Utrecht University

Contribution of Horizon 2020 to Artificial Intelligence

A research on publications and patents in artificial intelligence in the European Union.

Master Thesis
Human Geography – Economic Geography: Regional Development & Policy

Author: A. de Wilde
Student number: 6750907
Supervisor: Dr. P.A. Balland

Contribution of Horizon 2020 to Artificial Intelligence

A research on publications and patents for artificial intelligence in the European Union

Author: A. de Wilde
Student number: 6750907
E-mail address: a.dewilde@students.uu.nl
Master thesis

Study programme: Human Geography – Economic Geography: Regional Development & Policy
Organisation: Utrecht University
Supervisor: Dr. P.A. Balland
Graduation period: March – June 2021
Utrecht, 18 June 2021

Preface

Before you lies the thesis 'Contribution of Horizon 2020 to Artificial Intelligence.' The research for this thesis was conducted on the basis of quantitative data from the European Union and the OECD REGPAT database. This thesis was written as part of my graduation from the Master Human Geography, track Economic Geography: Regional Development & Policy, at Utrecht University. I worked on this thesis from March 2021 until June 2021.

I came up with this subject and research question myself. My supervisor Dr. P.A. Balland helped me to further specify my research question and also helped me with patent data from the OECD. This research was challenging at the beginning. I used the programming language R to answer my research question. During the thesis I learned a lot and greatly improved my R skills. Looking back at this period I am thankful for doing this because it is also helpful for after my graduation.

I would like to thank Dr. P.A. Balland for helping me during this thesis period. He gave me some good insights to improve my work and look at it from another perspective. I would also like to thank my family for feedback and support.

I hope you enjoy reading this thesis.

Anouk de Wilde

Utrecht, 18 June 2021

Summary

This research aims to provide insights into the role of Horizon 2020 funding on the obtaining of scientific publications and patents in artificial intelligence (AI). This is because the European Union is falling behind in the AI race. Europe is taking steps to upgrade the continent's ambitions for AI leadership and to avoid becoming a consumer of AI in the future. The goal of Horizon 2020 was to have more breakthroughs, discoveries and world-first by taking great ideas from the lab to the market. This programme had to close the innovation gap with China and the United States. The innovation gap in AI is often measured by the number of AI-related scientific publications and patents. Although, after this funding programme came to an end, literature still shows that Europe is falling behind. Therefore, the aim of this research is to investigate the relationship between Horizon 2020 funding and the number of AI-related scientific publications and patents.

Quantitative data for this research is used and collected with desk research. The data came from the European Union and OECD REGPAT database (January 2021). Before the contribution is measured, different maps are made to visualize the data. To estimate the contribution of Horizon 2020, the correlation is measured and different multivariate regression models are made.

This research shows that there is no significant relationship or effect between the number of AI-related patents and Horizon 2020 funding after controlling for GDP per capita, employment and population. Only the variables GDP per capita, employment and population contribute to the number of AI-related patents. When it comes to the AI-related scientific publications, the Horizon 2020 funding does contribute to the number of AI-related scientific publications. The three variables GDP per capita, employment and population do not contribute to the number of AI-related scientific publications. The conclusion is therefore that Horizon 2020 contributes to the number of AI-related scientific publications but it does not for the AI-related patents.

It must be said that for this research all AI-related patents are used, including patents that are not directly linked to Horizon 2020 funding. Because of this, the results for the AI-related patents are more reliable. However, for the AI-related scientific publications only Horizon 2020 AI-related scientific publications are used. This explains the strong relationship between Horizon 2020 funding and AI-related scientific publications. If this research is done again with non-Horizon 2020 and Horizon 2020 AI-related scientific publications, it can be expected that the results may be different. The significant relationship might not be as strong as it is now.

A suggestion for follow-up research is to look at the indirect effect of Horizon 2020 funding on the attraction of private investments in regions. Public funding is able to attract more private investments, and private investments are known to obtain more patents since they are more commercially orientated.

Table of Contents

Maps, Figures and Tables List

1. Introduction

Innovation and economic growth go hand-in-hand. Schumpeter (1934, 1942) and Nelson (1959) are both well-known economists who showed in their studies that innovation is a key driver of economic growth. Scientific knowledge, which is necessary for innovation, has economic value that stimulates economic growth (Nelson, 1959). Countries all over the world use innovation to stimulate their economic growth. By investing in Research and Development (R&D), countries generate new markets and build strong competitiveness (Bughin et al., 2019).

In 2000, at the Lisbon European Council, the European Union (EU) established the goal to make the EU the 'most competitive and dynamic knowledge based economy in the world'. In so doing, they explicitly acknowledged the gap separating the EU from the world leader, the United States (US). The EU intention was to catch up within a 10-year period (Crescenzi, Rodríguez-Pose & Stroper, 2007). However, the EU did not meet their goal on spending 3% of their GDP on R&D by 2010. Only 2% of the GDP was spend on R&D, which is 0.8 percentage points less than the US (Kekic, 2018). Since 2010, US companies have continued to increase their share of R&D in the world. In 2019 European companies still account for one-quarter of the total industrial R&D in the world. In the global innovation race, Europe seems to be well placed in some key areas and falling behind in others. Such as the amount of tech investments in artificial intelligence (AI), what is expected to transform economies and every aspect of human life over the next couple of decades (Bughin et al., 2019; Brattberg et al., 2020).

To overcome the innovation gap from the EU and to become more competitive, the EU has different funding programs to encourage this. European research and technological development activities have been defined and implemented by a series of multi-annual Framework Programmes (FPs). The FPs have been the main financial tools through which the EU supports research and development activities for almost all scientific disciplines (Collaboration in Research and Methodology for Official Statistics [CROS], n.d.). Horizon 2020 (H2020) is one of these FPs, to be specific, H2020 belongs to the FP8 (European Commission [EC], 2014a). This funding programme was the biggest research and innovation programme in 2014. The goal of H2020 was to have more breakthroughs, discoveries and world-first by taking great ideas from the lab to the market. This programme was in place from 2014 to 2020 (European Commission [EC], 2014).

The European Commission (EC) concluded that Europe is behind in private investments in AI. Which totalled around EUR 2.4-3.2 billion in 2016, compared to EUR 12.1-18.6 billion in the US. It is therefore important to create an environment that stimulates investments and uses public funding to leverage private investments in the EU. The EU invested under the H2020 research and innovation programme around EUR 1.1 billion in AI-related research and innovation during the period 2014-2017. The EU increased this investment with EUR 2.5 billion for 2018-2020. They did this to prevent the risks of losing out. This can result in a brain-drain and becoming a consumer of solutions in AI developed elsewhere (European Commission [EC], 2018). H2020 is the EU funding that fosters AI development. However, there are doubts to whether the volume of financing is really sufficient (Asgard & Roland Berger, 2018). The European aim with this investment was also to increase the private investments. Eventually, a total of EUR 20 billion per year should be invested by the public and private sector in AI by EU Member States (EC, 2018). After the funding programme H2020, there is an H2020's successor; Horizon Europe. Horizon Europe will mainly focus on the pillar addressing global challenges and European industrial competitiveness, which includes AI. This new FP has a larger budget compared to H2020 (European Commission [EC], n.d.). It is therefore important to know if investments from H2020 contributed to overcome the innovation gap, if the funding was well spend and to become more competitive in AI. In addition, the social relevance of this research lies in the fact that Europe tried to avoid becoming a consumer of AI.

Existing literature on the innovation gap for AI measures this gap differently. A study of OECD (2018) found that private equity investment in AI has accelerated since 2016. The US accounts for the majority of AI start-up equity investments worldwide and has two-third of the total value of investments since 2011. The EU on the other hand, accounted for just 8% of global AI equity

investment in 2017. This is an increase since the EU had an investment in 2013 of just 1%. However, member states varied widely in terms of investments levels. United Kingdom (UK) start-ups received 55% of the EU total investment over the period 2011 to mid-2018, followed by German (14%) and French ventures (13%). Another global player that is second in terms of the value of AI start-up equity investments is China. China went from just 3% in 2015 to 36% of global AI private equity investments in 2017. This is a dramatic increase compared to any other country (OECD, 2018).

Another way to measure the innovation gap in AI is by patents. According to Bughin et al. (2019) the ratio of AI-related patents granted in the US versus Europe has grown from less than 2.0 in 2005 to more than 2.5 in 2015. This shows that the gap is widening between the US and Europe. The percentage per capita patents in 2018 is 28% for Europe and 72% for the US (Bughin et al., 2019). According to Castro, McLaughlin and Chivot (2019) it is difficult to measure innovation with patents. This is because there is a difference in the national standards for granting patents. Patents issued by the Chinese Patent Office are from a relatively poor quality, and therefore patent counts from China cannot be compared easily with other patents offices. Therefore, Castro et al. (2019) look at the number of highly cited AI patent families between 1960 and 2018. The number of highly cited AI patents for China is 691, EU 2,985 and the US 28,031 (Castro, McLaughlin & Chivot, 2019). This again shows that there is a large innovation gap between the US and Europe. When it comes to cited AI patents, China lags behind Europe. This however won't take long since investments in China have increased dramatically. Although, within Europe the patent applications are mainly driven by the UK, Germany and France (Brattberg et al., 2020).

AI can also be measured by scientific publications. The World Intellectual Property Organization (WIPO) (2019) found that European researchers are much more active in publishing the results of their research than in patenting. Between 2015 and 2019, the US was the leader in publications with the UK and Germany as second and third. In this period China had an increase of 340% in scientific publications and got the second place in 2019. But, it is not always about the number of publications, it is also about the quality. This can be measured with the number of citations. Publications from China were cited about 20% less than the world average in 2019, and publications from the US were cited about 40% more than average (Savage, 2020). Castro et al. (2019) shows that in 2017, China published 15,199 AI papers, the EU 17,776, and the US 10,287. Historically, according to Castro et al. (2019), the EU has produced the most AI papers. From 1998 to 2017 EU researchers published 164,000 AI papers, compared with 135,000 and 107,000 by Chinese and US authors. From all the AI papers, the US produces the highest-quality research. And in 2016, researchers cited US authors 83% more than the global average. The EU was cited 20% more than the global average (Castro et al., 2019).

The EU is falling behind in AI from the US that is ahead in the AI competition (Bughin et al., 2019). European governments and EU institutions are taking steps to upgrade the continent's ambitions for AI leadership (Brattberg et al., 2020; Joint Research Centre [JRC] & European Commission [EC], 2018). The funding programme H2020 had to fill the innovation gap and was the funding to foster AI development (EC, 2014; Asgard & Roland Berger, 2018). However, the problem is that the literature still shows that Europe is falling behind in AI-related patents and scientific publications after the H2020 funding programme came to an end. It is unknown in the literature if investments from H2020 obtained patents and scientific publications and therefore closed the innovation gap for Europe in AI. Existing literature also shows doubts on whether European countries are able to transfer research input from the EU into research output. This raises the question if the EU contribution is well spend and contributed to the international AI race for Europe. Therefore, the aim of this research is to provide insights into the role of H2020 AI-related funding for the obtaining of AI-related patents and scientific publications. The research question for this research is:

To what extent did Horizon 2020 funding contribute to the obtaining of patents and scientific publications in the European Union for artificial intelligence?

To answer this research question quantitative methods are used and data from different resources is collected. The outline of this research is as follows; first the theoretical framework is presented. Secondly the data and methodology used in this research is described to provide validity and reliability. After this the descriptive statistics are given and results of this research are described. This is done with maps, a correlation matrix and multivariate regression models. The regression models are used to determine the effect of H2020 funding on patents and scientific publications.

2. Theoretical Framework

Investment in R&D has been regarded as one of the key strategies to secure technological potential and, therefore, innovation and economic growth (Bilbao-Osorio & Rodríguez-Pose, 2004). As said before, Schumpeter (1934, 1942) and Nelson (1959) showed in their studies that innovation is a key driver of economic growth. This also accounts for R&D investments, because it increases the possibility of achieving a higher standard of technology in firms and regions. This in turn will result in higher levels of income and economic growth (Bilbao-Osorio & Rodríguez-Pose, 2004). Bilbao-Osorio and Rodríguez-Pose (2004) did research on the impact of public investment in R&D on innovation and economic growth in Europe. There overall result was that there is a positive link between R&D activities and innovation. Although, not all research sectors are equally productive in terms of innovation production. Research activities done in the private sector have higher rates of return than research done in the public sector. This is because privately funded research tends to be more applied (Bilbao-Osorio & Rodríguez-Pose, 2004). Bilbao-Osorio and Rodríguez-Pose (2004) measured innovation with the number of application for patents. They found that privately funded research had a higher number of applications for patents, this is also because patents are strong commercial orientated (Bilbao-Osorio & Rodríguez-Pose, 2004).

In addition Bilbao-Osorio and Rodríguez-Pose (2004) also found that peripheral regions have a lack on private research and therefore university research with public funding investments compensate for this lack. For the efficiency of funding provided through H2020, this implies that the expected outputs from a given amount of resources will differ widely by country of destination. This gives important insights for the potential of H2020 to create technological breakthroughs. The H2020 plan will only be able to initiate the radical innovative renewal, its goal, if resources are concentrated in those countries that are best able to make efficient use of them (Veugelers et al., 2015). Eventually H2020 had to attract private investments to achieve the goal of EUR 20 billion every year (EC, 2018). As Bilbao-Osorio and Rodríguez-Pose (2004) and Veugelers et al. (2015) concluded, private investments have greater returns but will mainly happen in non-peripheral regions and not in peripheral regions. If the project selection process in the H2020 programme continues to allocate funding based not on excellence and expected output but on geographical considerations this might upgrade the innovation capabilities in the weaker countries in the very long run. But in the short run this will most likely not lead to a substantial increase in the production of new knowledge and technologies needed for regaining competitiveness and job growth in the long run (Veugelers et al., 2015).

The distribution of the research money for every project within the H2020 funding programme had large regional differences. The EU three biggest economies received almost 40% of the total funding. Researchers in Germany, France and the UK received a total of more than EUR 22 billion of H2020 funding. Relatively small countries also benefited from this funding programme. Sweden, Denmark and Finland received together EUR 4.8 billion (8% of total) in funding. These three countries together only account for just 4% of the total EU population. These high percentage in Western and Northern European countries resulted in an East-West divide in the European research landscape for the H2020 funding programme. This is because scientists and research institutions in Poland, Slovakia, Bulgaria and Romania were among the least successful participants in H2020, they received together just over EUR 1 billion (Schiermeier, 2020). Veugelers, Cincera and Frietsch et al. (2015) did research on the possible impact of H2020 on innovation in Europe. They show doubts on whether all countries are likely to benefit from H2020 funding. The overall performance of the countries can be split into an input component and an output component. The low R&D investment levels in many Southern and Eastern European countries are a response to the relatively low capabilities to transfer research investment into research output (Veugelers, Cincera & Frietsch, et al., 2015). Another reason for this East-West divide is that the level of national research in Eastern European countries is relatively low. Researchers in Eastern European regions are also less well integrated into informal networks of research that compete successfully for EU grants (Schiermeier, 2020).

The capability for countries to turn research input into research output is also dependent on a countries product space. AI is a specific technology and not every country or region will be able to specialize in this technology. According to Hidalgo et al. (2007), countries specializing in one product may or may not also specialize in the other. Countries move preferentially to related or 'nearby' goods. To see which goods are related or 'nearby', Hidalgo et al. (2007) introduced the product space. This is a network with connecting products that are likely to be produced at the same time. The product space shows what the probability of a country is, to start producing another product. The probability increases with the number of related products that this country already has. If two goods are related, because they require similar institutions, infrastructure, physical factors or technology then they will tend to be produced at the same time. Whereas highly dissimilar goods are less likely to be produced together. The similarity between products *i* and *j* is based on the conditional probability of having Revealed Comparative Advantage. This measures whether a country is more effective (RCA>1) in a given good *i* or not (RCA<1), given that the country has comparative advantage in good *j* at time *t*. If a country has an RCA<1 this country has disadvantage in a certain type of goods. Countries can develop an RCA by following a diffusion process of the product space. The structure of the product space limits the diffusion process by being non-traversable by jumps of any proximity. Poorer countries tend to have RCA mainly on peripheral products such as textiles, forest products and animal agriculture. A country's productive structure is not only constrained by its level of factor endowments, but also by how easily those product-specific factors can be adapted to alternative uses, as indicated by location in the product space. Convergence can only exist if countries have the ability to reach any area of the product space (Hidalgo et al., 2007).

The idea of the product space from Hidalgo et al. (2007) is one of the efforts that generalized the principle of relatedness. The principle of relatedness describes the probability that a region enters (or exits) an economic activity from the number of related activities present in that location, the related or 'nearby' goods. The relative cost of moving knowledge increased in the last decades. Knowledge is concentrated in a few places and is embodied in networks of people. Compared to knowledge, products moving around the world is easy. The principle of relatedness is important because policy makers cannot build "cathedrals in the desert" (Hidalgo et al., 2018). There have been multiple efforts of science, technology and innovation policies to promote inventive activities in regions. The goal was to upgrade and change their technological and industrial base. However, many of these efforts did not succeed and the public budget was used ineffectively. A reason therefore can be found in the fact that policy makers did not take into account the technological make up of a region (Uhlbach, Balland & Scherngell, 2017). Countries that want to specialize in AI need to have related or 'nearby' goods. The principle of relatedness is not for the short-term. The principle of relatedness also increases spatial inequality and can reduce the ability of peripheral cities to develop (Hidalgo et al., 2018). For instance, for Western Europe the principle is good news. For Eastern Europe the related and 'nearby' goods are more limited because these countries are less developed. Just as said by Veugelers et al. (2015) the H2020 funding programme has to take into account the principle of relatedness instead of geographical considerations.

Uhlbach, Balland and Scherngell (2017) found in their study that the probability of entry for a region in a particular technology increases with the level of relatedness density. The relatedness density of a region tells how close its existing set of technologies is to those technologies that are missing in the region. It gives an average score of the potential of a region to develop new technologies (Balland, Boschma, Crespo & Rigby, 2019). This is known as the principle of relatedness as described before. They also found that regions that participate in subsidized R&D projects are more likely to enter this technology. However, the relation between funding and entry is not uniform across the different levels of relatedness. Funding only seems to make a big difference for mid-level of relatedness. Regions that already have a high level of relatedness will enter this technology anyway. And for regions with a low level of relatedness it implies that there needs to be a critical mass of knowledge and capabilities for innovation policy to be effective (Uhlbach et al., 2017).

This theory shows that the capabilities of regions to join the AI race also depends on their level of relatedness. For the EU it is important to take into account what the level of relatedness for

countries and regions are. Only in this way the H2020 funding can be well spend and contributes to the innovation level of European regions in AI. Whenever a regions relatedness is to low, it is better to invest in technologies that are more related to their product space. It also tells that there will be regional differences in research output since not every country has the same product space. The compensation of public investments in peripheral regions will also make no difference for the global AI race, since their research output capabilities are low and their level of relatedness might be low. In the introduction it is shown that the three biggest economies of the EU (UK, Germany and France) are ahead of the AI race in Europe. Looking at the theory of Uhlback et al. (2017) it can be assumed that H2020 investments for AI in these countries are less effective than H2020 investments in mid-level of relatedness. However, the UK, Germany and France did receive almost 40% of the total H2020 budget (Schiermeier, 2020).

Economic activities are known to concentrate in space, and that concentration is increasing. One reason for this increase is the growth of complex economic activities. Complex economic activities require a deep division of knowledge and labour. Some activities require a large network of people with deep expertise in complementary knowledge domains. This also results in the fact that complex industries exhibit a much greater degree of spatial concentration than less complex industries. This explains the rise in importance of superstar cities and the growing spatial inequality (Balland, Jara-Figueroa, & Petralia, et al*.,* 2020). Balland et al. (2020) found an urban concentration of research papers, patents, occupations and industries in the US. In all four examples, the activities are highly concentrated, especially in large cities. There is also a scaling law for patents, research papers, industries and occupations. This means that there is a relation between the number of patents and population in metropolitan statistical areas. They scale with each other over a significant interval. They found a super linear relationship between the number of patents and the population in metropolitan statistical areas (Balland et al., 2020). An earlier study from Crescenzi, Rodríguez-Pose and Stroper (2007) also found that the spatial distribution of innovative output, by patents, in both Europe and the US concentrates in just a few locations. During the 1990s, 92% of all patents were granted to residents in US metropolitan areas (MSAs) (Carlino, Chatterjee & Hunt, 2001). Also in the EU, patenting is highly concentrated. The cumulative percentage of total patents recorded by the 100 most innovative EU-15 regions and US MSAs is similar in the two continents. And for both, the 20 most innovative regions account for around 70% of total patents. Agglomeration also increases the innovative output. The agglomeration and economic interactions are key for innovation (Crescenzi et al., 2007). The urban concentration in these theories show that it is difficult to move complex knowledge over space. This makes it hard for funding programmes to stimulate AI in “new” regions.

According to Balland and Rigby (2017) not all patents hold the same value. There are only a few metropolitan regions in the US that produce the most complex new technologies. This shows that there is a wide geographic variation in knowledge complexity (Balland & Rigby, 2017). Another finding from Balland and Rigby (2017) is that not all knowledge is spatially sticky. Complex knowledge tends to be produced in relatively few places and once this is produced, this knowledge is not easy to move. Knowledge that is less complex is easier to move over space. They found this by looking at patent citations. The production of low complexity knowledge provides an insecure foundation of competitive advantage. It is therefore important for regions to transform their knowledge cores into more complex knowledge (Balland & Rigby, 2017). The creative destruction from Schumpeter (1934, 1942) is important to transform the existing knowledge into more innovation and complex knowledge. This in turn will create economic growth.

It is questionable if H2020 funding made a difference in the distribution of AI across Europe and in the global AI race. Studies from Balland et al. (2020) and Balland and Rugby (2017) show that complex knowledge is spatially sticky. They also found that patents and scientific publications increase with the number of population in metropolitan areas. This can suggest that, since AI is already concentrated in a few places because of its complex knowledge, the number of patents and scientific publications do not increase because of H2020 funding but because of the number of population.

3. Methodology

In this chapter the data and method to estimate the contribution of H2020 funding on the obtaining of AI-related patents and scientific publications in European regions is described. As said before, this research will only focus on the H2020 funding programme (FP8) and AI. The time period for this Framework Programme was between 2014 and 2020 (EC, 2014). In the period that the H2020 funding programme ran, the UK was still an EU Member State. Therefore, the EU-28 instead of the EU-27 will be used in this research.

To estimate the contribution, different data is used and collected with desk research. The modelling language in this research to process the data and to do the analysis was R. The data about H2020 projects was extracted from the EU Open Data Portal (2021). This portal provides free access to all the H2020 data. Data that is used from this portal covers information such as title, date, description, total cost, EU contribution, coordinator and the obtained scientific publications from every H2020 project. Besides this data, patent data from the OECD REGPAT database (January 2021) is collected. There is no specific data that tells which patents came directly from H2020 funding. That is why all patents that belong to AI al used. As a result, this research includes H2020 and non-H2020 AI-related patents, and only H2020 AI-related scientific publications. Lastly, the remaining controlling variables on the regional level were retrieved from Eurostat (2021a; 2021b; 2021c) and Office for National Statistics (2019).

Whenever patents are discussed in the rest of this research, it refers to non-H2020 and H2020 AI-related patents. Furthermore, whenever publications are mentioned it refers specifically to H2020 AI-related scientific publications.

For this research the regional NUTS levels will be used. In total there are three NUTS levels. The level NUTS1 has large regions which include multiple cities. Thus, this NUTS level is too big to analyse. From a policy perspective the NUTS2 level is better. However, the smallest NUTS level 3 is better from an ecosystem perspective. This is because the NUTS3 level represents the metropolitan areas in Europe. Despite this, these regions are too small to use in the analysis and have limited observations per region. It is therefore decided to only use the NUTS2 level for this research. In total there are 92 NUTS2 regions in this research.

3.1 Data

For this research only the H2020 projects related to AI were of interest. In order to filter out non related projects 214 AI-related keywords and 134 AI-related IPC codes from Baruffaldi et al. (2020) are used. In this report they identify what developments belong to AI and what not. A project or publication is related to AI if it contains at least two out of the 214 AI-related keywords (see p. 66 and 67 in Baruffaldi et al., 2020). Baruffaldi et al. (2020) concluded in their research that writers use AI-related keywords in the abstract to make their publications more interesting. Although, when looking into these publications it is not related to AI at all. That is why it must have at least two AI-related keywords. If the description part of the H2020 projects contained at least two of these keywords it is related to AI. The list of 134 IPC codes from Baruffaldi et al. (2020) was used to identify which patents belong to AI and can therefore be used in this research. If a patent had an IPC code that was defined as AI-related it was used in this research (see page 68 in Baruffaldi et al., 2020).

But only selecting the H2020 projects and patents on AI was not enough. The H2020 funding programme was from 2014 until 2020. However, H2020 projects started or closed in 2020 might not have obtained patents or publications yet. This is because it takes time to apply for a patent or to release a scientific publication. So only the H2020 projects between 2014-2019 were relevant for this research. When the H2020 project data was filtered on AI, relevant years and EU-28 member states, there were 211 projects left. The publications that came from H2020 projects were joined with the 211 H2020 projects. This resulted in a total of 2,493 publications that belong to the 211 H2020 projects. These publications are from 2014-2020.

For the patent data, it is unlikely for a patent to be applied for in the first two years of the H2020 funding programme. That is why only the years 2016-2020 were used in this research. When

the patent data was joined with the NUTS2 regions that received EU contribution in the H2020 funding programme, 1,110 patents were left. This means that regions that have AI-related H2020 projects obtained 1,110 patents between 2016-2020. As part of the analysis the Revealed Comparative Advantage (RCA) (Balland, 2017) was measured. This was done with the patent data during the H2020 funding programme (2016-2020) and also with patent data from before the H2020 funding programme (2011-2015). To extract the latter patent data the same method as described before was used and was filtered for the years 2011-2015. This resulted in 1,715 patents obtained by the regions that received H2020 funding between 2014-2019.

Lastly, the remaining control variables on the regional level were added to the data set. These controlling variables were GDP (at current market prices), population and total employment. The three controlling variables are from 2018. This is because data for 2020 was not available yet and data from 2019 had too many missing values. The GDP information came from the data browser Eurostat (2021b). Because the UK uses another currency, information about the GDP came from the Office for National Statistics (2019). This information was in GBP instead of EUR, hence the GBP had to be calculated to EUR. This was done with the historical exchange rates from 31 December 2018 (Rabobank, n.d.). The next control variable added was population and total employment. This data came from Eurostat (2021a; 2021c) and also included data from the UK. The last control variable for this research was GDP per capita in 2018. This variable was calculated by dividing GDP in 2018 with Population in 2018.

The first step in this research was to make maps to see what the distribution of the different variables is over Europe. By comparing different regions it has to be taken into account that regions with a large population are more likely to obtain more publications and patents. It was therefore better to compare regions with the number of publications and patents per thousand capita. This resulted in two extra variables: H2020 AI-related scientific publications per thousand capita and AI-related patents per thousand capita. These variables were calculated with the independent variable Population in 2018.

The final variables for this research were:

- NUTS2 region code
- Total cost for H2020 AI-related projects (2014-2019)
- Total EU contribution for H2020 AI-related projects (2014-2019)
- H2020 AI-related publications (2014 – 2020)
- Total AI-related patents (2016 – 2020)
- GDP in 2018
- Population in 2018
- Total employment in 2018
- GDP per capita in 2018
- H2020 AI-related scientific publications per thousand capita (2014 – 2020)
- AI-related patents per thousand capita (2016 – 2020)

Variables to compute RCA:

- Total AI-related patents 2011-2015
- Total AI-related patents 2016-2020

The data has been handled with care to ensure the validity and reliability. The data used to obtain the data set has been provided by the OECD and the EU.

3.2 Research Method

First the descriptive statistics are given. As said before, the second step in this research was to visualize the data with maps. This was done for the following variables: Total cost for H2020 AI-related projects (2014-2019), Total EU contribution for H2020 AI-related projects (2014-2019), H2020 AI-related scientific publications per thousand capita (2014 – 2020), AI-related patents per thousand

capita (2016 – 2020), GDP per capita in 2018, Population in 2018 and Total employment in 2018. Next to these different maps, the RCA for before and during the H2020 funding programme was measured. The results of this research were also visualized in a map to show the RCA development of every region.

Thirdly, the correlation between the different variables was measured. The Pearson's correlation coefficient was used and is an association measure for the linear relationship between two variables. The correlation coefficient measures the strength and direction of the relationship, but does not say anything about the possible causality. Besides the correlation matrix that was made in the third step for this research, different scatterplots were made. The purpose of the scatterplots was to discover possible non-linear relationships. The correlation matrix was also used to make the multivariate regression models. The main method in this research was the multivariate regression models. The regression analysis was used to determine the effect of one or more independent variables on the dependent variable. In this research the dependent variables were publications and patents. Different models were made by adding variables step-by-step to check robustness and highlight changes when another variable is added. The order of the variables was determined by the strength of the correlation.

The two multivariate regression formulas were:

$$Y_{(pub.)} = \beta_0 + \beta_{(EU)}X_{(EU)} + \beta_{(empl.)}X_{(empl.)} + \beta_{(pop.)}X_{(pop.)} + \beta_{(GDPcap.)}X_{(GDPcap.)}$$

$$Y_{(pat.)} = \beta_0 + \beta_{(GDPcap.)}X_{(GDPcap.)} + \beta_{(empl.)}X_{(empl.)} + \beta_{(pop.)}X_{(pop.)} + \beta_{(EU)}X_{(EU)}$$

Where:
Pub. = publications
Pat. = patents
β_0 = intercept
EU = EU contribution
GDPcap. = GDP per capita in 2018
Pop. = population in 2018
Empl. = employment in 2018

4. Results

4.1 Descriptive Statistics

Before going into the correlation and the regression model for this research, the data will be described. To do this the descriptive statistics for the different variables are given in Table 1.

Table 1: Descriptive statistics

Statistic	*N*	*Mean*	*Median*	*St. Dev.*	*Min*	*Max*
Total Cost (2014-2019)	92	2,959,105	253,295	5,106,289	71,429	25,948,502
EU Contribution (2014-2019)	92	2,649,645	231,866	4,562,089	50,000	23,269,730
H2020 AI-related scientific publications (2014-2020)	92	33.59	6.50	54.79	0	233
AI-related patents (2016-2020)	92	12.07	2.83	42.87	0	355.67
GDP (2018)	92	87,646,128,554	61,027,415,000	94,865,406,153	4,718,000,000	724,980,460,000
Population (2018)	92	2,538,888	1,956,730	2,034,102	254,368	12,213,447
Employment (2018)	92	1,240,233	995,500	983,251.10	123,540	6,456,390
GDP per capita (2018)	92	34,600.79	31,514.71	20,950.12	8,082.00	161,404.30
H2020 AI-related scientific publications per capita (x 100,000) (2014-2020)	92	1.96	0.34	4.16	0	25.91
AI-related patents per capita (x 100,000) (2016-2020)	92	0.44	0.11	1.52	0	14.07

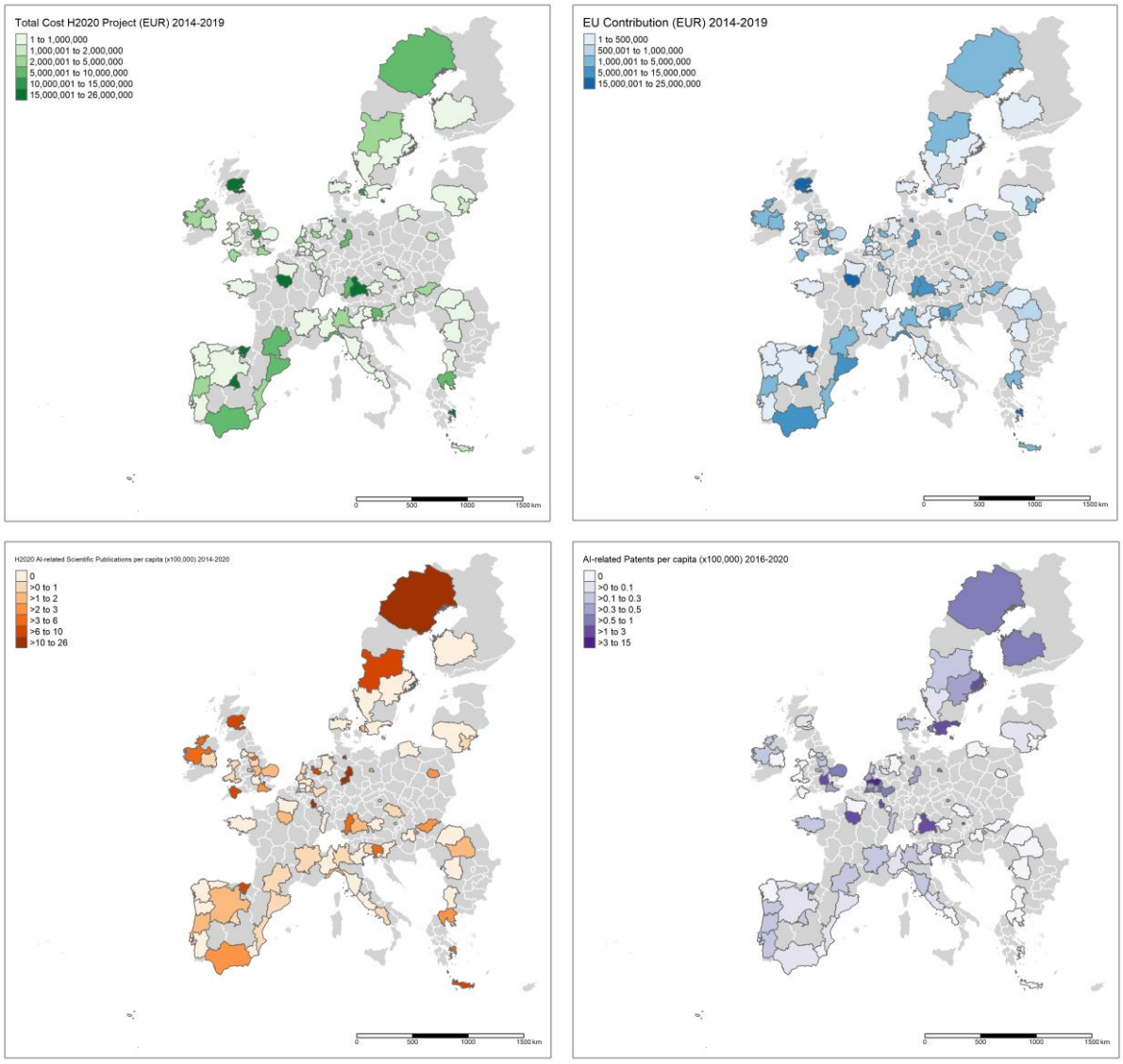
Source: made by author

4.2 Distribution in Europe

Different maps are made to show the distribution of the different variables over Europe. Map 1 shows the distribution of the four variables total cost, EU contribution, publications and patents. As it is shown in Map 1, almost every country in Europe received some H2020 funding. However, there are large regional differences that will be discussed later. Not every region obtained publications, to illustrate, 38 out of 92 NUTS2 regions have zero output in publications. On the other hand, only a small number of regions obtained more than 200 publications. Such as Andalusia (ES61) who has the highest number of publications with 233 publications. Nonetheless, Andalusia (ES61) only has 2.77 publications per thousand capita. The region Basque Country (ES21) in Spain has per thousand capita more publications (6.31). Although, in total numbers it looks like Andalusia (ES61) is the region with the most publications. Overall the number of publications is still well distributed over Europe and not concentrated in one place. For the patents this is a different story. Mainly Western and Northern European regions obtained patents. For example: île de France (FR10) has 184 patents and the region North Brabant (NL41) has the highest number of patents of 356. When comparing the number of patents per thousand capita, North Brabant (NL41) still has the highest number of patents per thousand capita (14.07). However, the region île de France (FR10) only has 1.50 patents per thousand capita in contrast to Stockholm (SE11) who has 2.78. To compare, Eastern European regions have zero research output in patents.

When looking at the research output of Europe in total numbers the following can be observed: regions that obtained only patents are mainly Western and Northern European regions, regions that only obtained publications are mainly Southern and Eastern European regions, regions that obtained both are Southern, Western and Northern European regions, and regions that obtained nothing are mainly Eastern European regions. With this observation the study from Crescenzi et al. (2007) can be confirmed. They found that the spatial distribution of innovation output, by total patents, in Europe concentrates in just a few locations. The 20 most innovative regions in the EU-15 account for around 70% of the total patents (Crescenzi et al., 2007).

Map 1: Distribution of total cost, EU contribution, publications and patents

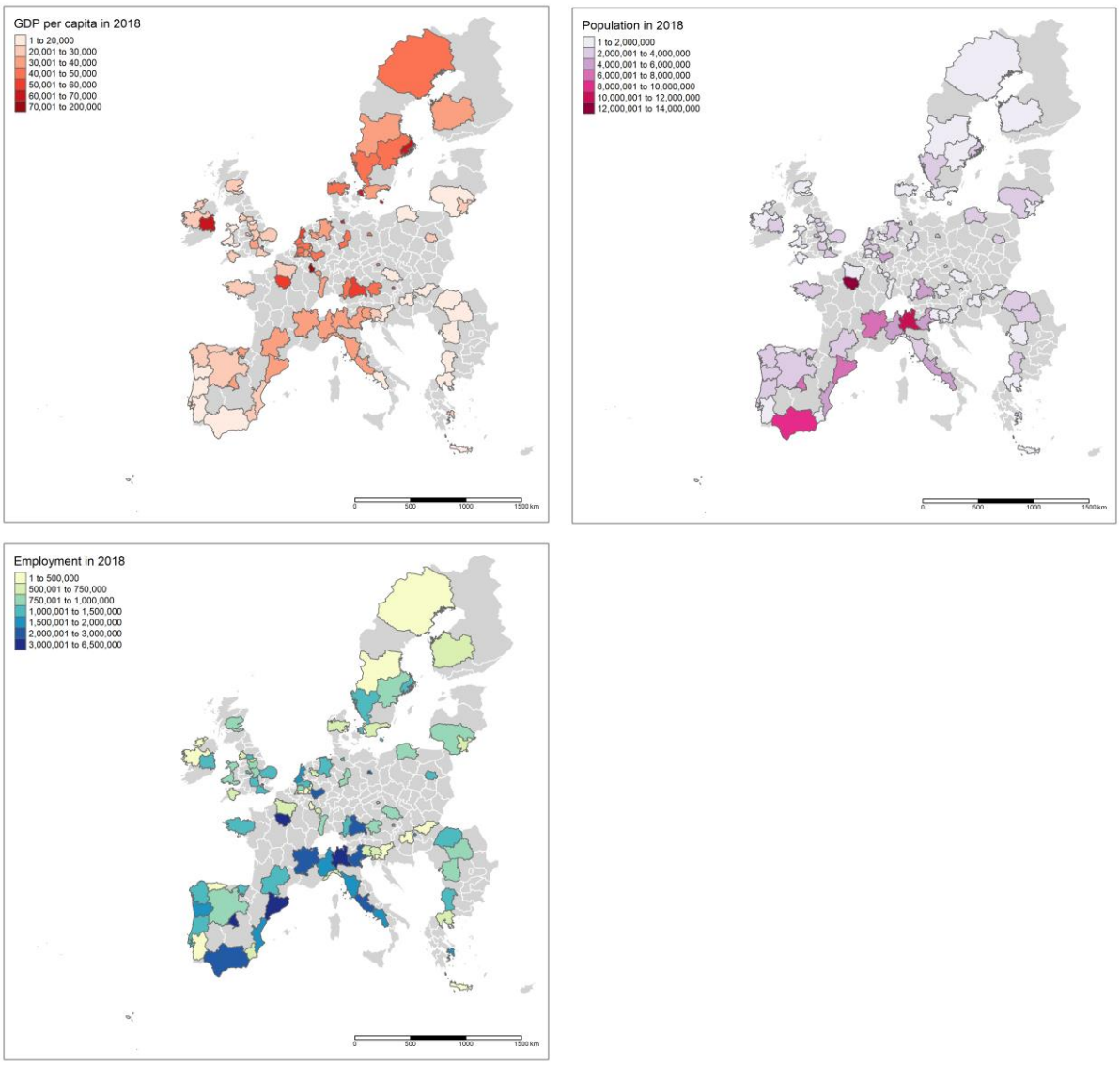


Source: made by author

The distribution of the EU contribution for every project within the H2020 funding programme had large regional differences. There is an East-West divide in the European research landscape. Eastern European regions received the least amount of H2020 funding. Given that the top five of countries that received the largest amount of funding are: Germany, UK, France, Spain and Italy. The first three are also the biggest economies of the EU. Small Northern countries, which are Sweden, Denmark and Finland, also received relatively a large amount of H2020 funding (Schiermeier, 2020). This further applies to the distribution of H2020 AI-related funding, Eastern European countries received less H2020 AI-related funding. The reason therefore is that the level of national research in Eastern European countries is relatively low. Researchers in these regions are in addition less well integrated into informal networks of researchers that compete successfully for EU grants (Schiermeier, 2020). However, Bilbao-Osorio and Rodríguez-Pose (2004) found in their research that less developed regions have a lack on private investments and therefore universities receive public investments to compensate for this lack. The private investments obtain more patents and public investments obtain more publications (Bilbao-Osorio & Rodríguez-Pose, 2004). According to the results in Map 1, less developed regions did not receive more H2020 AI-related funding as it was expected from Bilbao-Osorio and Rodríguez-Pose (2004).

In Map 2 the maps for the GDP per capita, population and employment for NUTS2 regions are shown. The observations from these maps show that NUTS2 regions with a higher GDP per capita did not receive more EU contribution nor had a higher total cost. Also the number of population and employment did not show a lower or higher output in publications or patents. Nevertheless, one observation that stood out is the region Upper Norrland (SE33) in Sweden. This region has only 519,760 inhabitants and 254,000 employees (48.87%), but it managed to have 109 publications and 3 patents. This is a relatively high number of outputs compared to Andalusia that has 8,410,095 inhabitants, 2,998,800 employees (35.66%), 233 publications and 2 patents. Andalusia has a population that is 16 times larger than Upper Norrland and an employment that is almost 12 times larger than Upper Norrland. To compare: Andalusia has one publication per 12,870 employees and Upper Norrland has one H2020 publication per 2,330 employees.

Map 2: Distribution of GDP, population and employment



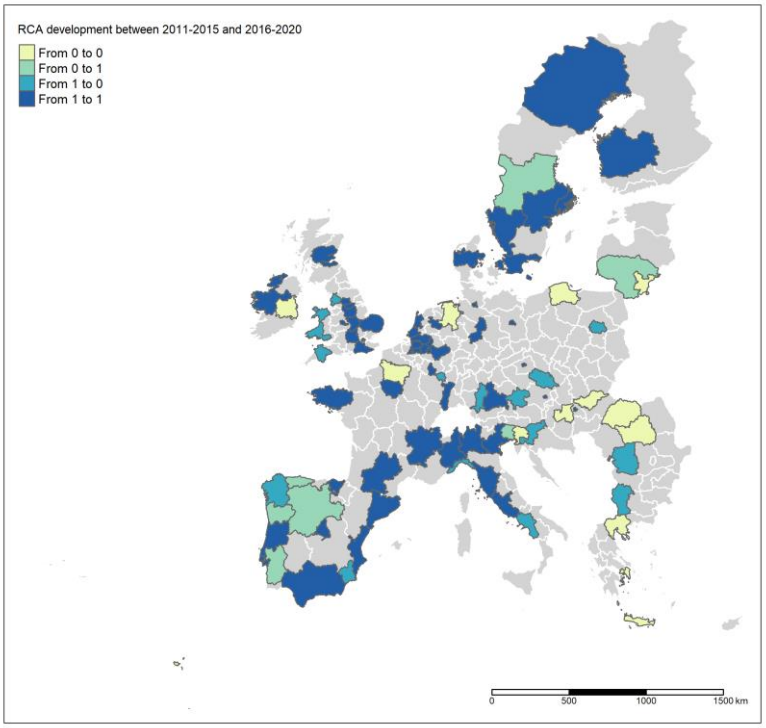
Source: made by author

4.3 Revealed Comparative Advantage

The different maps gave an overview of the distribution of the different variables in Europe. Although, these maps do not show if public funding provided comparative advantage or disadvantage of AI in European regions. When a region has comparative advantage it means that this region is more efficient in this technology than regions with comparative disadvantage. To measure

this, the Revealed Comparative Advantage (RCA) is used for the patent data. In different studies patent data is the most common used data and method to indicate industries and technologies in cities, regions or countries (e.g. Hidalgo, et al., 2007; Uhlbach, et al., 2017). To analyse the RCA patent data from before the H2020 funding programme (2011-2015) is compared to the patent data during the H2020 funding programme (2016-2020). Map 3 shows the result for this research. In total sixteen regions did not achieve comparative advantage with the H2020 funding programme. For both periods these regions had a comparative disadvantage in AI. In the same way, during the H2020 funding programme sixteen regions lost their comparative advantage. There RCA was below one (RCA<1) in the period 2016-2020. It is striking that mostly Eastern European countries have comparative disadvantage or lost their comparative advantage in AI during the two periods. Only the region Vidurio ir vakarų Lietuvos regionas (LT02) in Lithuania gained comparative advantage in AI.

Map 3: RCA development between 2011-2015 and 2016-2020



Source: made by author

Scientists and research institutions in Poland, Slovakia, Bulgaria and Romania were among the least successful participants in H2020, they received together just over 1 billion EUR. As a consequence for the relatively low level of national research in Eastern European countries there is an East-West divide (Schiermeier, 2020). The low R&D investments in the Southern and Eastern European countries is again a response to the relatively low capabilities to transfer research investment into research output, according to Veugelers, Cincera and Frietsch et al. (2015). The four Eastern European countries with the least successful participants in H2020 are also the countries with comparative disadvantage or lost their comparative advantage in AI. Southern European countries like Portugal, Spain and Italy are not in line with the theory from Veugelers et al. (2015). Since these regions mostly gained comparative advantage or kept comparative advantage in AI during the two periods.

Another explanation for the change in comparative advantage or disadvantage is the probability of entry from Uhlbach et al. (2017). Funding in R&D makes a big difference for regions with a mid-level of relatedness. In the Theoretical Framework the product space and the principle of relatedness are described. Regions move preferably to related or 'nearby' technologies. The relatedness density of a region tells us how close its existing set of technologies is to those

technologies that are missing in a region. It gives an average score of the potential of a region to develop new technologies (Balland et al., 2019). If the relatedness density to AI is too low the probability of that region to enter AI is unlikely. Regions need a critical mass of knowledge and capabilities, only than innovation policy can be effective (Uhlbach et al., 2017). Thus, regions that received H2020 funding but did not achieve comparative advantage in both periods might have a low level of relatedness density. Regions that gained comparative advantage might have mid-level of relatedness and provided from H2020 funding to enter AI.

However, it must be said that research activities done in the private sector have higher rates of return than research done in the public sector. This is because privately funded research tends to be more applied and thus have a higher return in patents. As said before, peripheral regions have a lack on private research; this can explain the comparative disadvantage in Eastern European countries. To compensate for this lack, university research receives public funding (Bilbao-Osorio & Rodríguez-Pose, 2004). This implies that the expected outputs from a given amount of resources will differ widely by country of destination. Eventually H2020 had to attract private investments (EC, 2018). The comparative disadvantage in AI for Eastern European countries suspects that investments in these regions did not attract private investments. For the Southern European regions that gained comparative advantage provided from the H2020 funding programme and might have attracted private investment, since private investments are related to the number of patents (Bilbao-Osorio & Rodríguez-Pose, 2004).

4.4 Relationship Between Variables

The third step in this research is to investigate the relationship between different variables. Table 2 shows the correlation between the different variables and their significance for the NUTS2 regions. All the variables have been transformed to their natural logarithm to meet the assumptions for the multivariate regression models. To do this the log1p function in R is used. The formula for this log is $\log_e(1+x)$. This type of log is used because the variable *H2020_Publications* and *TotalAIpatents_inregion* have zeros. With $\log_e(x)$ these zeros cannot be interpreted, however with log1p they can be interpreted. For larger numbers $\log_e(x)$ or log1p can both be used, the natural logarithm is the same for both $\log_e(x)$ and log1p. It is generally acceptable, when using log1p, to interpreting the estimates as if the variable were $\log_e(x)$ (Wooldridge, 2013). Accordingly, in this research log1p will be interpreting as if the variable were $\log_e(x)$.

One of the assumptions of the multivariate regression model is that there is no multicollinearity. If there is multicollinearity between two independent variables this can lead to unreliable and unstable estimates of the regression coefficients. A high correlation is bad for the regression and it increases the standard errors of their coefficients and may make those coefficients unstable (Allison, 2012). For a better regression model, multicollinearity should be avoided. When a correlation between two variables is >0.8 then severe multicollinearity may be present (Allison, 2012). By transforming the variables to their natural logarithm this assumption was met, as it is stated before. However, the independent variables *totalCost* and *ecMaxContribution* still have multicollinearity. It is therefore decided to only use the independent variable *ecMaxContribution*. Likewise, the independent variables GDP per capita, employment and population still has multicollinearity. To avoid this the *GDP_per_capita_2018* instead of *GDP_2018* will be used. Unfortunately, the variables *Population_2018* and *Employment_2018* still have multicollinearity after transforming them to their natural logarithm. The multivariate regression model will correct this in the analysis.

The Pearson's correlation coefficient is an association measure for the linear relationship between two variables. The correlation coefficient measures the strength and direction of the relationship, but does not say anything about the possible causality. When a correlation between two variables is significant (***, **, *) the null hypothesis is rejected. This means that there is a significant relationship between two variables. When the null hypothesis is not rejected there is no significant relationship between two variables. In that case, the correlation found is based on

coincidence. The purpose of the scatterplots in this chapter is to discover possible non-linear relationships.

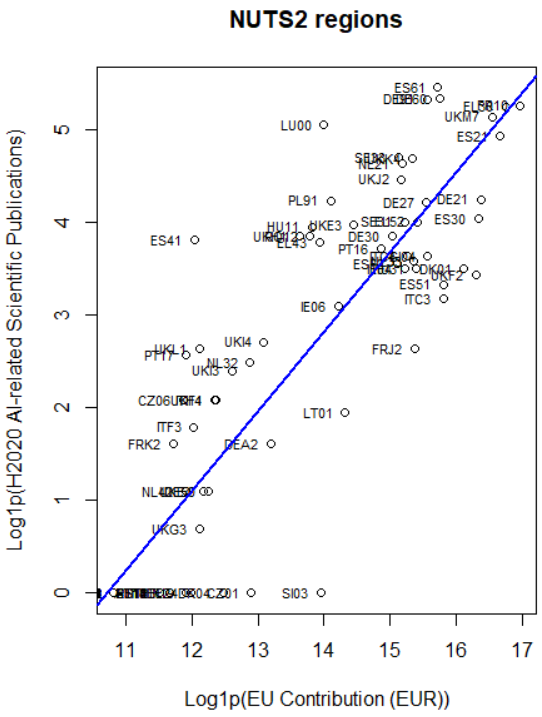
Table 2: Correlations between the different variables and their significance

NUTS2 regions		Log1p(totalCost)	Log1p(ecMaxContribution)	Log1p(H2020_Publications)	Log1p(TotalAIpatents_inregion)	Log1p(GDP_2018)	Log1p(Population_2018)	Log1p(Employment_2018)	Log1p(GDP_per_capita_2018)
Log1p(totalCost)	Pearson Correlation	1	0.997***	0.892***	0.080	0.208*	0.164	0.190	0.133
	P value (Sig. 2-tailed)		0.000	0.000	0.448	0.046	0.119	0.069	0.207
	N	92	92	92	92	92	92	92	92
Log1p(ecMaxContribution)	Pearson Correlation	0.997***	1	0.899***	0.075	0.208*	0.161	0.188	0.135
	P value (Sig. 2-tailed)	0.000		0.000	0.478	0.047	0.125	0.072	0.198
	N	92	92	92	92	92	92	92	92
Log1p(H2020_Publications)	Pearson Correlation	0.892***	0.899***	1	0.018	0.180	0.147	0.179	0.107
	P value (Sig. 2-tailed)	0.000	0.000		0.862	0.086	0.161	0.088	0.311
	N	92	92	92	92	92	92	92	92
Log1p(TotalAIpatents_inregion)	Pearson Correlation	0.080	0.075	0.018	1	0.582***	0.358***	0.420***	0.498***
	P value (Sig. 2-tailed)	0.448	0.478	0.862		0.000	0.000	0.000	0.000
	N	92	92	92	92	92	92	92	92
Log1p(GDP_2018)	Pearson Correlation	0.208*	0.208*	0.180	0.582***	1	0.793***	0.874***	0.628***
	P value (Sig. 2-tailed)	0.046	0.047	0.086	0.000		0.000	0.000	0.000
	N	92	92	92	92	92	92	92	92
Log1p(Population_2018)	Pearson Correlation	0.164	0.161	0.147	0.358***	0.793***	1	0.946***	0.024
	P value (Sig. 2-tailed)	0.119	0.125	0.161	0.000	0.000		0.000	0.819
	N	92	92	92	92	92	92	92	92
Log1p(Employment_2018)	Pearson Correlation	0.190	0.188	0.179	0.420***	0.874***	0.946***	1	0.224*
	P value (Sig. 2-tailed)	0.069	0.072	0.088	0.000	0.000	0.000		0.032
	N	92	92	92	92	92	92	92	92
Log1p(GDP_per_capita_2018)	Pearson Correlation	0.133	0.135	0.107	0.498***	0.628***	0.024	0.224*	1
	P value (Sig. 2-tailed)	0.207	0.198	0.311	0.000	0.000	0.819	0.032	
	N	92	92	92	92	92	92	92	92

Note: *** Correlation is significant at the 0.001 level (2-tailed), ** Correlation is significant at the 0.01 level (2-tailed), * Correlation is significant at the 0.05 level (2-tailed).
Source: made by author

To understand the different variables it is interesting to see what the correlation and significance between the dependent variables (*H2020_Publications* and *TotalAIPatents_inregion*) and the independent variables are. It is further interesting to see what the correlation and significance is between two independent variables. Table 2 shows that there is a significant relationship between de independent variable *H2020_Publications* and *ecMaxContribution* (r = 0.899***). This is also the only significant relationship for the independent variable *H2020_Publications.* Figure 1 shows the scatterplot for *H2020_Publications* and *ecMaxContribution*. The three regions Eastern Slovenia (SI03), Castile-Leon (ES41) in Spain and Île de France (FR10) are discussed below.

Figure 1: EU contribution (EUR) and H2020 AI-related publications



Source: made by author

The region Eastern Slovenia (SI03) is one of the outliers in Figure 1 Large cities in this region are Maribor and Celje. Eastern Slovenia has a total population of 1,089,717 and a total employment of 481,800. Compared to the other NUTS2 regions, Eastern Slovenia has one of the smallest total employment in 2018. With the significant relationship it would be expected that the number of publications increases with the amount of EU contribution. However, this region received a large amount of EU contribution (1,141,350 EUR) but did not obtain any publications. Despite the fact that Eastern Slovenia received a large amount of EU contribution, this region lost its comparative advantage in AI during the H2020 funding programme (Map 3). Before the H2020 funding programme, in 2011-2015, Eastern Slovenia had a comparative advantage in AI (RCA>1). The region Western Slovenia (SI04) on the other hand, which includes the capital Ljubljana, obtained 37 publications. But this region also received almost five times the amount of EU contribution as Eastern Slovenia (5,760,694 EUR). On the other hand, Western Slovenia did not achieve comparative advantage.

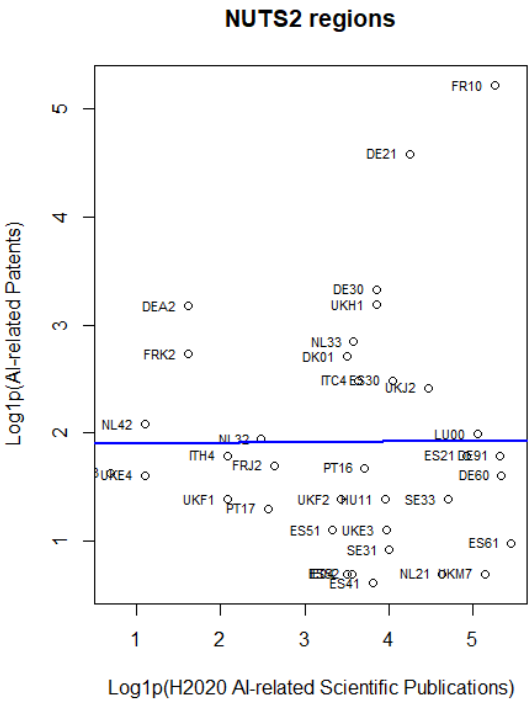
The opposite has happened in the region Castile-Leon (ES41) in Spain. The largest city in this region is Valladolid, in 2018 the total population in Valladolid was 298,866. This is only a small amount since the total population of the region Castile-Leon in 2018 was 2,418,556. This region received a low amount of EU contribution (170,121 EUR) but managed to obtain 44 publications. This region obtained more publications that it would be expected from the significant relationship in Figure 1 Castile-Leon also gained comparative advantage in AI during the H2020 funding programme (Map 3) by partly contributing to a patent.

One region that lies perfectly on the trend line is the region Île de France (FR10). This region is at the top right corner of Figure 1. This region received a large amount of EU contribution (23,269,729 EUR) but also obtained 192 publications and 183 patents. But when looking at the publications per thousand capita, Île de France is one of the regions with the lowest amount of publications per thousand capita (1.57). A reason for this large research output in patents might be the fact that île de France includes the capital Paris. Île de France is also the second region in the H2020 funding programme with the highest number of patents. As said before; France, Germany and the UK are the three countries in the EU-28 that are the main drivers of AI-related patents

applications (Brattberg et al., 2020). In the period 2011 to mid-2018 the amount of private equity investments in AI increased. French ventures received 13% of the total EU private equity investments. Only the UK and Germany received more private equity investments in AI (OECD, 2018). Since private investments are more related to the number of patents, it can be assumed that the number of patents increased with the private equity investments.

Figure 2 shows the scatterplot for the two dependent variables *H2020_Publications* and *TotalAIpatents_inregion* (r = 0.018). The correlation is very weak and only 36 out of the 92 regions managed to have research output in both publications and patents. This shows that, because there is no significant relationship, regions will probably obtain one or the other. This was also found in the study of Bilbao-Osorio and Rodríguez-Pose (2004) as discussed before. The less developed regions, Eastern European regions, receive public funding to compensate for the lack of private funding. This again results in different research output in different regions (Bilbao-Osorio and Rodríguez-Pose, 2004).

Figure 2: H2020 AI-related scientific publications and AI-related patents



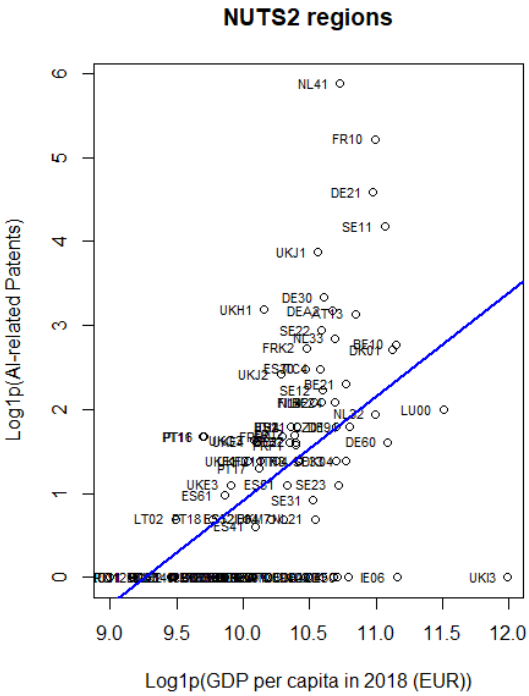
Source: made by author

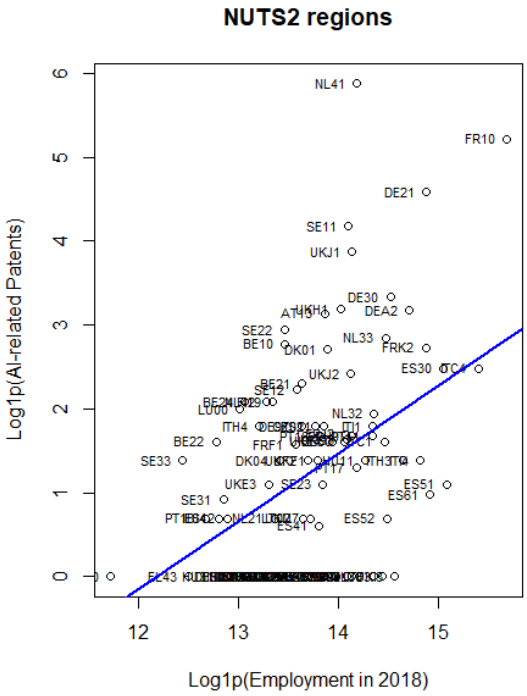
The dependent variable *TotalAIpatents_inregion* has a significant relationship with the independent variables *Population_2018* (r = 0.358***), *Employment_2018* (r = 0.420***) and *GDP_per_capita_2018* (r = 0.498***). The patents does not have a significant relationship with *ecMaxContribution* (r = 0.075). This means that there is no relationship between the amount of EU contribution and the obtained patents. EU contribution only makes a difference with the amount of publications.

Despite this, the independent variable that tells something about the size of a region has a significant relationship with the patents. If a region is larger it is also more likely to obtain patents. It is not a surprise that the independent variables *Population_2018*, *Employment* and *GDP_per_capita_2018* have a significant relation with each other. All three variables indicate the size of the region, if a region has a higher population this will also result in a higher total employment and a higher GDP per capita. Figure 3 and 4 show the scatterplot for GDP per capita and total employment with patents. Three regions will be described: Inner London-West (UKI3), North Brabant (NL41) in the Netherlands and Braunschweig (DE91) in Germany.

One of the outliers that is below the trend line in Figure 3 is the region Inner London-West (UKI3). This region is the central part of the British region Greater-London. Characteristics from this region are a high GDP per capita in 2018 (161,404 EUR) with a total population of 1,173,602 in 2018. Strikingly, the total employment (2,107,000) in 2018 is almost double the total population. This is because in Inner London-West people work in this region but live elsewhere. Despite the large employment this region only has 10 publications and zero patents.

Figure 3: GDP per capita in 2018 and AI-related patents
Figure 4: Employment in 2018 and AI-related patents





Source: made by author

Source: made by author

The region North Brabant (NL41) is an outlier in both scatterplots (Figure 3 and 4). This region has more patents than expected, looking at the GDP per capita and total employment. North Brabant is the region with the highest number of patents (356) but zero publications. This region is known for its innovative location Brainport Eindhoven (Brainport Eindhoven, 2021). The innovative location develops new technologies and solutions for the future, which explains the high number of patents. Other NUTS2 regions in The Netherlands that received H2020 funding also obtained patents, but the number of publications is higher. The high number of patents in North Brabant suggests that AI is mainly concentrated in this region.

Braunschweig (DE91) in Germany lies perfectly on the trend line in Figure 3. The number of patents (5) is the expected amount with a GDP per capita of 49,180 EUR. With a total employment of 832,980 this region also follows the trend line of Figure 4 Braunschweig is also the third region with the highest number of publications (205).

4.5 Multivariate Regression Model

The results of the regression analysis will now be discussed, this is also the last step in this research. The regression analysis is used to determine the effect of one or more independent variables on the dependent variable. In this case the dependent variables are publications and patents. Table 3 presents the results for the multivariate regression model for the dependent variable publications. Different model versions are made by adding variables step-by-step to check robustness and

highlight changes when another variable is added. The order of the variables is determined by the strength of the correlation (see Table 2). Model 1 shows that H2020 funding indeed seems to have a positive and significant impact on the number of publications. When the EU contribution increases by 1%, the number of publications increases by 0.8590%. With this first model 80.85% of the variance found in the publications variable can be explained by the EU contribution. When adding other variables the R^2 does not have a significant improvement or change. Also the controlling variables about employment, population and GDP per capita do not have a significant relationship with the number of publications. This might be because only H2020 publications are used.

Table 3: Regression model publications

Dependent variable: log1p(H2020 AI-related scientific publications)				
	Model 1	Model 2	Model 3	Model 4
log1p(EU Contribution)	0.8590*** (0.0000)	0.8571*** (0.0000)	0.8560*** (0.0000)	0.8591*** (0.0000)
log1p(Employment)		0.0293 (0.8290)	0.2098 (0.6160)	0.4765 (0.3730)
log1p(Population)			-0.1907 (0.6480)	-0.4421 (0.3977)
log1p(GDP per capita)				-0.1791 (0.4204)
Constant (intercept)	-9.2061*** (0.0000)	-9.5864*** (0.0000)	-9.2948*** (0.0000)	-7.5196* (0.0126)
Observations	92	92	92	92
R^2	0.8085	0.8086	0.8091	0.8105
Adjusted R^2	0.8064	0.8043	0.8026	0.8018
Residual Std. Error	0.8594	0.8640	0.8679	0.8696

Note: *** Correlation is significant at the 0.001 level (2-tailed), ** Correlation is significant at the 0.01 level (2-tailed), * Correlation is significant at the 0.05 level (2-tailed) and . Correlation is significant at the 0.1 level (2-tailed).
Source: made by author

The second multivariate regression model in Table 4 is for the patents. Also here different models are made by adding variables step-by-step to check robustness and highlight changes when another variable is added. Just as for the publications the order of the variables is determined by the strength of the correlation (see Table 2). The first variable that is added is the GDP per capita in Model 1. This variable seems to have a positive and significant impact on patents. When the GDP per capita increases with 1% the number of patents increases with 1.2339%. Compared to the regression model in Table 3, in this first model only 24.84% of the variance found in the patents can be explained by GDP per capita. However, in Model 2 the independent variable employment adds 10.03% extra. This means that 34.87% of the variance found can be explained by GDP per capita and employment. When adding the independent variable population in Model 3, the independent variable employment went from a positive and significant impact to a negative and no significant impact. The independent variable population has a positive and significant (at the 0.1 level) impact on patents. It is not surprising that the number of patents increased with population. Balland et al. (2020) found in their research that there is a relation between the number of patents and population in metropolitan statistical areas. They scale with each other over a significant interval. This means that the number of patents increases with the number of population (Balland et al., 2020). For this regression model, when the population increases with 1% the number of patents increases with 1.1532%.

Table 4: Regression model patents

Dependent variable: log1p(AI-related patents)					
	Model 1	Model 2	Model 3	Model 4	Model 5
log1p(GDP per capita)	1.2339***	1.0534***	1.3453***	1.3572***	
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
log1p(Employment)		0.6267***	-0.5099	-0.4970	
		(0.0004)	(0.4320)	(0.4459)	
log1p(Population)			1.1510 .	1.1532 .	
			(0.0719)	(0.0726)	
log1p(EU contribution)				-0.0302	0.0480
				(0.5893)	(0.4780)
Constant (intercept)	-11.4228***	-18.2087***	-22.2384***	-22.1775***	0.6730
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.4530)
Observations	92	92	92	92	92
R^2	0.2484	0.3487	0.3723	0.3744	0.0056
Adjusted R^2	0.2400	0.3340	0.3509	0.3457	-0.0055
Residual Std. Error	1.1430	1.0700	1.0560	1.0610	1.3150

Note: *** Correlation is significant at the 0.001 level (2-tailed), ** Correlation is significant at the 0.01 level (2-tailed), * Correlation is significant at the 0.05 level (2-tailed) and . Correlation is significant at the 0.1 level (2-tailed).
Source: made by author

The main question in this research is if H2020 funding contributed to the obtaining of patents. When looking at Model 4 it seems like the EU contribution did not have any effect on the number of patents. EU contribution does not have a significant effect on the number of patents. Also, when adding the independent variable EU contribution, only 0.21% extra of the variance found in the patents can be explained by EU contribution. If the EU contribution increased with 1% the number of patents would decrease with 0.0302%. Model 5 in Table 4 shows the regression for patents with only the independent variable EU contribution. When only looking at the EU contribution, it seems that this variable has a positive impact on patents. However, this impact is not significant and only 0.56% of the variance found can be explained by EU contribution. This is in stark contrast compared to Model 1 in Table 3, where the EU contribution explained 80.85% of the variance found in publications. This would suggest that EU contribution only contributes to the obtaining of publications. These regression models are in line with the fact that privately funded research obtains more patents (Bilbao-Osorio & Rodríguez-Pose, 2004). Private investments are not taken into account in this research. If the amount of private investments is many times greater than H2020 funding, it can be explained why there is no significant relationship between patents and H2020 funding.

5. Conclusion

With this research the following research question is answered: 'To what extent did Horizon 2020 funding contribute to the obtaining of patents and scientific publications in the European Union for artificial intelligence?' To do this, quantitative research is done for the possible relationship between the number of AI-related scientific publications and AI-related patents and H2020 funding.

The results show that not every region that received H2020 funding, was able to obtain publications. The same holds for patents. Regions that obtained patents are mostly Western and Northern European regions. Compared to publications, patents are concentrated in only a few places. While for the publications it is still well distributed over Europe and not concentrated in one place. What is also observed is that there is an East-West divide when it comes to the amount of H2020 AI-related funding. Eastern European regions received less H2020 AI-related funding. In line with the East-West divide, Eastern European regions mostly have comparative disadvantage or lost their comparative advantage in AI during the H2020 funding programme. In contrast; Southern European regions in Portugal, Spain and Italy gained comparative advantage or kept comparative advantage in AI.

Results from the correlation and regression models show that H2020 funding does not affect the number of patents, but it does positively impact the number of publications. The obtaining of patents is dependent on the variables GDP per capita, employment and population. When it comes to publications these variables do not play a significant role. Only the EU contribution played a significant role. During this research it was also found that there is no significant relationship between publications and patents.

From this quantitative research it can be concluded that H2020 AI-related funding mainly contributed to the obtaining of publications and not for patents. Since H2020 AI-related funding does not contribute to the number of patents, it was also not able to stimulate the Revealed Comparative Advantage in AI for European regions. There are more regions that lost their comparative advantage than regions that gained comparative advantage in AI.

6. Discussion

For this research different data from the European Union and the OECD is used. This research has a total of 92 NUTS2 regions with multiple observations. Therefore, this research is representative. On the basis of this it can be stated that if this research is repeated, the results would be the same and the results of this study are therefore valid.

The first results in this research showed that not every region was able to obtain publications or patents. The patents are more concentrated, compared to the publications. They are also mostly obtained in Western and Northern European regions. This research showed that there is an East-West divide. Not only for the research output but also for the research input, the H2020 AI-related funding. This is in line with the literature that shows that Southern and Eastern European countries have relatively low capabilities to transfer research investment into research output (Veugelers, Cincera & Frietsch, et al., 2015). The peripheral regions that do not receive private investments will be compensated with public funding. However, what is known in the literature is that private investments have greater research output. This is because private investments obtain more patents and is commercial orientated (Bilbao-Osorio & Rodríguez-Pose, 2004). This can explains why Western and Northern European regions do obtain the most patents, they might receive more private investments.

Eastern European regions also have comparative disadvantage or lost their comparative advantage in AI. This also confirms the East-West divide. What is striking is that Southern European regions in Portugal, Spain and Italy gained comparative advantage or kept comparative advantage in AI. The latter is not in line with the literature from Veugelers et al. (2015). The Southern European regions showed that they are able to transfer research investment into research output, while according to Veugelers et al. (2015) they could not. A reason for this can be found in the level of relatedness of regions. Uhlbach et al. (2017) found in their study that the probability of entry for a region in a particular technology increases with the level of relatedness density. Regions that receive public funding in R&D are more likely to enter this technology. Although, the relation between funding and entry is not uniform across the different levels of relatedness. It seems to make a big difference for regions that have mid-level of relatedness. Regions with a high level of relatedness will enter this technology anyway. Regions with a low level of relatedness are unlikely to enter this technology (Uhlbach et al., 2017). Thus, the Southern European regions that gained comparative advantage might have had a mid-level of relatedness and provided from H2020 funding to enter AI. While Eastern European regions might have had a low level of relatedness and were not able to provide from H2020 funding to enter AI.

Results from the multivariate regression models for publications show that H2020 funding (EU contribution) has a strong relation with the number of publications. The independent variables GDP per capita, employment and population do not have a significant relationship or effect on the number of publications. For the patents, the results show that H2020 funding (EU contribution) does not have an effect on the number of patents. Only the independent variables GDP per capita, employment and population do have a significant relationship and effect on the number of patents.

A possible explanation for the significant relationship between patents, GDP per capita, employment and population can be found in the study of Balland et al. (2020). Economic activities are more and more concentrating in space. Balland et al. (2020) found an urban concentration of patents in the United States. This activity is highly concentrated and especially in large cities. In their study the number of patents increases with the number of population in metropolitan statistical areas. When the population increases, the number of patents also increases (Balland et al., 2020). The results from this multivariate regression model show that if the GDP per capita, employment or population increases the number of patents also increases. This is in line with the study from Balland et al. (2020).

This research has limitations because only H2020 AI-related scientific publications are used. All non-H2020 AI-related scientific publications are left out of this research. This also is a possible explanation for the strong relationship between H2020 funding and the number of publications. It

can be expected that if this research is done again with also non-H2020 AI-related scientific publications the results may be different. The significant relationship might not be as strong as it is now.

For the patents used in this research there are no limitations like the publications. This is because all H2020 and non-H2020 AI-related patents are used. This makes the statements for AI-related patents more reliable. However, there are H2020 funding projects that started in 2020. These projects are left out of the research since they are not finished yet. It is possible that these projects obtain patents after the project is finished. If this research is done again in a few years, with additional H2020 funding projects and their results (publications and patents), there may be a difference in the results.

This current research supplements the existing literature on the contribution of H2020 to measure the innovation gap in AI. Earlier studies did not look at the possible contribution of H2020 on the obtaining of publications and patents. With this research, granting funding within Horizon Europe can be done more efficiently to bridge the innovation gap and to prevent becoming a consumer of AI. Funding can be spent better in the future. The advice for follow-up research is to look at the contribution that H2020 funding has on the attraction of private investments in regions. This is because H2020 had to attract private investments to achieve the goal of EUR 20 billion every year (EC, 2018). Bilbao-Osorio and Rodríguez-Pose (2004) showed that patents are more commercially orientated and a higher number of patents applications come from privately funded research. It can therefore be expected that H2020 funding has an indirect effect on the number of patents. Regions that have a lack on private research will be compensated with public funding investments. It is also known that these regions are more likely to obtain publications instead of patents (Bilbao-Osorio & Rodríguez-Pose, 2004). Another future research suggestion is the probability of entry for regions in AI. This research looked at the development of the Revealed Comparative Advantage (RCA). However, it did not look at the probability for a region, that received H2020 funding, to enter AI. The probability of entry can show if H2020 really made a difference for regions to be able to enter AI. This can also show if the H2020 funding is well spend.

Reference list

Asgard, & Roland Berger. (2018). *Artificial Intelligence: A strategy for European startups.* Munich: Roland Berger GMBH.

Allison, P. (2012, 10 September). When can you safely ignore multicollinearity? [website]. Retrieved from https://statisticalhorizons.com/multicollinearity

Balland, P.A. (2017). Economic Geography in R: Introduction to the EconGeo Package, *Papers in Evolutionary Economic Geography, 17*(9), 1-75.

Balland, P.A., Boschma, R., Crespo, J., & Rigby, D.L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies, 53*(9), 1252-1268. DOI: https://doi.org/10.1080/00343404.2018.1437900

Balland, P.A., Jara-Figueroa, C., Petralia, S.G. et al. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour, 4***,** 248–254. https://doi.org/10.1038/s41562-019-0803-3

Balland, P.A., & Rigby, R. (2017). The Geography of Complex Knowledge. *Economic Geography*, *93*(1), 1-23. DOI: 10.1080/00130095.2016.1205947

Baruffaldi, S., Van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, D., & Squicciarini, M. (2020). *Identifying and measuring developments in artificial intelligence: Making the impossible possible.* OECD Science, Technology and Industry Working Papers. DOI: https://dx.doi.org/10.1787/5f65ff7e-en

Brattberg, E., Csernatoni, R., & Rugova, V. (2020). *Europe and AI: Leading, Lagging Behind, or Carving Its Own Way?.* Washington: Carnegie Endownment for International Peace.

Bilbao-Osorio, B., & Rodríguez-Pose, A. (2004). From R&D to Innovation and Economic Growth in the EU. *Growth and Change, 4*(35), 434-455.

Bughin, J., Seong, J., Manyika, J., Hämäläinen, L., Windhagen, E., & Hazen, E. (2019). *Notes from the AI frontier: Tackling Europe's gap in digital and AI.* McKinsey Global Institute.

Brainport Eindhoven. (n.d.). Brainport Eindhoven [website]. Retrieved from https://brainporteindhoven.com/nl/ on 7 May 2021

Castro, D., McLaughlin, M., & Chivot, E. (2019). *Who is winning the AI race: China, the EU or the United States?* Center for Data Innovation.

Collaboration in Research and Methodology for Official Statistics. (n.d.). Research Projects under Framework Programmes [website]. Retrieved from https://ec.europa.eu/eurostat/cros/content/research-projects-under-framework-programmes-0_en on 18 March 2021

Crescenzi, R., Rodríguez-Pose, A., & Stroper, M. (2007). The territorial dynamics of innovation: a Europe – United States comparative analysis. *Journal of Economic Geography*, *7*, 673-709. DOI: 10.1093/jeg/lbm030

European Commission. (n.d.). Horizon Europe [website]. Retrieved from https://ec.europa.eu/info/horizon-europe_en on 29 March 2021.

European Commission. (2014). *Innovation Union Scoreboard 2014*. Belgium.

European Commission. (2014a). *Horizon 2020: The EU Framework Programme for Research & Innovation.* Luxembourg: Publications Office of the European Union. DOI:10.2777/3719. Retrieved from https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/H2020_inBrief_EN_FinalBAT.pdf

European Commission. (2018). *Artificial Intelligence for Europe* (SWD(2018) 137 final). Retrieved from https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

Eurostat. (2021a, 4 March). Employment (thousand persons) by NUTS 3 regions [dataset]. Retrieved from https://ec.europa.eu/eurostat/web/rural-development/data

Eurostat. (2021b, 10 March). Gross domestic product (GDP) at current market prices by NUTS 3

regions [database]. Retrieved from https://ec.europa.eu/eurostat/databrowser/view/nama_10r_3gdp/default/table?lang=en
Eurostat. (2021c, 11 March). Population on 1 January by age group, sex and NUTS3 region [dataset]. Retrieved from https://ec.europa.eu/eurostat/web/rural-development/data
EU Open Data Portal. (2021, 10 February). CORDIS - EU research projects under Horizon 2020 (2014-2020) [dataset]. Retrieved from https://data.europa.eu/euodp/en/data/dataset/cordisH2020projects
Hidalgo, C.A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The Product Space Conditions the Development of Nations. *Science, 317*(5837), 482-487. DOI: 10.1126/science.1144581
Hidalgo C.A. et al. (2018) The Principle of Relatedness. In: Morales A., Gershenson C., Braha D., Minai A., Bar-Yam Y. (eds), *Unifying Themes in Complex Systems IX* (451-457). Springer, Cham. https://doi.org/10.1007/978-3-319-96661-8_46
Joint Research Centre & European Commission. (2018). *Artificial Intelligence: A European Perspective* (EUR 29425 EN). Luxembourg: Publications Office of the European Union.
Kekic, L. (2018, 4 April). Why the EU lags the US in innovation [article]. Retrieved from https://www.headspringexecutive.com/why-the-eu-lags-the-us-in-innovation/
Nelson, R.R. (1959). The simple economics of basic scientific research. *Journal of Political Economy, 67*(3), 298-306. DOI: https://doi.org/10.1086/258177
OECD. (2018). *Private Equity Investment in Artificial Intelligence.* Paris: OECD. Retrieved from www.oecd.org/going-digital/ai/private-equity-investment-in-artificial-intelligence.pdf
OECD. (2021, January). REGPAT database.
Office for National Statistics. (2019, 19 December). Regional gross domestic product all NUTS level regions [dataset]. Retrieved from https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/regionalgrossdomesticproductallnutslevelregions
Rabobank. (n.d.). Historische koersen 2018 [website]. Retrieved from https://www.rabobank.nl/particulieren/betalen/service/buitenland/wisselkoers-vreemde-valuta/historische-koersen-2018/ on 15 March 2021
Savage, N. (2020, 9 December). The race to the top among the world's leaders in artificial intelligence [article]. Retrieved from https://www.nature.com/articles/d41586-020-03409-8
Schiermeier, Q. (2020, 22 December). Horizon 2020 by the numbers: how €60 billion was divided up among Europe's scientists [article from Nature]. Retrieved from https://doi.org/10.1038/d41586-020-03598-2
Schumpeter, J.A. (1934). *The Theory of Economic Development.* Oxford University Press: London.
Schumpeter, J.A. (1942). *Capitalism, Socialism and Democracy.* George Allen and Unwin: New York.
Uhlbach, W.H., Balland, P.A., & Scherngell, T. (2017). R&D Policy and Technological trajectories of Regions: Evidence from the EU Framework Programmes. *Papers in Evolutionary Economic Geography, 17* (22), 1-22.
Veugelers, R., Cincera, M., Frietsch, R., Rammer, C., Schubert, S., Pelle, A., Renda, A., Montalvo, C., & Leijten, J. (2015). The Impact of Horizon 2020 on Innovation in Europe. *Intereconomics, 50,* 4–30 (2015). https://doi.org/10.1007/s10272-015-0521-7
Wooldridge, J.M. (2013). Multiple Regression Analysis: Further Issues. In: J.M. Wooldridge (ed.), *Introductory Economitrics: A Modern Approach, Fifth Edition* (186-226). Mason: South-Western, Cengage Learning. Retrieved from https://economics.ut.ac.ir/documents/3030266/14100645/Jeffrey_M._Wooldridge_Introductory_Econometrics_A_Modern_Approach__2012.pdf
World Intellectual Property Organization. (2019). *Artificial Intelligence* (ISBN: 978-92-805-3007-0). Geneva: World Intellectual Property Organization