bias in the decision. That bias increases when the relocation distance increases because the uncertainty is higher, (De Bok and Sanders, 2005). Based on this, Arauzo made a research that emphasised the importance of the size (and in general the intrinsic characteristic of a firm) in the relocation in the Catalan context; in particular, labour-intensive firms seem more likely to be located in the city of Barcelona, (Arauzo Carod and Manjón Antolín, 2004).

The last approach is the institutional one. It starts from the assumption that economic activity is socially and institutionally situated. Moreover, this theory considers a dynamic environment while the two previous approaches a static one, (De Bok and Sanders, 2005). The final decision for a location (and a relocation) is a matter of negotiation with clients, suppliers, labour unions considering the local prices, infrastructure, wages etc., (Brouwer et al., 2004). According to the literature there are two types of institutions that affect the relocation decisions: a) governmental initiatives; b) real estate market, (De Bok and Sanders, 2005). Both elements will be discussed deeper in the following parts of this review. Papke

performed an analysis in which she affirms that "the estimates indicate that a high state marginal effective tax rate reduces the number of firm births for half of the industries examined", (Papke, 1991). According to the literature, the institutional approach is most suitable for big companies that can exploit a stronger bargaining power, (Brouwer et al., 2004).

The analysis made by Manjon-Antolín confirms that this theoretical framework could be used also for relocation but underlines that the determinants between relocation and location decisions are different. Moreover, the few determinants in common affect with a different intensity the decisions, (Manjón-Antolín and Arauzo-Carod, 2011). Similar conclusions are affirmed by De Bok and Sanders in their work underlining "that location theory is more concerned with locational pull factors, whereas relocation also deals with push-factors: the trigger to moving.", (De Bok and Sanders, 2005), meaning that neo-classical approach is more tailored to location theory while behavioural is more suitable for relocation. The concepts of pull factors and push factors will be described later in the review.

In addition to these three approaches, some scholars started to consider in their analysis a new theory called the evolutionary approach, (Musolino et al., 2020). It applies the concepts of the evolutionary economics in the context of economic geography providing alternative explanations of the main aspects of the subject. The starting point of this theory is that the decision-making process is based on the routine of the company that consist also in a large part of its knowledge because they learn from their own mistake through trial and error. Consequently, since companies are different in their routines, models cannot anymore rely on a "representative agent". Basically, when a routine does not work well, the firm starts to search for a new routine to survive, for instance by investing more money in R&D, (Boschma and Frenken, 2006). In a context characterised by a rapid technological change, such as the one in which start-ups usually operate, the new knowledge is strongly path dependent entails that a competitive advantage gained by a firm is quite difficult to be stolen by competitors, (Teece et al., 1997). This is quite evident when killer acquisitions are taken into consideration, indeed big companies find much easier buy and then close a start-up and its related product, rather than imitate the start-up's business model despite an availability of resources much bigger. Another important aspect of this theory is the importance given to agglomeration being the result of the studies on the spatial distribution of the routine over time. In other words, a concentration of routines in one place implies a concentration of knowledge in that place (following the learn-by-doing approach) that represents the precondition for an agglomeration economy, (Boschma and Frenken, 2006). A way through which that knowledge is transferred (spillover) from a company to another may be the exchange of personnel or a collaboration among

two or more companies in the area. Therefore, this theory does not explain regional economic gaps considering a macro-level perspective but through an inner perspective of the firms operating there, (Boschma and Frenken, 2006).

Going further with the literature review, many scholars identify two types of relocation: a) partial relocation, where the main unit is not eliminated but it is added a new one located elsewhere and they cooperate; b) complete relocation, it takes place when a firm changes its location closing the previous place and opening a new one elsewhere. The former is performed by multi-plans firms, which aims at exploiting the different local conditions and for that reason usually does not involve the entire process, but a single state (for instance a firm's plant is moved to a developing country where the labour costs is much less expensive). The latter is done by single site firms that decide to move because the new location fulfilled their needs in a better way, (Brouwer et al., 2004).

Factors that affect the migration decision may be classified in different way but the most diffused is the one that identifies three types of factors: a) push factors; b) pull factors; c) keep factors. The first type of factors includes all the reasons that push a firm to relocate itself. In that case, usually the current location has reached already (or is closed to) the so called "spatial margin of profitability", (van Dijk and Pellenbarg, 2000). It means that having the current position is not anymore profitable for the company due to the change of the context in which it operates. Many scholars, including Van Dijk and Pellenberg, identify as the lack of space the most common push factor that brings to a relocation and the second one is the accessibility. However, it is also possible that the firm still is in a profitable condition but decides to move from there because the decision-maker has information about other locations that ensure higher profits. Those factors are called pull-factor. The third type of factors are called keep factors and basically are composed by those reasons why a decision-maker should remain in the same site. Essentially, they represent costs (fixed and variable) related to a firm migration. These kinds of costs also include the sunk costs (for example for the plant), and the variable costs that are mainly related to the availability of skilled labour force, in fact after a relocation it might be necessary training new employees, sustaining its cost. The most important keep factor, by far, is the labour force, especially if the firm relies on highly specialised workers and faces high hiring, firing, and training costs, (van Dijk and Pellenbarg, 2000). In case the migration is needed for some push/pull factors but the firm, on the other hand, is facing the labour force issue, a widespread compromised is the minimization of migration distance, so that employees can maintain their work without changing the house (intra-regional firm migration). In the real world these three factors coexist and their interplay is extremely important in the location decision process, (van Dijk and Pellenbarg, 2000).

A different classification of the migration factors can be done clustering the same factors considering their degree of control over them. In other words, it is possible identify between: a) firm internal factors; b) firm external factors; c) location factors. This categorization was developed by (Lloyd , Dicken, 1977). The internal ones are composed by those factors that can be controlled by the management of a firm (organisational goals, employment etc.). Conversely, the external ones cannot be controlled by managers (e.g., regional policy, economic regional structure etc.). The location factors represent, instead, the characteristics of the current site such as the distance between the firm and the customers or the availability of space for an extension. Some of these factors can be modified only in the long run while others can be changed within a short period, (van Dijk and Pellenbarg, 2000). When a firm needs to relocate for internal factors usually is connected to the life cycle of the firm and, in case few types of products are produced there, to the product life cycle. This implies that the perfect location (assuming a neoclassical approach in which it exists) is a temporary concept, that will be modified when the needs will be different. For example, the volumes of production during a launch of a product and its mature period are totally different and may be necessary changing the location of the plant in order to have a plenty of space. In the previous example was considered an internal factor of production but external and location factors change over time as well, creating a sub-optimal condition and pushing towards a relocation.

Having discussed about the main aspects of relocation theories, it is time to talk about the destination of these migrations. The literature debated a lot whether it is better moving towards cities or less populated areas, however there is not a clear answer to this topic. Some scholars have underlined the importance of relocating in cities due to the possible presence of agglomeration economies or entrepreneurial ecosystems that might provide an important support for firms once there, nevertheless others stressed the fact that overall, the cons overcome the pros in moving to towns. Thus, here will be presented the main characteristics of these concepts considering both positive and negative aspects, without taking any position for the moment.

1.3. Agglomeration economies

As mentioned before, the Academic literature often underlines the importance of agglomeration economies as a key driver to assess the advantages of one location compared to another, thus becoming an important topic in the relocation discussion. Krugman's works were, and still are, fundamental to better understand the

importance of agglomeration economies; in particular agglomeration economies may occur without the presence of regional differences but just when "both firms and workers find more profitable to cluster in one region rather than to spread out over more regions" and a possible reason can be the drop of transportation costs, (Krugman, 1991). The main effect of agglomeration economies is the exploitation of external economies that positively affect the firms' productivity. According to Nilsen, (Nilsen et al., 2020) agglomeration effects can be classified into two categories: a) co-location of similar firms (localization economies); b) co-location with other firms (urbanization economies). Having similar firms close each other implies the diminishing of some costs because, for instance, people in a localization economy should not be trained; moreover, in this context the impact of knowledge spillovers, that may occur through the exchange of personnel, is very important. Urbanization economies entail the presence of network externalities that ensure important advantages for the firms like being close to suppliers or customers and the possibility of a better cooperation and collaboration along the different levels of the value chain, (Nilsen et al., 2020). Thus, urbanization economy may be seen as an important driver for relocation because having both suppliers & customers close ensures to control and have stable relations with them, (Risselada et al., 2013). Moreover, the agglomeration benefits diminish as we increase the distance but with a degree that varies between sectors, (Nilsen et al., 2020), suggesting that the intrinsic characteristics of a firm are the main driver of relocation. Usually, agglomeration effects take place in the big cities that are characterised by parameters such as high educational level or the population density.

Agglomeration economies lead to have a country with specific regional conditions, that is a common situation for the most important countries in the world, Italy included. These circumstances lead to differences in the average daily salary received by an employee, for instance, and that affects the firm's propensity to relocate. Thus, on one hand there could be the desire of a company to move to an area in which the salary, on average, is lower, (Kronenberg, 2013), and on the other hand, people may be attracted to move to area where the wages are higher creating a problem of availability of labour force. Some studies show that the propensity of the firm to go away from an agglomeration economy or remain there depends on some sector-specific drivers, such as the knowledge-intensity and the degree of technology in the business model/business processes of the firm. In other words, when a firm is less knowledge-intensive & low-tech presents a higher tendency to relocate, (Kronenberg, 2013). The reasons of this evidence will be discussed deeply later in a dedicated paragraph.

However, the topic of agglomeration economies is complex and somehow contradictory. Indeed, other authors have underlined that firms located in the most

densely populated areas in Europe are thinking to relocate in rural regions because they face a tough competition and have significant problems in terms of recruiting of the personnel. The reason behind this phenomenon is that labour market is sometimes saturated in cities and housing costs are very high so many workers (also with a high educational degree) prefer live outside the biggest cities to afford bigger house, save a lot of commuting time, and have in general a better life, (van Dijk and Pellenbarg, 2000). In addition, firms themselves would benefit from a relocation towards rural areas in case they will not face the problems related to hiring & training costs because they will save money because, there, the fixed costs, such as rent expense, are significant lower. Connected to this problem there is the issue regarding the lack of space in the metropolitan areas, that affects the decision regarding the relocation in a different way depending on the type of sector in which the firm runs its business. Indeed, manufacturing firms that need plants, offices, warehouse etc are likely to move outside the urban areas, however when are added other determinants the size of the firms becomes secondary in importance while service firms prefer to remain inside the agglomeration economy and exploit all the related advantages, (Kronenberg, 2013). Moreover, is not clear if an agglomeration economy has some thresholds in terms of size, beyond which the net benefits start to diminish, (Arthur, 1990) because of the exponential increasing of the problems related to the availability of space, labour force, congestion, competition etc. Considering the Silicon Valley case, it seems that this threshold does not exist because that agglomeration is constantly attracting new workers, firms, external capital becoming bigger and bigger. Vice versa, looking at the Randstad case, that is the most populated innovative area in the Netherlands, it appears having reached its threshold, (van Dijk and Pellenbarg, 2000). Reasons that can explain that difference can be many, such as a different approach used by the policymaker in order increase the purchasing power of the people in the area by reducing taxation. However, it is evident that finding a universal theory that describe perfectly all the different agglomeration economies, their evolutions and behaviours looks very difficult because of the complexity of the system analysed, and the presence of bounded rationality in all the players involved.

Beside the traditional topic of agglomeration economies, few scholars have investigated also "less-known factors such as the cultural factors (entrepreneurial culture/ecosystem) and historical factors (the impact of the presence of an organised crime)", (Musolino et al., 2020), that influence the spatial patterns of locational preferences. However, as the author themselves admitted, most of the literature has not considered yet those new factors, such as FDI (foreign direct investments), (Musolino, Mariotti and Brouwer, 2020) being focused on the traditional ones. It is worth of mention saying that COVID-19 introduced other new factors to this discussion, first of all the impact of smart-working in the relocation decision

considering the different necessity of offices for the companies. However, data regarding relocations during the pandemic do not allow yet a quantitative analysis and to discover potential differences in trends but, in the future, this might be an important and interesting new contribution to this topic.

1.4. Entrepreneurial ecosystems

Having cited entrepreneurial ecosystem, that is a concept directly related to the agglomeration economies, it might be important going further in the description of this significant phenomenon that has a growing impact on relocation processes. Companies are not distributed homogeneously in the countries, but this distribution is "a function of heterogeneous endowments in knowledge, institutions, resources, and demand.", (Fischer et al., 2022). The most important case in the world is, by far, The Silicon Valley. A possible definition of Entrepreneurial Ecosystem was given by Carayannis and Campbell: "agglomerations of human, social, intellectual and financial capital stocks and flows as well as cultural and technological artifacts and modalities, continually co-evolving, co-specializing, and co-opeting", (Carayannis and Campbell, 2009, p. 3). The key aspect in this ecosystem is the importance of the relations of the different actors involved and not just those among entrepreneurs meaning that the spatial proximity of people does not ensure the creation of that innovative environment, (Fischer et al., 2022), or in other words that not all the agglomeration economies become EE (Entrepreneurial Ecosystems). That concept of EE is related to the one of relocation because firms may be stimulated to move towards EEs themselves in order to stay into an innovative environment where developing new ideas is easier also due the interactions among different the players. According to the literature, there are two main classifications of the actors involved in an EE: the first one introduced by Isenberg (Isenberg, 2011) that identifies six components: a) an enabling culture; b) supportive policies and leadership; c) the availability of financing; d) human capital; e) markets open to new businesses; f) institutional and infrastructural support. The second one is provided by Spigel (Spigel, 2017) who individuates three categories: a) material; b) social; c) cultural. Roundy, (Roundy et al., 2018) affirmed that the main characteristics of EE are six: self-organization, open but distinct boundaries, complex components, nonlinear dynamics, adaptability through dynamic interactions, and sensitivity to initial conditions. Thus, EE looks like a business ecosystem where all the actors cooperate in the value creation process and firms can fully exercise their "capacity of creating and capturing value generated in the surrounding systems, producing

complementary products and services", (Cavallo et al., 2019), and therefore it seems to be an appropriate place where moving a business.

Many scholars have debated a lot regarding EE's boundaries, that is quite important in that thesis because is aimed at providing an assessment of the spatial trends of relocation therefore is very crucial being clear in what the academic literature considers inside the EE area and what is outside. Fischer and colleagues in their work highlighted three main approaches but each one has some drawbacks, and this is a proof of the fact that this topic has to be investigated deeper. The first approach considers the EE any kind of geographical region that facilitates the entrepreneurial activity, thus this approach leaves unsolved the problem of the boundaries because is too generic. The second approach is the one that considers the EE as the portion of earth around a variable radius considering a non-clearly defined central location, but it does not take into account the fact that EEs may be different one each other, in terms of shape, size etc, (Fischer et al., 2022). The last approach considers administrative units, but this implies the neglection tout court of the interactions and possible integrations between near regions, (Fischer et al., 2022). Moreover, EE is not something static but is constantly changing embracing the newest innovations and that phenomenon has consequences also on its boundaries that are moving consequently. In conclusion, EE boundaries are subjected to arbitrary decisions, (Fischer et al., 2022).

Thus, it is interesting spend few words on the evolution of these EEs in order to understand better, if possible, what is the role played by start-ups in this growth, and checking if there are some determinants, among the others, that have influenced the relocation start-up process. To study this evolution, Cloutier and Messeghem, (Cloutier and Messeghem, 2022) introduced the concept of EE path dependence based on a combination of evolutionary approach (discussed previously) and quantitative analysis. For seek of completeness, the study takes into consideration the EE of Montpellier, France. The evocative image used by these authors to describe the evolution is whirlwind, because "EE trajectories are sinuous, unpredictable and it underlines the logic of coevolution of the sub-ecosystems involved in the EE", (Cloutier and Messeghem, 2022), where each one represents an actor of the system itself, thus EE path dependence is "nonlinear subject to endogenous and exogenous influences and the result of the dynamic interactions between complex components", (Cloutier and Messeghem, 2022). Considering the concept of path dependence, there are three different phases that can be individuated: a) impulse; b) creation; c) structuring. Talking more in details about the case studied by Cloutier and Messeghem, the impulse phase was a political initiative sponsored by the major of Montpellier, Georges Frêche, that supported intensively the business creation, then the creation phase started with the

foundation, in the 1987, of the incubator to encourage economic activities to solve the significant problem of the unemployment of the area. During the structuring phase, many activities are created and assisted in their growth and in their first three years of life. In this example was quite clear the importance of the connections between all parties, but the policymaker seems to be the real determinants because, through its initiative, triggered the creation process of Montpellier's EE spurring the relocation phenomenon that was driven by the presence of the incubators (location determinant) being a clear and concrete example of the potential huge impact of this institutional infrastructure, that will be discussed deeper lately.

1.5. Innovation, knowledge spillover, tacit knowledge

One of the most important pros of being located in a city, is the proximity to innovation because of the presence of universities and research centres that are a constant source of knowledge and providing new potential employees. That physical proximity improves significantly the firms' innovation process, (Phelps et al., 2012), that is one of the most important source to create and maintain competitive advantage, (Carayannis, 2013). Moreover, innovation might be seen as an antecedent to EE, (Cavallo et al., 2019). Having said why innovation is important, it is useful describing a little bit more in details what it is and how it can affect relocation process. Innovation is not a process performed alone by a single company or entrepreneur, but it is the result of an interaction of different point of views, ideas, experiences etc. In other words, innovation is affected by the context, and this is the reason why in the academic literature there is the concept of innovation ecosystem. Roger, (ROGERS et al., 2008) emphasized that not only the process of creation but also the diffusion one is based on a social system that is the connections among the different innovators. The innovation ecosystem can be represented as a coin with two faces that are quite different one each other: one is the research economy, which is driven by fundamental research and the other "side" is the commercial economy, which is driven by the marketplace, (Oh et al., 2016). In the same article the author distinguishes the features of the innovation ecosystem are many, as underlined by: a) more explicitly systemic (innovation as interaction-based process); b) digitalization; c) open innovation (open-sourcing, crowdsourcing, etc.); d) mimetic quality and its appeal to the news media; e) many different innovators that are complementary one each other in the innovation process; f) greater importance of market forces.

Deciding to relocate in an innovative ecosystem might be fundamental to continue the growth of start-ups not only due to increase of knowledge through spillovers

and tacit knowledge but also because it may increase the attractiveness of the start-ups towards VC. Indeed, being in that system enables to take part of some programs tailored for start-ups such as those provided by incubators and accelerators, that means a signalling of quality for VCs when they decide where invest their money. A signalling in a context characterized by high uncertainty and opportunistic behaviour (moral hazard and adverse selection) is very powerful tool.

Moving to the topic of knowledge spillovers, they are extremely important in this discussion because they are one of the main ways to diffuse the knowledge and generate innovation in a system. It is possible defining spillovers as exchanges of ideas among individuals and they are important because they foster the entrepreneurship and improve the performance of firms. The Knowledge Spillover Theory of Entrepreneurship identifies two main streams of knowledge spillover as locational strategies: the first one focuses on the existence and distribution of universities spillovers (spatial bounded); the second one focuses on the impact of spillovers on the location decisions (not only those related to universities), (Audretsch et al., 2005). The theory affirms basically, that a context with more knowledge will provide more entrepreneurial opportunities, where each entrepreneurial venture can be seen as a channel of knowledge spillover (Audretsch and Keilbach, 2007). In that article, the authors affirmed that there is an empirical evidence that the entrepreneurial opportunities are not "exogenously generated but they are created by incumbent players that invest in innovation". This sentence underpins the theory according to the best place where founding a start-up is the metropolitan area where most incumbent players are located. Up to this moment, it was considered just the positive side of knowledge spillover, nevertheless there is an evident problem that is the misappropriation. Indeed, Shaver & Flyer, (Myles Shaver and Flyer, 2000) argued that just worse firms, with weaker technologies, human capital etc., will benefit on being inside the agglomeration while leader firms will face a problem of misappropriation of their competences, expertise and, sometimes, employees. The article seems to suggest that firm migration is performed, mostly, by "bad" companies because good ones will be negative impacted by spillover outflows and therefore cons overcome the pros regarding the relocation towards agglomeration economies. This result, probably, does not perfectly reflect the reality because there is a sort of entry barrier in accessing to the agglomeration economies, that are the significant transactions costs, (Williamson, 2016), that can be sustained just by profitable companies. In case, bad companies are founded in the agglomeration, then they face tough competition that brings to their failure. However, misappropriation is still an important problem that is solved (partially) through patents and intellectual property rights.

For seek of completeness, the problem related to the EE's boundaries also affected the spillover theme because they take place within the ecosystem and the authors suggest that these "spillovers have a sharp attenuation when distance increases", underlining the "hyperlocal character of EE". Also in that case, is arbitrary defining where is the reference point and therefore assess where the spillover effect arrives, (Fischer et al., 2022).

"The theory of localized knowledge spillovers suggests that profits will be greater in agglomerations and spatial clusters, since access to tacit knowledge is easier", (Audretsch et al., 2005). Tacit knowledge is a peculiar type of knowledge that is difficult to express and to transfer through an oral or a written way. Indeed, this transfer occurs due to intuition, experience etc. Tacit knowledge is important in order to be more profitable because increase the efficiency of exploitation of inputs, (Berger, 2013). This concept is connected to the learning curve, meaning that there is a cap in terms of improving if the company continues in using the same technology. A study conducted by Berger, (Berger, 2013) assessed that between the 50% and 90% of firm knowledge is tacit by nature. Moreover, this kind of knowledge is acquired by employees by their experiences (including observing and imitating), and it takes a significant amount of time to be acquired. However, another way to increase the tacit knowledge capital is through dialectical inquiry, that is developed, in nineteenth century, by the philosopher G.W.F. Hegel. Hegel sustained that a higher degree of knowledge is achieved when two opposite points of view are compared, having a moment of synthesis as a result, (Woods, 2019). Of course, deeper is the knowledge that underpins those initial points of view and better will be the final result, meaning that this approach is more useful when employees have a high educational level, that is agglomeration economy.

All these facts provide a robust corroboration to the hypothesis sustaining that the best location is agglomeration economies.

1.6. The relocation decision making process

Once having discussed about the main (re)location theories and the main concepts related to the topic, it is appropriate talking about the decision-making process that leads to a relocation. As mentioned, relocation decision is a crucial moment for a firm. The process that leads to the final choice is, as well, quite complex, involving many phases and considering a significant number of determinants that will be discussed deeply, further in the chapter. Despite this importance, scholars have overlooked to discuss about how a firm arrive to its final relocation decision. One

of the few scholars to be focused on the topic was Townroe, who introduced a first classification of it in 1973 (Townroe, 1973), identyfing five decision stages: 1) stimulus; 2) problem definition; 3) search; 4) formulation and comparisons of alternatives; 5) choice and action. Successively, other authors introduced some new classifications but all of them based on this one. Louw, (Louw, 1996) in his work, explained which are those factors that affect mostly each stage of the process. More in details, he affirmed that phase 3 & 4 (recalled by him orientation phase and selection phase) are driven by spatial factors such as geographical position, accessibility etc. The last phase, recalled negotiation phase, is driven by financial and contractual factors.

In 2021, Hassanain and Ibrahim (Hassanain and Ibrahim, 2020) published a framework that, theoretically, would facilitate the relocation decision to the managers of a firm, indeed the following decision-making process is described from a managerial perspective. As they underline, this work was quite innovative because "literature did not focus holistically on the relocation as a process, in a detailed manner in terms of covering and describing the activities and underlying the lifecycle of a workplace relocation, as a process, from the decision-making to the satisfactory occupancy". In order to understand which are the actual common practises for a relocation decision making process, a survey, based on the previous literature, was developed by the authors. They identified 3 different moments during the process: 1) initiate the pre-relocation activities; 2) implement the relocation activities; 3) conduct the post-relocation activities.

In the first phase, managers should analyse the business needs of the companies (for instance understand if the possible relocation is due to a growth of the business or a downsizing of the turnover). Then, is important assessing the condition of the current location checking its suitability with respect to the new needs of the firm. In case the as is situation does not fulfil all the needs, managers identify a set of relocation options individuated considering the needs, previously mentioned, and perform an analysis to define which is the best choice. At this point, managers can realize the relocation plan, setting some milestones that will be fundamental to control if the plan is developing correctly or not. At the end of this first phase, managers should allocate budget for relocation activities according to the framed milestones and design new premises.

The second phase implies the preparation of work packages that allows to send requests to targeted new partners (suppliers, providers etc.). The partner selection is, in fact, the very next step. Then is the moment to acquire and fit-out the new premises, and afterwards, according to the authors, is the right moment to announce the relocation activities, preparing a support for employees to allow them to

overcome possible hurdles and oppositions. At that point, relocation can be performed and, finally, managers should evaluate the compliance of the new location to the plan defined previously.

The last phase entails the initialization of the new location. Managers at this point should guarantee a healthy workplace and safety plan by implementing the actions designed formerly and carrying out the maintenance programs. A good practise is recording all the information transferred considering all the means of communication. Lastly, during that phase is important being constantly in contact with employees to check their satisfaction related to the new location and, in case, assist them if some problems arise.

This example of decision-making process is not presented here to show how an evaluation of relocation must be, but it is described to be a support for managers, using it as a benchmark.

Having discussed about the main concepts related to relocation (agglomeration economies, spillovers, etc.) and about the process that leads to the relocation decision, it is now important discussing about its determinants.

1.7. Relocation decision drivers

In this paragraph will be discussed the main determinants that drive relocation decisions. Along the previous paragraphs most of them are already cited but not discussed deeply. This section will be a sort of overview in which some elements will be deepened more than others, according to the core topic of the thesis. The first paragraph is dedicated to the comparisons between location and relocation drivers to assess if they are the same or not. Then, the discussion will focus on the important of the size of the firm and, more in general, the spatial problems that may arise along the life of a venture. Once clarified this topic, it is important to show how the actors involved in the business model affects the relocation decision. First, considering the impact of employees and partners in the relocation decision, then considering the external capital providers; in particular, the topic of the VC and start-ups will be discussed more in details because VCs, by far, are the most important source of finance for a start-up. At that point is possible to talk about the external environment and how it influences the decision-making process; here, will be discussed the importance of the policymakers, the infrastructure (physical and not), incubators &accelerators and, finally, universities and research centres in attracting relocations of start-ups. In particular, the point about universities and research centres will be deepened more, coherently with the aim of the thesis.

1.7.1. Location drivers vs Relocation drivers.

As mentioned before, location drivers and relocation drivers are not exactly the same because the processes are performed in a different period of the life of the firm or start-up. Indeed, many authors underlined that the first decision regarding where founding a firm is made under important constrictions that might be already solved later during the relocation phase. For example, the place where entrepreneurs have graduated from university is likely to be the location where they found their company, (Lee, 2022). Moreover, Pellenberg, (Pellenberg, 2002) introduced the concept of historical dependency, meaning that the previous location(s) affects the new decision process. Holl, (Holl, 2004), affirms that during a location decision, parameters such as the low costs of personnel are much more relevant than during a relocation. Vice versa, in relocation a fundamental aspect is the accessibility to important infrastructure such as motor ways in order to enlarge the market. A much more radical position is held by Stam, that affirms "In many cases, the choice of location is made at random, sometimes motivated by the entrepreneur's knowledge of locations or because premises are available through personal relationships", (Stam, 2007). The main reasons that underpin his position are financial because initially the new-born business cannot be profitable and therefore, in the large majority of cases, locate elsewhere is too expensive.

In another article, (Larsson et al., 2017), it was assessed that 63% of graduate students start their entrepreneurial activities in the region where the graduated, and only the 37% elsewhere. The percentage of entrepreneurs that found a company in the region of their graduation increases even more if also the family is located closely. This suggest that personal background affects deeply the first location choice. It is worth underling that universities are not distributed homogenously across the country but mainly concentrated in cities or town centres, and this might imply a further polarization of entrepreneurial ventures and the related consequences (see agglomeration economies & entrepreneurial ecosystems). Moreover, it is mandatory admit that universities are not the same indeed, according to the course provided and the relations with firms, the degree of encouragement towards an entrepreneurial career is different, (Rothaermel et al., 2007). It is not a coincidence that the two most important agglomeration economies (Silicon Valley & Technology Square) arise next to the most innovative and entrepreneurial universities, and similar reasoning can be applied to the England case, considering the area delimited by Oxford-London-Cambridge, (Smith and Ho, 2006).

What emerges seem to be the presence of a sort of local bias of the entrepreneurs, at least, in their very first step, and this bias becomes more and more important as the entrepreneur is more embedded in the local society, (Stam, 2007). In the same paper, the author underlined that this social connection becomes less important when the firm has grown but inter-regional relocations and openings of new branches in a different area, is not so common because entrepreneurs usually supposed to reach new markets remaining in the original position.

Although the local bias seems to be evident, there are also important benefits in being located in a familiar context. Indeed, the proximity with universities implies important reduction of the costs in accessing to the academic knowledge and in general resources (such as hiring students just graduated), (Heblich and Slavtchev, 2014). Another important benefit in knowing the context is the better-informed decision regarding the selection of the partners, because the information asymmetries are reduced, but not nullified, (Holl, 2004).

Indeed, a study conducted by Michelacci and Silva, stressed the positive consequences of these benefits. They found that, especially in the more developed regions, firms where the founders have local roots, are bigger, more profitable, and get a higher amount of external capital, (Michelacci and Silva, 2007). In particular, the latter point is important, also because the other two are, usually, direct consequences of it. In a context characterised by uncertainty and lack of information such as the one of the start-ups and new-born firms, social relation may be the key to the success, indeed theories underline the importance of Family and Friends (& Fool) and of a reputable name, especially in the very first stages of the firm.

1.7.2. Firm size, space requirements & real estate market

Having discussed how and why is different the first location decision compared to the relocation one, it is now possible focusing on the later stages. It is evident, indeed, that a firm grows or decreases over the time, and this has an impact on its needs related to the geographical position. However, firms are not the same, indeed they have a large variety in size, nature of activity, age etc. and therefore different needs to be tackled. This is the reason why the same factor affects in a different way different firms and may lead to a different decision. Nevertheless, those factors are interdependent and should be considered together, with a proper specific weight, by managers during the decision-making process. However, here in order to be clearer in the explanation, those factors will be grouped.

Probably, the first parameter that comes to mind as a relocation driver is the firm size and the related space requirements. The aspiration of entrepreneurs is seeing their entrepreneurial ventures grow in turnover and, especially, in profits. To perform that growth, an enlargement of the firm is, usually, a mandatory transition point that may represent a push factor that makes the current location unfit considering the needs, current and future. Moreover, parallelly to the growth, firms increase their knowledge and become more self-confident, and these may be the reason why they "become spatially more adventurous with age", (Taylor, 1975). A completely opposite viewpoint is held by other scholars (Brouwer et al., 2004) sustaining that older firms usually are larger and more connected with the environment; therefore the sunk costs are higher, including the opportunity cost of losing the relationship built over time with partners and for that reason they seem to be less inclined to relocate. The location theories have usually overlooked an aspect of the growth that is its distinction between internal and external growth, being focused just on the former. The study conducted by (Brouwer et al., 2004) affirms that external growth increases significantly the propensity of a firm to relocate. In addition, the same authors have noticed that larger is the market served by the firm and higher will be the propensity to relocate. It is also interesting underlining that two irrational factors may affect relocation decision that are the entrepreneur's ambition and lifestyle, (Risselada et al., 2013).

However, considering just the size disregarding the industry where the firm runs its business would be an enormous error. As mentioned before, manufacturing firms need significant space for their eventual new premises therefore they are likely to be located in less populated areas, in order to face enough availability of space and save money (less expensive areas compared to cities). On the other hand, service provider firms present an important tendency to agglomerate. It is very interesting the result of the analysis of Kronenberg, (Kronenberg, 2013), because he found out that even high-tech manufacturing firms prefer to locate themselves outside the municipalities while the most of the theories suggest the opposite in order to benefit of the knowledge spillovers, the proximity to universities etc,. suggesting a kind of "fear" of the competitors in those firms (misappropriation of knowledge).

The necessity of new spaces implies that relocation becomes a problem connected to the real estate market and more in general with regulations and planning regime. An important work in that field was published by Risselada, (Risselada et al., 2013), indeed she stressed the importance of regulations in the relocation choice. She provided an example considering the Dutch planning regime, where some environmental rules forbid to polluting firms to be located near neighbours. Moreover, in the same article is described the importance of property characteristics

that affect the relocation behaviour, indeed firms located in a commercial property are less likely to move than residential properties because the latter offer few opportunities of growth. Another important evidence, that comes out from this analysis, is that firms located in lease properties are more inclined to relocate than the ones that own the premises. Other studies were focused only on the distinction between young and established companies, sustaining that the latter having undertaken investments in fixed assets, like plants, offices etc., are less inclined to move, (PELLENBARG and KOK, 1985). These authors suggested that established company prefer opening new branches or acquiring firms in other regions than relocating. A third contribution, coming from Risselada et Al., is related to the analysis of the multi-tenant properties that show very clearly, in opposition to their initial hypothesis, that multi-tenant property firms tend to relocate much more than firms that are not, suggesting that this is not a flexible solution that allows to avoid or procrastinate investments. Moving to the real estate market, it appears very heterogeneous having zones much more expensive than others. Although this situation may suggest that the less expensive areas are the most suitable to locate a firm, the discussion has to consider other important factors. Taking into account the work of De Bok and Sanders, (De Bok and Sanders, 2005), firms can be grouped considering their mobility attributes such number and types of employees, age, etc.. In the same article, the authors provide an explanatory example, "if a firm has a high car dependency for its activities, it is likely that this firm has a preference for a location easily accessible by car". Until now, in this literature review are considered just factors directly related to the firm's property (size, etc.). However, in the last two decades, at least, the concept of liveability of the employees becomes fundamental. As already said, indeed, there is a new flows of firms that decide to move away from agglomeration economies not only for economic reasons (cities on average are more expensive), but especially for labour force availability because workers decide to live elsewhere, (van Dijk and Pellenbarg, 2000), and similar conclusions are identified by Risselada.

To conclude this part, the location decision is a complex choice and looking just at the features of the real estate market is not a correct decision because it is characterised by areas that are much more expensive than others and that situation may lead to an incorrect choice. So that, a right compromise is needed that take into account all the necessities mentioned.

1.7.3. Employees & partners.

Firms do not run their business alone, but they need personnel to perform the activities, suppliers to have the resources to be used and clients that make profitable

the business model. The selection of the right partners and employees is fundamental for the success of the firm and therefore it affects significantly even the relocation decision. It is quite evident that the necessities in terms of partners and employees are a function of the business performed by the firm. For instance, depending on the type of process performed by the company, it needs high skilled workers or not. Indeed, Kronenberg in his work, (Kronenberg, 2013) distinguish between "knowledge-intensive/high-tech and less knowledge-intensive/low-tech manufacturing and service sectors" while previous authors usually considered only manufacturing firm behaviour. This distinction suggests that relocation decision is sector-dependent, meaning that not only the determinants of migration are different between high-tech and low-tech firms, but also that the destinations are different. More in details, high wages are push factors for low-tech and less knowledge intensive firms, because moving away from there would imply an important reduction of the labour costs and the future training and hiring costs are lower compared to other sectors because there is no specific expertise required. Vice versa, a knowledge intense & high-tech firm needs qualified workers depending on their competences therefore moving away from agglomeration will imply a problem of availability of skilled work-force surmountable just through an important plan of training that is very expensive. However, some of his results seem to be a little bit contradictory because they show that even low-tech manufacturing firms are attracted towards regions where their sector is overrepresented, so written in other words, are attracted towards localization economies that are a peculiar type of agglomeration economies as said in the related paragraph. This result may be seen as contradictory because in localization economy salaries are, on average, higher than less populated area, also according to the pillar of the political economy that affirms that the higher is the demand (for workers) and the higher is the related price (wages require by employees). A possible explanation of this apparent controversy is given considering that production costs diminish when many firms share the suppliers, and larger is the number of firms and stronger will be the reduction of costs, (Duranton and Puga, 2001), but a comprehensive analysis is still missed in the literature.

However, the distribution of employees is not homogenous across the regions, both considering the percentage of unemployment of an area and the distribution of high skilled workers. Indeed, the latter tend to concentrate themselves in the municipalities and, more in general, in the agglomeration economies. This important hypothesis that is wide accepted and verified by many studies such as the one of Frenkel, (Frenkel, 2001) corroborates the idea that high-skilled firms are likely to be located in cities while low skilled firms try to avoid urban area. Universities in this distribution play an important role because they train students providing them a higher degree of knowledge.

The segregation of high-skilled workers is widely studied by scholars; the large majority of them stressed the negative side of this issue that are, first of all, the spreading of the income differences and then the missed opportunity of knowledge spillover that agglomeration economies usually provide and would benefit low skilled workers; however other authors (Diaz et al., 2021) sustain that this segregation has positive consequences in terms of productivity especially if the complementarity between high and low- skilled workers is not significant. In addition, the relation between unemployment and entrepreneurship is quite vexed, indeed authors are, more or less, equally splitted between the hypothesis that affirms that an high unemployment rate spurs entrepreneurship because people cannot find other possible solutions, and the hypothesis that low unemployment rate implies a better economic wealth of the environment and therefore is easier get financing and develop a business venture, (Audretsch et al., 2015a).

Talking to the importance of partners, Kronenberg (Kronenberg, 2013), found out that the propensity of firm to relocate does not decrease when the firm shares the sector specialization of the municipality where is located in. This is a result that goes in contrast to the academic literature that affirms that being located in area specialised in its own sector provides important benefits in terms of knowledge spillovers, tacit knowledge and existence of specialized and reliable suppliers, (Holl, 2004). On the other hand, Kronenberg's results show that for both knowledge intensive and low-tech manufacturing firms the sectoral diversity in the cities is a keep factor meaning that companies in those circumstances do not show the interest to relocate. These results are coherent with the concept of urbanization economies already introduced that emphasised the importance of the reliable relation with partners of the value chain, (Nilsen et al., 2020). Moreover, Kronenberg, (Kronenberg, 2013) underlined that firms belonging to the service sector, prefer much more being in cities because depend deeply on local demand and agglomerations guarantee a much larger local market. However, considering the start-up's point of view, opportunistic behaviour of its partners may be a real threat due to the higher specificity of its assets (common characteristics for start-ups), (Lee, 2022) indeed according to the literature "the higher the asset specificity is, the more opportunistically the firms' partners are expected to behave, because the invested resources are already locked in (or held up) for the use of the firms' invested partners", (Williamson, 1981).

Another important contribution comes from Duranton and Puga, (Duranton and Puga, 2001) because they make a study discovering that firms change their tendency, in terms of relocation, according to the stage of life in which they are. In the article, the authors identified two main moments of the firm's life: a first one, where the most important aspect is learning and a second one, where being

profitable becomes the priority. They suggest that companies are using an approach very close to the one described by the evolutionary economy that is based on trial and error, therefore the more is diversified the context (and so the local knowledge) and the more they learn, making possible find their proper business model. Thus, they argue that, in the first phase, companies prefer being located in urbanization economies. In the second phase, that starts when the proper business model is identified, companies prefer to locate themselves in localization economies and exploit the benefits previously mentioned.

1.7.4. Finance

How to finance an entrepreneurial venture is the one of the most important and discussed topic in the academic literature and of course it has an impact on the relocation decision. One of the main reasons is that firms usually do not have the money to successfully run their business thus they need to get money from others. Depending on the stage at which the firm is, different actors interact with the company; for instance, in the very first stage of a new-born start-up is fundamental, besides the entrepreneurs' savings, the economic support of the Family, Friend & Fools, because they are the only available to invest in such ventures. This evidence gives a support to the hypothesis that local entrepreneurs find less difficulties in developing their business.

It is quite clear that the lack of information is a huge problem in these phases, and this prevent other investors from giving a potential money provision. Indeed, they find difficulties in assess the new-born companies because of many possible reasons such no track records, no benchmark, no clear addressable market, defining the so called "funding gap" that is true even for the new technology based firms (NTBF), (Colombo and Grilli, 2007) that would have the potential to spur significantly the economic environment. Thus, the context is characterised by information asymmetry both ex-ante, the so-called adverse selection, meaning that is impossible for investors distinguish between good and bad start-ups (and companies in general), and ex-post, meaning that investors have difficulties in controlling the behaviour of the entrepreneur once received the money. The presence of ex-ante information asymmetry implies that costs of financing increases, (Myers and Majluf, 1984), especially when a new-born company relies on equity. Indeed, this milestone of the literature affirms that using internal resources is the cheapest solution for a firm, and in case its availability is not enough is better to issue debt than equity (Pecking Order Theory). The aforementioned Family, Friends & Fool can be considered as a type of internal resources, (Colombo and Grilli, 2007).

A possible way to avoid the external capital is performing bootstrapping that implies the reduction of the costs of operations even included investments and so important opportunity costs. However, the bootstrapping does not work with the same efficiency in all the sectors. Another creative way to avoid external capital is through crowdfunding, a modality of funding that exploits the new technologies. Some scholars have discovered that there is a local bias also in the crowdfunding meaning the proximity, between who gives money and who gets money, increases the probability that this happens, (Hornuf et al., 2022).

If these alternatives are not feasible, a start-up should issue debt, for instance asking for a loan. However, the probability that it occurs are extremely low. This is due to an inclination of entrepreneurs to keep secret the most important information to avoid leakages related to the start up, and, sometimes, due to an overclaiming about the commitment. Moreover, it is no ratings for start-ups on which banks can rely on. In the real world, this is translated into the so-called, credit rationing that can occur in two different ways: a) the entrepreneur receives a smaller amount of money at the same interest rate; b) not all the entrepreneurs get the loans despite they seem equal, (Stiglitz and Weiss, 1981).

Thus, at the end is quite common that start-ups are obliged to issue equity despite is the most expensive way. Here, the VCs are by far the most important actor involved.

1.7.4.1. Venture Capital availability

In the previous paragraph is written why the venture capital is the main source of external funds for start-up while in this paragraph will be analysed the connection between availability of VC and the relocation of start-ups. The topic regarding the determinants of VCs' investments is widely discussed in the literature, while the connections between their spatial distribution and the relocation is investigated but not with the same completeness.

In that case, it is important to present, briefly, the main features of VC. (I)VCs are professional investors with a huge availability of capital and specific competences, composed by the general partners (GP) and the limited partners (LP). Usually, GP is composed by a limited company, participated only by the fund promoters, that provides a small fraction of the overall fund to increase its credibility. LP is composed by different kind of investors (pension fund, insurance companies etc.), providing the great part of the investment fund. The investors belonging to LP have a priority in case of liquidation. VCs' investments are time bounded, meaning that they invest considering, since the beginning, an exit strategy through they should

realize a profit. There can be identified four different types of VC: a) Independent VC; b) Corporate VC, c) Bank-controlled VC; d) Governmental VC. IVCs (most common type of VC) are the only ones that pursue just financial purposes. Being backed by VCs provides significant benefits to start-ups not only because they ensure the fundings needed to run their business, but also because VCs provide networking and coaching support that usually turns out to be significant for the success of start-ups themselves. To give some information to better understand, VC-backed firms are more likely to survive in the following five years since the investment, and they have higher probability to be acquired having a better valuation compared to non-VC-backed firms, (Puri and Zarutskie, 2012). This shows why start-ups are so attracted to VCs.

Recently, some authors have underlined how local bias of VCs, (Cumming and Dai, 2010) affects the relocation decision of the start-ups, (De Prijcker et al., 2019). So, first of all is important understanding the characteristics of that bias. Cumming et Al, (Cumming and Dai, 2010), found out that distance increases the information asymmetry and the cost of monitoring, therefore VCs are likely to invest in local start-ups. The same authors, in a later article, uses an expression to describe this VCs' tendency that is the "twenty minutes rule" meaning that a start-up must be located in a place accessible within 20 minutes from the VCs' office, (Cumming and Dai, 2012). Moreover, they added that this bias diminishes when there is competition between VCs in the area, and that reputable VCs tend to have a larger investment range. Conversely, a higher degree of specialization of the VC increases the attitude of investing in local firms. Another important contribution was given by Powell and colleagues, that found out that elder VCs tend to increase their investments towards non-local firms, especially if these ventures are older and larger, (Powell et al., 2002). These results are very important because the proximity with the VCs can be seen for sure as a push factor, especially in the very first stages of a start-ups, and therefore included in the set of determinants for relocations. In fact in a recent study, De Prijcker and colleagues showed that start-ups located in areas where the availability of VCs is poor, are likely to move to the two most important VC hubs of the world, that are the Silicon Valley and Boston, and by doing so they have a higher probability to be financed by VCs, compared to companies that have been founded there, (De Prijcker et al., 2019). A possible explanation may be that being able to undertake the investments related to a relocation is a signalling of the wealthiness of start-ups. However, relocation should not be seen as a perfect solution, indeed the same authors underlined that the percentage of failing for relocated start-ups is higher compared to start-ups that remain in their home state, (De Prijcker et al., 2019). It is important stressing the fact that the VCs' decisions, regarding in which start-ups investing in, have a significant

human component and therefore "unsystematisable". Gut feel and noisy cues are very important at this stage, (Huang and Pearce, 2015).

However, the distribution of VCs is not homogenous, especially considering the U.S case, indeed the most important VCs are concentrated in few areas: Silicon Valley, Boston and New York Metropolitan area in U.S and London and Paris Metropolitan area in Europe, (Colombo et al., 2019). These areas are characterised by a significant number of start-ups implying the presence of tough competition among them that leads to an increase of the costs for start-ups. Moreover, the more is reputable the VC and the more the entrepreneur pays a higher premium to be backed by VC, (Hsu, 2004). That hypothesis leads other authors to sustain that the most suitable regions where founding an entrepreneurial venture are not the best considering the financial performance, (Stuart and Sorenson, 2003).

Some authors have underlined that these preferences of VCs for local investments may be caused by the demand side, because entrepreneurial ventures, located far away from VCs, are highly sceptical in get financed and therefore, they renounce to ask for capital, (Mason and Harrison, 2002). Arguably, the evidence of local investments is the result of a combination of local bias and reluctancy of seeking for external capital by distant start-ups.

Being located in the proximity of VCs is not only important to increase the probably of receive external capital, but some results suggest that this closeness positively affects the growth of start-ups, indeed Cumming and colleagues, found out that local start-ups have a higher probability of performing a successful exit strategy, that is represented by an IPO or M&A, (Cumming and Dai, 2010).

To conclude this part, other articles, previously mentioned regarding relocation, underline that is not straightforward the connection between migration and availability of VCs. Indeed, Colombo and colleagues (Colombo et al., 2019) found out that in Europe start-ups founded in areas characterised by a low availability of VCs does not show a higher propensity to relocate than those start-ups located near VCs' hubs. Also De Prijker and colleagues, in their work stress the fact the VCs' available is not the only determinants but it has a non-trivial weight in the relocation theories, (De Prijcker et al., 2019) also considering the impact of human component in the picking decision of VCs. Besides that, considering the European situation, where start-ups migrations occur even when these ventures are not seeking for external fund and where the VC market is much smaller than the American one, it might suggest that the role of VCs in relocation decisions is not so impacting.

1.7.5. Policymakers' initiatives.

Government, and more in general, policy institutions can have an important role in affecting the relocation decision of firm, both pushing away or attracting them. They cannot change the characteristics of a company, but they can modify the environment/context in which it operates in a way that does not allow the firm to continue its operations there. For instance, they can increase the taxation and make unprofitable the business. It can be identified two main areas of intervention: a) taxation, laws etc,; b) investment in infrastructure such as motor ways, airports, internet connection etc,.

Scholars have studied quite intensively the impact of taxation on relocation, and it is quite obvious that a low taxation would benefit larger firms that otherwise should pay significant amount of money; this issue is important also because has important political consequences. However, setting a low taxation rate can be problematic for the economic sustainability of the environment and some social issue may arise. Talking about some data, Chow and colleagues, (Chow et al., 2022), have conducted an important analysis that found out that "a one-percentage-point increase in the HQ state corporate income tax rate increases the likelihood of firms relocating their HQ out of the state by 16.8%, and an equivalent decrease in the HQ state rate decreases the likelihood of HQ relocations by 9.1%." In particular, taxation becomes extremely important when international relocation (or opening of new branches of the firms) is taken into account. Indeed, taxation is one of the main determinants when a firm has a set of foreign locations available but seems to be much less important in the decision whether performs an international relocation, (Devereux and Griffith, 1998).

Also, through laws, a policymaker can help to create an environment that supports entrepreneurs and therefore increases the local (or national) attractiveness to new businesses. An important example in this field, is the "Law on Innovation and Research to Promote the Creation of Innovative Technology Companies" adopted in July 1999 that enabled academics and researchers to participate in the creation of a private ventures. Moreover, it allowed Universities to set-up incubators to encourage spin-off creation, (Autio et al., 2014). Closely to the topic of laws, there is the one of the bureaucracy, indeed reducing the administrative barriers make easier run and start business therefore start-ups may decide to move in order to face a less complicated system. To give a concrete example, to be honest about location and not relocation, Djankov and colleagues (Djankov et al., 2002), showed that to open a start-up in Italy are needed 62 days, while in U.S only four. Moreover, their results showed that heavier entry regulation is a characteristic of the most corrupted countries, that is another parameter that push away firms. Another important way

to support the economic growth of a region, is developing the so-called Governmental Venture Capital, that is a very peculiar type of VC owned by governmental bodies. GVCs undertakes riskier investments because they are aimed at creating job and developing local economy rather than being profitable, therefore they provide capital earlier than other types of VCs; these investments are sector-specific, usually sectors are those related to new technology-based firm, and spatially concentrated, mainly within great metropolitan areas. The presence of GVC in a region is extremely attractive for a company if it has the same sector-specificity. The drawback of this option is that when a GVC-backed company fails, then it is a public wasted money. In addition, public authorities have understood that they can spur the economic growth of an environment, through a set of initiatives such the support to new venture R&D or providing grants to the firms that want relocate and by doing so they have created pull factor determinants. For instance, Germany has significantly encouraged the bio-technology industry by supporting the start-ups, (Autio et al., 2014), (Dohse, 2000).

Investments in infrastructure are very expensive, irreversible, and usually carry out by public institutions or a public-private partnership, involving hundreds or maybe thousands of workers, (Audretsch et al., 2015b) (for instance for the construction of an airport). The importance of the infrastructure is strictly connected to the importance of accessibility. It is possible to say that, nowadays, accessibility is not only a physical matter but, and especially, is related to internet connection, indeed, no firms can survive without it. This is true especially for start-ups as noted by (Audretsch et al., 2015a). Infrastructures are important because they allow, physically, the exchange of knowledge between near ventures and sometimes even further, (Audretsch et al., 2015b). Also considering the traditional civil infrastructure, there was a significant change comparing the last century with the current one. In the current days, having an airport becomes fundamental for ventures, especially when they have a large market to serve and far partners, that is a situation that might be realized when the company has grown. So, airports can be seen as important keep-factor in the relocation topic. Some authors have studied the role of airports in the regional economic development, (Florida et al., 2015) saying that they are a key component to provide connectivity to a place. The topic regarding the connection between cities and infrastructure is quite debated in the academic literature, in particular Neal and colleagues, said that the economic prosperity of a city depends on "structural advantage" (better infrastructure) and that "centrality drives employment" , (Neal, 2011). Indeed, cities can attract people that live in other cities (commuters) only in case of important and reliable infrastructure. Even motorways are considered in the discussion related to relocations, as highlighted by Targa and colleagues, indeed high-capacity roads affect the decision regarding location and relocation, (Targa et al., 2006) and more

in general, data show a significant correlation between availability of important infrastructures and the economic activity of the city. Thus, when a policymaker wants to create an entrepreneurial ecosystem, it has to invest in infrastructure because they seem to be a mandatory pre-condition to spur the local development attracting new employers and start-ups, (Targa et al., 2005).

1.7.6. Incubators & accelerators

Both incubators and accelerators can be seen as institutional infrastructures, or at least a portion of them, but they are discussed in a different paragraph because they are tailored in encouraging the entrepreneurial initiatives and the creation and support of the start-ups, while the other infrastructures, discussed previously, are more "general purpose". Incubators and accelerators differ because they intervene in a different stage of start-ups, indeed the former play a fundamental role in the very early stages while the latter in further stages. Both of them provide support only to selected start-ups and the programs provided are time bounded meaning that a firms must leave when it has grown or after a certain number of months. These programs are very attractive for start-ups also because attending that courses increases the possibility to receive capital from VC.

Starting from the incubators, The National Business Incubation Association (NBIA) defines business incubators as "a catalyst tool for either regional or national economic development", identifying five categories: a) academic institutions; b) non-profit development corporations; c) for-profit property development ventures; d) venture capital firms; e) a combination of the above. Incubators' programs are useful to reduce the obstacles faced by new-born start-ups, both the financial ones, through their connection with funds, and the operating ones through the provision of space/offices, computer services, coaching etc, (Aernoudt, 2004). Often the importance of sharing the space is underestimated, indeed this allows knowledge spillovers and experience sharing, (Bergek and Norrman, 2008). Incubators can be classified also considering how they pick the start-ups; more in details, there are four approaches: a) idea-based selection; b) entrepreneur-based selection; c) "picking the winner"; d) "survival of the fittest". In the first case, the focus is on the evaluation of the entrepreneurial vision, in the second case what matters is the characteristics of the entrepreneur and his/her team (education, previous experience etc.), in the third approach incubators are extremely selective stopping the large majority of the proposals, while in the last approach the selection ex-ante is not so rigid, but it occurs even during the programs, (Bergek and Norrman, 2008).

Arguably, the fact that incubators are related to the starting phase of a venture may imply that it affects more location decisions than relocation, indeed it does not seem reasonable choosing a location, and immediately after performing a relocation, sustaining the related costs, as the Montpellier's case suggests, (Cloutier and Messeghem, 2022).

Start-ups accelerators are educational programs for still young start-ups but little bit more mature compared to the ones attending incubators' programs; accelerators' programs include mentorship activities, coaching etc., and thus they should be seen as a relocation parameter. These programs last usually few months (at most 6) and is highly difficult be selected to take part of them. During the accelerators' program start-ups receive external funds, usually in exchange of equity, (Larsson, 2012). Talking about the possible connection between accelerators and migration firms, Brown and colleagues, highlighted that transnational entrepreneurs represent a non-negligible component of the accelerators program held in the Silicon Valley, and in that way, they can exploit, at once, the benefits of being in contact with a different context and the benefits of being embedded in their local context, (Brown et al., 2019). The result of this study might suggest that accelerators are not a significant factor in the relocation decision.

However, the academic literature overlooked the connections that exist between the presence of incubators and accelerators, usually in cities, and the theme of relocation. Indeed, most of the studies focus on the benefits related to their presence in the economic environment but do not show interest in understanding if there is, for instance, a local bias in the selection of start-ups, or if there is a migration of start-ups attracted by these programs. Further analysis should be done to clarify the topic.

1.7.7. Universities and Research Centres

This topic will be discussed deeply because the following analysis is aimed at assessing the interactions among relocations and knowledge institutions, such as universities. The current society is named in many ways and one of them is the "knowledge society", expression made well-known by Drucker in 1969, (Drucker, 1969) because all the players involved in the current economic activities, considering their broader meaning, have to deal with knowledge therefore universities and research centres, that are the institutions entitled to hold and diffuse that knowledge, become (or should) extremely important for the society. To perform completely this passage, universities have to add to their common services of teaching and researching, the ones of creating and applying their knowledge,

(Etzkowitz, 2003) (Etzkowitz, 2013), and becoming an "international know-how hubs", (Wissema, 2009). Scholars have intensified their studies regarding the intersections between universities and ventures especially after the publication of an important theory that is called "The Triple Helix" where it is stressed the benefits that arise from the collaborations between universities, entrepreneurial ventures, and policymakers. Indeed, the economy is becoming, year by year, more knowledge intensive and places such as universities and laboratories increase their importance, (Etzkowitz and Leydesdorff, 2000). For that reason, the authors concluded that the utility function of the firms cannot be based just on the profit maximization, but it has to take into account also the opportunity "structures".

As underlined by Larsson and colleagues, "many students move locations after graduation and the extent to which they start businesses close to their place of graduation or elsewhere has not been studied", (Larsson et al., 2017) and it is appropriate adding that is not discussed neither if being close to Universities is a parameter that influences the relocation decision of a venture. Larsson and colleagues have highlighted, in the same work, that is important considering graduates university because they are in their most "mobile life-stage", and so they can choose without many limitations that may arise in further moments of their life, for instance family, suggesting that the propensity to relocate is not just a matter of maturity of ventures but also of maturity of entrepreneurs. Moreover, university graduates show a stronger inclination to start a new business compared to university employees and this inclination is affected by the type of university attended, that means that students coming from ranked universities are more inclined to start a venture. In addition, according to the authors, students that graduate in universities located in metropolitan area tend to remain and start there their entrepreneurial ventures. An interesting topic related to university graduates is the so-called "brain drain" that consists in the coming back home of students graduated in a foreign country after having gained entrepreneurial skills and managerial experiences. Usually, this phenomenon takes place when the home country is a developing country, while the graduation country is a developed country, (Kenney et al., 2013). So, authors suggest that the relocation of ventures can be motivated not only by financial reasons but also by cultural and familiar reasons. Moving to the academic entrepreneurs (professors and PhD students at universities), there is an aspect that affects their location decisions, that is the role held by them in the start-ups, in fact when they are the founders, or they are preeminent then is common that the firm is located in their proximity while when they are members of the board or advisories the probability falls. Moreover, the proximity to knowledge assets affects their decision, meaning that they tend to locate their business close to those assets, (Kolympiris et al., 2015).

Audretsch and colleagues, proposed that the proximity to universities is a locational strategy because firms are located in order to exploit knowledge spillovers coming from the nearness to universities and in that way, they reduce the companies' knowledge acquisition costs. However, the decision of relocation has to consider also the higher costs (quite common) of being located close to university, (Audretsch et al., 2005). Moreover, in their results emerge that start-ups based on new knowledge/technologies tend to locate closer to universities and they argue this is since spillovers in that fields are very precious. In addition, The theory of localized knowledge spillovers suggests that," profits will tend to be greater in agglomerations and spatial clusters, since access to tacit knowledge is greater" and that transferring cost of that knowledge is a function of the distance to be covered, (Siegel et al., 2003), thus the benefit of being close to the universities diminishing with the distance. Another reason is the fact that this transfer of knowledge is performed through personal contacts. According to the Audretsch and colleagues, the knowledge spillover may occur in two ways: a) through publication, and in that case it does not affect the location choice; b) through mobility of "human capital", (Saxenian, 1996) that are people involved in universities as employee or graduate students, and this affects the location decision. Moreover, they sustain that larger is the output of university and the closer will be the location of a start-ups and that younger firms tend to locate themselves, closer to university compared to older firms, because they cannot afford important costs of R&D and therefore are more dependent on external knowledge and so spillovers. This last sentence is particularly important for this thesis because it says, basically, that proximity to university is an important driver for the initial location but reduces its importance later when relocation may take place.

Other authors in their work underlined the importance of social proximity, intended as people with some academic relationship, as an element that increases the interactions between universities and firms, filling the gap that exists because the universities and research centres are interested in a more theoretical knowledge while the entrepreneurs are interested in the application of it, (Colombo and Garcia, 2021). However, in their work is not considered if this social proximity may also fill a geographical distance between the entrepreneurs and the university, and this work is aimed at clarifying this issue. Moreover, the authors suggests that social proximity implies a higher degree of trust between the actors involved and this can be the reason why it might overcome the geographical distance issue, since frequent interactions are not anymore needed. Thus, the issue of the social proximity is linked to the one of "first match" that tries to explain which are the determinants that increase the percentage that an entrepreneur establishes a working relationship for the first time with a university, (Colombo et al., 2022). According to Colombo and colleagues, this first match is influenced by common background between the

entrepreneurs and personnel of the universities, distance and cultural proximity that is named "cognitive proximity". The first match does not occur naturally but it needs resources for the evaluation of the partner, but nor universities neither start-ups, or in general young ventures, have abundance of them therefore understanding if a parameter reduce that costs is fundamental to know how they choose, (Colombo et al., 2022). It is quite reasonable to say that collaborations occur when the benefits overcome the existing costs. More in details, for a university having contacts with many valuable firms/start-ups increases significantly its prestige while the start-ups can use students and academic personnel as additional employees, (Drejer and Østergaard, 2017) arguably in a first moment temporarily and then permanently. According to Colombo and colleagues, the importance of social proximity decreases when universities and entrepreneurs get closer, probably because trust in that case is based on interactions. Finally, they introduce a new element to the discussion that is that policymaker may undertake some initiatives to encourage the relocation towards universities for example establishing a university incubator/accelerator, (Colombo et al., 2022). Another important contribution to the discussion is brought by D'este and Iammarino (D'Este and Iammarino, 2010). From the analysis of their results, they conclude that engineering-departments are more affected by geographical proximity rather than science-based departments, in starting a collaboration. Moreover, they provide a hypothesis of the "structural features of the spatial profile of university-industry interactions", suggesting that: the higher is the concentration of universities, the larger is the number of local spillovers therefore the inclination in collaborating with local firms and that the more frequent are that collaborations and the more, both universities and venture, accept to interact with a further partners, (D'Este and Iammarino, 2010). To conclude this review, also D'este and Iammarino sustained that their work is one of the first steps of this area that must be deepened, and this is the aim of that thesis.

2   Hypotheses

In this chapter are presented the hypotheses elaborated after the study of the academic literature regarding the topic of start-ups relocation, shown in the previous chapter. This analysis wants to highlight the possible connections between the availability of social resources, in particular universities and research centres, and the relocation process of a firm. The reason that encourages to undertake this stream of study is the difference that emerges between U.S situation and Europe situation. Indeed, two important analyses were done in the same year, 2019, and provided very different conclusions. On one hand, there is the analysis conducted by De Prijcker and colleagues, (De Prijcker et al., 2019) that found out that in U.S., ventures tend to relocate towards the main two VCs' hubs (Silicon Valley & Technology Square) when the availability of VCs in their home state is low. On the other hand, Colombo and colleagues (Colombo, D'Adda and Quas, 2019), showed that in Europe "companies located in regions with less VC or that are more distant from VC hubs are not more likely to relocate than other companies" and when relocations take place they do not imply that ventures and VCs get closer. It is also fundamental stressing the intrinsic differences between U.S. and Europe, indeed moving from a state to another in U.S. is much easier than in Europe due to the common language and close culture. Moreover, the size of VCs market is completely different: according to KPMG, in U.S. during the 2021, VC investments were $330 billion, while in Europe were $123 billion, and these data are an empirical evidence of the different weight that VCs' availability has in relocation decisions between Europe and U.S. Thus, considering the European case, the debate about what drives relocation is very open. A possible explanation of the relocation process, and the different results among continents, might be the geographical distribution of universities, that in Europe are more homogeneously spreaded throughout the countries, because they are older institutions whilst in U.S. the most important VCs' hubs are located very closely to the most important universities of the country (for instance MIT for the Technology Square & Stanford for the Silicon Valley). That thesis focuses on the European situation, nevertheless some American studies are taken into account as reference points because in that way the academic literature is much larger and to point out eventual further dissonances between the systems.

The hypotheses that will be presented in this chapter have to take into account the characteristics of the sample used in the analysis. In other words, the database VICO 5.0 include companies that have received VCs' funding and therefore they can be

considered as innovative firms. The companies belonging to the control sample, that are not VC-backed, have to be as similar as possible to the others in order to be more accurate the final results of the analysis and therefore it is possible to say that even the latter are innovative firms. So, through the sample used is not possible to assess if there is any difference between knowledge intensive firms and non-knowledge intensive firms, for instance. Nevertheless, it is possible to check if high-tech companies prefer being located closer to universities, exploiting a boost in innovation and the spillovers (Frenkel, 2001), being the latter a sort of "know-how hub", (Wissema, 2009), by considering the NACE Rev.2 of the companies and their locations. However, this topic is vexed because other authors suggest that firms prefer to locate themselves outside the cities because of the misappropriation risk, thus this thesis will contribute to the debate.

Having said that, it is important saying that "University & relocation" is a topic not discussed deeply by the scholars and this is quite peculiar because, so far, they have considered and studied many aspects of universities such as their importance as a parameter for location, (Audretsch et al, 2015), or their entrepreneurial side (universities spillovers, academic entrepreneurship etc.), (Etzkowitz, 2003) (Etzkowitz, 2013), as well as the contribution of universities to the context (knowledge spillovers, higher degree of innovation, etc.), (Audretsch et al, 2005). However, none of them verify, for instance, if a start-up has moved to a place where there was/is a reputable university that provide each year many new high-skilled employees. This work is aimed at providing a contribution on the filling process of that gap. In making the hypotheses, particular attention was given to those studies that try to identify some spatial patterns for the location, for instance, D'este and Iammarino (D'Este and Iammarino, 2010), affirm, through their work, that science-based departments collaborate with firms/start-ups located further compared to the engineering-based departments, thus it would be interesting studying if those companies that collaborate with science-based department tend to relocate more, and in case if they relocate towards the university regions. It is also important highlighting, also in this chapter, that this thesis is focused specifically on the start-ups relocation while the large majority of the existing literature consider firms relocations, probably because there is a larger availability of information. However, focusing on mature firms may lead to a partial result because many authors affirm that elder firms are more embedded in the context and therefore less inclined to move, sustaining higher costs, (Brouwer et al, 2004). Another important aspect is the overlooking of the international relocation, while, as already mentioned before, for foreign entrepreneurs a migration of the venture may be caused by cultural aspect, (Kenney et al, 2013). The reason behind this choice is the difficulties in tracking those kinds of relocations, especially if they are intercontinental relocations. For that reasons scholars have studied relocations within a specific

country, and usually it is the one where they work and live, (De Bok and Sanders, 2005), (van Dijk and Pellenbarg, 2000) to give an example. This situation has important consequences because each country is analysed in a different way depending on the decisions made by the scholars. This implies that just comparing the results coming from several analyses is not a correct approach because it would not consider the ex-ante conditions. Moreover, it brings to a situation where some countries are much more studied than others, indeed the Netherlands and, in general, the northern Europe, are much more discussed in the literature, while, for example, Spain is barely considered. However, the real world is full of companies that relocate to another country to exploit the differences in terms labour costs, bureaucracy, and norms etc., (Djankov et al., 2002). The attention for this relocation comes out when an incumbent player decides to move away, usually leaving hundreds of employees without work; however, this thesis tries to clarify a little the start-ups' situation also considering international relocation, in case data allows it of course.

As said before, many scholars pointed out the importance of the proximity to university; indeed, it reduces the costs in accessing to the academic knowledge, (Heblich and Slavtchev, 2014), and transferring that knowledge becomes more expensive as the distance to be covered increases, (Siegel et al, 2003); on the other hand, some authors suggest that the first location decision is taken under important restrictions suggesting that is not uncommon that the location chosen is not able to fulfil all the needs, (Stam, 2007); to be more tailored on the aim of this thesis, the analysis is focused on the behaviour of companies (start-ups) located far from universities. Regarding the issue of relocation, two components can be identified: the first component where the object of analysis is understanding which factors affect the probability of relocating while the second component is aimed at identifying where they migrate and why.

So, it is now possible present the first and introductive hypothesis.

Hypothesis 1: companies that are located in regions where the availability of universities is lower, tend to relocate more than companies close to these institutions.

It is also possible to introduce another hypothesis referred to the probability that a venture relocates itself. More in details, it is interesting understanding if the specialization of a university affects firms' tendency to relocate. Indeed, the academic literature stressed a lot the importance of intrinsic characteristics of the universities, (Nilsen et al., 2020), (D'Este and Iammarino, 2010) in the location and relocation process, therefore this is a topic is going to be discussed further. Indeed, the fact that a university is specialized in a specific-field, or it is well-known for a

particular department, may be a signal of belonging to a localization economy where the availability of high-skilled labour is higher and therefore the training costs are significantly lower, (Nilsen et al., 2020). On the other hand, universities that do not present a strong specialization may be most suitable for urbanization economies where externalities are fundamental. So, the results coming from these hypotheses could provide some tips to analyse if companies are more attracted by localization economy or urbanization ones.

For what it is written up to now, it is possible to introduce the hypothesis dealing the topic of specialization. Of course, the specializations considered in the analysis must be coherent with the characteristics of the companies meaning that the companies and the specialised universities must belong to the same industry or area of study. Moreover, the comparison of the results of the first and second hypotheses can be a measure of that importance of these intrinsic characteristics on the relocation decision.

Hypothesis 2: companies that are located in regions where the availability of specialized university, in the area of interest of the company, is lower, tend to relocate more than companies close to these institutions.

Arguably, the effects of specialized universities on the decisions are even more intense when they are reputable. Reputable universities seem to attract significantly entrepreneurs and there are some empirical evidence that may corroborate this thesis, for instance, considering the areas around Oxford and Cambridge (Smith and Ho, 2006); it is true that a higher concentration of start-ups around reputable universities is also due to graduates students coming from those universities that decide to start a new venture, (Larsson et al., 2017), but understanding if there is a further contribution coming from relocated start-ups, and assess, if possible, to what extent, would clarify significantly this topic. Another aspect that may be seen as a confirmation of the quality of this hypothesis, is the fact that having collaborations with high ranked university could be a signalling of start-ups quality, and in that way, they can have a higher probability of receiving external funds, for instance from VCs. Therefore, it is worth analysing if reputable universities attract more than others.

Hypothesis 3: companies that are located in regions where the availability of reputable universities is lower, tend to relocate more than companies close to these institutions.

As said before, the issue of relocation can be divided into two components and these first three hypotheses are related to the first part that is aimed at understanding the factors that affect the probability of relocation. Thus, they are useful to introduce

the problem, underlining if there is a difference in the behaviours but without describing an eventual spatial pattern of that phenomenon because negative results might be the consequences of the wrong spatial pattern proposed rather than evidence of the fact that universities are not important in the relocation process.

The following three hypotheses will be related to the second part of the topic that is the identification of place where they migrate in case a relocation takes place. For that reason, these hypotheses will be more specific. Once said that other three hypotheses can be introduced:

Hypothesis 4: the availability of universities of the region where the company has relocated is higher than the one characterising the previous location.

Hypothesis 5: the availability of specialized universities, in the area of interest of the company, of the region where the company has relocated is higher than the one characterising the previous location.

Hypothesis 6: the availability of reputable universities of the region where the company has relocated is higher than the one characterising the previous locations.

In the literature review emerges the importance of VCs for start-ups as the main source of external funds and for their networking and coaching activities. Authors have also studied if their position affects the relocation decisions of firms, (De Prijcker et al., 2019), (Colombo, D'Adda and Quas, 2019), however the results do not highlight a clear pattern. The purpose of this thesis is not to solve this issue, nevertheless it may be useful in assessing if being backed by VC affects the propensity to relocate. Indeed, a VC-backed firm, through the coaching and networking activity performed by the VC, may be less dependent on external knowledge (Audretsch, Lehmann and Warning, 2005) and so having a lower propension to relocate towards universities. Moreover, moving away from a VC may negatively affect the impact on the service provided to the firm, so being backed may be seen as a keep-factor, (van Dijk and Pellenbarg, 2000). The hypothesis is defined as follows:

Hypothesis 7: VC-backed firms have a weaker inclination in relocating towards universities compared to non-VC-backed firms.

The last hypothesis, presented in this chapter, is introduced as a sort of verification of the previous ones. Indeed, due to the fact that many universities are located within cities, and so usually within agglomeration economies, is not so easy distinguish which is the real driver of the relocation. In order to check if the migration of firms towards universities is the driver rather than a side effects, the number of patents registered by firms after the relocation is taken into account. The

assumption made here is that if a firm is moving towards a university is because of the intention of collaborating with it. The assumption is considered solid because collaborating with a university is a direct way to exploit those knowledge spillovers, previously mentioned, and in addition it can be seen as a solution against the misappropriation issue because it "is an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something" (definition taken from WIPO website), so preventing the others to use the invention.

Thus, the last hypothesis is:

Hypothesis 8: after the relocation towards university, a firm increase its number of patents registered.

The very last point of this chapter is used to explain, briefly, why the hypothesis regarding the process of relocation towards entrepreneurs' alma mater was rejected. There are two main reasons: the first one is the difficulty in identifying the entrepreneurs belonging to the relocated firms and so to identify which is their alma mater; secondly, authors agree on the importance of the position of the attended university in the first location rather than in later moments, (Larsson et al., 2017).

3   Materials & descriptive analysis of the dataset

Now it is the moment to start the explanation of the empirical part of this thesis. First of all, it is necessary to describe how and where data have been collected. The databases used are two: a) VICO 5.0; b) ETER. The former was used to take into account firms and their eventual relocations, while the latter was used to consider universities. Moreover, the EUROSTAT database is used as a source of information but to understand the characteristics of the NUTS3 observed.

3.1. VICO 5.0 database

The VICO database contains information about firms founded starting from 01/01/1988 which have received at least one venture capital starting from 1/1/1998 up to 31/12/2020, operating in many European countries (Belgium, Finland, France, Germany, Italy, Spain, and the United Kingdom and others) and Israel. More in detail, that database contains information on more than 76000 distinct investors, of which 69855 venture capitalists (VCs) and 5050 business angels (BAs). The VICO 5.0 database uses data coming from other three database that are; a) Thompson One Private Equity; b) Zephyr; c) Crunchbase. In addition, accounting information from 2005 to 2018 have been collected from the Orbis database. The database is currently available in the Stata format (.dta) and the key aspect of that database is that both companies and investors are geographically defined using a Google API and then to each position is associated a NUTS code (Nomenclature of territorial units for statistics). The NUTS is a classification that divides European Union and United Kingdom in areas where overlappings do not occur. It is important to say that this classification presents 3 hierarchical levels:

NUTS 1: It defines macro socio-economic regions (ex: Northwest Italy). Each NUTS 1 is defined through a code of 3 capital letters in the majority of cases (considering the previous example, Northwest Italy is coded ITC), while in some cases the last letter is substituted by a number (for instance in Germany).

NUTS 2: it defines regions of a country (ex: Lombardy). Even the NUTS 2 is defined by a code, obtained adding to the NUTS 1 code of the macro area to which the region belongs, a number. For instance, the Lombardy's code is: ITC4.

NUTS 3: it defines a small area that is comparable to the Italian province. The NUTS 3 codification follows the same reasoning of the NUTS 2, indeed these small areas are "named" considering the related NUTS 2 code, at which it is added another number. For instance, the province of Brescia is named ITC47.

All the companies belonging to the sample used for the analysis present all of these three codes.

(https://ec.europa.eu/eurostat/web/nuts/background)

The database is structured in two main tables: the first one considers the deals, having as units of analysis considered the company that receives the founding, the investment deal itself and who provide the external capital, while the second table takes into account the accounting data of the firms, that are divided year by year.

Talking more in detail about the first table, the information can be clustered in three groups coherently to the classification of the units of analysis. Indeed, there are information regarding the companies (CompanyID, CompanyName, CompanyNACERev2Corecode, ZipCode, Nation, City etc.), regarding the investments (RoundNumber, InvestmentYear, InvestmentDate, TotalEquityInvested_round etc.), and regarding the investors (InvestorID, InvestorName, ZipCode, Nation, City, InvestorStatus, InvestorType, etc.). Moving to the second table, there are the accounting data regarding the same sample of firms. Here, it is possible looking at information on the Balance Sheet of the companies (FixedAssets, NonCurrentAssets, CurrentAssets, Equity, Debt, etc.), on the Income Statement (Sales, Gross Profit, EBITDA, R&D Expenses, etc.), on the Cash Flow Statement, and Number of Employees.

Some analyses have been performed on the overall database to highlight some of the main characteristics. First of all, it is computed the frequency of the different types of investors. The following tables confirms the existing theory that considers Venture Capitals as the main external source of finance both quantitatively and qualitatively, especially if it is considered the Independent VC type that accounts for 49.42% of the overall number of investors, takes part in the 64.85% of the investments (for whom is known the investors) and provides the 67.81% of the overall value of the total equity invested for those investments where the amount invested is known, as shown by the figure below. Regarding this information, the VICO 5.0 database appears to be quite complete indeed it provides information about 11334 out of 12015 companies considered in the database. To get the values of the second table (table 3.2) and of the pie chart (figure 3.1), two databases, both provided by VICO 5.0 are merged, and by doing so the unit of analysis is not, anymore, the investor but the investment, meaning that it is considered if an investor has made more than one investment even towards the same company. Moreover, the pie chart takes into account even the size of those investments. In these procedures, some variables have been dropped because they are not useful for the purpose of this thesis, and those observations, where it is not possible identify who performed the investments and the size, have been dropped too.

Investor Type	Frequency	Percentage	Cumulative
BA	2801	24.71	24.71

BVC	597	5.27	29.98
CVC	1583	13.97	43.95
GVC	328	2.89	46.84
IVC	5601	49.42	96.26
Other	333	2.94	99.2
UVC	91	0.80	100
Total	11334	100	

Table 3.1: Number of Investors per type

Type	*Frequency*	*Percentage*	*Cumulative*
BA	5050	6.56	6.56
BVC	4815	6.26	12.82
CVC	7357	9.56	22.38
GVC	6775	8.80	31.18
IVC	49912	64.85	96.03
Other	2062	2.68	98.71
UVC	996	1.29	100
Total	76967	100	

Table 3.2: Number of investments per type

% of Equity invested per type of investor

BA 6%
BVC 6%
CVC 9%
GVC 8%
IVC 68%

Legend: BA, BVC, CVC, GVC, IVC, Other, UVC
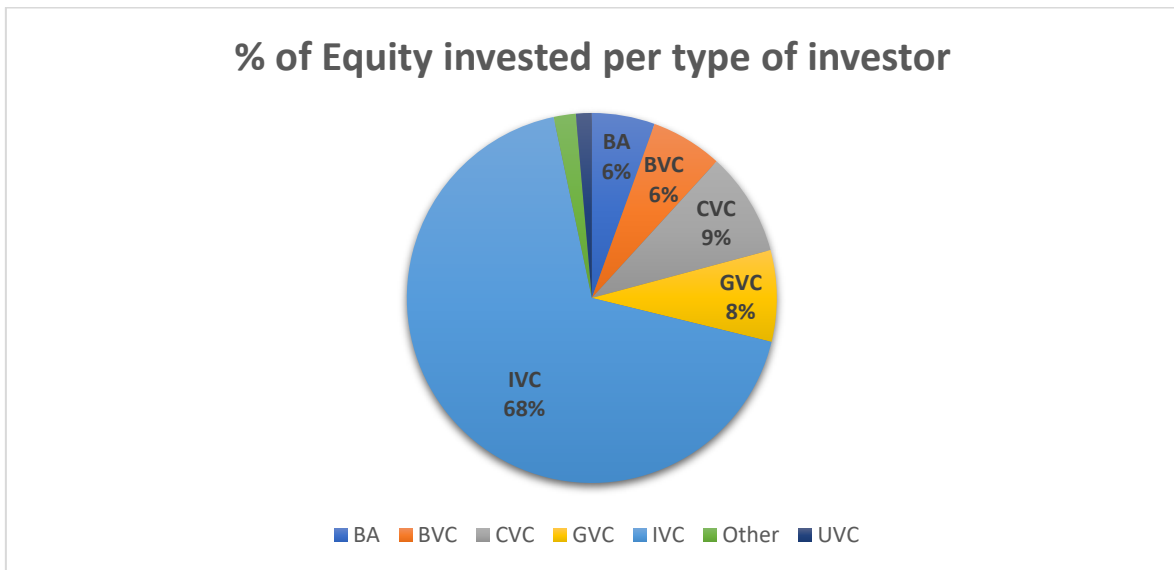
Figure 3.1: Percentage of equity invested per type of investor

Another information regarding the investors that can be shown is about the country of origin. This information is important to have a first idea of the European distribution of the investors that might provide some hints to identify possible spatial patterns on the relocation phenomenon, considering the presence of the local bias of VCs, (Cumming and Dai, 2010). Also in this case, VICO 5.0 is highly reliable, providing information regarding 10619 out of 12015 firms. In the fifth column of the table 3.3, are shown the number of Investments that can be associated to a specific country considering the investors' HQs. Actually, investments can be done from the investors' home countries (international direct investments) or through branches located in the same country of the firm that receives the finance. Unfortunately, VICO 5.0 does not provide specific information about this matter, therefore it is considered where is the investors' HQ in all the cases. In order to be both synthetic and comprehensive it was set a threshold of at least 30 investors in one country as a condition to be considered by its own, otherwise the values of the countries having less than 30 investors are included in the row dedicated to "Others". From the data emerges that investors are not homogenously distributed, indeed considering just 4 countries (U.S.A., U.K., France & Germany), they contain the 52.87% of the overall number of investors and provide the 59.93% of the total number of investments.

Country	Number of Investors	Percentage	Cumulative	Number of Investments
Australia	67	0.63	0.63	119
Austria	116	1.02	1.65	727
Belgium	248	2.34	3.99	1842
Canada	97	0.91	4.9	216
China	103	0.97	5.87	170
Czech Republic	56	0.53	6.4	245
Denmark	155	1.46	7.86	1618
Estonia	50	0.47	8.33	278
Finland	267	2.51	10.84	2152
France	1059	9.97	20.81	12545
Germany	932	8.78	29.59	8391
Greece	31	0.29	29.88	173
Hong Kong	62	0.58	30.46	203
Hungary	62	0.58	31.04	552
India	44	0.41	31.45	61
Ireland	115	1.08	32.52	1543
Israel	402	3.77	36.29	3145
Italy	369	3.47	39.76	1577

Japan	61	0.57	40.33	288
Luxembourg	89	0.84	41.17	374
Netherlands	352	3.31	44.48	2721
Norway	59	0.59	45.04	395
Poland	142	1.34	46.38	875
Portugal	61	0.57	46.95	673
Russia	103	0.97	47.92	374
Singapore	68	0.64	48.56	242
South Korea	30	0.28	48.84	73
Spain	667	6.28	55.12	3279
Sweden	551	5.19	60.31	3505
Switzerland	169	1.59	61.9	994
United Kingdom	1723	16.23	78.13	15286
United States of America	1900	17.89	96.03	9439
Others	422	3.97	100	2106
Total investors	10619	100		76181

Table 3.3: Number of Investors per Country

3.1.1. Descriptive statistics of the sample provided by VICO 5.0

Once presented the main characteristics of VICO and some numbers, underlining why it is suited for the purpose of this thesis, it is the time to describe the sample used to perform the analyses. First of all, it is important to specify that this thesis will be focused on few European Countries because of the limited availability of the data. More in detail, the countries are: Czech Republic, Denmark, France, Germany, Netherlands, and Portugal. The next table (table 3.4) shows the distribution of the companies (considering their observations) across the different countries considered.

Country	*Frequency*	*Percentage*	*Cumulative*
Czech Republic	275	0.56	0.56
Denmark	1525	3.12	3.69
France	15879	32.52	36.20
Germany	26186	53.63	89.8
Netherlands	4505	9.23	99.06
Portugal	461	0.94	100
Total	48831	100	

Table 3.4: Number of companies per Country

Beside the spatial information regarding firms is also important provide information regarding the temporal aspects of companies, that consists in the year of foundation. The literature, also in this case debates a lot without defining a definitive theory, because is not clear if being in a place for a while increases the tendency to move away, (Taylor, 1975), or it leads to an embedded condition, (Stam, 2007)

Foundation Year	Frequency	Percentage	Cumulative
y<1990	637	1.30	1.30
1990≤y<1995	2788	5.71	7.01
1995≤y<2000	7096	14.54	21.55
2000≤y<2005	10452	21.4	42.95
2005≤y<2010	14708	30.12	73.07
2010≤y<2015	11963	24.5	97.57
2015≤y≤2022	1187	2.43	100
Total	48831		

Table 3.5: Distribution of the foundation years

Variable	Observation	Mean	Std. deviation
Foundation Year	48831	2004.83	6.29

Table 3.6: Descriptive statistics of the Foundation years variable

The sample do not consider, on average, very young firms and this is reasonable due to the fact that it is aimed at providing empirical observations on relocations events, based on which identifying determinants of this phenomenon.

Through the database is also possible identifying at which age the relocations take place. To do so, it is used the command "tab relocationdummy" and to eliminate the overlapping data the command is restricted to a specific year and then repeated considering the following year. The relocationdummy is a dummy variable, meaning that it is a discrete variable that can assume only two values: 0 or 1, in the first case the event considered (relocation in this case) does not occur, while in the second case it has occurred.

Age	0	1	2	3	4	5	6	7	8
Relocationdumm y= 0	48831	47411	47409	47112	46619	45949	45092	43435	41046
Relocationdumm y= 1	0	1322	1034	927	859	800	725	658	603
% of relocated firms	0	2.71	2.13	1.93	1.81	1.71	1.58	1.49	1.45

Age	9	10	11	12	13	14	15	16	17
Relocationdumm y= 0	38274	35209	32022	29252	26095	23429	20898	18469	16559
Relocationdumm y= 1	578	449	424	341	304	257	214	200	183
% of relocated firms	1.49	1.26	1.31	1.15	1.15	1.09	1.01	1.07	1.09

Age	18	19	20	21	22	23	24	25	26
Relocationdumm y= 0	14880	13369	11545	9424	7126	5504	4505	3667	2973
Relocationdumm y= 1	143	149	120	102	81	49	50	36	29
% of relocated firms	0.95	1.10	1.03	1.07	1.12	0.88	1.10	0.97	0.97

Age	27	28	29	30	31	32	33	34	35

Relocationdummy= 0	2350	1847	1407	995	562	257	57	22	2
Relocationdummy= 1	23	17	15	7	4	2	0	0	0
% of relocated firms	0.97	0.91	1.05	0.70	0.71	0.77	0	0	0

Table 3.7: Age at relocation

Variable	Observation	Mean (Age)	Std. deviation (Age)	Min. Age	Max. Age
Age of relocation	10705	7.52	5.86	1	32

Table 3.8: Descriptive statistics of Age of relocation variable

From these last tables (table 3.7 & 3.8) one data stands out that companies relocate when they are young, indeed both in absolute terms and in percentage (with some exceptions interpretable as fluctuations) the numbers of relocation decrease the elder are the firms considered. The average year at which relocation takes place is 7.52 but there is a significant std. deviation.

Going further with the analysis of the data regarding relocation, it is also worth of checking if this phenomenon has somehow a trend along the period considered, indeed relocation might be easier in the recent years rather than decades ago due to, for instance, better transportation or a new business mentality. The period before 2000 is not considered because of the size of the sample analysed is much smaller than the one considered in the following years and so it would be useless to compare them. Moreover, in the 2021 and 2022 the number of observations is significantly lower than the ones of the previous years, and this diminishing is not justified by the amount of struck off occurred in that period, therefore in this table there are data up to 2020. The variables in the table 3.9 are:

Year, that indicates the year at which the values are referred.

Relocation, it considers the number of events that take place on the reference year. In order to be considered as relocation, the company has to change its NUTS3.

Active company, that is computed using data shown in table 4.9 (see the description below).

% of relocated firms, it assesses the likelihood of relocation considering that the sample of firms has grown up to 2014 and then diminished but in a less extent.

Mean (Age), provides information regarding age of the companies that have relocated in that year.

Std. deviation (Age), other information regarding the companies relocated in the year

Year	Active company	Relocation	% of relocated firms	Mean (Age)	Std. deviation (Age)
2000	13433	110	0.82	3.39	2.71
2001	15534	128	0.82	2.95	2.88
2002	16991	219	1.29	3.50	2.94
2003	18674	205	1.10	4.11	3.10
2004	20540	253	1.23	4.40	3.32
2005	22647	268	1.18	4.52	3.49
2006	25199	342	1.35	4.62	3.61
2007	28147	366	1.30	5.21	4.27
2008	31235	444	1.42	5.60	4.46
2009	33911	570	1.68	5.14	4.39
2010	36755	646	1.76	5.70	4.61
2011	38829	716	1.84	5.55	4.61
2012	40657	715	1.75	6.11	4.92
2013	41878	710	1.70	7.08	5.44

2014	42169	682	1.62	6.99	5.30
2015	41626	829	1.99	8.88	5.84
2016	40785	675	1.65	9.40	5.82
2017	39703	652	1.64	9.97	6.13
2018	38440	678	1.76	11.12	6.10
2019	36864	603	1.64	12.20	6.18
2020	35410	459	1.30	13.63	6.20

Table 3.9: Active Companies per Year

The first data that emerge from the table is the increasing of the percentage of relocation across the period (except for the 2020), that arguably might be seen as an indicator of lower difficulties in moving from a place to another. In this case the last two columns must be read together because in that way the relocations of the new-born companies are considered. Indeed, in the column "Relocation" are taken into account all the movements and therefore both the first relocations and the following relocations. The mean age is growing due to the fact that the sample is getting older and still some companies continue to relocate, especially in the last 5 years when the number of "Foundation" is not so remarkable (see table 3.15), while the Std. deviation grows too since even young companies relocate (see table 3.8). The 25th percentile of the relocation takes place within 3 years from the foundation, the 50th percentile within 6 years from foundation and the 75th percentile within 11 years from the foundation.

Moreover, one of the hypotheses shown in the previous chapter is aimed at identifying possible differences between VC-backed and non-VC-backed companies in terms of tendency to relocate. In order to do so, it was taken into consideration another dummy variable that is called "VC_Backed"; similarly, to the variable it assumed the value of 0 when the company is not backed by VC, while it assumes value 1 when it is or will be. Firms that have relocated more than once have been accounted each time. The command used is "tab relocationdummy VC_Backed".

	VC_Backed=0	VC_Backed=1	Total
Relocationdummy=0	628555	105048	733 603
Relocationdummy=1	8703	2002	10 705
Total	637258	107050	744 308

Table 3.10: Matrix Relocationdummy-VC_Backed

So, in 2002 out of the 107050 observations regarding VC-backed companies, a relocation takes place that represents the 1.87% of the overall number of VC-backed. On the other hand, considering non-VC-backed firms, there are 8703 observations of a relocation out of 637258 total observation that means the 1.36%. The variation in percentage is not so relevant since the probability of observing a relocation for a VC-backed firms is slightly higher than the one of non-VC-backed companies.

In addition, relocation events are clustered considered the different countries in order to identify some potential differences between them. The results, shown below (table 3.11), do not underline particular differences in terms of relocation considering the percentage value, despite the diverse contexts regarding the availability of investors. This aspect has been deepened more, by performing a further analysis considering only the relocation referred to VC backed companies. Unfortunately, data do not show a particular trend, but the results (table 3.12) are very different among countries without a clear reason

	Czech Republic	Denmark	France	Germany	Netherlands	Portugal
Relocationdummy=0	3684	18687	231259	424508	50045	5420
Relocationdummy=1	66	121	3835	5453	1119	111
% of relocation	1.76	0.64	1.63	1.27	2.19	2

Table 3.11: Relocation per Country

(relocationdummy =1)	Czech Republic	Denmark	France	Germany	Netherlands	Portugal
VC_Backed=0	44	101	2612	5025	868	53
VC_Backed=1	22	20	1223	428	251	58
% VC-Backed	33.3	16.53	31.89	7.88	22.43	52.25

Table 3.12: VC-backed companies per Country

To conclude the descriptive analysis of the relocation phenomenon, it is analysed if companies tend to relocate again after the first relocation

	1	2	3	4	5	6	7
VC_Backed=0	4984	940	153	26	4	2	1
VC_Backed=1	1217	344	33	4	3	0	0

Table 3.13: Number of relocations per company

The large majority of firms relocate just one, indeed 6201 have relocated only once out of 7711 that relocated at least once, and the number of firms that have change their site more than 3 times is negligible. Once shown how the probability of relocation varies considering different ages of the firms, it might be interesting performing a similar analysis considering when the struck off occurs.

Age	0	1	2	3	4	5	6	7	8
Struckoffdummy= 0	48742	48460	48053	47496	46801	45958	44872	43097	40624
Struckoffdummy= 1	89	273	390	543	677	791	945	996	1025
% of closed firms	0.18	0.56	0.81	1.13	1.43	1.69	2.06	2.26	2.46

Age	9	10	11	12	13	14	15	16	17
Struckoffdummy= 0	37828	34770	31640	28816	25749	23054	20641	18199	16343
Struckoffdummy= 1	1024	888	806	777	650	632	471	470	399
% of closed firms	2.64	2.49	2.48	2.63	2.46	2.67	2.23	2.52	2.38

Age	18	19	20	21	22	23	24	25	26
Struckoffdummy= 0	14658	13169	11326	9291	7030	5403	4432	3602	2925
Struckoffdummy= 1	365	349	339	235	177	150	123	101	77
% of closed firms	2.43	2.58	2.91	2.47	2.46	2.70	2.70	2.73	2.56

Age	27	28	29	30	31	32	33	34	35
Struckoffdummy= 0	2303	1812	1385	958	548	251	55	22	1

Struckoffdummy= 1	66	52	37	44	18	8	2	0	1
% of closed firms	2.78	2.79	2.60	4.39	3.18	3.09	3.51	0	50

Table 3.14: Age at the struckoff

The values on the table (3.14) go in contrast with the literature that assesses that the very first years for a new-born venture is the deadliest. Without considering the years beyond 30, that are data not so reliable since the sample is quite small, there is a constant, more or less, percentage of firms that face struck off since the "Year 7", (around 2.50% each year). This trend, arguably, is caused by the methodology used to build the VICO's sample. Indeed, VICO is a database mainly focused on VC-backed companies and the control sample is composed by similar firms. Since receiving external funds can be seen as a proxy of value, it is possible to affirm that the firms considered are, on average, reliable and effective companies and therefore the likelihood of surviving is higher.

It is also interesting discovering the average age of the companies and the standard deviation computed year by year along the period considered by the database. This information is important to check if the sample considered achieved somehow an equilibrium or it presents very volatile data. The period considered is between 2000 and 2020. The variable indicated in the following table are:

Year, that indicates the year at which the values are referred.

Foundation, in this column are accounted the numbers of companies founded in the reference Year

Struck off, similarly to the previous one, it considers the number of firms that stopped their activities.

Active company, it provides information regarding the size of the sample, and it is computed considering the value of the active company of previous year adding the value of Foundation and subtracting the value of Struck off of the reference year.

Mean (Age), it provides additional information about the companies considered in the entire sample.

Std. deviation (Age), as the previous variable and it adds information regarding the stableness of the sample.

Relocation, it considers the number of events that take place on the reference year.

Fortunately, the sample, after 2012, has stable data where the average year increases slightly less than 1 (due to the struck off of some old companies) each year and the std. deviation remains quite stable. The size is constantly growing up to the 2014 and then it starts to decrease. Arguably, this is caused by the lack of data regarding the variable Foundation because of a sort of temporal lag in collecting data of the firms, indeed is not realistic assume as true the diminishing of the value of new-born firms and in such extent.

Year	Foundation	Struck off	Active company	Mean (Age)	Std. deviation (Age)
2000	2924	4	13433	3.38	3.43
2001	2110	9	15534	3.79	3.53
2002	1563	106	16991	4.35	3.64
2003	1833	150	18674	4.83	3.80
2004	2022	156	20540	5.27	4.01
2005	2255	148	22647	5.65	4.25
2006	2719	167	25199	5.94	4.52
2007	3155	207	28147	6.17	4.79
2008	3420	332	31235	6.40	5.05
2009	3159	483	33911	6.72	5.27
2010	3450	606	36755	7.00	5.50
2011	2813	739	38829	7.43	5.69
2012	2563	735	40657	7.88	5.88
2013	1998	777	41878	8.44	6.04
2014	1139	848	42169	9.18	6.16

2015	547	1090	41626	10.04	6.24
2016	372	1213	40785	10.93	6.31
2017	153	1235	39703	11.88	6.36
2018	55	1318	38440	12.86	6.40
2019	20	1596	36864	13.85	6.42
2020	16	1470	35410	14.84	6.84

Table 3.15: Descriptive statistics of the sample

To conclude, it is considered the NACE Rev2 that is the "*statistical classification of economic activities in the European Community*" and its implementation is imposed by the European Union to all its Member States. The exploitation of NACE nomenclature implies a higher degree of comparability because it is a part of a *"integrated system of statistical classifications"*. The main characteristics of NACE are the presence of mutually exclusive categories, meaning that a company can belong just to one category, and the hierarchical classification. Indeed, to each company can be associated a NACE-2-digits and a NACE-4-digits. The latter represents a finer classification where the categories of the former classification are subdivided into subcategories, and in that way, it is possible to get the information at various levels of aggregation. The unit of observation in the NACE classification is any economic activity that takes place when resources (any kind of input) are combined, through a process, to realize products or services (output).

Talking about the data shown in the following table (table 3.16), it emerges that the 29.16% of the firms do not provide information regarding the type of their business and are account in the row "missing". However, the percentage of firms that provided information is assessed as satisfactorily. It is important to recall that the percentages obtained by each section do not necessarily match the real distribution because the sample of firms considered by VICO is composed by innovative firms that are VC-backed or non-VC-backed but still innovative in order to be used as control sample. Probably, this is the explanation for the fact that the 43.9% of firms belong to "Information and communication" section, and the same reasoning can be applied to "Professional, scientific, and technical activities" section as well as "Wholesale and retail trade; repair of motor vehicles and motorcycles" section.

Section	Title	Div.	Freq.	%
A	Agriculture, forestry and fishing	01-03	14	-
B	Mining and quarrying	05-09	3	-
C	Manufacturing	10-33	3740	10.8
D	Electricity, gas, steam and air conditioning supply	35	140	0.40
E	Water supply; sewerage, waste management and remediation activities	36-39	65	0.19
F	Construction	41-43	160	0.46
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	45-47	4658	13.4
H	Transportation and storage	49-53	171	0.49
I	Accommodation and food service activities	55-56	113	0.33
J	Information and communication	58-63	15186	43.9
K	Financial and insurance activities	64-66	1746	5.05
L	Real estate activities	68	114	0.33
M	Professional, scientific, and technical activities	69-75	6419	18.6

N	Administrative and support service activities	77-82	1082	3.13
O	Public administration and defence; compulsory social security	84	0	-
P	Education	85	98	0.28
Q	Human health and social work activities	86-88	161	0.47
R	Arts, entertainment, and recreation	90-93	130	0.39
S	Other service activities	94-96	591	1.71
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	97-98	0	0
Total			34591	
Missing			14240	

Table 3.16: Distribution of the companies per NACE SECTION

It is also possible combining the information presented in the previous tables to show additional information regarding the companies of the sample. In the following table (table 3.17), are analysed the relations between the section, defined using NACE classification, and the observations about relocations and VCs. The percentage of relocated firms that provides information regarding their NACE is 74.9% while considering the VC-backed firms, there are information in the 89.5% of the cases. The fact that the latter provide more information, on percentage, is reasonable because VCs require a significant amount of information before the potential investment and then want to be updated with recent and comprehensive data. There are some interesting data that come out form the comparisons of the second and third column: in the section C (manufacturing) the number of VC-backed companies is the 39.8% of the entire companies while the number of

relocations is less than the half. Moreover, the number of relocated firms, probably, is lower than the total number of relocations because some firms might have relocated more than once. So the two phenomena present quite different trend in this specific field. Similar reasoning can be applied to section M (Professional, scientific, and technical activities). On the other hand, only two sections show a higher number of relocations than VCs-backed companies, and they are Section J (Information and Communication) and Section S (other services activities), so basically two segments of the market concerning services towards customers.

Section	A	B	C	D	E	F	G	H
N° of relocation	1	1	704	41	12	17	954	27
Companies VC-backed	14	4	1849	62	29	80	998	82
Section	I	J	K	L	M	N	O	P
N° of relocation	11	3858	406	33	1476	285	0	28
Companies VC-backed	50	3503	487	71	2222	348	0	44
Section	Q	R	S	T		Total		Missing
N° of relocation	36	19	110	0		8019		2686
Companies VC-backed	69	43	94	0		9645		1129

Table 3.17: Number of relocation & VC-backed companies per NACE SECTION

3.2. ETER

In order to have information about the universities ETER database was used, that is aimed at clarifying the situation of the European higher education. ETER was chosen because nowadays is the most comprehensive and freely usable register in Europe about high education institutions. ETER (European tertiary education register) is founded by EU and collects quantitative and qualitative information about high education institutions (HEIs) in Europe that entails basic characteristics such as the students enrolled in a specific department, (https://www.eqar.eu/qa-results/synergies/european-tertiary-education-register-eter/). The fundamental point of this database is that provides information on HEI's activities and output both at country level and regional level provided by EUROSTAT, apart from descriptors and geographical information that have been collected mostly from National Statistical Authorities, thus ETER represents a useful tool considering the aim of this thesis.

In that register are considered those institutions that deliver degrees at ISCED level 5 (short-cycle tertiary education), levels 6 (bachelor), 7 (master) and 8 (PhD), providing information regarding 3439 HEIs, while as of Spring 2022 the countries taken into account are 41, considering more than 25 million of undergraduate students, more than 0.75million of PhD students, more than 1.62 million of academic personnel. That database is tailored to support analyses about the connections between HEIs and the firms' productivity and business model, being complementary to EUROSTAT. In other words, the database provides comparable data on European HEIs in order to compare them considering different variables.

ETER Coverage 2017-2019

Data Published ■ Data Delivered (Publication in Preparation) ■ Data Not Yet Delivered
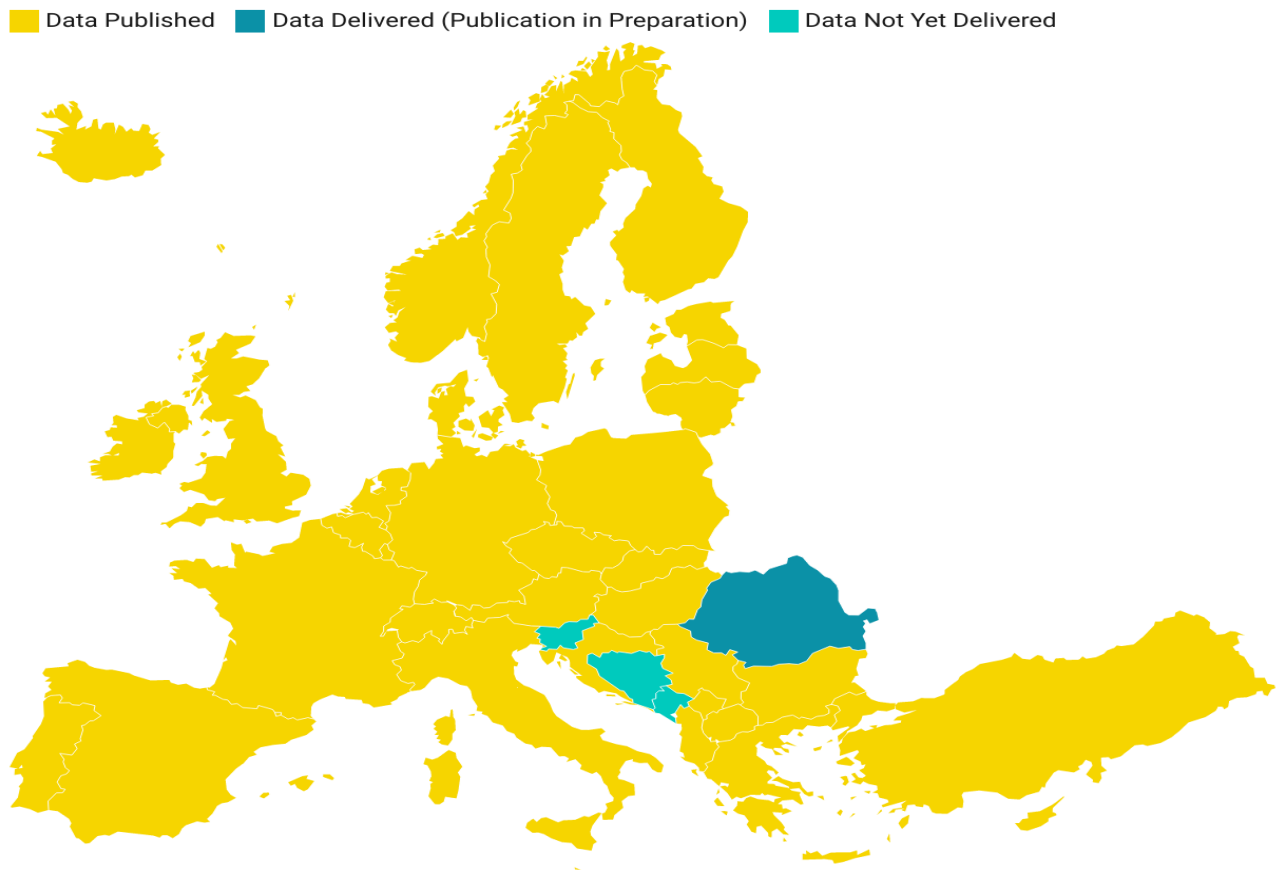
Map: ETER Coverage 2017-2019 • Source: www.eter-project.eu • Created with Datawrapper

Figure 3.2: ETER coverage.

The presence of R&D activities is not a mandatory condition to be in the resister so not all the universities provide this information. Moreover, that database has a threshold below which the universities are not taken into account that is the presence in each institution of at least 30 FTEs (Full Time Equivalent) of staff or at least 200 enrolled students in order to allow a better comparability among universities in the register. Exceptions might apply for institutions of national importance. Nevertheless, this limitation does not impact on the value of the database because it considers more than 85% students enrolled for a course of ISCED level 6, level 7, or level 8. Moreover, considering the goal of that thesis, information regarding the students enrolled or graduated with an ISCED level 5 course were not taken into account because is not directly related to universities' output. As of 2017, ETER contains information about 90% of students and graduate data, and only slightly lower for staff data. Conversely, financial data are available only for about half of the HEIs in the database.

In the document provided by the ETER's website information is clustered in seven "dimensions":

a) Descriptors and regulatory characteristics, like the foundation year, legal status, institutional type.

b) Geographical information, including the city, postcode, geographical coordinates of the main campus, the presence of branch campuses in other cities. This information is added in the database to find a correlation between the presence of universities and the regional or city-level of economy. For this thesis this data are highly valuable because allow to assess whether companies moving towards universities or not.

c) Students and graduates divided by level of study (diploma, bachelor, master), field of education, gender, nationality, and mobility. This information fully characterizes the educational profile of the HEI and is very important to find possible localization economies and to assess the presence of high-skilled workers.

d) Staff, divided between academic and non-academic, divided by gender, nationality and field of education are provided, as well as the number of full professors.

e) HEI expenditures, divided between personnel, non-personnel and capital, and revenues, divided by stream (core funding, third-party, tuition fees) and source (public, private, international).

f) Research and transfer activities, including the number of PhD students and graduates and R&D expenditures. Regarding data on research, they are less complete in ETER.

g) Some measure about the gender balance, international mobility etc.

The reliability of ETER is high for two main reasons: first because many data come from EUROSTAT or other well-known databases that are highly accurate, second because of the process of collecting data that is controlled deeply and standardized in order to reduce the probability of errors, indeed are used predefined excel templates. Collected data are then validated, for instance performing a statistical check on outlying data; in case there are some suspects cases data are matched with values of other databases such NSAs and adjusted. The time lag between the ETER data collection and the Source release date is reduced as much as possible.

(https://www.eter-project.com/wp-content/uploads/2022/02/Final_report_ETERII_EN.pdf)

Talking about the descriptive characteristics of ETER, here it will be shown tables providing information about the graduated students of the last 3 years in the 41 countries considered. However, the following econometric analysis will focus just on some of these countries but in this paragraph, data referring to all the 41 countries (in the most cases) in order to clarify better the as is educational situation in Europe.

Considering the first table presented here (table 3.18), it is shown the total number of graduated students divided by ISCED level and year. Moreover, it is presented the contribute provided by women and men regarding the numbers of graduation also in this case divided by year and ISCED level. It is clear, by summing the number of men and women that not all the universities provide detailed information regarding their students, indeed total graduates ISCED is always overcoming that sum. However, the percentage of missing data remains stable across the period and the level of education.

Looking more in detailed at the table, it seems to show a kind of trend where women are more than men considering both Bachelor and Master degree, while the relation is reversed considering the PhD. The data is confirmed in each year belonging to the period considered and it is somehow quite strange because there is no reason for thinking that women have a lower inclination in applying to PhD programs. However, it was decided to consider the graduated students rather than enrolled students because the former indicate in a better way the availability of high-skilled workers, indeed the latter takes into account also those students that will never complete their education process.

ISCED level	2017	2018	2019
ISCED 6 - men	882274	891125	1026349
ISCED 6 - women	1166393	1185499	1340224
Total Graduates ISCED 6	2474642	2507609	2788022
ISCED 7 - men	432773	443850	526954
ISCED 7 - women	560384	583819	705429
Total Graduates ISCED 7	1255040	1293411	1491293
ISCED 8 - men	47560	47310	51713

ISCED 8 - women	43519	43474	46935
Total Graduates ISCED 8	125289	124955	132052
Total Graduates	3854971	3925975	4411367

Table 3.18: Gender distribution per ISCED level

Through the database provided by ETER it was possible also find out more specific data, in particular regarding the areas where students get graduated. Indeed, ETER identifies 9 different categories of study:

a) Education

b) Arts and Humanities

c) Business, administration and low

d) Natural science, mathematics & statistics

e) Information & communications technologies

f) Engineering, manufacturing & construction

g) Agriculture, forestry, fisheries & veterinary

h) Health & welfare

i) Services.

There is also another category that is "generic programme & qualification" but it is not taken into account because it does not provide any kind of information about the preparation of the graduate. The next three tables show data regarding graduated students considering the period 2017-2019. The choice of the period derives form ETER, because that information has been collected and added to the database at the same time.

In the table 3.19 is presented the number of graduates students to Bachelor. It is quite clear that information about the attendance at the different courses is not provided by all the universities. Indeed, there are countries where information is

regarding graduated students are missing such Turkey, Serbia, Montenegro. In general, countries belonging to the western part of Europe provide much more data than the eastern part with the only exception of United Kingdom. This is probably caused by the "Brexit" because this project is financed by UE, and ETER provide data of English universities up to 2016. However, for the purpose of this thesis this

does not represent a significant issue because the countries that will be investigated are some of those that provide information. Considering only ISCED – 6, there is information regarding the 63.07% of students in 2017, 61.81% in 2018 and 63.14% in 2019, that is a percentage enough satisfying especially because is not homogenous across the different countries, as said before. For the ISCED – 7 and ISCED – 8 the percentage is even higher.

Looking at the data, the stream business, administration & law provides the major contribution to the overall number of graduates each year of the period for the bachelor. Its contribution in 2017 for ISCED – 6 graduated students was 21.38%, and this value grows in the following two years reaching the of 23.09% in 2019. The other main contributions arrived from the section "Engineering, manufacturing and construction" and "Health and welfare". The trends for these two areas are very similar because the number of graduates students remains more or less the same across the years with a small increasing between 2018 and 2019. There is not a clear explanation why the values of the 2017 are higher (in general) compared to those values of 2018. However, considering the entire period the overall number of graduates increases, and more in details numbers increase in each area apart from "Agriculture, forestry, fisheries and veterinary". On the other hand, the area that contributes to a lesser extent is "Agriculture, forestry, fisheries and veterinary" moving from 1.90% to 1.60% of the overall number of graduates at ISCED – 6 level. It is important to say that all the percentages are computed considering the number of students graduated in one specific area divided by the overall of students for whom is known the area of graduating. This is due to fact that it is impossible to assess the specialization of the students at those universities that do not provide data. The results are in line with the expectations of a constant growth of the numbers of graduates.

Area of graduating ISCED – 6	2017	2018	2019
Education	151192	146872	153296
Art & Humanities	178486	176051	217031
Business, administration & law	333767	338250	406469
Natural sciences, mathematics and statistics	92188	91347	121202
Information and Communication Technologies	65138	64462	75614

Engineering, manufacturing and construction	241319	238532	252968
Agriculture, forestry, fisheries and veterinary	29715	28106	28145
Health and welfare	212409	211669	220482
Services	81010	83025	85975
	1385224	1378314	1561182

Table 3.19: Areas of Study of the ISCED level 6

The considerations explained presenting the table 3.19 hold also for table 4.20, indeed countries such as Turkey, Montenegro and United Kingdom do not provide specific information regarding the areas of graduation. For the same reason, this does not affect the validity of the data with respect to this thesis. Also in that case, the area of "Business, administration & law" is the most important one in all the years considered, moving from the 24.61% of the 2017 to the 23.75% of the 2019. It is worth of mentioning saying that in ISCED – 7, there is a larger heterogeneity among the different areas meaning that "Services" plus "Agriculture, forestry, fisheries and veterinary", account only for the 4.33% of the overall number in 2019 while considering ISCED – 6 they count for 6.48%. Moreover, there is a significant drop in the value of "Health and welfare", and it is reasonable because of a selection test in several countries such as Italy. During the period considered, does not emerge any suspect value, indeed the growth of number between 2018 and 2019 is homogenous for all the sectors thus it is probably referred to the fact that some institutions start to provide data to the database in that year.

Area of graduating ISCED – 7	2017	2018	2019
Education	101948	100976	131994
Arts & Humanities	83720	84924	100013
Business, administration & law	217421	219293	267796
Natural sciences, mathematics and statistics	67177	65427	89132
Information and Communication Technologies	30292	31352	41977

Engineering, manufacturing and construction	153672	154338	179434
Agriculture, forestry, fisheries and veterinary	14572	15417	15465
Health and welfare	73669	82052	128709
Services	31256	32036	33381
	773727	785815	987901

Table 3.20: Ares of Study of the ISCED level 7

The table 3.21 shows the situation regarding the PhD students. Numbers are much smaller than the ones shown in the previous tables, and this is very reasonable. Indeed, after getting master degree the large majority of students start their work career rather than undertake PhD programs. Of course, this tendency is not the same for all the areas, indeed is evident considering "Natural science, mathematics and statistics". In 2019, the PhD graduated students in this field are 1 out of 4 of the master degree graduated students, while considering the areas the ratio is slightly larger than 1/20. This could have important consequences, because D'este and Iammarino (D'Este and Iammarino, 2010) in their results underlined that science-based departments are less effected by geographical proximity bias, meaning that they tend to start collaboration even when the distance with the company is larger. The heterogeneity among the areas is much more emphasized compared to ISCED – 7 situation, indeed "Natural sciences, mathematics and statistics" plus "Engineering, manufacturing and construction" and "Health and welfare" account for the 62.4% of the overall numbers of PhD graduated students. The PhD students represent the highest-level of high-skilled employees and in some fields (high-tech) this presence might significantly affect the relocation of firms.

Area of graduating ISCED - 8	2017	2018	2019
Education	2521	2388	2625
Arts & Humanities	8454	8189	9555
Business, administration & law	6771	6239	6731
Natural sciences, mathematics and statistics	19361	19416	22381

Information and Communication Technologies	2584	2301	2853
Engineering, manufacturing and construction	12549	12815	13530
Agriculture, forestry, fisheries and veterinary	3025	2586	2596
Health and welfare	15765	16394	17970
Services	817	892	737
	71847	71220	78978

Table 3.21: Areas of Study of the ISCED level 8

Once having shown the difference in terms of areas of studies, it is useful to present the contribution of the western Europe countries, among which will be chosen the subjects of the following countries in order to show the utility of this database for this thesis. The countries considered are: Germany; France; Italy; Portugal; Spain; The Netherland; Belgium; Denmark; Czech Republic. Unfortunately, not all the countries provide data for the entire period considered, for instance Czech Republic is a particular case because that country provides information up to the 2018, whilst for the others the last year in which there is the availability of information is the 2019.

Country	Total graduates at ISCED 6	Total graduates at ISCED 7	Total graduates at ISCED 8
Germany	258616	144367	27781
France	216316	215308	11134
Italy	233981	132854	8121
Spain	192629	131267	9349
Portugal	52470	18118	1941
Netherlands	107232	51864	4984

Belgium	70274	42790	2865
Denmark	40407	23823	2082
Czech Republic	35167	27437	2419
Total	1207092	787828	70676

Table 3.22: Total graduates per Country

There is a pattern that holds for all the countries taken into account, except for France, where the ISCED – 6 graduated students are much more than ISCED – 7 and ISCED – 8 graduated students, and this is highly reasonable. The French case is, arguably, explained by its particular educational system. An important information that emerges is that these countries provide a larger contribution to the overall number as higher educational levels are considered. Indeed, they account for the 43.16% of the ISCED – 6 graduated students, 52.44% of the ISCED – 7 graduated students and for the 53.34% the ISCED – 8 graduated students. This could have important consequences on international relocations, because, arguably, high-tech and high-skilled companies may be more inclined to move to one of those countries, especially in case of a particular specialization of the country itself (for instance Germany for automotive).

The next table (table 4.22)is introduced just as an additional proof of the validity of the database. Indeed, considering the German case (2019), there is specific information for basically all the students graduated in each ISCED level, especially if it is considered that the area "generic programme & qualification" is not accounted. By far there are two areas that emerge from the group that are "Business, administration and law" and "Engineering, manufacturing and construction" that together account for 48.89% for ISCED – 6, 50.55% for ISCED – 7, while for ISCED – 8 "Natural sciences, mathematics and statistics" and "Health & welfare" account for 55.17%.

Area of study	ISCED 6	ISCED 7	ISCED 8
Education	14819	10106	898
Arts and Humanities	23624	12124	2052

Business, administration and law	73897	35511	2375
Natural sciences, mathematics and statistics	18274	14894	7710
Information and Communication Technologies	17494	9888	1017
Engineering, manufacturing and construction	52516	37452	3661
Agriculture, forestry, fisheries and veterinary	3468	2339	798
Health and welfare	18929	3955	7613
Services	12002	3054	190
Total	235023	129323	26314

Table 3.23: Areas of Study

Lastly, in order to distinguish between high-tech industries and low-tech industries was used EUROSTAT, that is the official database of the European Union. Indeed, it provides schedules where each category of NACE Rev.2, considering the manufacturing industries, is clustered in four groups as follow:

Manufacturing Industries	NACE Rev. 2 codes – 2-digit level	
High-technology	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations;
	26	Manufacture of computer, electronic and optical products
Medium-high-technology	20	Manufacture of chemicals and chemical products;
	27 to 30	Manufacture of electrical equipment; Manufacture of machinery and equipment n.e.c. ; Manufacture of motor vehicles, trailers and semi-trailers; Manufacture of other transport equipment
Medium-low-technology	19	Manufacture of coke and refined petroleum products;
	22 to 25	Manufacture of rubber and plastic products; Manufacture of other non-metallic mineral products; Manufacture of basic metals; Manufacture of fabricated metals products, excepts machinery and equipment;
	33	Repair and installation of machinery and equipment
Low technology	10 to 18	Manufacture of food products, beverages, tobacco products, textile, wearing apparel, leather and related products, wood and of products of wood, paper and paper products, printing and reproduction of recorded media;
	31 to 32	Manufacture of furniture; Other manufacturing

Figure 3.3: NACE REV.2 codes for manufacturing industries

While for the services the classification is done as follow:

Knowledge based services	NACE Rev. 2 codes – 2-digit level	
Knowledge-intensive services (KIS)	50 to 51	Water transport; Air transport;
	58 to 63	Publishing activities; Motion picture, video and television programme production, sound recording and music publish activities; Programming and broadcasting activities; Telecommunications; computer programming, consultancy and related activities; Information service activities (section J);
	64 to 66	Financial and insurance activities (section K);
	69 to 75	Legal and accounting activities; Activities of head offices, management consultancy activities; Architectural and engineering activities, technical testing and analysis; Scientific research and development; Advertising and market research; Other professional, scientific and technical activities; Veterinary activities (section M);
	78	Employment activities;
	80	Security and investigation activities;
	84 to 93	Public administration and defence, compulsory social security (section O); Education (section P), Human health and social work activities (section Q); Arts, entertainment and recreation (section R).

Figure 3.4: NACE REV2 codes for knowledge services

Knowledge based services	NACE Rev. 2 codes – 2-digit level	
Less knowledge-intensive services (LKIS)	45 to 47	Wholesale and retail trade; Repair of motor vehicles and motorcycles (section G);
	49	Land transport and transport via pipelines;
	52 to 53	Warehousing and support activities for transportation; Postal and courier activities;
	55 to 56	Accommodation and food service activities (section I);
	68	Real estate activities (section L);
	77	Rental and leasing activities;
	79	Travel agency, tour operator reservation service and related activities;
	81	Services to buildings and landscape activities;
	82	Office administrative, office support and other business support activities;
	94 to 96	Activities of membership organisation; Repair of computers and personal and household goods; Other personal service activities (section S);
	97 to 99	Activities of households as employers of domestic personnel; Undifferentiated goods- and services-producing activities of private households for own use (section T); Activities of extraterritorial organisations and bodies (section U).

Figure 3.5: NACE REV.2 codes for knowledge services

Moreover, EUROSTAT provides useful information about employment grouping the data by NACE Rev2, NUTS 1 & 2. Unfortunately, is not possible getting information about NUTS3, that would be very important because most of the relocations are intra-regional. Conversely, other information used in the following analyses are available at NUTS3 level.

4 Methodology & variables

4.1. Merging the datasets and cleaning-up phase

Once having described the dataset used, it is appropriate showing how the analysis was carried out in order to verify the hypotheses defined in the dedicated chapter (see Chapter 3). First of all, the data coming from both VICO 5.0 and ETER are not ready to be used for an econometric analysis through the exploitation of STATA software. Indeed, initially data are collected in a way that does not allow the unique representation of the whole dataset in one table.

To do so, also data coming from ETER are clustered considering the NUTS3 classification, which is already used in VICO. In other words, it is possible to have information regarding the availability of graduated students (in the six different countries and along the period considered) at NUTS3 level considering the different universities located in the same NUTS. This is important also considering the fact that relocation is, mainly, an intra-regional phenomenon, therefore in that way it is possible to assess the importance, eventually, of the availability of students in the decision-making process related to the migration in a proper way. As introduced in the previous chapter, however it is fair to underline the presence of a non-negligible component of missing data. More in detail, 3 out of 6 countries (Germany, Netherlands, Portugal) provide information regarding their graduated students from 2011 to 2019 that the period considered in the database of ETER. Denmark provides information for the entire period apart from the 2014, while Czech Republic, as said in the previous chapter does not provide data related to the 2019. Unfortunately, French data are not as comprehensive as the ones of the other countries indeed France gives information regarding the 2011, 2012,2013,2014 and 2019.

At that point, ETER's data are not ready yet to be merged in the database of VICO but they need to be collapsed by NUTS3 and year in order to have one data per each year, NUTS and category of study. Afterwards, it is possible merging the data coming from the two different databases; specifically, the type of merge was 1:m, without the creation of new variables meaning that the same data coming from ETER would be associated to many firms in VICO depending on their locations and reference year. At that point, the data regarding the availability of graduated students per NUTS3 are converted to natural logarithm scale in order to have more robust input data for the analysis. After that, the database was cleaning-up, eliminating those observations related to the years before the 2011 and after the 2019

that are not considered in ETER. Then each observation that has a missing data regarding the availability of students was dropped too. This missing data is due to the fact that ETER does not provide information regarding the NUTS where the company is located in that specific year. Indeed, by performing a simple command "tab" in Stata of the NUTS3 it is possible to find out that, especially for the German case, the number of NUTS3 in VICO database is larger than the one in ETER.

At the end of this process, the number of observations based on which the analysis would be performed was 139002 while the initial number of observations was 744308 (considering only the six countries previously mentioned). Each observation presents information regarding the position (NUTS1, NUTS2,NUTS3), its state (VC-backed or not through a dummy variable), the occurrence of the relocation event in the year (even in this case through the exploitation of a dummy variable), where for relocation event is considered every migration that implies a change in the NUTS3 whatever is the reason, the reference Year that define the year at which the other values referred to, the age of the company, the NACE SECTION to which it belongs, the availability of graduates students (the overall values and considering each stream of study), the size of the firms (in terms of employees and total assets), the profitability (in terms of ROA and EBITDA, the value and the growth of the local GDP at NUTS3 level. At least, in order to allow the software to recognized as a panel the database, the command "xtset BvDIDnumber Year" was launched; it uniquely determines one observation, where "BvDIDnumber" is a code used by VICO to identify the firms. The information regarding the panel is important because the observation of independent errors is not possible.

4.2.   The econometric analysis

4.2.1.   The Cox with multiple events

The presence of dummy variables in the database reflects the characteristics of the phenomenon that is discrete. For that reason, also the econometric analysis must consider this peculiar aspect that does not allow to perform a linear regression. Moreover, the probability of relocation varies over the years of the exposition, that is the interval of time during which the firms are analysed. Indeed, in the very next year after a relocation, the probability of occurrence of another relocation is much lower than other year.

This analysis wants to assess the relocation propensity of a firm depending on the availability of graduated students; however, it is important to say that the risk of

being relocated in a different NUTS3 occurs multiple times therefore it is used a particular type of Cox model that is Cox with multiple events. It means that each company in the panel, after a relocation, still is at risk of another relocation, as well as (with a different percentage) a firm that has never relocated itself.

The Cox model is a survival model that underlines the connection between the time that passes and the probability of occurrence of some event that depends on one, or more, covariates that can assume continuous or discrete values. The initial relation relationship between the event and the covariates defines what it is called baseline, that is a failure rate where the failure is represented by the relocation event. More in detail, the baseline in the Cox is called hazard rate that is a continuous function that expresses the failure rate, and it is continuous because the interval of time for which it is computed the ratio approaches to 0. Once defined the baseline, the Cox analysis allows to check how other covariates, called control variables, affect the hazard ratio. They are used in order to clean up the phenomenon and they can be both discrete and continuous and they can affect the probability to relocate. There are two parameters that need to be considered in order to assess this impact: the p-value and the hazard rate. The former can be defined as the error committed in case it is considered that the control variables considered do not affect the initial baseline, so the lower it is and better it is. In other words, it expresses the statistical importance of the value. The latter is a measure of the impact on the relocation propensity of the firms.

The case analysed in this thesis is characterized by a lack of independency among the observations that define the dataset used. Indeed, each firm, listed on the database, is observed as many times as it is its time at risk; for instance a company founded in 2013 and acquired in 2017 will be observed 5 times, and this is why each observation is uniquely identified by a combination of BvDIDnumber Year. This is another reason why the Cox model is chosen because it allows to overcome the problem of collinearity, that arises from the presence of interdependent observations, by dropping one or more of the variables. For seek of clearness, the collinearity, or multicollinearity defines a condition where a variable can be predicted from the others with a substantial degree of accuracy. The collinearity implies that the results obtained through the computer may be inaccurate.

The Cox multiple events model is used to test the hypotheses 1,2,3 and 7.

4.2.2. T-test

The t-test is an inferential statistic that is used to compute the mean of two groups, describing how they are related, and which are the most important differences, for instance the std. error. Considering the thesis, the two groups identified are:

a) availability of graduates before the relocation

b) availability of graduates after the relocation

This analysis is based on three assumptions: the two groups are independent; they can be approximated as normally distributed; the variance within each group is similar. There are different types of t-test depending on the relation between the two groups determined. Indeed, if the two groups come from the same population, as it is in the analysis performed here, then it is the paired t-test; in case the groups come from two different populations then it is the so called two-sample t-test. Actually, there is a third type of t-test, named one-sample t-test where a group is compared against a standard value.

Specifically, the paired t-test is a blocking method, meaning that it is used to create groups (blocks) of observations, which are similar one each other.

The paired t-test is used to test hypotheses 4,5,6.

4.3. Variables of the model

After having described which econometric tools are used to perform the analyses, it is important specify which variables are used and the reasons behind these decisions. For each variable will be provided a brief description and some descriptive statistics. The data related to the variables that are going to be described cover a period of time between the 2011 and 2019. As introduced previously, there are two kinds of variables, the main variables, and the control variables. This distinction depends only on the relation between the variable itself and the hypothesis introduced in the previous chapter. In other words, if a variable is used to test a hypothesis, then it is considered a "main variable", otherwise if it used to provide further information then it is a control variable. The sources used, for both the categories of variables, are VICO 5.0, EUROSTAT and ETER. More in detail, VICO 5.0 is used for those variables related to the company (NUTS3, relocationdummy, Total Assets etc.), EUROSTAT is used to get information regarding the NUTS3 where the companies are located (for instance the employment) while ETER is used for the data related to the availability of students.

4.3.1. Main Variables

They are the variables used to test the hypothesis already presented, therefore they come, mainly, from ETER database. These variables are about the availability of graduated students, clustered by area of study, level of education, reputation level and NUTS3. Moreover, the hypothesis number 7 introduced another dimension that

is VC-backed status, therefore even the related dummy variable is a main variable, while hypotheses number 2 and 5 introduced the specialization of the company. Beside these variables, there is the one that signal the occurrence of the relocation event. Thus, the main hypotheses are:

- Relocationdummy: it is a dummy variable that is equal to 1 in the year when a relocation takes place and 0 otherwise.

- lnAvailability: it provides information regarding the availability of students in a specific NUTS3. In particular, the argument of the natural logarithm is obtained summing up the number of graduated students in each area of studying of each ISCED level. That number of students is Year-specific meaning that this data is yearly updated. Beside this variable, there are those variables obtained considering the natural logarithm of the graduated students in one specific area. They are lnAgricolture, lnArt, lnBusiness, lnEducation, lnEngineering, lnHealth, lnInformation, lnNaturalScience, lnService. It is possible having missing data for the ln related to specific area because have been dropped just those observation that had a total availability equal to 0, therefore the number of observations for them is lower. At that point, the distinction between the different ISCED level is abandoned because ISCED 8 graduates do not contribute in a significant way to the overall amount, while the other two ISCED level graduates have a very high correlation (0.72). Moreover, the data of (total) availability of graduated students is also split between students graduated from reputable universities and students graduated in less prestigious universities. In order to distinguish the reputable from less reputable universities is used QS rankings: if the university is ranked in the top 500th in the world, then it is considered reputable. Also, for these two data, is used the natural logarithm. The reason why it is used the natural logarithm rather than the absolute value is due to the heterogeneity (skewed distribution) of the latter that would negatively impact the robustness of the result. Moreover, in order to have data suited with the hypotheses to be checked, for each ln is computed also the lagged one, that basically consider the availability of the same NUTS but taking into account the number of graduated students of the year before. This is due to the fact that decisions regarding a relocation are based on information related to the previous year(s) because in the real world is extremely tough having real-time data. The lagged logarithms are indicated with a capital L at the begin (for instance LlnAvailability). All the data regarding the availability of students are retrieved from ETER.

- VC_Step: it is a dummy variable that assumes the value equals to 1 if the company has received at least one VCs' funding, while 0 if it has not. The other two dummy variables cannot be used in the analysis because one, called VC_Backed, is a dummy variable that assumes the value equals to 1 if the company has received or will receive a funding from a VC within the period under analysis and the other one, that is the VCDummy, instead, takes into account only the fundings, so it is 1 only when a round of fundings takes place, while it is 0 otherwise. In testing the hypothesis 7, it is necessary to distinguish between firms that have already been backed by VC and firms that have not disregarding if they will or not. This information is obtained from VICO 5.0

- NACE_SECTION: it is a variable that specify the area of interest of the company. It is obtained from the data NACEFirst2digits coming from VICO 5.0. The reason why it is created this variable is that NACEFirst2digits is too dispersive due to the presence of many categories; the introduction of the section rationalizes this classification through "just" eighteen categories. However, the distribution of the companies is not homogenous among the NACE SECTION, but very concentrated in 5, as it is possible to be observed in the previous chapter in the table 4.13; these sections are section C (manufacturing), section G (wholesale & retail trade), section J (Information & Communication), section K (Financial Services) and section M (Professional, scientific & technical activities) that account for about the 90% of the total observations and for this reason the hypotheses number 2 and 4 are tested only considering them. The information regarding the Nace Sections are obtained from EUROSTAT.

4.3.2. Control variables

After having presented the variables used to test the hypotheses, it is the proper moment to describe the control variables that are aimed at underpinning, if possible, the results coming from the analysis having used the main variables. The following control variables, or at least a large majority of them, have been used by other researchers in order to verify their hypotheses regarding the determinants of the relocation process. Therefore, the control variables are chosen in order to provide a comprehensive description of the phenomenon of relocation. In other words, control variables can be clustered into two different groups: the first group is composed by those variables that provide further information regarding the company observed, and more in details, it is possible to divided even further these variables between financial variables and non-financial variables. The second group of control variables are those that give information regarding the context and

therefore the NUTS3. Variables belonging to the first have been identified from VICO 5.0 while the variables of the second group have been taken from EUROSTAT, that provides information per NUTS3. Similarly, to the Availability case, the variables here are considered with a time gap of one Year. So, the control variables used in the analysis are:

- TotalAssets: it is a variable that is used to provide information regarding the size of the company. The Academic literature, many times, has underlined the importance of this characteristics in the relocation process. Indeed, as explain in the Literature Review, there is a stream of scholars that suggests that the larger is the amount of total assets and the lower will be the propensity to relocate. Even in this case the absolute value is converted into a logarithm scale in order to homogenize the data. However, the availability of data regarding the total assets, which are taken from VICO 5.0, is not so significant, and therefore there is an important number of missing observations. At the end, the total number of observations that have information regarding the TotalAssets of the company are about 13000, more or less, equally split between VC-backed and non-VC-backed.

- Employees: it is a variable that, similarly to the previous one, provides information regarding the size of the company. It is chosen, among the other variables related to size, because it might be seen as complementary to the one previously described, because it is a measure that is not particularly affected by the presence of significant intangible assets, meaning that in some Nace Sections, such as the one of Information & Communications, can provide the real "weight" of the size of the company. Indeed, moving intangible assets is, theoretically, quite easy and cheap. However, many of these firms with large amount of intangible assets have many employees that, instead, are difficult to be replaced in particular if they are high skilled. Also in this case it is used the natural logarithm for the same reasoning. In the formula used to compute the ln, it is added one to the argument, by default, because there are some firms with no employees (but with positive of total assets). The data are retrieved from VICO 5.0 with a percentage of missing data similar to the one that characterized the variable TotalAssets.

- EBITDA: the variable is used to assess the profitability of the company. The reason why it has been chosen this measure rather than the NetProfit is for matter of availability, indeed from VICO 5.0 the data regarding the NetProfit is very rare. The EBITDA can add some important information to the discussion regarding the relocation, indeed any firm that decides to migrate elsewhere has to sustain some costs, that can be higher or lower considering

the type of activity carried out by the company. That variable used in the cox can tell something the reason after the relocation; in other words, relocation can be an "extrema ratio" practise, that is performed because otherwise is not anymore possible carry on with activity (maybe due to the high rental costs), or actually can be a practise perform just from those companies that are wealthy and therefore they can undertake such costs. In this case it was not possible using the logarithm scale because of the presence of negative value. The data are retrieved from VICO 5.0 and its availability is in line with the one of the first two control variables.

- ROA: it is another parameter to express the profitability of a company and it is computed as the ratio between EBITDA and TotalAssets. The advantage of this measure compared to the EBITDA is that is a ratio and therefore independent from the size of the company. In addition to the reasoning described for the former variable, it will be interesting underlining some eventual discrepancies in values regarding the interactions with the phenomenon analysed in order to have a first view regarding the hierarchy of the determinants of relocation decisions. Unfortunately, the availability of these data is even lower than the other variables retrieved from VICO 5.0 due to the fact that is a ratio, therefore in case of missing of just one data between the EBITDA and TotalAssets, then the observation is missed.

- GDP: it is the first variable presented here that is referring to the description of the NUTS. It is one of the most important parameters, it assesses the entity of the economic activities located within the borders of the NUTS3. This variable is related to the discussion regarding the relocation choice because can be considered as a measure through which identifying urbanization areas, which play a massive role in the location and relocation theory thanks to the presence of the externalities (Nilsen *et al.*, 2020) on one hand and the potential leakages and misappropriation on the other (Myles Shaver and Flyer, 2000) in these urbanization areas. In that case the data is obtained through EUROSTAT and the availability of data was highly comprehensive covering each NUTS3 considered in this thesis.

- GDP growth: it is strictly connected to the previous parameter by being the percentage variation between two consecutive values of GDP in one specific NUTS3. It is considered because it adds, to the GDP variable, a temporal dimension that otherwise it would be omitted. As shown in the following table, the correlation between GDP and GDP growth is very low. It was not possible compute the GDPgrowth values for the year 2011 because it was not

available, on STATA, the information regarding the population of the previous year. The value is expressed in percentage.

- Population: it is another variable providing information regarding the NUTS3. It can be seen as a measure of the availability of workforce, although without distinguishing between the "quality" of the potential employees. Moreover, this variable affects the variable of the GDP, because, theoretically, larger is the population and larger would be the GDP, and a growth of population should be translated into a growth of the GDP, however in the real world this connection is much less straightforward. The data are retrieved from EUROSTAT therefore it was possible to get information for each NUTS3, however it was not possible to obtain data about Population per NUTS3 for the years 2011, 2012 and 2013, while it was possible for the following years of the period.

- EmploymentperNUTS3: this parameter considers only the active part of the population that contributes to the economic activity of the NUTS3, and the comparison between Employment and Population variables provides a hint on the economic wealth of the region observed. It was not possible to compute the unemployment as the difference between the Population and the EmploymentperNUTS3, both referring to the same observation, because by performing this calculation both students and retires would be considered unemployed. The data are retrieved from EUROSTAT, and for each year of the period examined it was possible get information.

- GDPpercapita: this variable is obtained by dividing the GDP for the Population of the related NUTS3. The reason why it is decided to consider even this parameter is due to the fact that it is not impacted by the size of NUTS3, that is not homogenous across the different European Countries. Therefore, this measure would assess, in more accurate way, the difference, in terms of "richness" among the different regions. The GDP is computed in millions of EUROs this is why, in the table below, each value of the GDPpercapita is lower than 1. As shown in the Literature Review, the relation between areas with a better economic condition (meaning higher GDPpercapita) and relocation is a quite vexed topic, indeed on one side it is true that higher GDPpercapita is obtained in metropolitan areas, where a company can benefit from spillovers etc (Audretsch and Keilbach, 2007), but at the same time the rental costs and the cost of personnel, at least, increase significantly, and even for the employees living within a metropolitan areas implies higher life cost (van Dijk and Pellenbarg, 2000). It was possible to

compute the GDPpercapita since 2014 due to the lack of the data regarding the Population variable in the previous years.

- CommunityDesignRights: it considers the number of applications for an industrial design rights per NUTS3 and it is a proxy of the degree of innovativeness of the region taken into account. Also in this case, the implication of being located into an innovative context has already been discussed in the chapter related to the Literature Review. The values are converted into a natural logarithm scale to stabilize the observed values and it was possible to retrieve information regarding the application up to the 2016, included. The source was, also in this case, EUROSTAT.

- d_Year: it is a dummy variable that assumes the value of 1 when the observation is referred to a specific year, and 0 for the others. Each year considered in the interval analysed has its own dummy variable d_Year.

- d_NACE_SECTION: it is a dummy variable too that assumes the value of 1 when the company of the observation belongs to a specific NACE SECTION, and 0 in the other cases. As the previous variable, each NACE SECTION has its own dummy variable.

At the end of this chapter, three tables are shown where the first summarizes the main statistical characteristics of the aforementioned variables, while the second and the third define the level of correlation among them. Since now, all the control variables, except for the dummy, are considered in a lagged form meaning that the values are referred to the previous year.

Variable	Obs	Mean	Std. dev.	Min	Max
LlnTotalAs~s	10,092	7.272385	2.160621	.0732876	18.3151
LlnEmployees	10,092	2.648639	1.621424	0	12.04706
LEBITDA	10,092	7093.349	201591.5	-565277	1.23e+07
LROA	10,092	.1019457	11.56679	-54.73903	1156.702
LGDP	106,853	46265.94	42280.84	1312.3	156208.8
LGDPgrowth	84,660	.8361845	65.42244	-6229.033	98.26171
Population	80,823	1033974	945087.2	39839	3644826
Employeesp~3	139,002	572893.5	489059.7	33560	2074230
LGDPpercap~a	50,711	.0519001	.0236094	.0101151	.1324824
LlnCommuni~s	67,655	3.363476	1.27729	.6931472	5.313206

Table 4.1 Descriptive statistics of the control variables

There are some of these variables that are negatively affected by some observations like the GDPgrowth variables that has a minimum value of -6229,033 (%). This heterogeneity of the results enhances the decision to use, where is possible, the logarithm scale.

In the following table will be shown the level of correlation among the control variables described previously. Some of the variables' names have been abbreviated for a matter of space.

	LlnTot~s	LlnEmp~s	LEBITDA	LROA	LGDP	LGDPgr~h	Popula~n	Employ~3	LGDPpe~a	LlnCom~s
LlnTotalAs~s	1.0000									
LlnEmployees	0.8035	1.0000								
LEBITDA	0.2369	0.2715	1.0000							
LROA	0.1283	0.0453	0.0228	1.0000						
LGDP	0.2743	0.2371	-0.0383	-0.0409	1.0000					
LGDPgrowth	0.0115	0.0067	-0.0000	0.0064	0.0409	1.0000				
Population	0.0776	0.0635	-0.0501	-0.0838	0.8472	0.0370	1.0000			
Employeesp~3	0.1524	0.1298	-0.0503	-0.0707	0.9309	0.0412	0.9768	1.0000		
LGDPpercap~a	0.4172	0.3636	0.0302	0.0728	0.3542	0.0294	-0.0664	0.0943	1.0000	
LlnCommuni~s	0.2458	0.2407	0.0266	-0.0173	0.8293	0.0573	0.6879	0.7732	0.4090	1.0000

Table 4.2 Correlation of control variables

It is quite clear that there are some variables with a remarkably high degree of correlation. The large majority of these relations were expected but not with this intensity. For example, the correlation between lnCommunityDesignRights and GDP is 0,8293, that is higher than expected. However, after the observation of the table 5.2 it was decided to drop lnEmployees, Population and EmployeesperNUTS3 as a control variable because they cannot add useful information to the discussion. Therefore, the table that shows the correlation between the control variables is the following one. The variable lnCommunityDesignRIghts was kept because it provides information regarding level of innovation of a specific NUTS, which is a topic very connected to the one discussed by this thesis that is the role of the University in the relocation decisions.

	LlnTot~s	LEBITDA	LROA	LGDP	LGDPgr~h	LGDPpe~a	LlnCom~s
LlnTotalAs~s	1.0000						
LEBITDA	0.2369	1.0000					
LROA	0.1283	0.0228	1.0000				
LGDP	0.2743	-0.0383	-0.0409	1.0000			
LGDPgrowth	0.0115	-0.0000	0.0064	0.0409	1.0000		
LGDPpercap~a	0.4172	0.0302	0.0728	0.3542	0.0294	1.0000	
LlnCommuni~s	0.2458	0.0266	-0.0173	0.8293	0.0573	0.4090	1.0000

Table 4.3 Correlation of control variables after the drop

5 Results

In this chapter will be presented the results of the econometric analysis performed. Here, it will be discussed the processes and the decisions that underpin the tests of the different hypotheses, the signs and the significancy of the results obtained and a first comment regarding the relation between these results and the existing literature. This last part, however, will be deepened in the next and last chapter.

5.1. Results of the Hypothesis 1

The first hypothesis is the most generic one, in which it is evaluated if there is sort of connection between relocation and availability of students, without introducing other variables. In order to recall it, the first hypothesis is the following one:

Hypothesis 1: companies that are located in regions where the availability of universities is lower, tend to relocate more than companies close to these institutions

In order to test this hypothesis, it is used the Cox with multiple events, where the events are recorded through the dummy variable "relocationdummy". Then, to enhance the results obtained are added into the Cox analysis the control variables remained. Before launching the analysis, for the variable of "Year" and "NACE_SECTION" were created a dummy variable in order to be used in the Cox to get more accurate results. In this first case, it will be shown the results with and without these dummy variables, but in the following the analysis are performed considering always both. It was decided to avoid the creation of a dummy variable related to the NUTS1 (macro-regions) because this geographical variable would change significantly the results of the analysis, because the variable "Availability" is already considering a geographical division. For this first hypothesis, the only main variable considered is LlnAvailability. Before showing the results, it is important to say that the software STATA uses a listwise deletion by default, therefore in case of a missing data among the variables involved in a specific test, then that observation is excluded from the analysis and does not affect the data.

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]	
LlnAvailability	.9250144	.0170894	-4.22	0.000	.8921191	.9591227

Table 5.1 Descriptive statistics of LnAvailability

This result is obtained by doing a simple cox considering the lagged natural logarithm of the availability. The two most important results are the p value and the hazard ratio. In this case the p value assumes 0 as value, meaning that the significancy of the result is very important. In other words, it means that LlnAvailability impacts on the tendency of relocation. To assess the intensity of this impact it is important looking at the hazard ratio. In this case, the value assumes by the hazard ratio is significant meaning that an increase of the availability of students reduces the inclination of the firms to relocate. Indeed, the further is the hazard ratio from the value of 1 and the stronger is the impact on the failure event analysed, that is the relocation. The last two values of the row identify a range within it there is the 95% of accuracy that the hazard ratio belongs to one of the values included. In this case both of the value that define the interval are below 1.

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]	
LlnAvailability	.9294207	.0172083	-3.95	0.000	.8962976	.9637678
d_Year1	1	(omitted)				
d_Year2	.7871005	.096668	-1.95	0.051	.6187136	1.001315
d_Year3	.7324106	.0894131	-2.55	0.011	.5765534	.9304001
d_Year4	.7856043	.0941611	-2.01	0.044	.6211269	.9936361
d_Year5	1.095922	.1362306	0.74	0.461	.8589534	1.398266
d_Year6	.8606587	.1134688	-1.14	0.255	.6646745	1.11443
d_Year7	.907159	.1194895	-0.74	0.459	.7007523	1.174363
d_Year8	.8787367	.1174869	-0.97	0.334	.6761659	1.141995
d_Year9	1	(omitted)				

Table 5.2 Cox analysis with d_Year

This table is obtained by considering the dummy variable related to the years belonging to the interval of time analysed. By comparing the two tables, it is possible to affirm that the significancy of the result remains very high, and also the impact of the parameter LlnAvailability is very similar to the previous case (slightly higher). The hazard ratios in the rows that start with d_Year assess if in the related

year there was a higher (>1) or lower (<1) number of relocations compared to the year omitted to overcome the issue of the collinearity, described in the previous chapter. The hazard ratio assumes always values smaller than 1, except for the d_Year5, however not always these values have the sufficient statistical significancy, that means having a p value larger than 0.1, to be taken into consideration as reliable results.

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]	
LlnAvailability	.9219624	.0171784	-4.36	0.000	.8889006	.9562539
d_Year1	1	(omitted)				
d_Year2	.7958439	.097926	-1.86	0.063	.6253026	1.012897
d_Year3	.7388627	.0903847	-2.47	0.013	.5813487	.9390546
d_Year4	.7916999	.0950674	-1.95	0.052	.6256742	1.001781
d_Year5	1.097289	.1365177	0.75	0.456	.8598448	1.400303
d_Year6	.8614553	.113631	-1.13	0.258	.665203	1.115607
d_Year7	.9058667	.1193733	-0.75	0.453	.6996723	1.172827
d_Year8	.877783	.117373	-0.97	0.330	.6754116	1.14079
d_Year9	1	(omitted)				
d_NACE_SECTION1	3.081548	3.142303	1.10	0.270	.4176239	22.73801
d_NACE_SECTION2	7.09e-19	.	.	.	.	.
d_NACE_SECTION3	.9580174	.2073486	-0.20	0.843	.6268209	1.46421
d_NACE_SECTION4	.5072585	.3722916	-0.92	0.355	.1203672	2.137719
d_NACE_SECTION5	.4937977	.503301	-0.69	0.489	.0669832	3.640259
d_NACE_SECTION6	.7177387	.4379455	-0.54	0.587	.2170628	2.373271
d_NACE_SECTION7	1.156077	.2456668	0.68	0.495	.7622652	1.753345
d_NACE_SECTION8	.9568648	.4336004	-0.10	0.922	.393669	2.325787
d_NACE_SECTION9	1.32e-20	.	.	.	.	.
d_NACE_SECTION10	1.32648	.2652771	1.41	0.158	.8963393	1.963038
d_NACE_SECTION11	1.276449	.3094046	1.01	0.314	.7937372	2.052723
d_NACE_SECTION12	2.377677	.9634801	2.14	0.033	1.074558	5.261093
d_NACE_SECTION13	1.190849	.2469648	0.84	0.400	.7931036	1.788066
d_NACE_SECTION14	1.588454	.3713745	1.98	0.048	1.004539	2.511784
d_NACE_SECTION15	2.086254	.9452729	1.62	0.105	.8584034	5.070409
d_NACE_SECTION16	2.139243	.7274786	2.24	0.025	1.098493	4.166033
d_NACE_SECTION17	.9642187	.5181267	-0.07	0.946	.3363404	2.764216
d_NACE_SECTION18	1	(omitted)				

Table 5.3 Cox analysis with d_Year and d_NACE_SECTION

Instead, this table is achieved by adding the dummy variable related to the NACE SECTION in the Cox analysis. The considerations to be done are quite similar to the ones described previously with the other table: the results keep their statistical significancy even considering the NACE SECTION, and the impact of the main variable LlnAvailability remains, basically, the same, meaning that higher is the availability of students in a region, lower will be the tendency to relocate elsewhere for the firms. The hazard ratios related to the NACE SECTION rows indicate if the companies belonging to that field have a higher (>1) or lower (<1) tendency to relocate than the NACE SECTION omitted (NACE SECTION18) due to the collinearity issue. In some cases, the p values are very high also because the number of observations for that specific section is low.

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf. interval]	
LlnAvailability	.9769278	.0223046	-1.02	0.307	.9341753	1.021637
d_Year1	1	(omitted)				
d_Year2	.8945997	.1158777	-0.86	0.390	.6940207	1.153148
d_Year3	.8305235	.1063599	-1.45	0.147	.6461657	1.06748
d_Year4	.8764793	.1099144	-1.05	0.293	.6854839	1.120692
d_Year5	1.093729	.1367362	0.72	0.474	.8560398	1.397415
d_Year6	.8474144	.112101	-1.25	0.211	.6538737	1.098241
d_Year7	.8887054	.1173881	-0.89	0.372	.6859991	1.15131
d_Year8	.8686496	.1163258	-1.05	0.293	.6681215	1.129364
d_Year9	1	(omitted)				
d_NACE_SECTION1	2.758714	2.824187	0.99	0.322	.3709401	20.5168
d_NACE_SECTION2	6.08e-19	.	.	.	.	.
d_NACE_SECTION3	.9786635	.2161233	-0.10	0.922	.6348311	1.50872
d_NACE_SECTION4	.5601797	.4118425	-0.79	0.431	.1325948	2.366619
d_NACE_SECTION5	.4567292	.4662384	-0.77	0.443	.061764	3.3774
d_NACE_SECTION6	.7270409	.4468956	-0.52	0.604	.2179438	2.425343
d_NACE_SECTION7	1.192019	.2558192	0.82	0.413	.7827209	1.815347
d_NACE_SECTION8	.9236897	.4195814	-0.17	0.861	.3792033	2.249987
d_NACE_SECTION9	1.45e-20	.	.	.	.	.
d_NACE_SECTION10	1.322023	.2705763	1.36	0.173	.8851668	1.974482
d_NACE_SECTION11	1.332526	.3285291	1.16	0.244	.8218934	2.16041
d_NACE_SECTION12	2.280876	.9437089	1.99	0.046	1.01372	5.13198
d_NACE_SECTION13	1.200281	.2523303	0.87	0.385	.7949474	1.81229
d_NACE_SECTION14	1.549105	.3635375	1.87	0.062	.9779672	2.453789
d_NACE_SECTION15	1.813474	.8343208	1.29	0.196	.7360394	4.468087
d_NACE_SECTION16	2.040483	.6977412	2.09	0.037	1.043916	3.988417
d_NACE_SECTION17	.9497597	.5126799	-0.10	0.924	.3297125	2.735848
d_NACE_SECTION18	1	(omitted)				
d_NUTS1_1	.7942194	.4893838	-0.37	0.708	.2373807	2.65727
d_NUTS1_2	.5818052	.3413638	-0.92	0.356	.1842257	1.837405
d_NUTS1_3	.6008811	.350286	-0.87	0.382	.1916802	1.883648
d_NUTS1_4	.3998743	.2379418	-1.54	0.123	.1245728	1.283582
d_NUTS1_5	1.15988	.7072325	0.24	0.808	.3510721	3.832039
d_NUTS1_6	1.009487	.6858062	0.01	0.989	.2665799	3.822733
d_NUTS1_7	.4825895	.2913413	-1.21	0.227	.14781	1.575621
d_NUTS1_8	.6571893	.3905202	-0.71	0.480	.2050608	2.106193
d_NUTS1_9	.8642729	.5479423	-0.23	0.818	.249457	2.994374
d_NUTS1_10	.5238444	.3176033	-1.07	0.286	.1596033	1.719344
d_NUTS1_11	.6108533	.3580361	-0.84	0.400	.1936541	1.926846
d_NUTS1_12	.5899211	.3738042	-0.83	0.405	.1703841	2.042484
d_NUTS1_13	.1171391	.1354779	-1.85	0.064	.0121406	1.130222
d_NUTS1_14	.4903849	.298012	-1.17	0.241	.1490229	1.613693
d_NUTS1_15	.2987977	.2116524	-1.71	0.088	.0745472	1.197631
d_NUTS1_16	.8177893	.50562	-0.33	0.745	.243424	2.747384
d_NUTS1_17	.1413171	.1155725	-2.39	0.017	.0284486	.7019849
d_NUTS1_18	.3071013	.2058534	-1.76	0.078	.082549	1.142487
d_NUTS1_19	.7234241	.4270986	-0.55	0.583	.2274346	2.301068
d_NUTS1_20	.5500203	.3813662	-0.86	0.389	.1413133	2.140791
d_NUTS1_21	.1873198	.1717129	-1.83	0.068	.0310672	1.129446
d_NUTS1_22	.2635663	.2025751	-1.73	0.083	.0584336	1.188823
d_NUTS1_23	.3852137	.24592	-1.49	0.135	.1102289	1.346194
d_NUTS1_24	.1164278	.0893905	-2.80	0.005	.0258538	.5243109
d_NUTS1_25	.454912	.2921802	-1.23	0.220	.1291865	1.601908
d_NUTS1_26	.1997059	.1535559	-2.10	0.036	.0442479	.90134
d_NUTS1_27	.3160025	.2077415	-1.75	0.080	.0871177	1.146237
d_NUTS1_28	.197758	.1314374	-2.44	0.015	.0537519	.7275696
d_NUTS1_29	.2402992	.1511826	-2.27	0.023	.0700204	.8246693
d_NUTS1_30	.4157041	.2567561	-1.42	0.155	.123893	1.394831
d_NUTS1_31	3.99e-20	.	.	.	.	.
d_NUTS1_32	.2123212	.2458805	-1.34	0.181	.0219408	2.054637
d_NUTS1_33	1.101436	.6605831	0.16	0.872	.3399801	3.56833
d_NUTS1_34	.9845606	.595293	-0.03	0.979	.3010108	3.220348
d_NUTS1_35	.6028192	.3576032	-0.85	0.394	.1884686	1.928125
d_NUTS1_36	.7394465	.4410298	-0.51	0.613	.2297326	2.380077
d_NUTS1_37	1.02e-19	.	.	.	.	.
d_NUTS1_38	1	(omitted)				

Table 5.4 Cox analysis with d_Year , d_NACE_SECTION, d_NUTS1

This is the last table shown of this type, and it is added here in order to clarify why it was decided to drop the dummy variable of NUTS1. Indeed, due to the fact that

LlnAvailability is already a variable that considers the geographical divisions of the NUTS, adding as a control variable the dummy NUTS1 reduces significantly the statistical significance of the results (0.307>0.1 and so not reliable), despite the hazard ratio suggests a coherency with the previous cases in terms of impact on the relocation events by the LlnAvailability. The reason why this has occurred when it is added the NUTS1 dummy variable and not with the dummy variable related to Year is because LlnAvailability has a higher homogeneity across the year rather across the NUTS1. In other words, the values of the main variable remain similar in each NUTS1 along the period considered while they change, even significantly, considering different NUTS1 in the same year. Moreover, the number of years, and therefore of dummy variables created, is much smaller than the one of the NUTS1.

The next table will summarize the results obtained in the first 3 tables, and since this moment the results will be always shown in that way in order to be more synthetic and, hopefully, clearer.

	(1) _t b/se	(2) _t b/se	(3) _t b/se	(4) _t b/se
LlnAvailability	-0.078***	-0.073***	-0.081***	-0.023
	(0.018)	(0.019)	(0.019)	(0.023)
Year Dummies	No	Yes	Yes	Yes
Industry Dummies	No	No	Yes	Yes
Geography Dummies	No	No	No	Yes
N	106866	106866	106866	106866
N_sub	21874	21874	21874	21874
N_fail	1365	1365	1365	1365

Table 5.5 Summary of the Cox analyses with LlnAvailability

It is important to say that in this last table there are not shown the hazard ratios, previously seen, but coefficients, which is mathematically connected to the latter by this formula:

$$(5.1) \quad e^{\wedge}(\text{coefficient} * \text{std. deviation}) = \text{hazard ratio}$$

The std. deviation in the formula is the std. deviation related to the variable LlnAvailibility; the main statistical characteristics of the variable are shown below in the table. The stars near the coefficients indicate the significancy of the results, where the three stars condition represents the highest level of significancy, while one star condition represents the lowest. A coefficient without star, as the one related to the analysis considering the NUTS1 variable, means no significancy of the results. Finally, the last three rows of the table indicate, respectively: the number of observations analysed, the number of subjects (companies) involved in the cox analysis, and the number of failure events (relocations) detected.

Variable	Obs	Mean	Std. dev.	Min	Max
LlnAvailab~y	106,866	8.32768	1.390894	1.609438	10.28387

Table 5.6 Descriptive statistics of LnAvailability part.2

Once defined the dummy variables to be included in the Cox analysis, then the impacts of the control variables, on the relation LlnAvailability-Relocation, are analysed in order to understand deeply the phenomenon. Here, it is shown the summarizing table that is obtained adding, one by one, the control variables.

	(1) _t b/se	(2) _t b/se	(3) _t b/se	(4) _t b/se	(5) _t b/se	(6) _t b/se	(7) _t b/se
LlnAvailability	-0.081***	-0.129*	-0.131*	-0.131*	-0.159*	-0.241**	-0.225*
	(0.019)	(0.069)	(0.069)	(0.069)	(0.082)	(0.095)	(0.131)
LlnTotalAssets		-0.080	-0.076	-0.076	-0.083	-0.072	-0.039
		(0.058)	(0.059)	(0.059)	(0.060)	(0.068)	(0.079)
LEBITDA			-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LROA				-0.001	-0.001	-0.001	0.043
				(0.009)	(0.009)	(0.012)	(0.135)
LGDP					0.000	0.000	0.000
					(0.000)	(0.000)	(0.000)
LGDPgrowth						-0.001***	0.008
						(0.000)	(0.014)
LlnCommunityDesignRights							0.018
							(0.198)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	106866	10092	10092	10092	10092	7899	4009
N_sub	21874	4777	4777	4777	4777	4023	1999
N_fail	1365	101	101	101	101	76	52

Table 5.7 Cox analysis with control variables

Thus, looking at the table the first and most important result is that by adding the control variables the relation between LlnAvailability-relocation is kept, indeed coefficients in the first row are always negative meaning that there still is an inverse relation between the presence of the students in a NUTS and the firms' tendency to relocate. Deepening the comments about this table, it can be observed that significancy of the results is reduced through the addition of the control variables but remaining acceptable. Only the introduction of LGDP growth does not impact particularly in terms of significancy. In terms of coefficient, instead, there is a further

reduction compared to the original one obtained without control variables. It means that the availability of students becomes even more impacting on the reduction of the firms' relocation tendency. Talking about the coefficient of the other control variables, unfortunately the results are not so significant statistically, this is also caused by the important reduction of the number of observations analysed due to the listwise selection by default. Only the variable GDPgrowth has statistically significancy but the coefficient is, basically, null meaning that it does not provide further information to this topic.

Thus to conclude, it is possible to affirm that the first hypothesis is corroborated by the results of the analysed performed.

5.2. Results of the hypothesis 2

The second hypothesis is designed to deepen the results obtained from the first hypothesis. Thus, once defined that exists a relation between the presence of students and the relocation phenomenon, it is important linking the areas of interest of the firms with the typologies of students that are more attractive for the firms themselves. So, first of all, it is important to recall the hypothesis mentioned:

Hypothesis 2: companies that are located in regions where the availability of specialized university, in the area of interest of the company, is lower, tend to relocate more than companies close to these institutions.

Even this hypothesis is tested through the use of a Cox with multiple events, considering since the beginning both dummy variables of Year and NACE SECTION. Before starting to discuss about the relation between the typologies of students and the areas of interest of firms, it might be useful see how each typology of students affect the relocation phenomenon in general and finding out whether there are some streams that impact more on the relocation phenomenon in general. Empirically, it was repeated the analysis done for the first hypothesis but considering the lagged natural logarithm of the availability of each stream of study.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	_t	_t	_t	_t	_t	_t	_t	_t	_t	_t
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
LlnAvailability	-0.081***									
	(0.019)									
LlnAgricolture		0.105***								
		(0.036)								
LlnArt			-0.101***							
			(0.023)							
LlnBusiness				-0.078***						
				(0.020)						
LlnEducation					-0.086***					
					(0.023)					
LlnEngineering						-0.083***				
						(0.022)				
LlnHealth							-0.047**			
							(0.021)			
LlnInformation								-0.089***		
								(0.024)		
LlnNaturalScience									-0.057***	
									(0.021)	
LlnService										-0.099***
										(0.023)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	106866	49804	85332	103475	91437	97267	83760	96744	85248	88808
N_sub	21874	11274	18773	21457	19754	20415	18534	20277	19006	19642
N_fail	1365	593	1027	1308	1150	1220	1028	1207	1039	1143

.

Table 5.8 Summary of the Cox analyses of the different typologies of graduates

The results of this analysis are quite clear, indeed there is a high significancy for all the parameters and the coefficients are always negative except for the agriculture students' availability. This result is not totally unexpected because it is reasonable affirms that agriculture's graduates are more concentrated in rural areas and many scholars have affirmed that in those areas most firms want to relocate away towards metropolitan areas to exploit the advantages related to the agglomeration economies, (Nilsen et al., 2020), (Risselada et al., 2013).

A bit unexpected is the coefficient related to the availability of students graduated in art. Indeed, considering the number of firms in this stream is not so significant the expectation was a coefficient closer to 0. The other coefficients are in line with the expectations, where students graduated in Business, Education, Engineering, Information & Service seem to be more attractive towards companies.

In a specular way, it is worth to be done an analysis that shows if there are some NACE SECTIONs that are more influenced by the general availability of students. As said in the presentation of the variables in the previous chapter, the analysis is limited to the five NACE SECTIONs that account for about the 90% of the companies observed in the database used.

	(1) _t b/se	(2) _t b/se	(3) _t b/se	(4) _t b/se	(5) _t b/se
LlnAvailability	-0.096*** (0.019)	-0.096*** (0.020)	-0.042 (0.027)	-0.082*** (0.019)	-0.069*** (0.020)
1.d_NACE_SECTION3	-1.362** (0.580)				
1.d_NACE_SECTION3#c.LlnAvail~t	0.167** (0.067)				
1.d_NACE_SECTION7		-0.993* (0.546)			
1.d_NACE_SECTION7#c.LlnAvail~t		0.140** (0.061)			
1.d_NACE_SECTION10			0.920** (0.361)		
1.d_NACE_SECTION10#c.LlnAvai~i			-0.078** (0.037)		
1.d_NACE_SECTION11				0.109 (0.816)	
1.d_NACE_SECTION11#c.LlnAvai~i				0.017 (0.096)	
1.d_NACE_SECTION13					0.797* (0.447)
1.d_NACE_SECTION13#c.LlnAvai~i					-0.077 (0.050)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes
N	106866	106866	106866	106866	106866
N_sub	21874	21874	21874	21874	21874
N_fail	1365	1365	1365	1365	1365

Table 5.9 Analysis of the relationship LlnAvailability-NACE SECTION

Also in this case, the results are mostly significant. The availability of students is a parameter that deters, in different ways, firms to relocate depending on their area of interest. Indeed, considering the NACE SECTION 3 (manufacturing) the coefficient is positive (0.167) and therefore has an opposite sign compared to the base case. This implies that for Manufacturing the availability of students does not deter the relocation choice. To compute the hazard ratio of this result it can be used the formula shown before where the coefficient is the sum of the coefficients shown in the table (-0.096+0.167). The hazard ratio computed is equal to 1.1037 meaning that in the NUTS with higher availability of students there were more relocations of manufacturing companies. This coefficient has a good statistical significancy

therefore, the related result is dependable. The negative coefficient -1.362 tells that manufacturing companies relocate more than others type of firms; therefore, the availability of students is not a good determinant to describe their migration. Similar considerations can be done for the NACE SECTION 7 (wholesale & retail trade). Considering the NACE SECTION 10 (Information & Communication), the coefficient has a sign that is coherent with the one of the base case, meaning that, for these categories of firms, the availability of graduates is even more impacting on the relocation decision. The hazard ratio computed is equals to 0.846. For the last two NACE SECTIONs analysed the statistical significancy is not sufficient to affirm something.

Now, it is the moment to really test the second hypothesis, and to do so the lagged natural logarithms of the different availabilities are used as terms of moderation to find out if they have a tighter connection with the relocation effects compared to the general case (LlnAvailability). Once again, the categories of firms analysed are the same 5.

	(1) _t b/se	(2) _t b/se	(3) _t b/se	(4) _t b/se	(5) _t b/se	(6) _t b/se	(7) _t b/se	(8) _t b/se	(9) _t b/se	(10) _t b/se
LlnAvailability	-0.096*** (0.019)									
1.d_NACE_SECTION3	-1.362** (0.580)	-0.590 (0.826)	-0.493 (0.588)	-0.770 (0.521)	-0.433 (0.492)	-0.641 (0.577)	-0.182 (0.636)	-1.294** (0.541)	-1.280** (0.618)	-0.711 (0.497)
1.d_NACE_SECTION3#c.LlnAvail~t	0.167** (0.067)									
LlnAgricolture		0.087** (0.037)								
1.d_NACE_SECTION3#c.LlnAgric~e		0.250* (0.135)								
LlnArt			-0.117*** (0.024)							
1.d_NACE_SECTION3#c.LlnArt			0.161** (0.078)							
LlnBusiness				-0.088*** (0.021)						
1.d_NACE_SECTION3#c.LlnBusin~s				0.112 (0.070)						
LlnEducation					-0.097*** (0.024)					
1.d_NACE_SECTION3#c.LlnEduca~n					0.108 (0.075)					
LlnEngineering						-0.091*** (0.023)				
1.d_NACE_SECTION3#c.LlnEngin~g						0.100 (0.079)				
LlnHealth							-0.056** (0.022)			
1.d_NACE_SECTION3#c.LlnHealth							0.130 (0.085)			
LlnInformation								-0.110*** (0.026)		
1.d_NACE_SECTION3#c.LlnInfor~n								0.244*** (0.088)		
LlnNaturalScience									-0.084*** (0.022)	
1.d_NACE_SECTION3#c.LlnNatur~e									0.291*** (0.081)	
LlnService										-0.109*** (0.024)
1.d_NACE_SECTION3#c.LlnService										0.131 (0.087)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	106866	49804	85332	103475	91437	97267	83760	96744	85248	88808
N_sub	21874	11274	18773	21457	19754	20415	18534	20277	19006	19642
N_fail	1365	593	1027	1308	1150	1220	1028	1207	1039	1143

Table 5.10 Cox Analysis NACE SECTION3

The coefficients related to the rows concerning the availability of students provide information already discussed through the table 6.7. The coefficients of the rows tell that disregarding the term of moderation, manufacturing companies tend to relocate more than other types. It is also true that in the most cases the reliability of the results is not high. However, the main goal of this table is showing the impact of the specific availability of students on the manufacturing companies. Honestly, the results are not satisfying because in most cases there is not the statistical significancy even considering engineering and business graduates that should be strictly connected to the manufacturing world. The only results in line with the expectation, and statistically dependable, are that agriculture and art students are

not attractive for manufacturing firms. The fact that even the availability of Information (ICT) and NaturalScience students increases the probability of relocation is unexpected.

	(1) _t b/se	(2) _t b/se	(3) _t b/se	(4) _t b/se	(5) _t b/se	(6) _t b/se	(7) _t b/se	(8) _t b/se	(9) _t b/se	(10) _t b/se
LlnAvailability	-0.096*** (0.020)									
1.d_NACE_SECTION7	-0.993* (0.546)	0.323 (0.646)	0.457 (0.564)	-0.493 (0.506)	0.044 (0.485)	-0.471 (0.561)	0.431 (0.541)	-0.169 (0.479)	0.121 (0.544)	-0.040 (0.461)
1.d_NACE_SECTION7#c.LlnAvail~t	0.140** (0.061)									
LlnAgricolture		0.092** (0.038)								
1.d_NACE_SECTION7#c.LlnAgric~e		0.105 (0.099)								
LlnArt			-0.105*** (0.024)							
1.d_NACE_SECTION7#c.LlnArt			0.036 (0.071)							
LlnBusiness				-0.087*** (0.021)						
1.d_NACE_SECTION7#c.LlnBusin~s				0.094 (0.065)						
LlnEducation					-0.091*** (0.024)					
1.d_NACE_SECTION7#c.LlnEduca~n					0.043 (0.073)					
LlnEngineering						-0.092*** (0.023)				
1.d_NACE_SECTION7#c.LlnEngin~g						0.095 (0.076)				
LlnHealth							-0.053** (0.023)			
1.d_NACE_SECTION7#c.LlnHealth							0.050 (0.065)			
LlnInformation								-0.096*** (0.026)		
1.d_NACE_SECTION7#c.LlnInfor~n								0.068 (0.077)		
LlnNaturalScience									-0.069*** (0.023)	
1.d_NACE_SECTION7#c.LlnNatur~e									0.099 (0.067)	
LlnService										-0.103*** (0.024)
1.d_NACE_SECTION7#c.LlnService										0.044 (0.077)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	106866	49804	85332	103475	91437	97267	83760	96744	85248	88808
N_sub	21874	11274	18773	21457	19754	20415	18534	20277	19006	19642
N_fail	1365	593	1027	1308	1150	1220	1028	1207	1039	1143

Table 5.11 Cox analysis NACE SECTION7

The comments to be done to this table are similar to the ones of the previous table. It means that availability, both the total and the specific one, deters the relocation in

general. It is not possible to add something regarding the wholesale & retail trade because there is not statistical significancy.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	_t	_t	_t	_t	_t	_t	_t	_t	_t	_t
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
LlnAvailability	-0.042									
	(0.027)									
1.d_NACE_SECTION10	0.920**	1.881***	1.073**	0.535	0.378	0.897**	0.971**	0.748**	1.751***	0.484
	(0.361)	(0.526)	(0.435)	(0.339)	(0.350)	(0.374)	(0.431)	(0.345)	(0.423)	(0.322)
1.d_NACE_SECTION10#c.LlnAvai~i	-0.078**									
	(0.037)									
LlnAgricolture		0.210***								
		(0.050)								
1.d_NACE_SECTION10#c.LlnAgri~r		-0.208***								
		(0.066)								
LlnArt			-0.072**							
			(0.033)							
1.d_NACE_SECTION10#c.LlnArt			-0.054							
			(0.046)							
LlnBusiness				-0.060**						
				(0.028)						
1.d_NACE_SECTION10#c.LlnBusi~s				-0.036						
				(0.039)						
LlnEducation					-0.085***					
					(0.032)					
1.d_NACE_SECTION10#c.LlnEduc~n					-0.003					
					(0.045)					
LlnEngineering						-0.033				
						(0.033)				
1.d_NACE_SECTION10#c.LlnEngi~n						-0.092**				
						(0.044)				
LlnHealth							-0.030			
							(0.031)			
1.d_NACE_SECTION10#c.LlnHealth							-0.032			
							(0.042)			
LlnInformation								-0.047		
								(0.034)		
1.d_NACE_SECTION10#c.LlnInfo~o								-0.086*		
								(0.049)		
LlnNaturalScience									0.023	
									(0.032)	
1.d_NACE_SECTION10#c.LlnNatu~i									-0.153***	
									(0.043)	
LlnService										-0.076**
										(0.033)
1.d_NACE_SECTION10#c.LlnServ~e										-0.045
										(0.046)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	106866	49804	85332	103475	91437	97267	83760	96744	85248	88808
N_sub	21874	11274	18773	21457	19754	20415	18534	20277	19006	19642
N_fail	1365	593	1027	1308	1150	1220	1028	1207	1039	1143

Table 5.12 Cox Analysis NACE SECTION10

The results shown here are much more interesting because, first of all, this sector presents a lower tendency to relocate compared to the other sectors as it is observable looking at the coefficients of the second row that are all positive and, most of them, reliable. Therefore, it is not connected to a particular data of availability, but it is a characteristic of the companies of this sector (ICT). Moving to the analysis of the terms of moderation, there are four reliable results that, only in part, have been expected. Indeed, the fact that the presence of ICT students decreases the probability of relocation for this segment is highly reasonable. This is also true considering Engineering students, especially Computer Science and Electronics ones. However, the relations between the availability of Agriculture & Natural Science students and relocations are unexpected, especially looking at these coefficients' values. Computing the LnAgricolture's ratio for this specific sector, the result is 1,0028 so basically this kind of availability does not impact on any trend. The other hazard ratios cannot be computed because at least one of the two parameters is not statistically reliable.

	(1) _t b/se	(2) _t b/se	(3) _t b/se	(4) _t b/se	(5) _t b/se	(6) _t b/se	(7) _t b/se	(8) _t b/se	(9) _t b/se	(10) _t b/se
LlnAvailability	-0.082*** (0.019)									
1.d_NACE_SECTION11	0.109 (0.816)	-0.676 (0.921)	1.014 (0.899)	0.080 (0.734)	0.664 (0.726)	-0.657 (0.948)	0.769 (0.712)	0.285 (0.764)	0.433 (0.797)	-0.227 (0.744)
1.d_NACE_SECTION11#c.LlnAvai~i	0.017 (0.096)									
LlnAgricolture		0.090** (0.037)								
1.d_NACE_SECTION11#c.LlnAgri~r		0.339** (0.164)								
LlnArt			-0.099*** (0.023)							
1.d_NACE_SECTION11#c.LlnArt			-0.048 (0.126)							
LlnBusiness				-0.079*** (0.020)						
1.d_NACE_SECTION11#c.LlnBusi~s				0.023 (0.098)						
LlnEducation					-0.085*** (0.023)					
1.d_NACE_SECTION11#c.LlnEduc~n					-0.045 (0.116)					
LlnEngineering						-0.086*** (0.022)				
1.d_NACE_SECTION11#c.LlnEngi~n						0.133 (0.138)				
LlnHealth							-0.046** (0.022)			
1.d_NACE_SECTION11#c.LlnHealth							-0.007 (0.097)			
LlnInformation								-0.088*** (0.025)		
1.d_NACE_SECTION11#c.LlnInfo~o								-0.004 (0.135)		
LlnNaturalScience									-0.059*** (0.022)	
1.d_NACE_SECTION11#c.LlnNatu~i									0.061 (0.113)	
LlnService										-0.101*** (0.023)
1.d_NACE_SECTION11#c.LlnServ~e										0.086 (0.134)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	106866	49804	85332	103475	91437	97267	83760	96744	85248	88808
N_sub	21874	11274	18773	21457	19754	20415	18534	20277	19006	19642
N_fail	1365	593	1027	1308	1150	1220	1028	1207	1039	1143

Table 5.13 Cox Analysis NACE SECTION11

Unfortunately, the results obtained in this table are not dependable except for the connections between the several availabilities and the relocation decisions, that have been already discussed. The only comment that can be done, it is regarding the moderation effect between LlnAgricolture and the relocation of companies belonging to the NACE SECTION 11 (financial services); indeed, these companies are not interested at all to graduates in Agriculture's studies, that is highly

reasonable. Surprisinlgy, the reliability of the results releted to business area is not sufficient to achieve some conclusions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	_t	_t	_t	_t	_t	_t	_t	_t	_t	_t
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
LlnAvailability	-0.069***									
	(0.020)									
1.d_NACE_SECTION13	0.797*	0.827	1.268**	0.918**	1.077***	0.266	0.784	0.272	0.815	0.268
	(0.447)	(0.578)	(0.512)	(0.404)	(0.412)	(0.460)	(0.491)	(0.415)	(0.502)	(0.381)
1.d_NACE_SECTION13#c.LlnAvai~i	-0.077									
	(0.050)									
LlnAgricolture		0.108***								
		(0.040)								
1.d_NACE_SECTION13#c.LlnAgri~r		-0.015								
		(0.084)								
LlnArt			-0.085***							
			(0.025)							
1.d_NACE_SECTION13#c.LlnArt			-0.103*							
			(0.062)							
LlnBusiness				-0.059***						
				(0.022)						
1.d_NACE_SECTION13#c.LlnBusi~s				-0.107**						
				(0.051)						
LlnEducation					-0.063**					
					(0.025)					
1.d_NACE_SECTION13#c.LlnEduc~n					-0.142**					
					(0.060)					
LlnEngineering						-0.081***				
						(0.024)				
1.d_NACE_SECTION13#c.LlnEngi~n						-0.013				
						(0.060)				
LlnHealth							-0.043*			
							(0.023)			
1.d_NACE_SECTION13#c.LlnHealth							-0.020			
							(0.056)			
LlnInformation								-0.085***		
								(0.027)		
1.d_NACE_SECTION13#c.LlnInfo~o								-0.021		
								(0.065)		
LlnNaturalScience									-0.052**	
									(0.023)	
1.d_NACE_SECTION13#c.LlnNatu~i									-0.034	
									(0.061)	
LlnService										-0.095***
										(0.025)
1.d_NACE_SECTION13#c.LlnServ~e										-0.024
										(0.061)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	106866	49804	85332	103475	91437	97267	83760	96744	85248	88808
N_sub	21874	11274	18773	21457	19754	20415	18534	20277	19006	19642
N_fail	1365	593	1027	1308	1150	1220	1028	1207	1039	1143

Table 5.14 Cox analysis NACE SECTION13

Moving to the last section analysed (professional, scientific & technological services), it has a lower tendency to relocate compared to others. Moreover, data show that this condition is dampened when the availability of Business and

Engineering graduates is significant. These two results are coherent with the expectations.

5.3. Results of the hypothesis 3

Also the third hypothesis is aimed at digging into the results obtained through the testing of the first hypothesis. Indeed, it is reasonable thinking that prestigious universities are more attractive for companies; in the real world there are some evidences that seem to validate this statement, for instance the Silicon Valley. For this hypothesis, the specialization of the graduated students is not anymore taken into account. Indeed, the availability of students is divided between availability of reputable students and availability of less reputable students based on their alma mater. The analysis performed is still the Cox analysis with multiple events. Thus, recalling the hypothesis:

Hypothesis 3: companies that are located in regions where the availability of reputable universities is lower, tend to relocate more than companies close to these institutions.

The analysis performed is still the Cox analysis with multiple events. In the first table, the coefficients are obtained by performing the Cox, but considering the three variables regarding the availability in three different analyses to highlight that all of them reduce the inclination to relocate of the start-ups. However, it is not possible compared coefficient obtained in different analysis, despite the other conditions are the same.

	(1) _t b/se	(2) _t b/se	(3) _t b/se
LlnAvailability	-0.081*** (0.019)		
LlnReputStudents		-0.307*** (0.065)	
LlnLessReputStudents			-0.044** (0.021)
Year Dummies	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes
N	106866	39462	96271
N_sub	21874	7452	19845
N_fail	1365	520	1180

Table 5.15 Cox analysis of the Reputable students

In order to assess if the availability of reputable students affects more, a Cox analysis considering both the variables has been performed. Then, through a Chi2 test is tested the hypothesis.

```
                                              (1)
                                              _t
                                             b/se
LlnReputStudents                          -0.302***
                                           (0.089)
LlnLessReputStudents                      -0.094*
                                           (0.052)
Year Dummies                                 Yes
Industry Dummies                             Yes

N                                           28867
N_sub                                        5160
N_fail                                        335
```

```
. test LlnReputStudents=LlnLessReputStudents

 ( 1)  LlnReputStudents - LlnLessReputStudents = 0

          chi2(  1) =     3.26
        Prob > chi2 =    0.0712
```

Table 5.16 Results Chi2-test Reputable Less Reputable

The results coming from this test are quite evident. Indeed, the availability of reputable students are deterring the relocation of companies much more compared to what is the effect of the availability of less reputable students, therefore the hypothesis is verified. This analysis could be deepened even more by clustering the graduates from reputable universities considering their areas of study, similarly to what it is done for the second hypothesis, as it will be explained in the next chapter.

5.4. Results of the hypothesis 4

The target of this hypothesis is identifying a possible spatial pattern for the relocation phenomenon. At this point, it is possible to affirm that the lower is the availability of graduates and the higher is the probability that relocation occurs. Conversely, by now it is impossible sustaining which could be the destination of these migrations. In order to tackle this issue, the fourth hypothesis is defined:

Hypothesis 4: the availability of universities of the region where the company has relocated is higher than the one characterising the previous location.

This hypothesis can be seen as the prosecution of the first and to be tested it was used the t-test, performed by comparing the lagged values to the "normal" ones. In order to have a reliable result, the analysis is limited to the case in which relocation occurs, that means considering those observations that have the relocatiodummy equals to 1, therefore the availabilities matched are referring to two different NUTS3: the lagged value is referred to the previous NUTS3 while the "normal" value is referred to the new NUTS3. In order to verify the hypothesis, the mean of the availability defined by the "normal" values should be higher than the one indicated by the lagged ones.

```
Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnAvai~y	1,365	8.201091	.0394383	1.457085	8.123725	8.278457
LlnAva~y	1,365	8.169616	.0379569	1.402354	8.095156	8.244077
diff	1,365	.0314748	.0578566	2.137563	-.0820227	.1449723

```
        mean(diff) = mean(lnAvailability - LlnAvailability)            t =    0.5440
H0: mean(diff) = 0                                          Degrees of freedom =      1364

Ha: mean(diff) < 0              Ha: mean(diff) != 0              Ha: mean(diff) > 0
Pr(T < t) = 0.7067          Pr(|T| > |t|) = 0.5865          Pr(T > t) = 0.2933
```

Table 5.17 Paired t-test LlnAvailability-lnAvailability

The result of this paired t test is not in line with the expectation, indeed the difference of means is not so significant (despite it is a logarithm), and, moreover, the p-value is 0.5865, that represents a value that cannot be defined as reliable.

5.5. Results of the hypothesis 5

The hypothesis number five is directly connected to the second hypothesis taking up the distinction between the areas of studies previously analysed. The reasoning underpinning this fifth hypothesis is the same of the fourth, that means understanding the possible target of the migration, in this case, by considering the characteristics of the companies (NACE SECTION) and the availability of the different typology of graduates. In order to considering only the observations where the relocationdummy==1 for the companies that belong to a specific NACE SECTION, all the observations with relocationdummy==0 are dropped, and then it is set the condition to belong to a specific NACE SECTION. The fifth hypothesis is:

Hypothesis 5: the availability of specialized universities, in the area of interest of the company, of the region where the company has relocated is higher than the one characterising the previous location.

The first NACE SECTION to be analysed is the manufacturing.

```
Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnAgri~e	25	4.737468	.2542511	1.271256	4.212719	5.262216
LlnAgr~e	25	5.30891	.2243272	1.121636	4.845921	5.771898
diff	25	-.5714422	.2814974	1.407487	-1.152424	.0095398

```
     mean(diff) = mean(lnAgricolture - LlnAgricolture)          t =  -2.0300
H0: mean(diff) = 0                                 Degrees of freedom =       24

Ha: mean(diff) < 0            Ha: mean(diff) != 0              Ha: mean(diff) > 0
Pr(T < t) = 0.0268        Pr(|T| > |t|) = 0.0536            Pr(T > t) = 0.9732
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnArt	64	6.232828	.1655129	1.324103	5.902077	6.563579
LlnArt	64	6.193069	.1524674	1.219739	5.888387	6.49775
diff	64	.0397595	.1871553	1.497243	-.3342407	.4137596

```
     mean(diff) = mean(lnArt - LlnArt)                         t =   0.2124
H0: mean(diff) = 0                                 Degrees of freedom =       63

Ha: mean(diff) < 0            Ha: mean(diff) != 0              Ha: mean(diff) > 0
Pr(T < t) = 0.5838        Pr(|T| > |t|) = 0.8324            Pr(T > t) = 0.4162
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnBusi~s	108	6.700829	.1334042	1.386377	6.43637	6.965287
LlnBus~s	108	6.740067	.1254732	1.303956	6.491331	6.988803
diff	108	-.039238	.1769743	1.839171	-.390069	.3115929

```
     mean(diff) = mean(lnBusiness - LlnBusiness)               t =  -0.2217
H0: mean(diff) = 0                                 Degrees of freedom =      107

Ha: mean(diff) < 0            Ha: mean(diff) != 0              Ha: mean(diff) > 0
Pr(T < t) = 0.4125        Pr(|T| > |t|) = 0.8250            Pr(T > t) = 0.5875
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnEduc~n	81	5.92952	.1338968	1.205072	5.663057	6.195983
LlnEdu~n	81	5.533552	.1447133	1.30242	5.245563	5.821541
diff	81	.3959682	.1889916	1.700924	.019863	.7720734

```
     mean(diff) = mean(lnEducation - LlnEducation)             t =   2.0952
H0: mean(diff) = 0                                 Degrees of freedom =       80

Ha: mean(diff) < 0            Ha: mean(diff) != 0              Ha: mean(diff) > 0
Pr(T < t) = 0.9803        Pr(|T| > |t|) = 0.0393            Pr(T > t) = 0.0197
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnEngi~g	89	6.679379	.1197598	1.129811	6.441381	6.917376
LlnEng~g	89	6.631892	.1331067	1.255726	6.36737	6.896414
diff	89	.0474867	.1975496	1.863679	-.3451015	.440075

```
    mean(diff) = mean(lnEngineering - LlnEngineering)        t =    0.2404
H0: mean(diff) = 0                               Degrees of freedom =        88

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.5947         Pr(|T| > |t|) = 0.8106         Pr(T > t) = 0.4053
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnHealth	63	6.292401	.1695288	1.345593	5.953518	6.631284
LlnHea~h	63	6.17341	.1683561	1.336285	5.836871	6.509949
diff	63	.1189912	.2158505	1.71326	-.3124878	.5504701

```
    mean(diff) = mean(lnHealth - LlnHealth)                  t =    0.5513
H0: mean(diff) = 0                               Degrees of freedom =        62

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.7083         Pr(|T| > |t|) = 0.5834         Pr(T > t) = 0.2917
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnInfo~n	87	5.437231	.1293178	1.206196	5.180156	5.694306
LlnInf~n	87	5.576273	.115466	1.076996	5.346734	5.805812
diff	87	-.1390421	.1971471	1.838866	-.5309576	.2528734

```
    mean(diff) = mean(lnInformation - LlnInformation)        t =   -0.7053
H0: mean(diff) = 0                               Degrees of freedom =        86

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.2413         Pr(|T| > |t|) = 0.4825         Pr(T > t) = 0.7587
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
lnNatu~e	63	5.899248	.1910446	1.51637	5.517355	6.28114
LlnNat~e	63	6.404088	.1413394	1.121847	6.121554	6.686621
diff	63	-.5048402	.2471656	1.961816	-.9989171	-.0107632

```
    mean(diff) = mean(lnNaturalScience - LlnNaturalScie~e)   t =   -2.0425
H0: mean(diff) = 0                               Degrees of freedom =        62

Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
Pr(T < t) = 0.0227         Pr(|T| > |t|) = 0.0454         Pr(T > t) = 0.9773
```

```
Variable    |     Obs        Mean     Std. err.    Std. dev.    [95% conf. interval]
------------+-----------------------------------------------------------------------
lnServ~e    |      73     4.982551    .1667707      1.42489     4.650099    5.315002
L1nSer~e    |      73     5.091965    .1528838      1.30624     4.787196    5.396733
------------+-----------------------------------------------------------------------
diff        |      73    -.1094139     .23474      2.005619    -.5773595    .3585317
-----------------------------------------------------------------------------------
    mean(diff) = mean(lnService - L1nService)                    t =   -0.4661
H0: mean(diff) = 0                                Degrees of freedom =        72

Ha: mean(diff) < 0               Ha: mean(diff) != 0              Ha: mean(diff) > 0
Pr(T < t) = 0.3213           Pr(|T| > |t|) = 0.6425           Pr(T > t) = 0.6787
```

Table 5.18 Paired t-test regarding the NACE SECTION 3

These ttests certify that the availability of agriculture's graduates is not a determinant for manufacturing firms, indeed after the relocation their availability is reduced. Similar conclusion can be ascribed to the availability of "NaturalScience" graduates. Instead, considering the availability of "Education" graduates it is possible to notice that the availability increases after the relocation. The data are quite unexpected, probably because manufacturing is group too broad of firms. However, the data have been discussed have always a statistical reliability. It is worth of mention, however, that observations, in each case, are not so many.

For seek of simplicity, for the following sections will be shown only those ttests that have a statistical significance (p-value>0.1). Unfortunately, considering the remaining NACE SECTIONs there are not results that have a statistical importance apart from the following one that is related to the NACE SECTION 13 (professional, scientific & technological services). This is probably due to the fact that the observations are not enough to perform the ttests in a proper way.

```
Paired t test

Variable    |     Obs        Mean     Std. err.    Std. dev.    [95% conf. interval]
------------+-----------------------------------------------------------------------
lnAgri~e    |      31     4.699459     .256428     1.427731     4.175763    5.223155
L1nAgr~e    |      31     4.212767    .2471039     1.375816     3.708113    4.717421
------------+-----------------------------------------------------------------------
diff        |      31     .4866919    .2685199     1.495055    -.0616989    1.035083
-----------------------------------------------------------------------------------
    mean(diff) = mean(lnAgricolture - L1nAgricolture)            t =    1.8125
H0: mean(diff) = 0                                Degrees of freedom =        30

Ha: mean(diff) < 0               Ha: mean(diff) != 0              Ha: mean(diff) > 0
Pr(T < t) = 0.9600           Pr(|T| > |t|) = 0.0799           Pr(T > t) = 0.0400
```

Table 5.19 Paired t-test regarding the NACE SECTION 13

5.6. Results of the hypothesis 6

The sixth hypothesis is tested through the last ttest performed. In that case the target is to assess if reputable universities can be seen as destination for companies when they relocate. So, the relation between this hypothesis and the third one is very tight. The idea behind this hypothesis is that reputable universities might be seen as location where it is easier performed innovation, and so targeted as destination. To assess this, it is checked if there is an increase of availability of reputable students. Indeed, similarly, to the third hypothesis, here it is not considered the differences between the areas where students are graduated, but there is just the same division between reputable and "lessreputable" graduates. The sixth hypothesis is:

Hypothesis 6: the availability of reputable universities of the region where the company has relocated is higher than the one characterising the previous locations.

In order to perform this ttest, it is put into relation the lagged availabilities of reputable students with the "normal" ones, and likely to what it is done for the hypothesis fourth it is added the condition "if relocationdummy==1" in order to limit the analysis to those observations where there is a change of NUTS.

```
Paired t test

Variable        Obs         Mean     Std. err.    Std. dev.    [95% conf. interval]

lnRepu~s        133     8.554251     .0464531     .5357231     8.462362      8.64614
LlnRep~s        133     8.276091     .0493364     .5689748     8.178499     8.373683

diff            133     .2781603     .0633212     .7302555     .1529047     .4034158

     mean(diff) = mean(lnReputStudents - LlnReputStudents)           t =    4.3928
H0: mean(diff) = 0                                       Degrees of freedom =      132

Ha: mean(diff) < 0             Ha: mean(diff) != 0                Ha: mean(diff) > 0
Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000              Pr(T > t) = 0.0000
```

Table 5.20 Paired t-test LlnReputable-lnReputable

Despite the number of observations is not incredibily high, the p-value of the result is 0, therefore it is remarkable important. The result confirms the theory, indeed it can be observed that the availability (its mean) of reputable students grows after the occurrence of the relocation.

5.7.   Results of the hypothesis 7

This is hypothesis is different from the previous one because it introduces another dimension that is the possibility, or not, for a firm to be backed by a Venture Capital. The literature has started to discuss about the impact of VC on relocations' decisions, as explained in the literature review, and this hypothesis try to provide a contribution by assessing if a VC-backed start-up is more or less responsive regarding the availability of students in terms of tendency to relocate. In this hypothesis, there is not introduced any type of classification of graduates. To test the hypothesis is used again the Cox analysis. The seventh and last hypothesis tested is:

Hypothesis 7: VC-backed firms have a weaker inclination in relocating towards universities compared to non-VC-backed firms.

To test this hypothesis is used the VC_Step that is described in the previous chapter. This analysis is performed considering the lnAvailability rather than the lagged value because in this case is not so significant controlling which is the condition before the VCs' entrance but assess if, once backed by a VC, the companies have a different inclination to relocate.

	(1) _t b/se
lnAvailability	-0.061***
	(0.014)
1.VC_Step	1.946***
	(0.428)
1.VC_Step#c.lnAvailability	-0.216***
	(0.051)
Year Dummies	Yes
Industry Dummies	Yes
N	133296
N_sub	23517
N_fail	2634

Table 5.21 Cox analysis VC-backed firms

The result disputes hypothesis defined previously, indeed being backed by a VC increases the tendency to relocate, that is just partially dampened by the availability of students. The result is highly dependable

6   Conclusions

This chapter is aimed at drawing conclusions regarding the topic discussed in this thesis. It is structured in four parts: contribution to the existing literature, limitations of the analysis, avenue for future research and the conclusion.

6.1.   Contribution to the existing literature

The topic that has been tackled in this thesis represents a novelty in the existing literature, indeed it wants clarifying the relationship between the distribution of universities across some European Countries, and the relocation events. The latter, it is a phenomenon that has many implications in terms of demography, economy etc, (van Dijk and Pellenbarg, 2000) and for that reason it was decided to investigate into it. Despite this "importance", the literature regarding the migration phenomenon is poor, especially the European one if compared with the American one. Moreover, as underlined in the Literature Review, most of the studies and the theories are about the location decisions rather than the relocation ones, and these two themes have important differences therefore the determinants are not the same or affect with a different intensity the two choices, (De Bok and Sanders, 2005), (Manjón-Antolín and Arauzo-Carod, 2011). The existing literature, recently, has started to study the impact of VC availability on the relocation events, (Colombo et al., 2019), (De Prijcker et al., 2019), but the results of the analyses are still very ambiguous and, numerically, scares. However, there is not a clear connection between the two elements that might really suggest that the VCs' availability can be considered as a determinant in the relocation decision process. These results encourage researchers to test if there are other possible determinants that describe the relocation phenomenon in a more appropriate way. This is why it is decided to analysed universities, and specifically, the availability of graduated students as a potential relocation determinant. The topic regarding the connection between universities and the entrepreneurial world, however, is not so common in the academic literature and is mainly focused on the analysis of how universities can contribute to innovation and the related potential spillovers (Audretsch et al., 2005), (Etzkowitz, 2003) (Etzkowitz, 2013). That is not the only one that allows to define this work quite peculiar compared to the other studies. First of all, this thesis is not country-specific, but it considers 6 countries: Czech Republic, Denmark, France, Germany, Netherlands and Portugal. Moreover, the previous studies regarding the relocation mainly focus on central and northern European countries (German,

Netherland, Denmark, Sweden in particular), while the Latin countries are basically excluded.

Before talking about specifically of the conclusions obtained from the results of each hypothesis, it is important recall what has been considered relocation in that thesis. Relocation is every migration from one place to another that involves a change in the NUTS3 (Nomenclature of territorial units for statistics), where a NUTS3 is the lowest level of the NUTS hierarchy and represents, approximately, the area of a province. The decision to neglect relocation intra-province is based on two reasons: the first one is the difficulty in getting the data useful for evaluating the hypothesis, indeed even EUROSTAT and ETER provides data at most at NUTS3 level. The second reason is the fact that relocation intra-province can be caused by several reasons such as the needs of broader spaces, and it would be difficult identify any kind of spatial pattern considering even this kind of relocation. So, this has had, for sure, an impact on the results shown in the previous chapter.

Another important comment is regarding the decision to translate the availability of universities, defined in the hypotheses, into the availability of graduates. First of all, it is decided to consider the graduated students rather than the enrolled students because the latter would also entail students that would have not completed their educational path. Then, it is considered the students rather than the number of universities per NUTS3 in order to take into account even the size of the universities. Moreover, it was much easier clustering the students considering their areas of activity compared to what would take clustering universities per type, in case it would be possible. The potential drawback of this choice is the possibility that students, after the graduation, move away. According to the literature, most students tend to remain and, in case, to start their activities in the area where they have graduated, (Larsson et al., 2017), therefore accounting the graduates was assessed as the best solution for this thesis.

Moving to the discussion of the first hypothesis, that is designed in order to check if the availability of graduated students deter, or not, relocations, can be confirmed by the results obtained from the analysis. The result keeps its significance even adding covariates of "Year" and "NACE_SECTION". The results of this hypothesis cannot be compared to others results in a classical way, because, as said, scholars have not analysed those relations. However, in this chapter will be presented possible matches between the results themselves and the literature. In the chapter 2, it was introduced an interesting classification (van Dijk and Pellenbarg, 2000), that affirms that factors affecting the relocation can be grouped in three categories: push factor, pull factor and keep factor. In their classification, graduates' availability is not considered as a factor that influences the relocation, however these results

suggest conversely. In particular, keep factors are defined as those reasons that convince the decision-maker to remain in the same place. According to the authors, they are costs (fixed and variable), however by looking at the statistics of this study it is possible to affirm that even the availability of students is a keep factor. Moreover, considering another classification of factors, introduced by (Lloyd, Dicken, 1977), the graduates' availability can be seen as a locational factor that defines the characteristics of the current location. Talking about the control variables used to test this hypothesis, the results, in terms hazard ratio and or coefficients, are not satisfying because they do not add other information except for the fact that increases the amount of Total Assets reduces the probability of relocation, that is coherent with what sustained by some scholars, (PELLENBARG and KOK, 1985). This is reasonable because a larger amount of fixed costs to be moved should imply higher migration costs. Unfortunately, the measure does not have the sufficient significancy and therefore this reasonable result cannot be taken. To conclude the discussion about the first hypothesis, the expectations regarding the control variables were that EBITDA would have positively affected the tendency of relocating because profitable companies could sustain the relocation costs, and same reasoning was ascribed for the ROA variable; considering the GDP, it would have negatively affected the firms' inclination because firms prefer to be located in metropolitan areas where the GDP is, usually, higher than rural areas. The GDPgrowth, instead, non-necessarily is linked to metropolitan areas and therefore the expectation was more "neutral." The last control variable is regarding the number of applications for the community design rights (intellectual property right), that would have had a negatively coefficients, similarly to what occurred for the availability of students. Indeed, as said even in this chapter, universities contribute significantly to the innovation of a specific region, and the output of this should be the application for an intellectual property, therefore the higher is the availability of universities and the higher should be the number of applications.

Talking about the second hypothesis, the results look a bit ambiguous and somehow contradictory. It is aimed at identifying possible areas of studies where the connections with the relocation events should have been stronger. The literature underpinning this is, for instance, the article published by D'Este and Iammarino, in 2010, where they sustained that engineering departments suffer of a local bias, in starting collaboration with firms, compared to science departments for instance. This would have been translated into a higher tendency to relocate towards engineering graduated students than "NaturalScience's" graduates, especially for those companies' operating in similar fields (manufacturing, professional scientific and technological services for instance). This result is confirmed looking at the impact of the availability of the single typology of students on the relocation in general (table 6.6), but not considering the manufacturing companies and those

related to the technological services that do not show a preference towards engineering departments. Each NACE_SECTION has some unexpected data or unreliable, meaning that probably it is needed to enlarge the sample analysed and to downsize the width of the sections considered in order to find some dependable spatial pattern.

The third hypothesis is developed based on the theory regarding the spillovers generated by the universities, (Audretsch et al., 2005), and the fact that the most important technological and successful hubs raised next to the most innovative universities (Silicon Valley and Stanford, for instance), (Smith and Ho, 2006). That evidence should entail not only that reputable universities are targeted in the location decisions, but also later in the relocation decision in case in a previous moment was not possible to set there. To be more precised these universities should attract younger (and innovative) start-ups because the latter depend more on the availability of external knowledge since they cannot sustain significant R&D costs (Audretsch et al., 2005). However, in this hypothesis is tested if the presence of reputable universities decreases the probability of relocation, and the result of the analysis confirms it. It is important to highlight that reputable universities are not concentrated in few cities (and thus few NUTS3), but they are distributed in the different regions across the countries taken into account. Only France and Czech Republic present a very high concentration around their capital, but overall, the distribution is evaluated satisfactory. This is an important data that, conversely, might negatively affect the results of the studies conducted about the VCs' availability. Indeed, VCs' hubs are extremely concentrated, and it might be difficult understanding if a specific result is caused by the presence of the VC's hub, or by the specific characteristics of the NUTS3.

Considering the fourth hypothesis, it affirms that area with a higher availability of graduates would be preferred compared to other regions during a relocation. The analysis performed is a ttest that seem to reject the assumption sustains by the hypothesis; actually, there is not statistical significancy, but this occurred since the two mean values, compared through the ttests, are too close. However, this condition is not in line with the expectation, because the availability of graduates after the relocation should be clearly higher rather than similar to before. Arguably, the availability of universities, and the related benefits in terms of innovation spillover etc, is a parameter that deters relocating elsewhere but is not a parameter that seems to "drive" relocation; in other words, companies do not relocate to have a larger availability of students, therefore it is not neither push nor pull factor, using the terminology introduced by van Dijk and Pellenbarg.

The fifth hypothesis is, probably, the most difficult to be discussed. The large number of results that are not reliable suggests that the analysis performed is not the proper one. As said in the previous chapter, dividing the sample considering the different NACE_SECTION implies that the number of observations has been reduced significantly and, statistically, it is not a good condition. Somehow, the results are coherent with those regarding the second hypothesis, meaning that there is not clear evidence that some areas of studies are more interesting for companies than others.

Instead, the hypothesis number five is satisfied completely, similarly to what occurred for the first and the third hypothesis. Differently to what it has been said for the availability of graduates, it seems that companies are highly interested in high reputable students, meaning that they can be seen as pull factor of relocation. The reasoning done to corroborate the results of the third hypothesis related to the distribution of reputable universities, can be perfectly ascribed even to this hypothesis. Considering for example the German case, there are 47 universities that belong to the top 1000 and 29 that are in the top 500. Munich and Berlin, two of the three largest cities in Germany are the only to have more than one reputable university in their NUTS3, respectively 2 and 3. The remaining universities are distributed across the countries; very similar conclusion can be done for Denmark and Netherlands, while Portugal has only 4 reputable universities, 2 of them are located in Lisbon and then one in Coimbra and Porto.

The last hypothesis is the one related to the VC-backed condition. The reason why the hypothesis affirms that being backed by a VC increases the "geographical stability" of a company is, firstly, due to the presence of a VCs' local bias, (Cumming and Dai, 2010); this means that in the selection of star-ups, those that have been chosen, usually, are closer to VCs' hubs and, therefore, already located within the metropolitan areas; as said before, it is unlikely that companies go away from there, in particular being backed by VC requires a direct contact among the two actors involved and therefore is unreasonable thinking about relocation for a VC-backed start-ups that would lead to an increase of the monitoring costs (Cumming and Dai, 2010). Moreover, the fact that VCs do not provide only economical support but also coaching & networking services, determines a lower degree of dependency on external sources, such as the universities' knowledge and R&D, for a VC-backed firm. Summing up all these elements, it is reasonable to make the seventh hypothesis, that however, is rejected.

6.2. Limitations of the analysis

There are some choices made during the elaboration of the thesis that can be seen as limitation. They are referred to the data collection and data exploitation. First of all, there is not a unique source of data but, as said, there are three databases that have been used: ETER, EUROSTAT and VICO 5.0. These databases do not cover the same geographically regions during the same interval of time, therefore many observations have been dropped. For instance, ETER do not provide information regarding French students in the 2015, 2016, 2017 and 2018, while VICO and EUROSTAT (partially) do. In these cases, many useful observations are lost because it is decided to keep only those that present a comprehensive set of data. Further limitation is caused by the fact that on ETER not all the universities provide information regarding the number of graduates per area of study, and therefore the database used does not represent the real availability of students. Moreover, as said in the chapter 4, ETER does not consider universities that are smaller than a certain threshold (200 students enrolled) except for entities of national interest. Similarly, even VICO's database does not perfectly match the real distribution of European firms because it is aimed at monitoring the VC-backed firms and the latter tend to be innovative (condition to be picked by VCs); moreover, the control sample of VICO is, by definition, composed by companies (non VC-backed) that are similar to the former and therefore some groups might be underrepresented while other groups the opposite. In addition, even the data retrieved from EUROSTAT sometimes are not available for the entire length of the temporal interval analysed. Finally, it is worth of mention the fact that the data referred to the "lessreputablestudents" are slightly undervalued because the percentage of reputable universities that provide information regarding the stream of studies of their graduates is higher. This implies that, basically, all reputable students are considered in the analyses whilst not all "lessreputablestudents" are.

6.3. Avenues for future research

This third part of this chapter is dedicated to the areas that have not been analysed or for lack of data, or because the idea is originated from the analysis of the results. The lack of availability of data regarding the patents' application, per companies, has prevented to tackle the eighth hypothesis. As said during its presentation in the third chapter, it was designed to be an additional verification of the previous hypothesis. Indeed, it wants to control if, after the relocation, companies would have a higher innovation rate thanks to the exploitation of universities spillovers

and in general universities collaborations. By now, relocation events have been investigated only in a quantitative way, but it would be interesting understanding even the qualitative part. Hopefully, in the future these data will be available.

The other two hints for the future are coming from the analysis of the data obtained during the tests. First of all, the data regarding the importance of the reputation in the relocation decision are encouraging, therefore it is worth of a possibility analysing the singular areas of studies just for the reputable universities; secondly, in these analyses the age of the firms, when they relocate, has not been considered. However, the academic literature debates widely regarding this topic and it would be nice clarify this issue.

6.4. Conclusions

This work provides a peculiar contribution within the academic discussion regarding the relocation of start-ups, by putting into relation the firms' migration with the availability of universities (graduated students). The analysis involves 6 European Countries for an interval of time from 2011 to 2019. Some of the countries considered have been ignored by the previous literature that focuses only few countries, usually the national country of the scholar. The main results achieved are the evidence that the availability of graduates deters relocation, and this effect is emphasised when the university considered is reputable. Instead, the availability of students in those areas afferent to the firms' business model does not seem having an impact, but the results are too ambiguous to assess this surely. Probably, the tests done were not the most proper, the groups analysed are too broad to identify some pattern and the observations are not numerically sufficient. Lastly, being backed by VC, increases the tendency to relocate, but still the availability of students deter relocation. To be precise, the deter-effect is even more intense, but the inclination to relocate remains higher for VC-backed firms compared to non-VC-backed even in case of high availability of graduates.

Bibliogaphy

Aernoudt, R., 2004. Incubators: Tool for Entrepreneurship? Small Bus. Econ. 23, 127–135. https://doi.org/10.1023/B:SBEJ.0000027665.54173.23

Arauzo Carod, J.M., Manjón Antolín, M.C., 2004. Firm size and geographical aggregation: An empirical appraisal in industrial location. Small Bus. Econ. 22, 299–312. https://doi.org/10.1023/b:sbej.0000022216.09083.76

Arthur, W.B., 1990. "Silicon Valley" locational clusters: when do increasing returns imply monopoly? Math. Soc. Sci. 19, 235–251. https://doi.org/10.1016/0165-4896(90)90064-E

Audretsch, D.B., Dohse, D., Niebuhr, A., 2015a. Regional unemployment structure and new firm formation. Pap. Reg. Sci. 94, S115–S138. https://doi.org/10.1111/pirs.12169

Audretsch, D.B., Heger, D., Veith, T., 2015b. Infrastructure and entrepreneurship. Small Bus. Econ. 44, 219–230. https://doi.org/10.1007/s11187-014-9600-6

Audretsch, D.B., Keilbach, M., 2007. The Theory of Knowledge Spillover Entrepreneurship*. J. Manag. Stud. 44, 1242–1254. https://doi.org/10.1111/j.1467-6486.2007.00722.x

Audretsch, D.B., Lehmann, E.E., Warning, S., 2005. University spillovers and new firm location. Res. Policy, The Creation of Spin-off Firms at Public Research Institutions: Managerial and Policy Implcations 34, 1113–1122. https://doi.org/10.1016/j.respol.2005.05.009

Autio, E., Kenney, M., Mustar, P., Siegel, D., Wright, M., 2014. Entrepreneurial innovation: The importance of context. Res. Policy 43, 1097–1108. https://doi.org/10.1016/j.respol.2014.01.015

Bergek, A., Norrman, C., 2008. Incubator best practice: A framework. Technovation 28, 20–28. https://doi.org/10.1016/j.technovation.2007.07.008

Berger, S., 2013. Making in America: From innovation to market. Mit Press.

Boschma, R.A., Frenken, K., 2006. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. J. Econ. Geogr. 6, 273–302. https://doi.org/10.1093/jeg/lbi022

Brouwer, A.E., Mariotti, I., van Ommeren, J.N., 2004. The firm relocation decision: An empirical investigation. Ann. Reg. Sci. 38, 335–347. https://doi.org/10.1007/s00168-004-0198-5

Brown, R., Mawson, S., Lee, N., Peterson, L., 2019. Start-up factories, transnational entrepreneurs and entrepreneurial ecosystems: unpacking the lure of start-up accelerator programmes. Eur. Plan. Stud. 27, 885–904. https://doi.org/10.1080/09654313.2019.1588858

Carayannis, E.G., 2013. The innovation ecosystem, in: The Innovation Union in Europe: A Socio-Economic Perspective on EU Integration. pp. 3–41. https://doi.org/10.4337/9780857939913.00007

Carayannis, E.G., Campbell, D.F.J., 2009. "Mode 3" and "Quadruple Helix": Toward a 21st century fractal innovation ecosystem. Int. J. Technol. Manag. 46, 201–234. https://doi.org/10.1504/ijtm.2009.023374

Cavallo, A., Ghezzi, A., Balocco, R., 2019. Entrepreneurial ecosystem research: present debates and future directions. Int. Entrep. Manag. J. 15, 1291–1321. https://doi.org/10.1007/s11365-018-0526-3

Chow, T., Huang, S., Klassen, K.J., Ng, J., 2022. The Influence of Corporate Income Taxes on Investment Location: Evidence from Corporate Headquarters Relocations. Manag. Sci. 68, 1404–1425. https://doi.org/10.1287/mnsc.2020.3906

Cloutier, L., Messeghem, K., 2022. Whirlwind model of entrepreneurial ecosystem path dependence. Small Bus. Econ. 59, 611–625. https://doi.org/10.1007/s11187-021-00553-x

Colombo, D.G. e., Garcia, R. de C., 2021. The role of the academic relations of former graduate students in university-firm collaboration. J. Technol. Transf. https://doi.org/10.1007/s10961-021-09881-2

Colombo, M.G., D'Adda, D., Quas, A., 2019. The geography of venture capital and entrepreneurial ventures' demand for external equity. Res. Policy 48, 1150–1170. https://doi.org/10.1016/j.respol.2018.12.004

Colombo, M.G., Grilli, L., 2007. Funding Gaps? Access To Bank Loans By High-Tech Start-Ups. Small Bus. Econ. 29, 25–46. https://doi.org/10.1007/s11187-005-4067-0

Colombo, M.G., Guerini, M., Rossi-Lamastra, C., Bonaccorsi, A., 2022. The "first match" between high-tech entrepreneurial ventures and universities: the role of founders' social ties. J. Technol. Transf. 47, 270–306. https://doi.org/10.1007/s10961-021-09848-3

Cumming, D., Dai, N., 2012. The Role of Geographic Proximity in Venture Capital, in: The Oxford Handbook of Venture Capital. https://doi.org/10.1093/oxfordhb/9780195391596.013.0031

Cumming, D., Dai, N., 2010. Local bias in venture capital investments. J. Empir. Finance 17, 362–380. https://doi.org/10.1016/j.jempfin.2009.11.001

De Bok, M., Sanders, F., 2005. Firm relocation and accessibility of locations: Empirical results from the Netherlands. Transp. Res. Rec. 35–43. https://doi.org/10.3141/1902-05

De Prijcker, S., Manigart, S., Collewaert, V., Vanacker, T., 2019. Relocation to Get Venture Capital: A Resource Dependence Perspective. Entrep. Theory Pract. 43, 697–724. https://doi.org/10.1177/1042258717739003

D'Este, P., Iammarino, S., 2010. The spatial profile of university-business research partnerships. Pap. Reg. Sci. 89, 335–350. https://doi.org/10.1111/j.1435-5957.2010.00292.x

Devereux, M.P., Griffith, R., 1998. Taxes and the location of production: evidence from a panel of US multinationals. J. Public Econ. 68, 335–367. https://doi.org/10.1016/S0047-2727(98)00014-0

Diaz, R., Garrido, N., Vargas, M., 2021. Segregation of high-skilled workers and the productivity of cities. Reg. Sci. Policy Pract. 13, 1460–1478. https://doi.org/10.1111/rsp3.12355

Djankov, S., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2002. The Regulation of Entry*. Q. J. Econ. 117, 1–37. https://doi.org/10.1162/003355302753399436

Dohse, D., 2000. Technology policy and the regions — the case of the BioRegio contest. Res. Policy 29, 1111–1133. https://doi.org/10.1016/S0048-7333(99)00077-3

Drejer, I., Østergaard, C.R., 2017. Exploring determinants of firms' collaboration with specific universities: employee-driven relations and geographical proximity. Reg. Stud. 51, 1192–1205. https://doi.org/10.1080/00343404.2017.1281389

Duranton, G., Puga, D., 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. Am. Econ. Rev. 91, 1454–1477. https://doi.org/10.1257/aer.91.5.1454

Etzkowitz, H., 2013. Anatomy of the entrepreneurial university. Soc. Sci. Inf. 52, 486–511. https://doi.org/10.1177/0539018413485832

Etzkowitz, H., 2003. Innovation in innovation: The Triple Helix of university-industry-government relations. Soc. Sci. Inf. 42, 293–337. https://doi.org/10.1177/05390184030423002

Etzkowitz, H., Leydesdorff, L., 2000. The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Res. Policy 29, 109–123. https://doi.org/10.1016/S0048-7333(99)00055-4

Fischer, B., Meissner, D., Vonortas, N., Guerrero, M., 2022. Spatial features of entrepreneurial ecosystems. J. Bus. Res. 147, 27–36. https://doi.org/10.1016/j.jbusres.2022.04.018

Florida, R., Mellander, C., Holgersson, T., 2015. Up in the air: the role of airports for regional economic development. Ann. Reg. Sci. 54, 197–214. https://doi.org/10.1007/s00168-014-0651-z

Frenkel, A., 2001. Why high-technology firms choose to locate in or near metropolitan areas. Urban Stud. 38, 1083–1101. https://doi.org/10.1080/00420980120051666

Hassanain, M.A., Ibrahim, A.M., 2020. A framework for modeling the knowledge of workplace relocation. J. Corp. Real Estate 23, 20–39. https://doi.org/10.1108/JCRE-01-2020-0001

Hayter, R., 1997. The dynamics of industrial location: the factory, the firm and the production system, The dynamics of industrial location: the factory, the firm and the production system.

Heblich, S., Slavtchev, V., 2014. Parent universities and the location of academic startups. Small Bus. Econ. 42, 1–15. https://doi.org/10.1007/s11187-013-9470-3

Holl, A., 2004. Start-ups and relocations: Manufacturing plant location in Portugal. Pap. Reg. Sci. 83, 649–668. https://doi.org/10.1007/s10110-004-0218-y

Hornuf, L., Schmitt, M., Stenzhorn, E., 2022. The local bias in equity crowdfunding: Behavioral anomaly or rational preference? J. Econ. Manag. Strategy 31, 693–733. https://doi.org/10.1111/jems.12475

Hsu, D.H., 2004. What Do Entrepreneurs Pay for Venture Capital Affiliation? J. Finance 59, 1805–1844. https://doi.org/10.1111/j.1540-6261.2004.00680.x

Huang, L., Pearce, J.L., 2015. Managing the Unknowable: The Effectiveness of Early-stage Investor Gut Feel in Entrepreneurial Investment Decisions. Adm. Sci. Q. 60, 634–670. https://doi.org/10.1177/0001839215597270

Kenney, M., Breznitz, D., Murphree, M., 2013. Coming back home after the sun rises: Returnee entrepreneurs and growth of high tech industries. Res. Policy 42, 391–407. https://doi.org/10.1016/j.respol.2012.08.001

Kolympiris, C., Kalaitzandonakes, N., Miller, D., 2015. Location choice of academic entrepreneurs: Evidence from the US biotechnology industry. J. Bus. Ventur. 30, 227–254. https://doi.org/10.1016/j.jbusvent.2014.02.002

Kronenberg, K., 2013. Firm relocations in the Netherlands: Why do firms move, and where do they go? Pap. Reg. Sci. 92, 691–713. https://doi.org/10.1111/j.1435-5957.2012.00443.x

Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Econ. 99, 483–499. https://doi.org/10.1086/261763

Larsson, G., n.d. Accelerating Success: A Study of Seed Accelerators and Their Defining Characteristics [WWW Document]. URL https://www.acceleratorstudy.com/ (accessed 9.17.22).

Larsson, J.P., Wennberg, K., Wiklund, J., Wright, M., 2017. Location choices of graduate entrepreneurs. Res. Policy 46, 1490–1504. https://doi.org/10.1016/j.respol.2017.07.004

Lee, I.H. (Ian), 2022. Startups, relocation, and firm performance: a transaction cost economics perspective. Small Bus. Econ. 58, 205–224. https://doi.org/10.1007/s11187-020-00406-z

LOCATION IN SPACE: A THEORETICAL APPROACH TO ECONOMIC GEOGRAPHY, 1977. . Tijdschr. Voor Econ. En Soc. Geogr. 68, 58–59. https://doi.org/10.1111/j.1467-9663.1977.tb01468.x

Manjón-Antolín, M.C., Arauzo-Carod, J.-M., 2011. Locations and relocations: determinants, modelling, and interrelations. Ann. Reg. Sci. 47, 131–146. https://doi.org/10.1007/s00168-009-0360-1

Mason, C.M., Harrison, R.T., 2002. The geography of venture capital investments in the UK. Trans. Inst. Br. Geogr. 27, 427–451. https://doi.org/10.1111/1475-5661.00064

Michelacci, C., Silva, O., 2007. Why So Many Local Entrepreneurs? Rev. Econ. Stat. 89, 615–633. https://doi.org/10.1162/rest.89.4.615

Musolino, D., Mariotti, I., Brouwer, A.E., 2020. Stated and revealed locational preferences of firms: a national scale perspective. Ann. Reg. Sci. 64, 493–499. https://doi.org/10.1007/s00168-020-00982-8

Myers, S.C., Majluf, N.S., 1984. Corporate financing and investment decisions when firms have information that investors do not have. J. Financ. Econ. 13, 187–221. https://doi.org/10.1016/0304-405X(84)90023-0

Myles Shaver, J., Flyer, F., 2000. Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. Strateg. Manag. J. 21, 1175–1193. https://doi.org/10.1002/1097-0266(200012)21:12<1175::AID-SMJ139>3.0.CO;2-Q

Neal, Z.P., 2011. The causal relationship between employment and business networks in U.S. cities. J. Urban Aff. 33, 167–184. https://doi.org/10.1111/j.1467-9906.2010.00538.x

Nilsen, Ø.L., Díez-Gutiérrez, M., Andersen, S.N., Tørset, T., Cherchi, E., 2020. Where and why do firms choose to move? Empirical evidence from Norway. J. Transp. Land Use 13, 207–225. https://doi.org/10.5198/jtlu.2020.1424

Oh, D.-S., Phillips, F., Park, S., Lee, E., 2016. Innovation ecosystems: A critical examination. Technovation 54, 1–6. https://doi.org/10.1016/j.technovation.2016.02.004

Papke, L.E., 1991. Interstate business tax differentials and new firm location. Evidence from panel data. J. Public Econ. 45, 47–68. https://doi.org/10.1016/0047-2727(91)90047-6

PELLENBARG, P.H., KOK, J.A.A.M., 1985. SMALL AND MEDIUM-SIZED INNOVATIVE FIRMS IN THE NETHERLANDS' URBAN AND RURAL REGIONS. Tijdschr. Voor Econ. En Soc. Geogr. 76, 242–252. https://doi.org/10.1111/j.1467-9663.1985.tb01032.x

Phelps, C., Heidl, R., Wadhwa, A., 2012. Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda. J. Manag. 38, 1115–1166. https://doi.org/10.1177/0149206311432640

Powell, W.W., Koput, K.W., Bowie, J.I., Smith-Doerr, L., 2002. The spatial clustering of science and capital: Accounting for biotech firm - Venture capital relationships. Reg. Stud. 36, 291–305. https://doi.org/10.1080/00343400220122089

Puri, M., Zarutskie, R., 2012. On the Life Cycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms. J. Finance 67, 2247–2293. https://doi.org/10.1111/j.1540-6261.2012.01786.x

Risselada, A., Schutjens, V., Van Oort, F., 2013. Real estate determinants of firm relocation in urban residential neighbourhoods. Tijdschr. Voor Econ. En Soc. Geogr. 104, 136–158. https://doi.org/10.1111/tesg.12010

ROGERS, E.M., SINGHAL, A., QUINLAN, M.M., 2008. Diffusion of Innovations, in: An Integrated Approach to Communication Theory and Research. Routledge.

Rothaermel, F.T., Agung, S.D., Jiang, L., 2007. University entrepreneurship: a taxonomy of the literature. Ind. Corp. Change 16, 691–791. https://doi.org/10.1093/icc/dtm023

Roundy, P.T., Bradshaw, M., Brockman, B.K., 2018. The emergence of entrepreneurial ecosystems: A complex adaptive systems approach. J. Bus. Res. 86, 1–10. https://doi.org/10.1016/j.jbusres.2018.01.032

Saxenian, A., 1996. Regional Advantage: Culture and Competition in Silicon Valley and Route 128, With a New Preface by the Author. Harvard University Press. https://doi.org/10.2307/j.ctvjnrsqh

Siegel, D.S., Westhead, P., Wright, M., 2003. Assessing the impact of university science parks on research productivity: exploratory firm-level evidence from the United Kingdom. Int. J. Ind. Organ., The economics of intellectual property at universities 21, 1357–1369. https://doi.org/10.1016/S0167-7187(03)00086-9

Smith, L.H., Ho, K.W., 2006. 'Measuring the performance of Oxfordshire's spin-off companies.' Res. Policy 35, 1554–1568.

Spigel, B., 2017. The Relational Organization of Entrepreneurial Ecosystems. Entrep. Theory Pract. 41, 49–72. https://doi.org/10.1111/etap.12167

Stam, E., 2007. Why Butterflies Don't Leave: Locational Behavior of Entrepreneurial Firms. Econ. Geogr. 83, 27–50. https://doi.org/10.1111/j.1944-8287.2007.tb00332.x

Stiglitz, J.E., Weiss, A., 1981. Credit Rationing in Markets with Imperfect Information. Am. Econ. Rev. 71, 393–410.

Stuart, T., Sorenson, O., 2003. The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms. Res. Policy, Special Issue on Technology Entrepreneurship and Contact Information for corresponding authors 32, 229–253. https://doi.org/10.1016/S0048-7333(02)00098-7

Targa, F., Clifton, K.J., Mahmassani, H.S., 2006. Influence of Transportation Access on Individual Firm Location Decisions. Transp. Res. Rec. 1977, 179–189. https://doi.org/10.1177/0361198106197700121

Targa, F., Clifton, K.J., Mahmassani, H.S., 2005. Economic activity and transportation access : An econometric analysis of business spatial patterns. Transp. Res. Rec. 61–71. https://doi.org/10.3141/1932-08

Taylor, M. j., 1975. Organizational growth, spatial interaction and location decision-making. Reg. Stud. 9, 313–323. https://doi.org/10.1080/09595237500185361

Teece, D.J., Pisano, G., Shuen, A., 1997. Dynamic capabilities and strategic management. Strateg. Manag. J. 18, 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

THE RISE OF THE KNOWLEDGE SOCIETY [WWW Document], n.d. URL http://archive.wilsonquarterly.com/essays/rise-knowledge-society-0 (accessed 9.19.22).

van Dijk, J., Pellenbarg, P.H., 2000. Firm relocation decisions in The Netherlands: An ordered logit approach. Pap. Reg. Sci. 79, 191–219. https://doi.org/10.1007/s101100050043

Williamson, O.E., 2016. The transaction cost economics project: Origins, evolution, utilization, in: The Elgar Companion to Ronald H. Coase. pp. 34–42.

Williamson, O.E., 1981. The Economics of Organization: The Transaction Cost Approach. Am. J. Sociol. 87, 548–577. https://doi.org/10.1086/227496

Wissema, J.G., 2009. Towards the third generation university: Managing the university in transition, Towards the Third Generation University: Managing the University in Transition.

Woods, J.G., 2019. From Decreasing to Increasing Returns: the Role of Tacit Knowledge Capital in Firm Production and Industrial Growth. J. Knowl. Econ. 10, 1482–1496. https://doi.org/10.1007/s13132-016-0351-2

Sitography

https://home.kpmg/xx/en/home/media/press-releases/2022/01/global-venture-capital-annual-investment-shatters-records-following-another-healthy-quarter.html

https://www.wipo.int/patents/en/

https://www.eqar.eu/qa-results/synergies/european-tertiary-education-register-eter/

https://www.eter-project.com/wp-content/uploads/2022/02/Final_report_ETERII_EN.pdf

https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hrst_st_nsec2&lang=en

https://ec.europa.eu/eurostat/web/main/data/database)

https://www.topuniversities.com/university-rankings/world-university-rankings/2020

List of Figures

List if Tables