UNIVERSITÀ DEGLI STUDI DI BRESCIA

DIPARTIMENTO DI ECONOMIA

Corso di Laurea Triennale
in Economia e Azienda Digitale

Relazione Finale

THE ROLE OF R&D INVESTMENT, PATENT ACTIVITY AND LABOR FORCE IN PRODUCTIVITY GROWTH: AN EMPIRICAL ANALYSIS FOR 10 EUROPEAN ECONOMIES IN THE INNOVATION AND GROWTH FIELD

Relatore:

Chiar.mo Prof. Mattia Guerini

Laureanda:
Morandi Arianna

Matricola n. 732411

Anno Accademico 2022/2023

SUMMARY

INTRODUCTION

At the basis of this thesis there is the aim to analyze, and consequently understand, the role of innovation on productivity growth, taking as a sample 10 European countries with different histories and economies. To study such a relationship, the innovation dimensions will be proxied by three variables: R&D investment, labor force and patent activity.

The main motivation that led me to choose, and deepen, the theme mentioned above is the importance that it has in our constantly changing world. Lately, one hears every day about research and development capital, about the workforce and the problems of unemployment and about innovations and new patented inventions. But, will these variables really have a positive and significant effect on a country's productivity growth? In literature, the ideas set forth in this regard are sometimes very conflicting.

Therefore, the objective of the analysis carried out in this document is precisely to try to give new answers to this very wide and complicated question, through the application of different econometric methodologies.

At first a graphical analysis of the individual variables will be done. Then, few functions that involve the three regressors that are individually related with the variable productivity will be introduced. Subsequently, they will be merged in a single model. Finally, the multiple regression model will be further implemented by adding the entity and time fixed effects and a lag for each one of the regressors.

The paper is divided into three chapters. The first focuses on analyzing the literature of the topic, providing a general review in the innovation and growth field. The second chapter describes the data collected from the OECD site and the econometric methodologies used. With this dataset, we will finally proceed to an empirical analysis, carried out in the third chapter. In this last chapter, the impact of the above-mentioned factors on productivity growth is verified, and the results of the estimated models are reported - with the help of figures and tables.

In light of this, it will be possible to answer the initial question, which guided the entire drafting of the thesis.

CHAPTER I

A LITERATURE REVIEW IN THE INNOVATION AND GROWTH FIELD

1.1. INNOVATION AND GROWTH

Innovation is the driving force behind economic growth. In our ever-changing world, it is increasingly essential for businesses to try to embrace and drive innovation. The intrinsic link between growth and innovation is manifested through the introduction of new technologies and new ideas. Innovation is not only about creating new products, but also about changing classic business models and adopting dominant strategies. In this context, it becomes essential to explore the relationship between innovation and growth to understand how today’s society can keep up with the continuous transformations.

1.1.1. DEFINITIONS AND CLASSIFICATIONS

Every day, innovation is increasingly talked about. To fully understand what this term represents, a distinction between innovation and invention, should be explained. These two words, although closely connected represent two different passages often taking place in two different moments. These phases can occur subsequently or after a duration of time. The invention represents the initial concept of a new product being introduced to the market, while the innovation defines precisely the moment in which this concept is commercialized for the first time.

To better understand the difference between the two, it is important to underline how unlike inventions that can occur anywhere, innovations usually take place within companies. These companies often have at their disposal various types of structures, financial resources, skills and knowledge.

Furthermore, the transition from invention to innovation is not a leap, but rather a continuous process. In fact, most of the significant inventions that characterize our society today were initially crude and unreliable concepts. Only later, thanks to the continuous incorporation of new innovations, we have reached today's conditions.

Innovations are classified according to 'type'. Schumpeter was able to distinguish five different ones, which are in order: new products, new production methods, new sources of supply, exploitation of new markets and new ways of organizing business.[1] Out of the five types, two require additional specifications: "process innovation" and "product innovation". Product innovation being the introduction of new products has a positive effect on income, employment and growth. While process innovation reduces costs and has an adverse effect on employment due to the ever-faster introduction of machinery that replaces human labor. Regarding process innovation, we can distinguish it into two further categories, as did Edquist et al. (2001), in "technological process innovation" and "organizational process innovation", which are related to, respectively, new types of machinery and new methods of work organization.[2]

A critical parenthesis is the relationship between product and process innovation and unemployment. Understanding whether technical change is beneficial or detrimental to employment is at the heart of political debate, especially in times of economic recession. When we talk about technological unemployment, we are talking about job losses due to technological change. This change usually involves the introduction of technologies that reduce the workload of operators and the introduction of automation. In any case, so far, the effects of innovation - in its many forms and intrinsic complexity - on labor demand have proved to be not unequivocal.

A last important classification to take into consideration also comes from Schumpeter, who classifies innovations with respect to the level they are rooted in the existing structure. He speaks of "marginal" innovations by contrasting them with "radical" innovations or "technological revolutions" as these innovations completely revolutionize the environment in which they are inserted. When you have these types of innovation, companies must consider the social and economic implications. The more radical an innovation is, the more significant the impact is to the project investments, and the social and organizational project requirements. However, this does not mean that marginal innovations should be forgotten, since compounding marginal revolutions, still has important and impactful outcomes. Ignoring marginal innovations can create problems as incremental changes are an important component of economic growth.

Finally, a concept introduced by G.Dosi, is that of technological paradigm together with the related notion of technological trajectory. The technological paradigm is a set of knowledge, codified and tacit, that includes research procedures and scientific notions related to the birth and development of a technology. It also works as a shared model for a community of researchers involved in solving

[1] Joseph A. Schumpeter in his work "*The Theory of Economic Development*", published in 1912 explains his idea about innovation and how he classifies different type of innovations.
[2] C.Edquist, L.Hommen, and M.D.McKelvey.

problems that emerge in the advancement of a technology. The origin of the paradigms lies in the interaction between scientific advances, economic factors, institutional variables and unresolved difficulties on established technological paths. In the initial (preparadigmatic) phase, different knowledge bases and scientific principles compete with each other, and much of the process of discovery and innovation proceed through trials and errors, following a multiplicity of paths.

Moving on to the definition, when it comes to innovation, the dictionary is not sufficient to explain the complexity of this process, so it is necessary to address key points that define the fundamental characteristics.[3]

First of all, innovation is an ongoing process. A company that decides to innovate, and therefore to undertake a specific path, must consider that, although it will initially enjoy the advantages that others will not enjoy, there is the risk of getting stuck on the initial path. Complexities related to costs and/or time could prevent moving onto more performative paths. For this reason, it is essential that during the first phase of an innovation project the variables are studied and planned for. In all cases the project managers should be agile to be open to new ideas and solutions, even changing strategy in the middle of a phase. This is because innovations themselves can be defined as a recipe of capabilities, resources, ideas, money and skills.

One of the most striking facts about innovations is their variability in space and time. They tend to concentrate in certain sectors, but not only, even in certain geographical areas and in certain time periods. As history shows us, the United Kingdom, around the middle of the nineteenth century, was for many years the center of the world for innovation. It then moved towards the end of the century to Germany, and now it is divided between the United States, China and South Korea.Therefore, innovation has always been defined as the engine of economic development. In fact, a company that successfully introduces an innovation into its environment will certainly enjoy a higher rate of profit. This result then acts as a signal for other companies, the so-called "imitators", which will try to derive benefits from it, until the effects of innovation on growth, and growth itself, will slow down. However, it is only thanks to this mechanism that the innovation-dissemination creative process is activated, where an innovation initially implemented by a company, then, passing through "imitator" companies, is modified, improved and increased and lays the foundations for other later innovations.

---

[3] The Oxford languages dictionary definition for innovation is: "*a new method, idea, product, etc*.".

1.1.2. THE SCHUMPETER PARADIGM

An interesting theory to understand is the one explained by Joseph Schumpeter, in which he explains the relationship between competition and innovation. There are three main ideas in his model. In the first one he argued that innovations were the main promoter of long-term growth and that without innovations the other factors would not have been sufficient. The second idea, he argued that all innovations, whether marginal or substantial, took place only thanks to the investments that entrepreneurs made in various fields: training, research and development, technologies, materials, etc. In turn, the entrepreneurs who make these investments respond to economic incentives that come from the economic institutions that drive and shape (even innovate) economic policies. In this sense, therefore, Schumpeter underlined the fundamental role played by growth institutions and policies within the innovation process emphasizing the impacts to society. Indeed, in those environments where there is a situation of hyperinflation or poor protection of property rights, growth will be discouraged due to damage to the profitability of innovation. Lastly, in the third idea he explained the so-called “creative destruction”. Old technologies are replaced by new innovations, in fact growth, as described by Schumpeter, is the result of continuous conflicts between the new and the old. The “old” companies that are already established in the market are motivated to prevent or delay the entry of new competitors into their sector, and in this case, we speak of the "political economy of growth".

In the baseline Schumpeterian, the companies who are in the process of selling innovative goods are not “active companies” i.e., the current technological leaders, but rather they are “inactive companies”. In this model this portrays a process of a firm moving from a pre-innovation stage to a post-innovation stage by growing its profit from zero to positive. In fact, it is precisely for this reason that competition discourages innovation. As the net profit of the innovation, which is equal to the post-innovation profit, it is reduced by competition.

Nevertheless, in most market sectors, we usually find at least two different types of firms, which react to competition and its increase in different ways. The "frontier companies" and the "latecomers". Frontier companies are close to their sector’s technological frontier, and in fact are active and make significant profits. The latecomers on the contrary are companies which are far below the current technological frontier as they follow the innovative path set forth by the companies on the frontier.

The "frontier companies" are motivated to drive innovation to escape the competition that exists in their sector. While the "latecomers" will be those who must try to recover the differences that divide them from frontier companies. Due to the high degree of competition, latecomers could be discouraged and their innovation could decrease.

Overall, the effect of competition on innovation and productivity is an inverted U, since, it has a positive and negative effect on the two types of firms.

This basic Schumpeterian model was extended in various ways (see Aghion et al. [1997]; Aghion et al. [2001]). What is called the "escape competition effect" was studied. When a company innovates to catch up with a sector's leader, an interaction takes place where this companies will go head-to-head in competition. This results in greater difficulties for all parties due to increased competition in the marketplace. Subsequently the aforementioned effect will take place i.e., these companies are encouraged to continue innovation in order to achieve better product in market results and an advantage over their rival.

On the other hand, in sectors where companies do not find themselves going head-to-head, there will be the opposite effect, which is called the "Schumpeter effect". The increase in competition results in lagging companies to be discouraged from innovating, the challenges that come with trying to catch up with the leader could lead to a noticeable decrease in extra profit in the short term.

In this extended model the relationship between innovation and competition is represented by an inverted U. In fact, when competition is low in head-to-head sectors, there is a correlation to the low intensity of innovation. Conversely, when competition is high, in head-to-head sectors the intensity of innovation is high, and therefore most sectors will be unlevelled creating clear market leaders, and the Schumpeterian effect will dominate.

The prediction of the inverted U-shaped relationship between competition and innovation was also confirmed by Aghion et al. [2005].[4]

1.1.3. GROWTH THEORY: FROM THE SOLOW MODEL TO ENDOGENOUS GROWTH THEORY

The Solow Growth model is an exogenous model of economic growth that studies the variables to productivity in an economy over time. The three variables are: the rate of population growth, the rate of technological progress, and the rate of savings.

[4] Angion et al. (2005) carried out an empirical analysis using data from British companies.

This model was developed in 1956 by Nobel Prize winner in economics Robert Solow and was built upon the Harrod-Domar Keynesian model.[5] It is considered the first neoclassical growth model and the basis of modern economic growth theory.

The model is based on fundamental concepts. The first concerns the accumulation of capital. In the model, capital is one of the most important engines for growth and can be both human and physical. In human capital, investments can take place through training and education. In physical capital, investing in infrastructure and machinery. The increase of human and physical capital leads to an increase in the production of goods and services, and therefore to an increase in the growth rate.

Secondly, the model explains the law of diminishing returns. When there is an increase in the rate of capital accumulation there will also be marginal returns which can decrease. Hence, as capital increases, the effect of accumulation on growth becomes less significant i.e., each new unit of capital generates a smaller and smaller increase on output.

In the model, another important factor is technology. It represents both innovation and technological progress which contribute to economic growth. The introduction of innovative technologies makes it possible to generate more sustainable long-term growth, thanks to the more efficient use of capital. Furthermore, innovation makes it possible to avoid and overcome the law of diminishing returns explained above.

Solow also speaks of the stationary state, and in his model it represents the equilibrium point towards economy growth in the long run.[6] The growth rates of population and output are constant in the steady state, while income per capita and capital per worker also remain stable. Technological innovation and capital accumulation manage to balance the decrease in marginal returns.

Finally, the model suggests that economies that grow faster are the less developed ones, while the more developed ones tend to grow at a slower rate. This process is called "convergence". Countries that benefit from high returns given by investments and advanced technologies from developed economies and countries, are precisely the least developed countries, and therefore those with less technology and less capital.

---

[5] The Harrod-Domar model is based on the Keynesian hypothesis that saving is a constant proportion of income and allows the identification of the 'guaranteed growth rate', that is, the rate that simultaneously ensures the balance between savings and investments and the fulfilment of expectations about the change in income produced.

[6] Steady-state is defined as the hypothetical situation of an economic system that remains constant, that neither progresses nor regresses.

In summary, this neoclassical theory explains how technological innovation, capital accumulation and the law of diminishing returns combine to determine long-term economic growth in a nation. However, it assumes that technological progress depends on a scientific process that is independent and separate from other economic forces. Therefore, the long-term growth rate is to be considered as exogenous data.

The theory of endogenous growth was born to refute the theories of neoclassical models, which did not take into account technological changes in providing forecasts and to demonstrate that long-term economic growth is determined by endogenous factors, therefore internal and not external to the economy.

Endogenous growth is defined as long-term economic growth whose rate is defined by the forces that move opportunities to create new technological knowledge. These are also internal to the economic system. The start of growth for per capita production depends on productivity levels, which in turn has a dependency on the technological progress based on human capital and innovation. The innovations that lead to technological progress manifest themselves as new processes, new markets, new products, and are always the result of various economic activities.

For example, a higher pace of economic activity can increase the pace of innovation. This is due to firms predominantly learning how to be more efficient in production through experience. This increase in production gives them a wider experience. Furthermore, it is important to underline that many innovations are born thanks to investments by companies in research and development. When they seek profit, it is possible the various economic policies implemented affect costs. The impacts can be felt in the context of competition, trade, education and intellectual property. Therefore, the benefits of research and development activities, influence the rate of innovation.

That said, there is a need for private sector enterprises to be incentivized through the provision of government subsidies or grants. These incentives would provide monetary support to their research and development to continue to innovate and drive economic growth. These incentives also support entrepreneurs by creating new jobs or business opportunities. This is of fundamental importance for the country's growth. An example of incentives that push companies to expand their economic activities can be copyrights, patents, and all kinds of intellectual property rights.

The value that comes from investing in human capital through training and instruction programs increase return to scale and improvement to the quality of work. Resulting in human capital investments leading to an increase in productivity, the improvement of production processes, and infrastructure which support new innovations.

As explained so far, increase in productivity, in the theory of endogenous growth, derives from various factors within the economic system. There are various models, by various theorists, which give greater importance to investments in different areas.

The Uzawa-Lucas model explains how, long-term economic growth, comes from increased investment in the accumulation of human capital and therefore in education. It is assumed that human capital is the only input into education. In addition, both human and physical capital are used to develop economic output. Therefore, the ratio between the two types of capital is the measurement used to establish the total capital within an economic system.

A different view is explained through Romer's model, which considers every technological change endogenous. In doing so, this considers them as bearers of the greatest economic improvements. Furthermore, this model assumes that by combining existing knowledge and important investments in research and development, new innovative ideas can be created to improve production within the economy.

1.2. PRODUCTIVITY AND R&D

A well-organized economy, which aims high and has excellent growth prospects, must be capable, through adequate investments in research, of reaping the fruits of technical progress. The second part of the last decade will be remembered, on an economic level, for the development of the so-called «New Economy».

Thanks to the emergence of new information and communication technologies, which constitute a determining factor for growth, a new enthusiasm has been observed in the markets. This has laid the foundation for a theoretical lasting increase in productivity. Despite this, however, it is more difficult to identify the channels through which innovation manifests itself. This is because it is difficult to understand to what extent it is the result of a natural evolutionary process in knowledge, or to what extent the contribution actually comes from research and investments aimed at exploiting the results.

Moreover, even when existent, technologies produce low yields. These trends are particularly worrying for Italy. Much depends on the poor quality of the human capital of workers and managers. The spread of unequal digitalization among enterprises has increased the dispersion of productivity, contributing to the slowdown in economic development and the increase in wage inequalities.

Enterprises with more skilled workers (which are identified as graduate ones) show higher adoption rates of digital technologies and achieve higher productivity gains. The quality of managers is also crucial: companies with graduate executives have higher returns associated with digital technologies and stronger complementarities between technologies and the skilled workforce. More qualified managers are therefore better able to jointly manage all inputs involved in digital transformation.

Nonetheless, it is never put in doubt the belief that our economy, like many others and consequently, in a certain sense, the society in which we live, are based on knowledge and its applications is not called into question.

From an observation of the trend of GDP growth rates in the 24 OECD countries, a clear slowdown can be seen in the 1990s (average growth rate of 1.8%) compared to the 1980s (average growth rate of 2.1%) and the 1970s (average growth rate of 2.6%).[7] Instead, in the second half of the last decade there was a significant acceleration, and an average growth rate of 2.6% was reached again, as in the 1970s.

In recent years, as is natural to think, the differences in the results obtained by the various countries have widened considerably, in some of them such as the United States, Canada and Australia they have increased; while in others such as notably Japan, they have decreased. These developments reflect the joint evolution of employment rates and labor productivity: countries in which the capacity to use the labor force has declined have recorded productivity developments that have proved to be too low to counteract the deceleration in productivity.

The so-called New Economy is associated with the investment and diffusion of new information and communication technologies (ICT). Thanks to the revolutions in research and how it is used on an industrial level, there has been an exceptional improvement in computing capabilities that have led to the reduction of costs to produce computer machinery. Furthermore, the huge investments in new technologies and cutting-edge software have made it possible to replace all the old and obsolete production processes with new ones that are much more efficient. The main takeaway is that for the same resources used in production, the yield has visibly increased.

These new technologies, as well as simplifying the organization at various levels, are very efficient because, independently of the component of technical progress inside the goods used in the production cycle, they broaden the potential and reduce the costs of communication. This benefit,

[7] OECD is the Organisation for Economic Co-operation and Development founded on December 14, 1960.

like many others, is the output of the increase in total productivity, in addition to the output determined by the use of: more skilled workforce, better performing capital goods and more production factors.

1.2.1. THE DECISIVE FACTORS OF PRODUCTIVITY GROWTH

From a study conducted over the last twenty years on a panel of 16 OECD countries it turns out that, on average, a 1% increase in firms' accumulation of investment in research and development (R&D) would produce a 0.13% increase in total productivity. [8]

Summarizing, the differences observed in growth rates are a direct consequence of different factors. First of all, of the different evolution of investment rates (both in capital goods and in the accumulation of human capital). In addition, of macroeconomic stability and of investments in R&D. And lastly, of openness to international trade. Furthermore, even the weight and composition of public intervention in the economy play a very important role, with both negative and positive effects.

The more R&D activities are developed at the level of private companies, the more they can absorb all the progress that comes from both academic and public research. This explains the support that OECD country governments give to private R&D spending. This support consists of direct subsidies and tax incentives. Tax incentives are subject to fewer bureaucratic obstacles and leave the market free to move according to its natural strengths. Finally, with reference to R&D expenditure of foreign origin, it is noted that their yield is very high. Moreover, they require lower investment costs than those given by national production.

In conclusion, the smallest countries are the ones with the greatest need for these investments. The reason is that on one hand, smaller countries have fewer researchers. And on the other hand, they tend to obtain a high degree of specialization in a limited number of sectors.

The question is how and why certain countries compared to others have managed to make the most of all the growth opportunities offered by technological progress and to adopt new technologies. Thus, allowing industrial sectors with higher productivity rates to develop to the fullest.

An in-depth analysis, carried out on the various sectors, highlights the importance of having regulated markets and institutional structures. It is also fundamental to have a high rate of

[8] Cfr. D. Guellec e B. Van Pottelsberghe, R&D and Productivity Growth: Panel Data Analysis of 16 Oecd Countries, Oecd Economic Studies, No. 33, 2001.

accumulation of human capital and investment in R&D. The outcome will be an increased productive potential.

Focusing on productivity and its growth, it is essential to analyze its distance from the technological frontier (represented by the highest level of total productivity). Productivity growth is higher the further a country's industries are from this frontier, except for high-tech manufacturing industries.

Furthermore, the role of market regulation appears to be more important the greater the distance of a sector of a country, from the leading sector of the country itself.

A faster rise in productivity occurs in those systems where: state interventions in the economy are modest, the barriers posed by administrative constraints to business activity are low and competition from the state is favored. It is also noted that in countries where there is no central coordination of the wage bargaining system, a high degree of employment protection will have negative effects on productivity. Except for sectors characterized by multiple and rapidly changing technologies. Conversely, where there is centralized coordination, technological shocks will be addressed through the retraining of employed workers. This solution is cheaper than training a new workforce.

This also applies to R&D activity, as it tends to grow as coordination in wage negotiation increases. At the same time, it tends to decrease as worker protection increases. Furthermore, investments in R&D depend on various factors. Positively it depends on the opening to international markets and on the size of the companies. While negatively it depends on the weight of the state in the economy and on the barriers placed on trade.

Finally, small businesses try to enter the market at low costs. The once that are able to get in the market, will test all their possibilities within it. And then, the successful ones will be able to grow and become large enough to be competitive. The likelihood of seeing it happen occurs only when access costs are low, and the degree of market regulation is modest.

On the other hand, in some markets the costs of entering the market and the cost due to technological shocks are higher. In this case, the entering firms will find themselves having lower chances of success due to the lower chances of experimentation. And however, even in the case of successful results, they will tend to have lower returns.

1.3. PATENTED INVENTIONS

1.3.1. AN HISTORICAL OVERVIEW

The history of patents and patented inventions begins hundreds and hundreds of years ago. It reflects humanity's innate and ongoing quest to progress and innovate. Patents have played and are playing a vital role in encouraging and promoting new technological advances that lead to economic growth.

Patents are defined as legal protections that are granted to inventors by the responsible entities. The main concept is based on the one of intellectual property. This concept can be traced back to ancient civilizations, particularly ancient Greece, where those who created new inventions sought rewards and recognition. The scientist and philosopher Archimedes, well known for his inventions, is considered one of the first to give due importance to intellectual rights.

Later, in the medieval period various associations began to arise that functioned as a community. In this communities the knowledge was shared only with the participants and they were concerned with protecting specific technologies and secrets from outsiders. Thus, this associations gave way to the birth of real knowledge protection systems.

Another predecessor of patents were exclusive rights. Those rights were granted in England in the 16th and 17th centuries by the king and queen. The inventors who received it used it as a symbol of prize.

The first New World patent was granted in North America in 1641 to Samuel Winslow.[9] This grant definitively opened the door to patents protecting intellectual property as known today. In fact, the Age of Enlightenment and the 18th century represent a period of intellectual awakening. It led to the industrial revolution and therefore to many technological advances. Patents became a real incentive, which spurred inventors to produce new ideas to enjoy exclusive rights.

In the United States, the official recognition of the importance of patents came in 1787. In this year the Constitution authorized Congress to "*promote the progress of science and the useful arts, by assuring for limited times to authors and inventors the exclusive right to their respective writings and discoveries*."

However, only in 19th century the birth of real national offices to grant patents happened. The process became more organized and systematic. In particular, in 1883, the Paris Convention was signed.[10] It was the first international treaty that united the patent legislation of all the participating countries,

[9] He invented a new process to make salt.

[10] Modified on October 2, 1979.

and which was aimed at protecting industrial property. This was followed by the Berne Convention signed in 1886. The aim was to protect artistic and literary works and deal with copyright protection.

Then came the 20th century, which as we all know, was a century of unprecedented technological progress. This progress brought a series of inventions that are now consider essential. In this environment patents played an even more important role in protecting these innovations and encouraging more investments research and development.

The American patent office is called USPTO (US Patent and Trademark Office). It was born in 1975 and deals with all applications for obtaining patents, examining them and determining whether to grant them. While WIPO is the World Intellectual Property Organization. It was founded in 1967 and is responsible for setting global standards for the granting of patents.[11]

Subsequently, the third industrial revolution and the advent of the digital age have represented many new challenges for the world patent systems. The number of applications to obtain patents has increased exponentially because of the internet revolution and all the new technologies that have brought to countless new inventions.

In this socio-political context, debates also arose on the limits that establish where the free exchange of information ends and where the protection of intellectual property begins. Some argued that patents hinder the sharing of knowledge for continuous scientific progress.

Finally, in the contemporary panorama of patents, it is evident that there is a continuous evolution. Nowadays, patents no longer cover only classic inventions, but a wider spectrum of inventions and new technologies. Some of these are in the fields of biotechnology, artificial intelligence, blockchain and renewable energy. Furthermore, countries from all over the world are collaborating to modify and adapt their laws to the progress made by new technologies. These countries are trying to go all in one common direction and share information, without only focusing on exclusive rights.

Concluding this overview, it’s possible to note how, in every historical period, there is a common thread. This thread is the natural drive that man has towards the continuous search for new progress and new innovations. From ancient times to the modern era, the development of patent systems has protected intellectual property, incentivized inventors and driven new advances.

What is important for the future of intellectual property and patents is to always find a balance. On one hand, granting inventors exclusive rights. On the other hand, fostering a collaborative

[11] Lately, in 1994 was found the European Union Intellectual Property Office as well, as seen in paragraph 2.1.4.

environment to encourage the free exchange of knowledge. Finding this balance will lead to continuous improvements of living conditions of mankind.

1.3.2. PATENTS: THE ECONOMIC VALUE

The economic value of patented inventions becomes an increasingly central issue. In fact, inventions are not only a key source of value in many new products and processes. More and more companies use the value of their results to evaluate the internal performance of their departments employed in research and development. This means that the value created during the process and development is largely explained by the quality and size of downstream assets.

When it comes to inventions, patents are considered an excellent indicator. Although not all inventions always receive a patent, they are a great source of wealth for companies. A first question to be addressed to understand the economic value of patents is the value of patent rights.

To be "maintained", a patent must be paid. The patent holders must annually renew the protection of their patent. This is already considered an indicator of the value of the right, as the more valuable a patent will be considered by the owner plus this will have an incentive to renew it for a greater number of years. On this issue various scholars, at the end of the 90s, carried out research to try to numerically establish this economic value. For example, Serrano (2012) combined information regarding both trade and patent renewals, to try to get more accurate results. His model was based on the idea that those patents that are exchanged and then renewed have a greater value than patents that are only renewed and not exchanged. In addition, those with even less value are expired patents. This happens because the patent holder also takes into account the returns that the buyers have from the invention. In fact, he noticed that in his sample the volume of traded patents is very high, almost equal to 50%.

However, all the studies conducted over the years lead to the conclusion that the ideal way to study the value of patent rights is to compare the value of an invention with and without a patent. Although not easily observable, Arora et al. (2008) created a model in which they considered the propensity that firms have to patent.[12] Conclusions show that firms, when they patent, gross of the cost of the application, expect to earn 47% more of what they would have earned if they had not patented the invention. Estimates also vary from sector to sector. The results show that the highest performing sectors are biotechnology (58%), pharmaceuticals (57%) and medical instruments (62%). It is

[12] "R&D and the patent premium."

interesting to observe that, however, just for some inventions the prize of the patent is high. While for the average invention, patent is just a net cost.

Although the original justification of patents is legal protection, they also perform other functions. A very important one, that recent literature has focused on, is patents value as a signal of quality. On the subject, a leading study is that of Hsu and Ziedonis.[13] They underline the "*advantage of the double resource*" that patents have i.e., as protection and as a signal of quality. It is also argued that the latter is even more important when between inventor and market there is a wider information gap. Their research findings suggest that the value of patents as property rights is greatest at later stages, so when products are closer to the commercialization stage. They also explain that technology companies grow over time and acquire new assets, as well as patents, which allow them to protect their inventions. Moreover, as companies grow, they can be more controlled. While when a company is new the controls that econometricians can rely on are very few if not zero. For this reason, since patents are easily observable, they can choose other characteristics that affect the value.

Another issue is the cost of patents. Given the high costs involved in acquiring patents, these can function as signals of quality. Companies considered to be of higher quality can afford to face higher costs. In the way that for them the opportunity cost of time and information dissemination may be a more important concern. At the same time, they often have better technologies and inventive capabilities at their disposal, which allows them to write patents more easily.

Conversely, lower quality companies may find themselves forced into a more difficult path to get to their patent because they lack a quality invention at the base. They may also find themselves forced into more lengthy and costly interactions with lawyers in order to succeed identify the subject matter of the patent and their claims.

Finally, the study by Hsu and Ziedonis explains how signals are much more valuable to younger companies, because there is less information about them. Young companies don't use patents as a signal of quality, but the number of claims in the patent. This works as a signal for companies with advanced technology because they will be able to write patents that contain many claims. On the other side, those companies with obsolete technology will face higher costs.

However, in conclusion, it is correct to explain how in literature there are also many studies for which patents have no (or even negative) effects on the economic system. Over the past 20 years, the flow of patents has more than quadrupled.

[13] 2013.

However, neither expenditure on innovation, nor expenditure on research and development, nor expenditure on factor productivity have shown any particular upward trend. According to the Bureau of Labor Statistics, the annual growth of total factor productivity in the decade 1970-1979 was about 1.2 percent, while in the decades 1990-1999 and 2000-2009 it was a little below 1 percent.[14]

In Boldrin and Levine (2008b), a meta-study was conducted collecting 24 studies carried out in 2006 which examined whether the introduction or strengthening of patent protection leads to greater innovation. Summing up, the two authors state that studies find little or no evidence that strengthening patent regimes increases innovation. They also find evidence that strengthening the patent regime increases patenting. And finally, also evidence that, in countries with initially weak IP [intellectual property] regimes, the strengthening of intellectual property increases the flow of foreign investment in areas where patents are often used.

1.4. LABOR FORCE AND PRODUCTIVITY

Labor force and productivity are two of the fundamental pillars for the economic success of any organization and nation. The labor force is the beating heart of productive activities and is composed of individuals who have different experiences and skills. However, work results in observable and economically relevant results only through increased productivity. An increase in productivity is achieved by adopting innovative technologies, optimizing processes to increase efficiency and promoting a working environment that stimulates the mind and encourages collaboration. Therefore, the interaction between productivity and labor is crucial to achieve sustainable levels of economic growth.

1.4.1. THE EFFECT OF AGEING

Population aging is a key issue when it comes to global demographics. The United Nations has projected that there will be 2 billion more people over 60 by 2050.[15] Additionally, the number of over 80 is expected to continue to rise at an even faster rate.

[14] The Bureau of Labor Statistics (BLS) is a unit of the United States Department of Labor. It is the main survey agency of the US government in the broad field of labor economics and statistics.
[15] Today there are 680 million people over 60.

The question is: how fast the aging process is? The answer is not heterogeneous between countries, because it depends on many different variables. One thing is certain: all the demographic changes that are currently taking place in the world are radically changing the size of the different age groups of the working population. To simplify, we divide the working-age population into 3 groups, young adults (15-29 years); adults of first age (30-49 years); and the elderly (50-64 years).

Since the end of the 1980s there has been a clear reduction in young adults. At the same time an increase in adults of first age happened. Furthermore, fertility and the proportion of the population under the age of 30 in most countries of the world are declining. This leads to changes in the composition and growth of the workforce as well.

There is evidence that the relationship between a worker income and his or her age is not a pure productivity effect. Instead, it is the result of wage policies put in place to reward good performance and encourage retention. Of course, those pay structures that reward for seniority will lead to higher wage costs as the workforce ages and may also face additional pressure to adjust.

Young workers tend to change their job more frequently than older workers, who tend to have more lasting employment relationships. As a result, there would be a decline in labor mobility, which would lead to both positive and negative outcomes. For employers, rotation costs and therefore recruitment and training costs would be reduced. On the other hand, the labor market would certainly become less flexible. This means that those countries with a high rate of mobility will be better able to adapt to changes in both labor and markets. While countries with a low rate of mobility will be less able to exploit the benefits that these changes bring.

Rates of migration within countries decrease with advancing age and have peaked among young adults. This trend reflects some natural life-course events that cause young adults to move more frequently. From leaving the parents' house to study, to finding a job and having children. These events happen early in life and decrease whit increasing age. Indeed, increasing the share of adults and older people in the population leads to a decrease in national migration rates. This could disadvantage a nation, as more migration would allow workers to be paired with jobs where their knowledge and skills are best exploited, which would increase productivity and employment.

An older workforce will obviously have a higher level of work experience, but skills do not only depend on age. Skills depends also on the stock of knowledge that the worker has acquired before entering the world of work.

Therefore, there is the risk that this wealth of skills will become increasingly outdated with the increase in the average age of workers. This will lead to negative effects on productivity and on innovation.

Relatively few adults over the age of 25 acquire new formal qualifications. The causes are various: the financial incentives to acquire these qualifications decrease and undertaking a new educational path often requires reducing the time dedicated to work and therefore reducing earnings. Additionally, the remaining working lifetime to obtain concrete benefits (in the form of better wages or opportunities) is less. Therefore, older workers will have to face significantly higher opportunity costs if they decide to undertake new training courses.

Another important issue concerns the health problems that increase with increasing age. Poor health can sharply reduce the productivity of older workers and often lead to early retirement. So, aging workforce will also have negative consequences in this respect.

In conclusion, there is evidence that, despite the influence of changes in demographic composition on labor market trends, there are factors such as: changes in demand, technological change, changes in pension policies and the performance of the economy which, in part, counterbalance the negative effects, just listed, of aging.

CHAPTER II

DATA AND ANALYSIS METHODOLOGIES DESCRIPTION

As mentioned previously, the aim of this thesis is to quantify the effect of innovation on the productivity of a set of European economies. The innovation in this analysis is represented by three highly relevant variables that relates to the three dimensions discussed in Chapter 1 of the thesis: investment in research and development, patented inventions and the overall workforce of a country. In this chapter, both data and methodologies used to carry out the analysis will be studied.

2.1. DATA

The first step of the study is data collection. The survey was carried out on a list of 10 European countries, 5 considered core, and 5 considered peripheral. The chosen countries are respectively: France, Germany, Italy, Spain and Netherlands (the core); and Greece, Slovak Republic, Poland, Portugal and Romania (the periphery). The time span that has been taken into consideration covers the 1995-2018 period.

A first dataset was built using the OECD "Productivity" database.[16] The data was collected and then filtered by country and by year of interest. The table obtained is an unbalanced panel composed of 240 observations. An "unbalanced panel" is a panel data that has some missing data for at least one time period for at least one entity. Even the datasets for the three independent variables R&D investment, patented inventions and labor force were collected and filtered from the OECD databases. The four datasets were then merged together to create a single one as shown in table 2.1.

Table 2.1. *Head of the table "FINAL.DATASET"*

STATE	YEAR	PRODUCTIVITY	R&D	PATENTS	LABFORCE
FRA	1995	52.13705	13397.686	3745.3	25356.760
FRA	1996	52.70515	13356.336	3791.5	25568.490
FRA	1997	53.57723	12956.144	4044.0	25731.050
FRA	1998	54.90250	13211.500	4235.1	26053.510
FRA	1999	55.69367	13531.609	4553.6	26480.290

The table shows the first 5 rows ordered by "STATE" and "YEAR" of the final dataset.

[16] The databases are located in the "statistics" area of the website.

2.1.1. PRODUCTIVITY GROWTH

In economics, productivity can be defined, as a first approximation, as the ratio between the quantity of output and the labor input used in the production process (this is also known as “apparent labor productivity”. It is calculated with reference to an individual company, a productive sector or more generally a nation.

In the analysis carried out in this context, the data on labor productivity are taken with reference to the nation, for each of the years taken into consideration. Productivity is computed as:

$$\textit{GDP/labor force}$$

2.1.2. R&D INVESTMENT

Investment in Research and Development represents the economic expenditure incurred by an organization, that can be a company, a government body or a research institute. The purpose is to conduct scientific research and experiment to develop new technologies, products or processes. These investments may include funding for the training of specialized personnel, the acquisition of laboratory equipment and tools, as well as the allocation of resources to conduct studies, experiments and prototyping activities.

As mentioned in the previous chapter, investment in R&D is considered crucial for innovation and long-term competitiveness in an increasingly technologically advanced environment. This analysis uses data on R&D investment rates for each country, considering all sectors jointly, during each single year taken into account.

2.1.3. LABOR FORCE

The "labor force" indicates the set of people who are available and able to carry out work activities in a given economic and geographical context. This category includes all people who are employed in paid jobs or who are looking for active employment. The workforce represents a key aspect of a country's economy, as it directly influences production, employment and the performance of the labor market. Fluctuations in the size and composition of the workforce can reflect economic, social and demographic changes.

In the analysis carried out in Chapter 3, the labor force is used as an independent variable. It is taken the number of subjects forming the workforce, of the 10 countries studied, for each year taken into consideration.[17]

2.1.4. PATENTED INVENTIONS

"Patented inventions" refers to inventions that have been officially recognized and registered through the patent system. A patent is a legal right that is granted by a government to an inventor or owner of an innovative idea for a specified period of time. This right gives the inventor a temporary monopoly on the use, production and marketing of the invention. Patented inventions can cover a wide range of fields, such as technology, medicine, manufacturing and many others.

The patents taken into consideration are all those issued by the European Union Intellectual Property Office (EUIPO).[18] The analysis here proposed employs the number of patents obtained over the years, by the 10 countries taken into consideration, in all sectors jointly.

2.2. METHODS

As explained in the previous section the analysis studies the effect, on productivity, of: the investment in research and development, the number of workers and the number of patents; of 10 European countries, over timeline ranging from 1995 to 2018.

Before discussing the results of the analysis, however, it is important to understand the econometric techniques that have been used.[19]

2.2.1. LINEAR REGRESSION MODEL WITH MULTIPLE REGRESSORS

The multiple regression model is the extended version of the single variable regression model, in which instead of having a single regressor, new variables are added as regressors. This model allows to study the effect on Y_i of the change of the variable X_{1i}, keeping the other regressors X_{2i}, X_{3i} etc. constant.

[17] In Section 3.1. the variable is studied in detail.

[18] Since its foundation in 1994, in Alicante, EUIPO has been managing the registration of EU brands and registered designs and models.

[19] This section heavily relies on the book "*Introduction to Econometrics*" by James H. Stock and Mark W. Watson.

To explain the model, let's take for example a linear multiple regression with only two independent variables X_{1i} and X_{2i}. The relationship between the dependent variable Y, and these two independent variables, is represented by the function

$$E(Y_i|X_{1i} = x_1, X_{2i} = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

where $E(Y_i|X_{1i} = x_1, X_{2i} = x_2)$ represents the conditional expectation of Y_i given that $X_{1i} = x_1$ and $X_{2i} = x_2$.

The equation represents the population regression function in the regression model. The intercept is represented by the coefficient β_0. The slope coefficient of X_{1i} is represented by the coefficient β_1 and, similarly, the slope coefficient of X_{2i} is represented by the coefficient β_2.

When switching from simple to multiple linear regression, the interpretation of the β_1 coefficient changes. In this case it predicts the difference in Y between two observations with a unit difference in X_1, holding the other regressor X_2 constant. The interpretation comes from comparing the predictions of two observations with an identical value of X_2, but values of X_1 that differ by ΔX_1. The first observation will have values of X (X_1, X_2), while the second observation will have values of X $(X_1 + \Delta X_1, X_2)$. Therefore, the value of Y will be predicted by the equation $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ for the first observation, and by the equation $Y + \Delta Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2$ for the second observation. Subtracting the two equations we obtain $Y = \beta_1 X$ and rearranging the terms we obtain $\beta_1 = \Delta Y / \Delta X_1$ keeping X_2 constant.

Regarding the β_0 intercept, the interpretation in the multiple regression model is similar to that in the single regression model. It is defined as the expected value of Y_i when X_{1i} and X_{2i} are equal to 0. In fact, it determines where on the Y-axis the population regression line begins.

The only problem with this model is that it does not take into account all the other factors which, in addition to the regressors already included, affect the dependent variable. Therefore, the population regression function must be augmented to incorporate these additional factors as error terms u_i. As a result, we will have

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

where the subscript i indicates the th of the n observations in the sample.

In addition, so far, the analysis has focused on the case of a single additional variable, X_2, but it is common to have k number of regressors $X_{1i}, X_{2i}, \ldots, X_{ki}$. In this new model, as in the previous one, β_1 is the slope coefficient of X_1, β_2 is the slope coefficient of X_2, and so on for each regressor.

The new final regression function will be:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_k X_{ki} + u_i,\ i = 1, \ldots, n$$

2.2.1.1. OLS ESTIMATES

The method of ordinary least squares (OLS) allows to estimate the coefficients of the unknown population $\beta_0, \ldots, \beta_k$. The key idea is that these coefficients can be estimated by minimizing the sum of the squared prediction errors. The process is based on the choice of a vector β containing k+1 parameters solving the following minimization problem:

$$\underset{b}{argmin} \sum_i [Y_i - \beta_0 - \beta_1 X_{1i} - \ldots - \beta_k X_{ki}]^2$$

The OLS estimators of the coefficients $\beta_0, \beta_1, \ldots, \beta_k$ which minimize the sum of squared errors in the expression are denoted $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$. The OLS terminology used in the linear multiple regression model is the same as that used in the single regressor model.

The OLS regression line is the straight line constructed using the OLS estimators: $\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}$. The predicted value of Y_i given $X_{1i}, \ldots, X_{ki}$ is $\hat{Y}_k = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}$. Finally, the OLS residual for the i^{th} observation is the difference between Y_i and its OLS predicted value. That is the OLS residual i.e., $\hat{u}_i = Y_i - \hat{Y}_i$.

2.2.1.2. LAD ESTIMATOR

Under certain conditions the OLS estimator can be sensitive to outliers. In case there are extreme values in the regression the OLS estimator can lose its efficiency. Then other estimators can be used to produce more reliable inferences. One of them is the least absolute deviation (LAD) estimator. The LAD estimators of β_0 and β_1 are the values of β_0 and β_1 obtained by solving a minimization problem of the function:

$$\sum_{i=1}^{n} |Y_i - \beta_0 - \beta_1 X_i|.$$

The LAD estimator is less sensitive to large outliers in u than the OLS estimator. This estimator can also be used to check against the results obtained with the OLS estimator.

2.2.1.3. MEASURES OF FIT

There are three summary statistics that are commonly used in multiple regression: the *standard error* of the regression, R^2 regression, and *adjusted* R^2. These statistics are used to measure how well the OLS estimate of the multiple regression line represents and describes the data.

The first statistic, the standard error of the regression (*SER*), estimates the standard deviation of the error term u_i. It represents a measure of the distribution of Y around the regression line. In multiple regression it is

$$SER = S_{\hat{u}} = \sqrt{s_{\hat{u}}^2}\text{ , where } s_{\hat{u}}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{SSR}{n-k-1}\text{ .}$$

SSR is the sum of the residuals squared, SSR = $\sum_{i=1}^{n}\hat{u}_i^2$.

Here, the divisor $n - k - 1$ adjusts to the downward bias introduced by estimating $k + 1$ coefficients. Using $n - k - 1$ instead of n is called: degrees of freedom adjustment. If there is only one regressor, then $k = 1$. The effect of adjusting the degrees of freedom is negligible when n is large.

The second statistic, the R^2, is the fraction of the sample variance Y_i explained by the regressors. Equivalently, R^2 is 1 minus the fraction of the variance of Y_i not explained by the regressors. The formula,

$$R^2 = ESS/TSS = 1 - SSR/TSS,$$

is the same used for the regression with a single regressor. The unfolded sum of squares is $ESS = \sum_{i=1}^{n} (\hat{Y}_I - \bar{Y})^2$ and the total sum of squares is $TSS = \sum_{i=1}^{n} (Yi - \bar{Y})^2$.

In multiple regression, the R^2 increases each time a regressor is added (unless the estimated coefficient added is exactly 0). Therefore, an increase in R^2 does not mean that adding a variable actually improves the fit of the model. In this sense, the R^2 provides an inflated estimate of how well the regression fits the data.

The third statistic, the *adjusted* R^2 is used to correct the increase that the R^2 undergoes. The *adjusted* R^2, or $\bar{R}^2$, does not necessarily increase when a new regressor is added.

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{s_{\hat{u}}^2}{s_y^2}.$$

The factor $(n - 1) / (nk - 1)$ is always greater than 1, so $\bar{R}^2$ is always less than R^2. Furthermore, adding a regressor has two opposite effects on R^2. On the one hand, $(n - 1) / (nk - 1)$ increases. On the other hand, the SSR falls, and the R^2 increases. Whether R^2 increases or decreases depends on which of these two effects is stronger. Obviously the higher the R^2 value, the better the data is explained.

2.2.1.4. OVB: OMITTED VARIABLE BIAS

Omitted variable bias means that the first least squares hypothesis for causal inference, $E(u_i|X_i) = 0$, does not hold. This is because in the linear regression model with a single regressor, the error term u_i represents all factors that determine the dependent variable Y_i, that are different from the dependent variable X_i, and which are omitted from the analysis. If one of these factors correlates directly with X_i, it means that the error term (containing this factor) is correlated with X_i. If there is a correlation between u_i and X_i, the conditional mean of u_i given X_i is non-zero. The consequence of the correlation and the violation of the first least squares assumption is that the OLS estimator is biased. Thus, the estimator is biased and inconsistent and this bias does not vanish even in very large samples.

The multiple regression model can help avoid, at least partially, this issue. If there are variables that are useful to explain Y_i, that can be correlated with the independent variable X_i and that are observable, it is valuable to include them in the regression model. In this way it is possible to estimate the causal effect of interest and control other variables.

2.2.2. NONLINEAR REGRESSION FUNCTION

It is important to study how, in nonlinear specifications, the dependent variable Y changes when the independent variable X_1 changes by the quantity ΔX_1, holding constant other independent variables $X_2, \ldots, X_k$. This effect is easy to calculate when the population regression function is linear. In fact, the expected variation in Y is $\Delta Y = \beta_1 \Delta X$ where β_1 is the regression coefficient of the population multiplied by X_1. Difficulties arise when the regression function is non-linear. In this case the expected variation in Y may also depend on the values of the independent variables and is therefore more difficult to calculate. The nonlinear regression function can generally be written as:

$$Y_i = f(X_{1i}, X_{2i}, \ldots, X_{ki}) + u_i, \quad i = 1, \ldots, n,$$

where $f(X_{1i}, X_{2i}, \ldots, X_{ki})$ represents the nonlinear regression function of the population and therefore a possible nonlinear function of all the independent variables. While u_i represents the error term.

The expected change in Y, represented by ΔY, associated with the change in X_1, represented by ΔX_1, holding $X_2, \ldots, X_k$, constant, is the difference between the value of the population regression function before and after the change in X_1, holding $X_2, \ldots, X_k$ constants. This means that the expected change in Y is:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots, X_k).$$

The estimator of this unknown population difference is the predicted difference in Y for two observations. The two observations have both the same values of $X_2, \ldots, X_k$, but have different values of X_1, namely $X_1 + \Delta X_1$ and X_1. Since the regression function f is unknown, this population causal effect is also unknown. To estimate this effect, we first estimate the function f. Let $\hat{f}(X_1, X_2, \ldots, X_k)$ be the predicted value of Y based on the population regression function estimator $\hat{f}$. Therefore, the expected change in Y is:

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \ldots, X_k) - \hat{f}(X_1, X_2, \ldots, X_k).$$

2.2.2.1. LOGARITHMIC REGRESSIONS

One possible way to specify a nonlinear regression function is to use the natural logarithm of Y and/or X. Logarithms allowed to convert variables changes into percentage changes. The natural logarithm is the inverse of the exponential function.[20] The base of the natural logarithm is e. The logarithm function is defined only for positive values of x and has a steep slope at the beginning which then flattens out (even if it continues to increase). The slope of the logarithmic function $ln(x)$ is $1/x$.

There are three different logarithmic regression models. The interpretations of the coefficients in the three regressions are different. In the linear-log model, X is logarithmically transformed while Y is not. In this case, the regression model is $Y_i = \beta 0 + \beta_1 ln(X_i) + u_i,\ i = 1, \ldots, n.$ A 1% change in X is associated with a change in Y of 0.01 β_1. In the log-linear model, Y is logarithmically transformed, while X is not. In this case, the regression model is $ln(Y_i) = \beta_0 + \beta_1 X_i + u_i,\ i = 1, \ldots, n.$ A one unit change in $X(\Delta X = 1)$ is associated with a (100 * β_1)% change in Y. In the log-log model both X and Y are logarithm transformed. In this case, the regression model is $ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + u_i,\ i = 1, \ldots, n.$ A 1% change in X is associated with a $\beta_1\%$ change in Y.

[20] The natural logarithm is the function for which $x = \ln(e^x)$ or, equivalently, $x = \ln[\exp(x)]$.

2.2.3. FIXED EFFECT REGRESSION

2.2.3.1. ENTITY FIXED EFFECT REGRESSION

Fixed effects regression is a method used to control for omitted variables in panel data when the omitted variables vary across entities (states) but do not change over time.

This regression model has *n* different intercepts, one for each entity. Intercepts can be represented by a series of binary variables. These variables absorb the influence of all omitted variables that differ from one entity to another but are constant over time.

Let's start with the regression model:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_i,$$

where Z_i is an unobserved variable that varies from state to state but does not change over time. We want to estimate β_1, i.e., the effect of X on Y, keeping the characteristics of the unobserved state Z constant. Z_i varies from one state to another but is constant over time so the regression model can be interpreted as having *n* intercepts, one for each state. The equation becomes:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}.$$

In the regression model, $\alpha_1, \ldots, \alpha_n$, known as entity fixed effects, are treated as unknown intercepts to be estimated. The slope coefficient of the regression line, β_1, is the same for all states. In contrast, the intercept of the population regression line varies from state to state.

State-specific intercepts can also be expressed using binary variables D to denote individual states. To do this we take D_{1i} as a binary variable equal to 1 when $i = 1$ and equal to 0 otherwise. Similarly, D_{2i} equals 1 when $i = 2$ and equals 0 otherwise, and so on. We cannot include all *n* binary variables plus a common intercept, as the regressors will be multicollinear. Therefore, we arbitrarily omit the binary variable D_{1i} for the first entity.

Consequently, the model, with *n-1* binary variables, can be equivalently written as:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \ldots + \gamma_n D_{ni} + u_{it},$$

where $\beta_0, \beta_1, \gamma_2, \ldots, \gamma_n$ are unknown coefficients to be estimated.

2.2.3.2. TIME FIXED EFFECT REGRESSION

Time-fixed-effects regression allows to control for variables that are constant across entities but evolve over time.

The population regression of *section 2.2.3.1.* can be changed to:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 S_t + u_{it},$$

where *St* replaces *Zi* and is unobserved.

The single subscript *t* represents variables that change over time, but are constant across states. Omitting S_t from the regression leads to omitted variable bias. The goal is to estimate β_1, controlling for S_t. The presence of S_t leads to a model where each time period has its intercept. The time fixed effects regression model with a single regressor *X* is:

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}.$$

This model has a different intercept, λ_t, for each time period.

Just like the entity fixed effects regression model, the time fixed effects regression model can also be represented using *t - 1* binary indicators. The regression will become:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2\, B2_t + \ldots + \delta_T\, BT_t + u_{it},$$

where $\delta_2, \ldots, \delta_t$ are unknown coefficients and where $B2_t = 1$ if $t = 2$ and $B2_t = 0$ otherwise, and so on. In this version of the time effects model the intercept is included, and the first binary variable $(B1_t)$ is omitted to prevent perfect multicollinearity.

It is possible to combine both entity and time fixed effect. It is useful when both variables that are constant over time but vary between states, and constant between states but vary over time, are omitted. The combined entity and time fixed effect regression model is:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}.$$

α_i is the entity fixed effect and λ_t is the time fixed effect. This model can be represented equivalently using the binary indicators *n - 1* entities and the binary indicators *T - 1* times, together with an intercept:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \ldots + \gamma_n Dn_i + \delta_2 B2_t + \ldots + \delta_T BT_t + u_{it}.$$

$\beta_0, \beta_1, \gamma_2, \ldots, \gamma_n,$ *and* $\delta_2, \ldots, \delta_T$ are unknown coefficients.

2.2.3.3. CLUSTERED STANDARD ERRORS

The usual standard errors for regression are invalid if the regression errors are autocorrelated, because they were derived under the false assumption of no serial correlation. There are standard errors that are valid if u_{it} is potentially heteroskedastic and time-correlated within an entity. These are the standard errors of heteroscedasticity and robust autocorrelation (HAR). One type of HAR standard errors are the clustered standard errors.

"Clustered" because these standard errors allow regression errors to have arbitrary correlation within a cluster, or grouping, but assume that regression errors are uncorrelated across clusters. In the context of the data used for this analysis, panel data, each cluster consists of an entity. Therefore, these errors allow for heteroskedasticity and arbitrary autocorrelation within an entity, but treat errors as uncorrected across entities. Finally, the clustered standard errors hold whether or not there is heteroskedasticity, autocorrelation, or both.

2.2.4. DYNAMIC CAUSAL EFFECT

It is possible to estimate the effect on *Y* now and in the future, of a change in *X*. It is the dynamic causal effect on *Y* of a change in *X*. To estimate dynamic causal effects, the so-called distributed delay regression model is taken. In this model, Y_t is expressed as a function of the current and past values of X_t.

Since dynamic effects necessarily occur over time, the model used to estimate dynamic causal effects must incorporate lags. Therefore, Y_t can be expressed as a distributed lag of past and present values of X_t:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \dots + \beta_{r+1} X_{t-r} + u_t.$$

Where u_t is the error term including the measurement error in Y_t and the effect of the omitted Y_t regressors.

The coefficient β_1 on the contemporary value of X_t is the contemporary or immediate effect of a unitary variation in X_t on Y_t. In turn, the coefficient β_2 on X_{t-1} is the effect on Y_t of a unitary variation in X_{t-1} or, equivalently, the effect on Y_{t+1} of a unitary variation in X_t. Therefore, β_2 is the effect of a unit change in *X* on *Y* over a subsequent period.

In general, the coefficient on X_{t-h} is the effect of a unit change in *X* on *Y* after *h* periods. Dynamic causal effect is the effect of a change in X_t on Y_t, Y_{t+1}, Y_{t+2}, and so on. Therefore, it is the sequence of causal effects on the present and future values of *Y*. Thus, in the context of the distributed lag model, the dynamic causal effect is the sequence of coefficients $\beta_1, \beta_2, \dots, \beta_{r+1}$.

CAPITOLO III

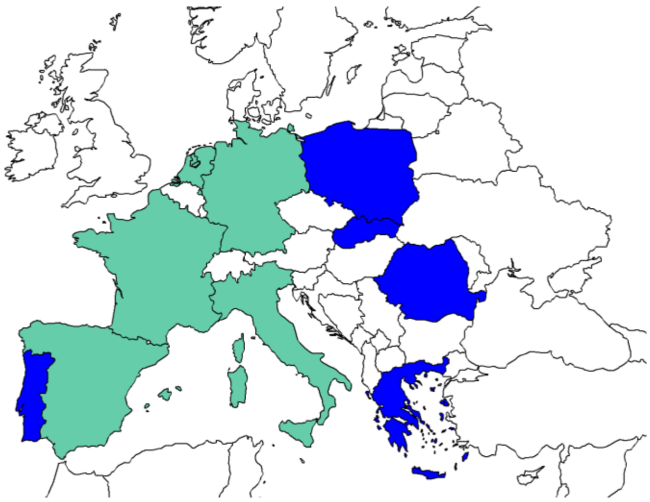
PRODUCTIVITY, R&D AND PATENTS: AN EMPIRICAL ANALYSIS

In this chapter, the analysis and its results will be presented in detail. As mentioned before the aim of the analysis is to study the role of R&D investment, patent activity and labor force on productivity growth. The statistical methodologies implemented are those seen in the previous chapter.

3.1. MAIN VARIABLES: A GENERAL AND GRAPHICAL ANALYSIS

The dataset was built using four variables: one dependent and three independents. The observations were taken for 24 years (1995-2018) and for 10 European countries divided in two groups, as shown in Figure 3.1.

Figure 3.1. *Map of the 10 European countries studied.*



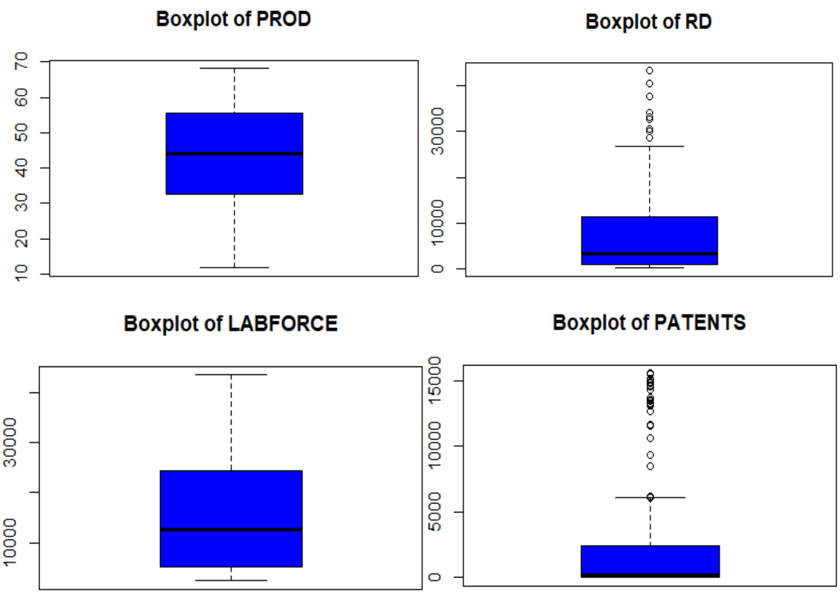
The 5 core countries, colored in turquoise, are: France, Germany, Italy, Spain and Netherlands. While the 5 peripheral countries, colored in blue are: Portugal, Romania, Greece, Poland, Slovak Republic.

For each variable a general and graphical analysis was carried out, through indexes and charts. Two relevant indexes for the shape of the distribution of a variable are the skewness and the kurtosis. The skewness indicates if the distribution is symmetric respect to the median value. If it is positive, it indicates a distribution with an asymmetric tail extending towards most positive values. While if it is negative the tail extends toward the opposite side. The results obtained are: -0.16 for productivity, 1.67 for R&D investment, 0.73 for labor force and 2.22 for patent activity. This means that all the variables are slightly asymmetric with respect to the median value but with different length of the tails. Productivity's tail extends in the direction of negative values, while the other three variables in the opposite direction.

The kurtosis of a distribution reveals instead how much a distribution moves away from normality (Gaussian distribution). If the value of kurtosis is positive the distribution is Leptokurtic i.e., more pointed than a normal curve. If it is negative the curve is less pointed than a normal curve and it's called Platykurtic. The results obtained are: -1.01 for productivity, 2.94 for R&D investment, -0.56 for labor force and 3.96 for patent activity. This means that productivity and labor force are Platykurtic i.e., less pointed than a Gaussian curve. While R&D investment and patent activity are Leptokurtic i.e., more pointed.

Similar conclusions, and more information, can be drawn using a graphical tool, called box-plot.

Figure 3.2. *Box-plots of productiviy, R&D investment, labor force and patent activity.*



The plots show the characteristics of each variable include in the model.

The box-plots in Figure 3.2. show the shape of the distribution and are formed by three elements. The elements are: a line, that indicates the median's position, a rectangle(box), whose width depends on a measure of variability and by two segments, starting from the upper and lower sides of the box, that indicates the variability outside the first and third quartile.[21] The points that are outside the box's

[21] The quartile is a position index that provides information about the structure of the distribution. Once the data is sorted, the quartiles are the three values that divide the data set into four ranges of equal number.

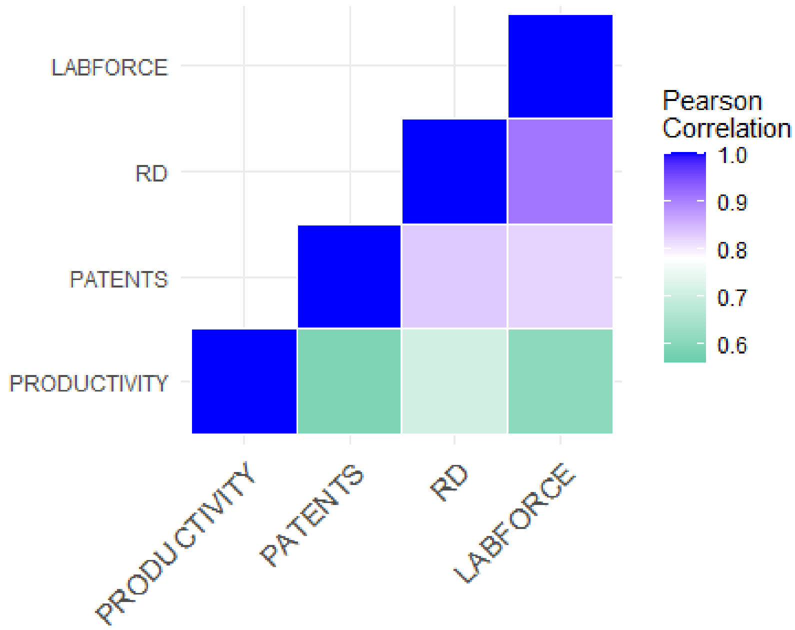
whiskers are called outliers. Out of the four, only R&D investments and patent activity's box-plots show to have outliers.

In addition, the box-plots show that the distribution is not symmetric for none of the variables because the whiskers are not equidistant from the box. This confirms what said before: productivity has a light negative asymmetry while the other three variables a positive one.

3.2. LINEAR REGRESSION MODEL WITH ONE REGRESSOR

In this paragraph the dependent variable will be studied in pair with each one of the independent variables as a single regressor. The first step is to compute the value of the Pearson correlation coefficient between each variable. The purpose is to verify the presence of a linear relationship. The results are collected in a single graph, a correlation heatmap matrix.

Figure 3.3. *Correlation heatmap matrix.*



The scale of correlation intensity goes from 0.6 to 1.0.
As shown in the bar next to the chart, the darker is the square,
the stronger is the degree of correlation.

In figure 3.2. it's clear how productivity is more strongly correlated with R&D investment, than labor force and patent activity. In addition, the heatmap shows that the three independent variables are correlated to each other because the squares' color is closer to blue.

3.2.1. PRODUCTIVITY AND R&D

The first linear regression model studied is the one that relates productivity and R&D investment. To verify the direction of the linear relationship, the covariance index is used. As a matter of fact, the covariance is equal to 91641.07. The positive value indicates that the two variables vary together i.e., when a variable increases the other one increases too.

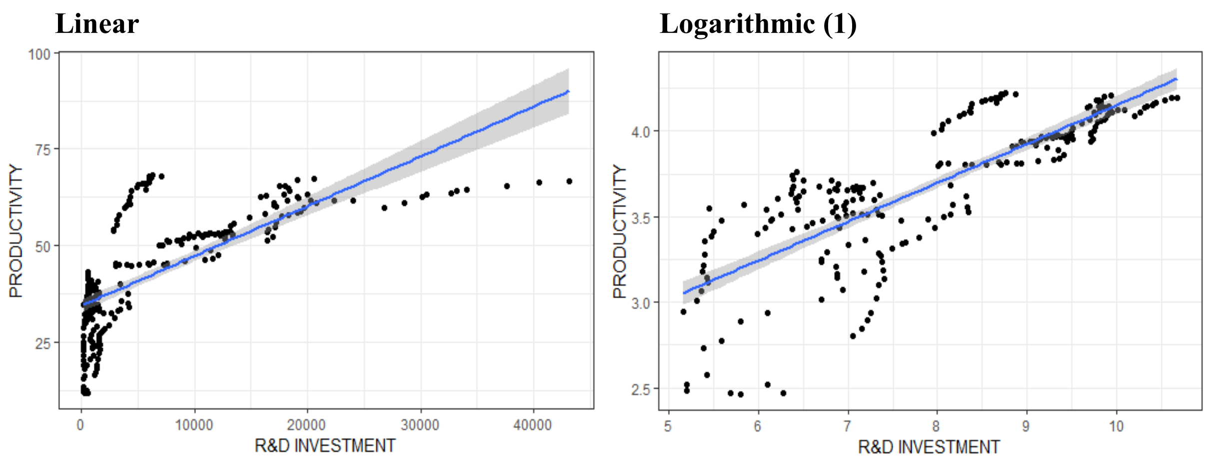
The simple linear regression model is built as:

$$\widehat{Productivity}_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R\&Dinvestment_{i,t},$$

where *i* represents states and *t* represents years.

Estimating it with OLS, $\hat{\beta}_0$ results 34.356, $\hat{\beta}_1$ results 0.001 and the R^2 is equal to 0.510. The value of $\hat{\beta}_1$ means that any increase to R&D investment leads to an almost constant output in productivity growth. Hence is important to convert the regression in a non-linear function, in particular in a logarithmic regression.[22]

Figure 3.4. *Relationship between productivity and R&D investment.*



The plot shows the relationship between productivity and R&D investment for the 10 states included, from 1995 to 2018. The left graph represents the estimated linear function, while the right one plot the estimated logarithmic function. The blue line is the estimated regression line.

As shown in Figure 3.4., the log-log regression function (1) better graphically represents the relationship between the two variables. Thus, the new model built modifying the linear model is:

---

[22] There are three different logarithmic regression models. The model used is the log-log model, where both the dependent and independent variables are transformed into logarithms.

$$\log(\widehat{Productivity_{i,t}}) = \hat{\beta}_0 + \hat{\beta}_1 \log(R\&Dinvestment_{i,t}).$$

In order to verify that, even numerically, the logarithmic regression is better, it is necessary to look at the regression's statistics, including coefficients significance and fits. Coefficients estimations are reported in table 3.1.

Table 3.1. *Regression statistics of the impact of R&D investment on productivity.*

	(1)
(Intercept)	1.878***
	(0.100)
log(RD)	0.228***
	(0.011)
Num.Obs.	240
R2	0.654
R2 Adj.	0.653
AIC	5.1
BIC	12.0
RMSE	0.24
Std.Errors	Heteroskedasticity-robust

* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression model estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

The results in table 3.1. validate what was stated previously. It can be noticed that each coefficient is significant and the coefficient $\hat{\beta}_1$ has increased from 0.001 to 0.228. This raise means that any increase to R&D investment leads to an increase in productivity growth. Since it is a log-log model, a 1% change in R&D investment is associated with a 0.23% change in productivity. In addition, the value of the R^2 has increased too, from 0.510 to 0.654. It means that this model better explains the data. 65% of the sample variance productivity is explained by the regressor.[23]

[23] The number of observations vary because all the models included in the analysis does not consider missing values.

3.2.2. PRODUCTIVITY AND LABOR FORCE

As done for the R&D investment variable, the linear relationship between the independent variable labor force and the dependent variable productivity, is studied. The covariance index of the two variables is equal to 111881.2. The positive value represents a direct linear bond. The two variables vary in the same direction.

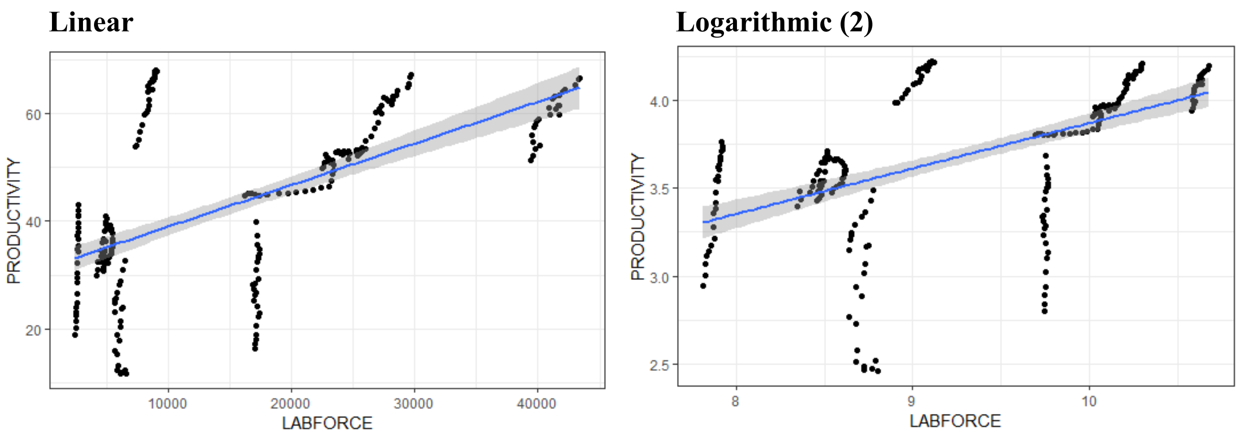
The simple linear regression model is built as:

$$\widehat{Productivity}_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 labor\ force_{i,t},$$

where *i* represents states and *t* represents years.

Estimating it with OLS, $\hat{\beta}_0$ results 31.264, $\hat{\beta}_1$ results 0.00, standard errors result 0.00 and the R^2 is equal to 0.273. The value of $\hat{\beta}_1$ means that any increase to labor force leads to an almost constant output in productivity growth. Hence, as done before, is important to convert the regression in a non-linear function, in particular in a logarithmic regression.

Figure 3.5. *Relationship between productivity and labor force.*



The plot shows the relationship between productivity and labor force for the 10 states included from 1995 to 2018. The left graph represents the estimated linear function, while the right one plot the estimated logarithmic function. The blue line is the estimated regression line.

In this case, Figure 3.5., the situation is slightly different, as it is difficult to understand from the comparison between the two graphs which one is the best model that represents the data. However, the logarithmic model (2) seems to be the best, and since the coefficient values showed that the linear model did not explain the data well, it is crucial to look for another model that explains them better.

Thus, the new model built modifying the linear model is:

$$\widehat{log(Productivity_{i,t})} = \hat{\beta}_0 + \hat{\beta}_1\ log(labor\ force_{i,t}).$$

In order to verify that numerically the logarithmic regression is actually better, it is necessary to look at regression's statistics, including coefficients significance and fits. Coefficients estimations are reported in table 3.2.

Table 3.2. *Regression statistics of the impact of labor force on productivity.*

	(2)
(Intercept)	1.288***
	(0.182)
log(LABFORCE)	0.258***
	(0.018)
Num.Obs.	240
R2	0.295
R2 Adj.	0.292
AIC	176.2
BIC	183.1
RMSE	0.35
Std.Errors	Heteroskedasticity-robust
* p < 0.1, ** p < 0.05, *** p < 0.01	

The table shows a summary of the regression model estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

Through table 3.2. it is possible to confirm that the logarithmic model (2) explains the data better than the linear model. All the coefficients are significant and the labor force coefficient $\hat{\beta}_1$has significantly increased. It moved from 0.001 to 0.258. In the log-log non-linear regression model it means that a 1% change in labor force is associated with a 0.26% change in productivity.

The R^2 value has slightly increased from 0.273 to 0.295. It means that this model explains the data better than the linear model.

3.2.3. PRODUCTIVITY AND PATENT ACTIVITY

The last independent variable studied is patent activity, as the number of patents obtained by a state in each year. The first step is to calculate the covariance index to verify the direction of the linear bond. The result is equal to 35396.39. The positive value indicates that when a variable increases the other one increases too, because the two variables vary in the same direction.

The simple linear regression model is built as:

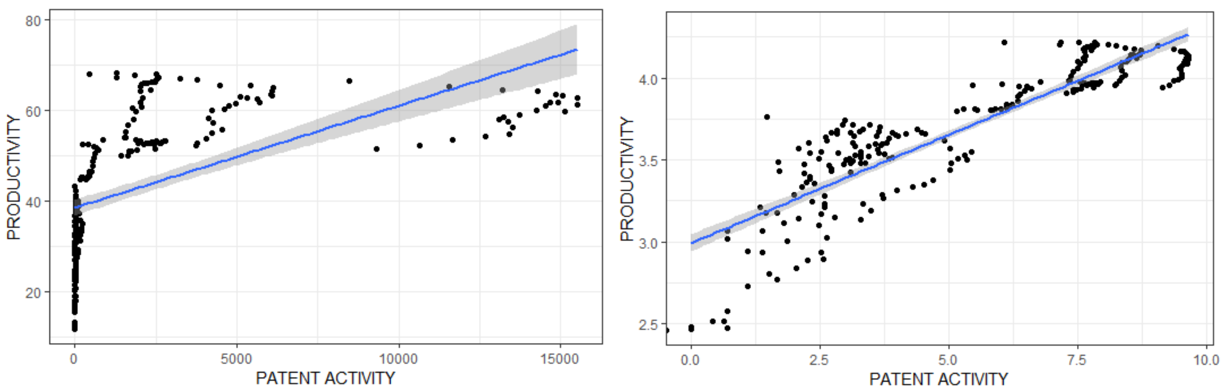
$$\widehat{Productivity}_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 patent\ activity_{i,t},$$

where *i* represents states and *t* represents years.

Estimating it with OLS, $\hat{\beta}_0$ results 38.451, $\hat{\beta}_1$ results 0.002, standard errors result 0.000 and the R^2 is equal to 0.343. The value of $\hat{\beta}_1$ means that any increase to labor force leads to an almost constant output in productivity growth. Hence, as done before, is important to convert the regression in a non-linear function, in particular in a logarithmic regression.

Figure 3.6. *Relationship between productivity and patent activity.*

Linear **Logarithmic (3)**



The plot shows the relationship between productivity and patent activity for the 10 states included from 1995 to 2018. The left graph represents the estimated linear function, while the right one plot the estimated logarithmic function. The blue line is the estimated regression line.

As it happened in the R&D investment model, the Figure 3.6. shows how the log-log regression function (3) graphically represents better the relationship between the two variables than the linear regression. Thus, the new model built modifying the linear model is:

$$log(\widehat{Productivity_{i,t}}) = \hat{\beta}_0 + \hat{\beta}_1\, log(patent\ activity_{i,t}).$$

In order to verify that even numerically the logarithmic regression is better, it is necessary to look at regression's statistics, including coefficients significance and fits. Coefficients estimations are reported in table 3.3.

Table 3.3. *Regression statistics of the impact of patent activity on productivity.*

	(3)
(Intercept)	2.994***
	(0.039)
log(PATENTS)	0.132***
	(0.006)
Num.Obs.	239
R2	0.783
R2 Adj.	0.782
AIC	−114.4
BIC	−107.4
RMSE	0.19
Std.Errors	Heteroskedasticity-robust

* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression model estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

Table 3.2. shows that the coefficients value confirm that the logarithmic regression model better fits the data. All the coefficients are significant and the coefficient $\hat{\beta}_1$ of the patents' log increased from

0.002 to 0.132. This means that the patent activity now better explain the variation of the productivity. In the log-log model it means that a 1% change in patent activity is associated with a 0.13% change in productivity. The R^2 value has significantly increased from 0.343 to 0.783. It means that 78% of the sample variance productivity is explained by the regressor.

3.3. LINEAR REGRESSION MODEL WITH MULTIPLE REGRESSORS

In previous paragraphs the independent variables R&D investment, labor force and patent activity, were taken and examined in 3 different models. However, since each independent variable explains the dependent variable and they are all positively correlated (as observed in the correlation matrix in Figure 3.3.), it is important to create a new linear regression model including all of them to jointly study their effect.

Thereby, the new multiple regression model will be:

$$\widehat{Productivity}_{i,t} = 36.198 + 0.002\ R\&D\ investment_{i,t} + 0.000\ labor\ force_{i,t}\ + \\ 0.000\ patent\ activity_{i,t}.$$

In this case $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ are almost equal to zero. The intercept and R&D investment coefficients are the only significant ones. While labor force and patent activity coefficients are not significant. This means that any change in labor force and patent activity will not have consequences on productivity. The estimated $\hat{\beta}_1$ is equal to 0.002, and indicates that an additional unit in R&D investment, keeping constant labor force and patent activity, leads to an increase in productivity of 0.002. Instead, the estimated $\hat{\beta}_0$ equals 36.198, is defined as the expected value of productivity when the 3 regressors coefficients equal 0. Finally, the R^2 value, equal to 0.518, indicates that the model has a moderate ability to adapt to the data, but there is still a significant amount of unexplained or residual variation that is not considered by the model.

As seen before, the logarithmic transformations of all the variables better explain the data in the simple regression model. Given that, those transformation will be used to build also a multiple regression model:

$$log(\widehat{Productivity}_{i,t}) = \hat{\beta}_0 + \hat{\beta}_1\ log(R\&D\ investment_{i,t}) + \hat{\beta}_2\ log(labor\ force_{i,t}) + \\ \hat{\beta}_3\ log(patent\ activity_{i,t}).$$

Coefficients estimations are reported in table 3.4.

Table 3.4. *Regression statistics of the impact of R&D investment, labor force and patent activity on productivity.*

	(4)
(Intercept)	5.025***
	(0.111)
log(RD)	0.300***
	(0.028)
log(LABFORCE)	−0.455***
	(0.024)
log(PATENTS)	0.098***
	(0.011)
Num.Obs.	239
R2	0.925
R2 Adj.	0.924
AIC	−365.3
BIC	−351.4
RMSE	0.11
Std.Errors	Heteroskedasticity-robust

* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression model estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

It is interesting to notice that, all the coefficients have drastically changed. $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ became all significant. The intercept coefficient has decreased to 5.025. This value is the expected value of productivity when the 3 regressors coefficients equal 0. Conversely, $\hat{\beta}_1$ has increased to 0.300. This means that a 1% increase in R&D investment, keeping constant labor force and patent activity, leads to an increase in productivity of 0.3%. In the same way, $\hat{\beta}_3$ has increased to 0.098. A 1% increase in patents leads to a 0.09% increase in productivity, keeping the other regressors constant.

Differently, what happened to labor force was a change of direction. It became negative, equal to negative 0.455. It means that a 1% increase in labor force leads to a decrease in productivity of 0.45%. This could be explained by the fact that introducing the three regressors at the same time the effect that the individual has on the dependent variable is reduced. Part of the positive effect that was explained in the linear model by the variable labor force is captured by the other regressors. The productivity variable, as explained in paragraph 2.1.1., is the result of the fraction between GDP and labor force. Therefore, since labor force is at the denominator, the coefficient $\hat{\beta}_2$ will be negative. Finally, the R2 adjusted has truly increased, going up to 0.924, showing that the model has an excellent ability to adapt to the data.

3.3.1. ENTITY FIXED EFFECT MODEL

Even if model (4) properly represents the data, it's important to study the regression controlling for fixed effect.[24] Entity fixed effect is a method used to control for omitted variables when they vary across entities (states) but do not change over time.

The equation becomes:

$$log(\widehat{Productivity_{i,t}}) = \hat{\beta}_1\ log(R\&D\ investment_{i,t}) + \hat{\beta}_2\ log\big(labor\ force_{i,t}\big) + \hat{\beta}_3\ log(patent\ activity_{i,t}) + \alpha_i.$$

In the regression model, α_i known as entity fixed effects, are treated as unknown intercepts to be estimated, that varies from state to state. State-specific intercepts can also be expressed using binary variables D to denote individual states. In order to do it, D_{1i} is taken as a binary variable equal to 1 when $i = 1$ and equal to 0 otherwise, and so on. The binary variable D_{1i} for the first entity is arbitrarily omitted to avoid multicollinearity.

Consequently, the model can be equivalently written as:

$$log(\widehat{Productivity_{i,t}}) = \hat{\beta}_0 + \hat{\beta}_1\ log(R\&D\ investment_{i,t}) + \hat{\beta}_2\ log\big(labor\ force_{i,t}\big) + \hat{\beta}_3\ log(patent\ activity_{i,t}) + \hat{\gamma}_1 D(FRANCE)_i + \ldots + \hat{\gamma}_{10}\ D(ROMANIA)_i,$$

where $\beta_0, \beta_1, \beta_2, \beta_3, \hat{\gamma}_1, \ldots, \hat{\gamma}_{10}$ are unknown coefficients to be estimated.[25]

[24] See paragraph 2.2.3.1.

[25] The States included are the 10 listed at the beginning of the chapter.

Table 3.5. *Regression statistics of the impact of R&D investment, labor force and patent activity on productivity controlling for entity fixed effect.*

	4	5
(Intercept)	5.025***	
	(0.111)	
log(RD)	0.300***	0.272***
	(0.028)	(0.033)
log(LABFORCE)	−0.455***	−0.411***
	(0.024)	(0.158)
log(PATENTS)	0.098***	0.103***
	(0.011)	(0.018)
Num.Obs.	239	239
R2	0.925	0.948
R2 Adj.	0.924	0.946
R2 Within		0.684
R2 Within Adj.		0.680
AIC	−365.3	−435.5
BIC	−351.4	−390.3
RMSE	0.11	0.09
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust
FE: STATE		X

* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression models estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

It is clear that with respect to model (4) the results did not drastically change. R&D coefficient has slightly decreased, while labor force and patent activity coefficients has slightly increased. In addition, R^2 adjusted value has increased only of 0.022. Finally, when comparing different models it is possible to use additional metric or indexes. For example, the AIC and BIC indexes work in the opposite way to R^2.[26] The lowest those values are the most reliable is the model. Last, RMSE, a farther common metric used to evaluate the performance of regression model, works as AIC and BIC, therefore, the lower it is, the better is the model. Thus, it is possible to conclude that model (5) seems to be the best choice so far.

[26] The Information Criteria (AIC, BIC) are used to evaluate the goodness of fitness of a model, have a comparative value and are not normalized.

3.3.2. TIME FIXED EFFECT MODEL

Time fixed effect is another instrument to control for omitted variable. Unlike entity fixed effect, it controls for variables that vary across years but remain constant over states. The equation becomes:

$$log(\widehat{Productivity_{i,t}}) = \hat{\beta}_1 \, log(R\&D \; investment_{i,t}) + \hat{\beta}_2 \, log(labor \; force_{i,t}) + \hat{\beta}_3 \, log(patent \; activity_{i,t}) + \lambda_t.$$

This model has a different intercept, λ_t, for each time period. Just like entity fixed effect, intercepts can be expressed using binary variables B to denote each year.

$$\log(\widehat{Productivity_{i,t}}) = \hat{\beta}_0 + \hat{\beta}_1 \log(R\&D \; investment_{i,t}) + \hat{\beta}_2 \, log(labor \; force_{i,t}) + \hat{\beta}_3 \, log(patent \; activity_{i,t}) + \hat{\delta}_1 B(1995)_t + \cdots + \hat{\delta}_{24} B(2018)_t,$$

where $\beta_0, \beta_1, \beta_2, \beta_3, \hat{\delta}_1, \ldots, \hat{\delta}_{24}$ are unknown coefficients to be estimated.

Table 3.6. *Regression statistics of the impact of R&D investment, labor force and patent activity on productivity controlling for entity fixed effect and time fixed effect.*

	4	5	6
(Intercept)	5.025***		
	(0.111)		
log(RD)	0.300***	0.272***	0.249***
	(0.028)	(0.033)	(0.035)
log(LABFORCE)	−0.455***	−0.411***	−0.422***
	(0.024)	(0.158)	(0.028)
log(PATENTS)	0.098***	0.103***	0.116***
	(0.011)	(0.018)	(0.012)
Num.Obs.	239	239	239
R2	0.925	0.948	0.934
R2 Adj.	0.924	0.946	0.926
R2 Within		0.684	0.928
R2 Within Adj.		0.680	0.927
AIC	−365.3	−435.5	−349.3
BIC	−351.4	−390.3	−255.4
RMSE	0.11	0.09	0.10
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust
FE: STATE		X	
FE: YEAR			X

* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression models estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

Comparing model (4) with the new model (6) it is clear that how variables change and the differences are very similar to those observed between model (4) and (5). The R&D investment coefficient has decreased of 0.051, while the labor force and patent activity coefficients have increased of, respectevely 0.033 and 0.018. The R^2 has increased, but only by 0.002. In addition, in the model with time fixed effect, AIC and BIC indexes are higher (and therefore worse) than in models (4) and (5). Finally, the RMSE index is better than the model without fixed effect but not better than the model with entity fixed effect. Hence, it is possible to conclude that the model with time fixed effect did not improve the results, compared to linear model.

The last option is to control for both entity and time fixed effect simultaneously. It means merging together models (5) and (6). The combined entity and time fixed effects regression model eliminates omitted variables bias arising both from unobserved variables that are constant over time and from unobserved variables that are constant across states. The time fixed effects model and the entity fixed effects model are both variants of the multiple regression model.

The combined regression model is:

$$log(\widehat{Productivity_{i,t}}) = \hat{\beta}_1\ log(R\&D\ investment_{i,t}) + \hat{\beta}_2\ log\big(labor\ force_{i,t}\big) + \hat{\beta}_3\ log(patent\ activity_{i,t}) + \alpha_i + \lambda_t.$$

Term α_i represents the entity fixed effect and term λ_t represents the time fixed effect. As in entity fixed effect model and time fixed effect model, this model can be estimated by OLS. It can be equivalently written using the binary indicators n - 1 entities and T - 1 times.

The equation becomes:

$$log(\widehat{Productivity_{i,t}}) = \hat{\beta}_0 + \hat{\beta}_1\ log(R\&D\ investment_{i,t}) + \hat{\beta}_2\ log\big(labor\ force_{i,t}\big) + \hat{\beta}_3\ log(patent\ activity_{i,t}) + \hat{\gamma}_1 D(FRANCE)_i + \ldots + \hat{\gamma}_{10}\ D(ROMANIA)_i + \hat{\delta}_1 B(1995)_t + \cdots + \hat{\delta}_{24} B(2018)_t.$$

The unknown coefficients of the regression to be estimated through OLS are: β_0, β_1, β_2, β_3, $\hat{\gamma}_1$, ..., $\hat{\gamma}_{10}$, and $\hat{\delta}_1$, ..., $\hat{\delta}_{24}$.[27]

Coefficients estimations are reported in table 3.7.

[27] In both fixed effect regression models the intercept is absent because it has been eaten by the fixed effect estimation.

Table 3.7. *Regression statistics of the impact of R&D investment, labor force and patent activity on productivity, controlling for entity fixed effect, time fixed effect and both combined.*

	4	5	6	7
(Intercept)	5.025***			
	(0.111)			
log(RD)	0.300***	0.272***	0.249***	0.174***
	(0.028)	(0.033)	(0.035)	(0.035)
log(LABFORCE)	−0.455***	−0.411***	−0.422***	−0.725***
	(0.024)	(0.158)	(0.028)	(0.185)
log(PATENTS)	0.098***	0.103***	0.116***	0.116***
	(0.011)	(0.018)	(0.012)	(0.016)
Num.Obs.	239	239	239	239
R2	0.925	0.948	0.934	0.961
R2 Adj.	0.924	0.946	0.926	0.955
R2 Within		0.684	0.928	0.486
R2 Within Adj.		0.680	0.927	0.478
AIC	−365.3	−435.5	−349.3	−458.5
BIC	−351.4	−390.3	−255.4	−333.4
RMSE	0.11	0.09	0.10	0.08
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust
FE: STATE		X		X
FE: YEAR			X	X

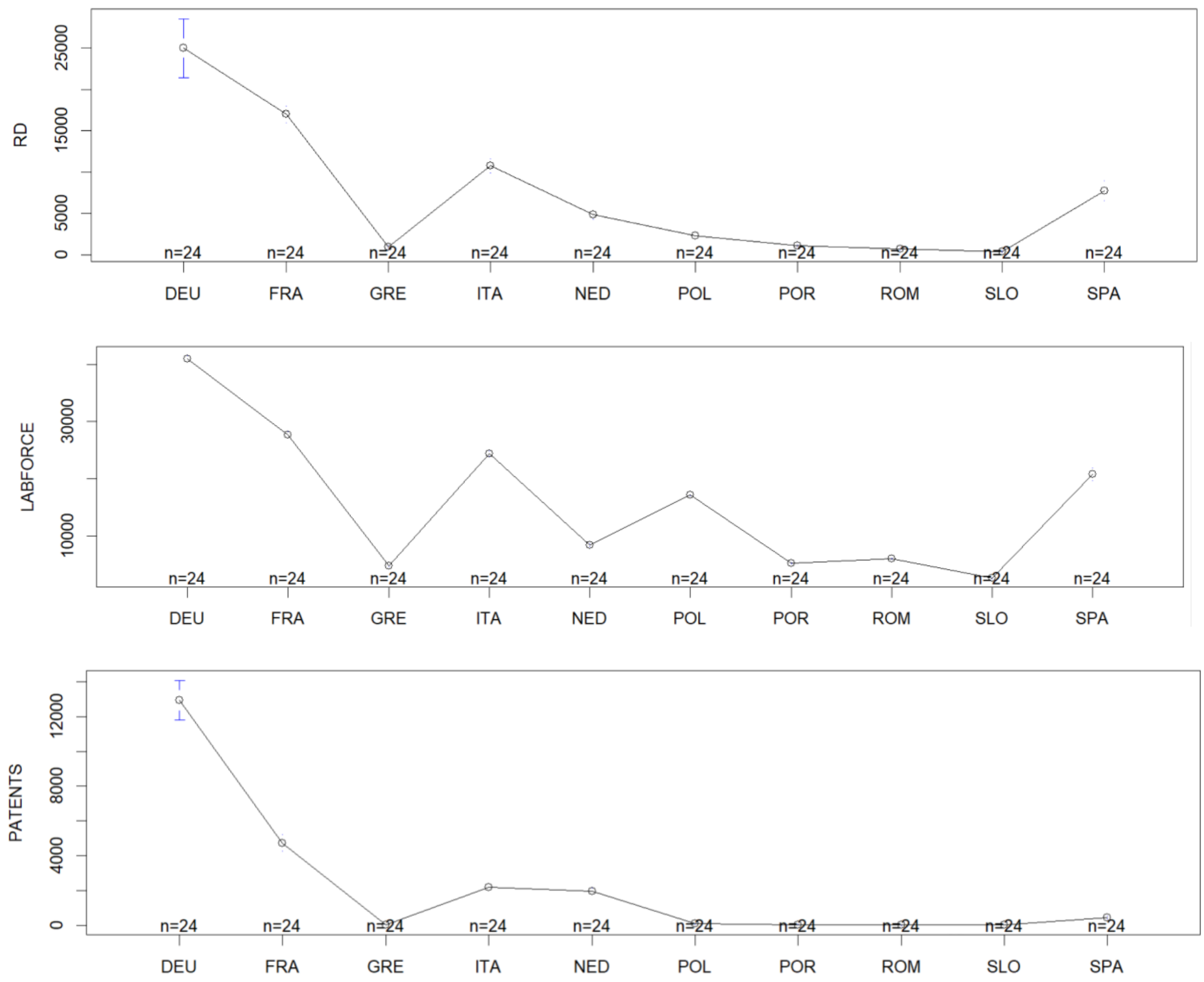
* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression models estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

The last model (7) includes both entity and time fixed effect. The three regressors coefficients slightly changed. R2 has increased and it reached its highest value, so it means that this model is the one that better fits the data. The same conclusion can be drawn by looking at index AIC and RMSE values that reachest their lowest point. Therefore, model (7) is the chosen one.

To conclude this section of the analysis is interest to study the variables for heterogeneity. This index will be study across states.

Figure 3.7. Heterogeneity of R&D investment, labor force and patent activity across states.



The graphs show the heterogeneity of the R&D investments, labor force and patent activity across states from 1995 to 2018.

Studying figure 3.7. it is clear how Germany results to be the richest country in the period from 1995 to 2018 and it has a really high value of all the three regressors. On the contrary, even if, both Italy and Spain have a medium level of R&D investment and labor force, the numer of patents that they gratend was definitely low.

3.3.3. DYNAMIC CAUSAL EFFECT

The following step, is to estimate the effect on productivity now and in the future, of a change in R&D investment, labor force and patent activity. This means to estimate dynamic causal effect. The model used is the distributed delay regression model. Lags must be incorporated. The new model is built as:

$$\widehat{log(Productivity_{i,t})} = \hat{\beta}_0 + \hat{\beta}_1\, log(R\&D\ investment_{i,t}) + \hat{\beta}_2\, log(labor\ force_{i,t}) + \hat{\beta}_3\, log(patent\ activity_{i,t}) + \hat{\beta}_4\, log(R\&D\ investment_{i,t-1}) + \hat{\beta}_5\, log(labor\ force_{i,t-1}) + \hat{\beta}_6\, log(patent\ activity_{i,t-1}) + \hat{\gamma}_1 D(FRANCE)_i + \ldots + \hat{\gamma}_{10}\, D(ROMANIA)_i + \hat{\delta}_1 B(1995)_t + \cdots + \hat{\delta}_{24} B(2018)_t.$$

Table 3.8. *Regression statistics of the impact of R&D investment, labor force and patent activity on productivity controlling for fixed effect and including lags.*

	7	8
log(RD)	0.174***	0.078
	(0.035)	(0.063)
log(LABFORCE)	−0.725***	0.604
	(0.185)	(0.597)
log(PATENTS)	0.116***	0.053
	(0.016)	(0.033)
log(RDL)		0.092
		(0.063)
log(PATENTSL)		0.076**
		(0.031)
log(LABFORCEL)		−1.145**
		(0.486)
Num.Obs.	239	228
R2	0.961	0.965
R2 Adj.	0.955	0.959
R2 Within	0.486	0.525
R2 Within Adj.	0.478	0.510
AIC	−458.5	−474.8
BIC	−333.4	−344.5
RMSE	0.08	0.07
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust
FE: STATE	X	X
FE: YEAR	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression models estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

The coefficient $\hat{\beta}_1$ is the contemporary or immediate effect of a unitary variation in R&D investment on contemporary value of productivity. In turn, the coefficient $\hat{\beta}_4$ is the effect on productivity in t of a unitary variation in R&D investment in t-1. Similarly, $\hat{\beta}_2$ and $\hat{\beta}_5$ represent the effects on productivity in t of a unit change in labor force, respectively in t and in t-1. Finally, in the same way $\hat{\beta}_3$ and $\hat{\beta}_6$ are the effect on productivity in t of a unit change in patent activity, respectively in t and in t-1.

Table 3.8. shows how introducing lags changes all the coefficents. The three regressors' coefficients lose their significance. It happens because it is captured by the same regressors in t-1. It means that the results of the effect of R&D investment, labor force and patent activity on productivity growth, is not immediate and is manifested over time. Moreover, the variable labor force in t-1 turns negative because, as explained in paragraph 3.3., it is the denominator of the fraction GDP/labor force that is the definition of productivity. Thus, since labor force is on the other side of the equation with respect to productivity, the relationship between the two variables is inverse.

Finally, the variable R&D investment in t-1 loses its significance. This could be explained by the fact that R&D investment's significance is captured by patents. To verify if, and how, R&D investment explain patents, a linear regression model was run. The coefficient of R&D investment is equal to 0.499 and it is significant. In addition, the R2 is equal to 0.980. Given these results, it is possible to conclude that the regressor R&D investment explains and has an effect on the dependent variable, patent activity.

3.3.4. CLUSTERED STANDARD ERRORS

As said in paragraph 2.2.3.3., the standard errors used for the regression are considered invalid if the errors are autocorrelated. The standard errors of heteroscedasticity and robust autocorrelation (HAR) could be used. HAR allow regression errors to have arbitrary correlation within a cluster, or grouping, but assume that regression errors are uncorrelated across clusters.

Results of model (8) using clustered standard errors are shown in model (9). The observations are clustered by state.

Coefficients estimations are reported in table 3.9.

Table 3.9. *Regression statistics of the impact of R&D investment, labor force and patent activity on productivity, controlling for fixed effects, including lags and using HAR.*

	8	9
log(RD)	0.078	0.078
	(0.063)	(0.072)
log(LABFORCE)	0.604	0.604
	(0.597)	(0.897)
log(PATENTS)	0.053	0.053***
	(0.033)	(0.010)
log(RDL)	0.092	0.092**
	(0.063)	(0.038)
log(PATENTSL)	0.076**	0.076**
	(0.031)	(0.028)
log(LABFORCEL)	−1.145**	−1.145*
	(0.486)	(0.605)
Num.Obs.	228	228
R2	0.965	0.965
R2 Adj.	0.959	0.959
R2 Within	0.525	0.525
R2 Within Adj.	0.510	0.510
AIC	−474.8	−474.8
BIC	−344.5	−344.5
RMSE	0.07	0.07
Std.Errors	Heteroskedasticity-robust	by: STATE
FE: STATE	X	X
FE: YEAR	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01

The table shows a summary of the regression models estimated by OLS. The stars next to the coefficients represent the level of significance of each coefficient according to values described in the last row. The standard errors are the values under coefficients in parentheses.

Comparing model (8) and (9), in table 3.9., it is possible to conclude that there is correlation within clusters, that are, groups created by countries. This conclusion is justified by the fact that the three regressors have effects not only on the productivity variable in year t and subsequent years, but also on the variables themselves in the years following t. So, using clustered standard error improves the analysis results. Indeed, patent activity in t and R&D investment in t-1 have recovered their significance. While, patents in t-1 and labor force in t-1 remained significant. Finally, R&D investment and labor force in t remained not significant as their effect is captured by the other variables. In the end, it is possible to conclude that it results to be the best version of the initial model.

CONCLUSION

This study aims to answer the question posed in the introduction: Does innovation have a positive and significant effect on a country's productivity growth?

In this analysis, innovation is proxied by three variables: R&D investment, labor force and patent activity. The goal is to answer this open question. The analysis that has been carried out in this thesis includes the collection of data, from the OECD site, relating to a sample of 10 European economies, including 5 cores and 5 peripherals, for a period of 24 years, from 1995 to 2018. The empirical analysis has instead been structured by the estimation of different econometric models.

The second chapter describes both the data and the econometric methodologies, while the third chapter presents and discusses all the results of the analysis.

As regards to the simple regression function between R&D investments and productivity, the function has been transformed into logarithmic (log-log model), as this model explains the data better than the linear one. The effect on productivity of this variable is positive and significant and the R^2 is equal to 0.65.

Similarly, simple regression models that include individually labor force and patent activity are transformed to logarithmic functions. The estimated coefficients are both positive and significant, and the R^2 are, respectively, 0.29 and 0.78.

However, since each independent variable explains the dependent variable and the variables are all positively correlated, a new multiple linear regression model has been created, to jointly study the effect of the three independent variables on productivity. The results show that all the regressors remain significant, but with the labor force coefficient having a negative sign. The reason behind this estimate could be mechanical since the productivity has been measured as the ratio between GDP and labor force. In addition, the value of R^2 increases and reaches 92%.

The inclusion of fixed effect to the model does not significantly change the estimates. When you simultaneously control for omitted variables that change over time but are constant between states, and change between states but are constant over time, the results improve, since the regressors retain the same significance, but the R^2 reaches a value of 96%.

Given the nature of the variables included in the analysis, which tend to have a prolonged effect over time, the model has been implemented including lags. The results change a lot, and all the variables lose their significance at time t, while only patent activity and labor force (which remained negative) are significant at time t-1.

For this reason, and to make sure the interpretation of the data is correct, errors must not be left out. In fact, the standard error used for the simple regression are considered invalid if the errors are autocorrelated. For this reason, the regression model is modified, and clustered standard errors are included. The consequence is that the results change, and indeed improve. This means that there is a correlation within clusters, that are, groups created by countries. This conclusion may be justified by the fact that, as mentioned above, the three regressors also affect the variables themselves in the years following t, and not only productivity. Indeed, the coefficients of the number of patents in t and in t-1 are both positive and significant; the labor force coefficient in t-1 is significant and negative (for the mechanical inverse relationship previously discussed) and the R&D investment coefficient in t-1 is positive and significant. While the coefficients of R&D investment and labor force remain insignificant in t, because their effect is captured by the other variables. In addition, the R2 of the final model is equal to 0.96 and therefore this means that the model explains the data well.

In light of these results, it can be concluded that, for the sample analyzed, the model fits the data and that the three regressors have significant effects on a country's productivity growth.

However, in our ever-changing world, it is essential to carry out continuous analysis and never stop studying the relationship between all these variables.

BIBLIOGRAPHY

Aghion, Philippe (2018). "Innovation and growth from a Schumpeterian Perspective". *In Revue d'économie politique Volume 128, Issue 5, pages 693 to 711.* *https://www.cairn-int.info/article.php?ID_ARTICLE=E_REDP_285_0693*

Aghion, Philippe; Akcigit, Ufuk; Howitt, Peter (2015). "The Schumpeterian Growth Paradigm". *Annual Review of Economics 7:1, 557-575.* https://www.annualreviews.org/doi/abs/10.1146/annurev-economics-080614-115412

Bassanini, Andrea; Scarpetta, Stefano; Visco, Ignazio (2000). "Knowledge, technology and economic growth: recent evidence from OECD countries". *National Bank of Belgium Working paper No.6.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1705109

Bloom, David E.; Sousa-Poza, Alfonso (2013). "Ageing and Productivity: Introduction". *IZA Discussion Paper No. 7205.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2219130

Boldrin, Michele; Levine, David K. (2013). "The case against patents". *Journal of Economic Perspectives, 27(1), 3-22.* https://www.aeaweb.org/articles?id=10.1257/jep.27.1.3

Calvino, Flavio; Virgillito, Maria Enrica (2018). "The innovation-employment nexus: a critical survey of theory and empirics". *Journal of Economic surveys, 32(1), 83-117.* https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12190

Dalgaard, Carl-Johan; Strulik, Holger (2013). "The history augmented Solow model". *European Economic Review, 63, 134-149.* https://www.sciencedirect.com/science/article/abs/pii/S0014292113000871

Dixon, Sylvia (2003). "Implications of population ageing for the labour market". *Labour Market Trends, 111(2), 67-76.* *http://www.re-integrate.eu/resources/webre--implications-of-population-ageing-for-the-labour-market.pdf*

Dosi, Giovanni (1982). “Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change”. *Research policy, 11(3), 147-162.* https://www.sciencedirect.com/science/article/abs/pii/0048733382900166

Durlauf, Steven N.; Kourtellos, Andros; Minkin, Artur (2001). “The local Solow growth model”. *European Economic Review, 45(4-6), 928-940.* https://www.sciencedirect.com/science/article/abs/pii/S0014292101001209

Fagerberg, Jan; Mowery, David; Nelson, Richard (2003). “Innovation: A Guide to the Literature”. *The Oxford Handbook of Innovation, 1-26.* https://www.researchgate.net/publication/24134940_Innovation_A_Guide_to_the_Literature

Gambardella, Alfonso (2013). “The economic value of patented inventions: Thoughts and some open questions”. *International Journal of Industrial Organization, 31(5), 626-633.* https://www.sciencedirect.com/science/article/abs/pii/S0167718713000180

Gambardella, Alfonso; Harhoff, Dietmar; Verspagen, Bart (2005). “The value of patents”. *Universita' Bocconi, Ludwig-Maximiliens Universitaet, and Eindhoven University, Working Paper.* *https://ftp.zew.de/pub/zew-docs/veranstaltungen/inno_patenting_conf/GambardellaHarhoffVerspagen.pdf*

Guellec, Dominique; Van Pottelsberghe de la Potterie, Bruno (2001). “R&D and productivity growth: panel data analysis of 16 OECD countries”. *OECD Economic Studies 2001(2), 103-126.* https://www.oecd.org/economy/growth/1958639.pdf

Johnson, Björn; Edquist, Charles; Lundval, Bengt-Åke (2003). “Economic Development and the National System of Innovation Approach”. In *The First International Globelics Conference: Innovation systems and development strategies for the third millenium, Rio de Janeiro, November 2-6, 2003.* https://vbn.aau.dk/en/publications/economic-development-and-the-national-system-of-innovation-approa

Joseph, Federico Pasquale (1929). “Origin and early history of patents”. *J. Pat. Off. Soc'y, 11, 292.* https://heinonline.org/HOL/LandingPage?handle=hein.journals/jpatos11&div=62&id=&page=

Kortum, Samuel S. (1997). “Research, patenting, and technological change”. *Econometrica: Journal of the Econometric Society, 1389-1419.* https://www.jstor.org/stable/2171741

Maestas, Nicole; Kathleen, J. Mullen; Powell, David (2016). “The effect of population ageing on economic growth, the labor force and productivity”. *National Bureau of Economic Research, No. w22452.* https://www.nber.org/papers/w22452

Moser, Petra (2013). “Patents and Innovation: Evidence from Economic History”. *Journal of Economic Perspectives, 27(1), 23-44.* https://www.aeaweb.org/articles?id=10.1257/jep.27.1.23

Schaafsma, Paul E. (1997). “An Economic Overview of Patents”. *J. Pat. & Trademark Off. Soc'y, 79, 241.* *https://heinonline.org/HOL/LandingPage?handle=hein.journals/jpatos79&div=35&id=&page=*

Visco, Ignazio (2003). “Perché non si può crescere senza ricerca”. *Banca d’Italia, 34-42.* https://www.bancaditalia.it/chi-siamo/funzioni-governance/direttorio/ignazio-visco/visco-pubblicazioni/Perche-non-si-puo-crescere-senza-ricerca.pdf

SITOGRAPHY

https://stats.oecd.org/

RINGRAZIAMENTI

Vorrei riservare questo spazio finale della mia tesi di laurea ai ringraziamenti verso tutti coloro che hanno contribuito, con il loro supporto, ad arricchire il mio percorso di crescita universitario e personale.

Un primo ringraziamento va al mio Relatore, il professore Mattia Guerini, per la grande disponibilità e professionalità dimostratami in questi mesi di lavoro: i suoi suggerimenti e il suo sostegno sono stati preziosi per il raggiungimento di questo traguardo.

Ai miei genitori, che mi hanno aiutato a spiegare le ali e con tanto amore mi hanno insegnato a inseguire i miei sogni. A loro che credono nella persona che sono ora e in quella che diventerò.

Alle due metà del mio cuore, le mie sorelle, Carolina e Matilde. A loro che, ogni giorno, senza sapere di farlo, mi spronano a diventare la versione migliore di me per loro.

Ai miei nonni, per credere in me ed essere sempre i miei primi sostenitori.

A Chiara, alla mia compagna di vita sin dall'asilo, la mia terza sorella, alla persona con cui sono cresciuta e che è presenza costante, in ogni successo e insuccesso.

Alle mie amiche di una vita, a quelle 8 pazze, con le quali negli anni ho creato un numero infinito di bei ricordi che porto nel cuore, e che tra alti e bassi, mi hanno sempre dimostrato di poter contare su di loro in qualsiasi momento.

A Robi, alla sua capacità e generosità nell'ascoltarmi, supportarmi e spronarmi, anche quando i miei sogni mi portano lontano da casa. Al suo cuore grande e ai nostri sogni ancora più grandi.

All'Erasmus, al Portogallo, ad Aveiro, e ai miei "Pastinhos". All'esperienza più bella, tosta, intensa, ricca, formativa, e stimolante. A tutti i pianti e le risate, a tutte le emozioni vere e a tutto ciò che mi ha insegnato.

A tutti i miei compagni di università, con cui ho condiviso le gioie e i dolori di questo percorso, e che, grazie al loro sostegno, hanno reso questi anni di studio più leggeri.

A tutte le persone che non ho citato ma che, per un lungo o breve periodo, in questi anni, hanno condiviso tempo, consigli e risate con me.

Infine, un ringraziamento speciale va a mia mamma, che esattamente un anno fa mi scrisse: "Viaggiare, allontanarsi dalle certezze, dalle comodità e dagli affetti non è facile… andare in contro all'ignoto è sempre un'avventura, che ti porta a capire te stessa ed a crescere. Sono così orgogliosa del tuo coraggio. Della tua capacità di uscire dalla comfort zone per entrare più in profondità nella vita e crescere."

A te Mamma, che mi hai insegnato a volare.

Arianna