LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY

School of Business and Management

Business Analytics

EFFECT OF MACHINE LEARNING ON COMPETITIVENESS OF REGIONS IN FINLAND

Author: Artur Mossolainen

1st Examiner: Professor Kaisu Puumalainen

2nd Examiner: Professor Heli Arminen

ABSTRACT

Author:	Artur Mossolainen
Title:	Effect of machine learning on competitiveness of regions in Finland
Faculty:	School of Business and Management
Master's Programme:	Business Analytics (MBAN)
Year:	2021
Master's Thesis:	LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY
	62 pages, 9 figures, 18 tables
Examiners:	Professor, Kaisu Puumalainen
	Professor, Heli Arminen

Keywords:	machine learning, artificial intelligence, competitiveness, automation

The purpose of this research is to examine the effect of Machine Learning on competitiveness of 310 municipalities in Finland and find out the economic implications using data from 2010-2018. To achieve this, a formative index was created to measure competitiveness of municipalities. This index consisted of four dimensions, which were social equity, innovativeness, centralization, and reachability. To measure Machine Learning impact, a score was used for each municipality called the Suitability for Machine Learning. This measure scores every occupational task based on how well it can be automated with Machine Learning. The score is then calculated for every occupation based on the tasks it contains and applied on employment data of Finnish municipalities to get a final score for each municipality. The competitiveness index score and Suitability for Machine Learning score were then used in panel regression analysis to see their impact on Gross Domestic Product per capita, and yearly mean salary, which were used as the economic performance measures in this study. The study concludes that Machine Learning does have a positive impact on economy, but a negative impact on competitiveness, while competitiveness has a negative short-term impact on GDP and salary, and a positive long-term impact.

TIIVISTELMÄ

Tekijä:	Artur Mossolainen
Tutkielman nimi:	Koneoppimisen vaikutus alueelliseen kilpailukykyyn Suomessa
Tiedekunta:	Kauppatieteellinen tiedekunta
Maisteriohjelma:	Business Analytics (MBAN)
Vuosi:	2021
Pro gradu -tutkielma:	LAPPEENRANNAN–LAHDEN TEKNILLINEN YLIOPISTO LUT
	62 sivua, 9 kuvaajaa, 18 taulukkoa
Tarkastajat:	Professori, Kaisu Puumalainen
	Professori, Heli Arminen
Avainsanat:	koneoppiminen, tekoäly, kilpailukyky, automaatio

Tämän pro gradu -tutkielman tarkoitus on selvittää koneoppimisen vaikutusta Suomen 310 paikkakunnan kilpailukykyyn sekä talouteen hyödyntäen tilastoja vuosilta 2010–2018. Tutkimusongelman ratkaisemiseksi luotiin formatiivinen kilpailukykyindeksi mittaamaan Suomen kuntien kilpailukykyä. Kilpailukykyindeksi koostui neljästä alaindeksistä, joita olivat sosiaalinen pääoma, innovatiivisuus, keskittyminen ja saavutettavuus. Koneoppimisen vaikutuksen mittaamista varten tutkimuksessa käytettiin koneoppimisen soveltuvuus -mittaria. Mittari pisteyttää olemassa olevat työtehtävät sen perusteella, miten helposti niiden suorittaminen onnistuu koneoppineelta ohjelmalta. Työtehtävistä saaduilla tuloksilla pisteytettiin jokainen ammatti sen sisältämien työtehtävien perusteella. Lopuksi jokaiselle paikkakunnalle laskettiin painotettu keskiarvo ammattien määrän perusteella. Koneoppimisen soveltuvuus -mittarin sekä kilpailukykyindeksin vaikutusta talouteen mitattiin regressioanalyysillä. Selitettävinä muuttujina toimivat paikkakuntien BKT per asukas sekä keskimääräinen vuosipalkka, jotka oli tutkimuksessa valittu talouden kehityksen mittareiksi. Tutkimuksessa selviää, että koneoppimisella on negatiivinen yhteys kilpailukykyyn ja positiivinen yhteys talouteen. Kilpailukyvyllä todetaan olevan talouteen negatiivinen vaikutus lyhyellä aikavälillä.

ACKNOWLEDGEMENTS

First, I want to thank Kaisu Puumalainen and Heli Arminen for their continuous help and feedback on my journey to write a master's thesis. I cannot imagine how long it would have taken me to complete this project if it were not for your guidance. I learned a lot during this process, which I am grateful for.

I want to thank my friends, family, and my girlfriend Eira, who supported me throughout the duration of my studies.

Lappeenranta University of Technology for the opportunity to study and learn valuable skills that helped me create this thesis and succeed in my career.

And a special thanks to VITAKO Ry, who provided me with the opportunity to conduct this thesis on such an intriguing subject.

Table of contents

List of abbreviations

AI	Artificial Intelligence
ML	Machine Learning
SML	Suitability for Machine Learning
GDP	Gross Domestic Product
PCA	Principal Component Analysis
KMO	Kaiser-Meyer-Olkin
OLS	Ordinary Least Squares

List of Appendices

List of Tables

List of Figures

1. INTRODUCTION

One of the most significant technological breakthroughs of its time happened in the late 1700's, when Watt's steam engine was invented. This happened as a result of breakthroughs in fields, such as chemistry, mechanical engineering, and metallurgy. This set off what is now known as the Industrial revolution. Industrial Revolution acted as the first step towards modern life, as it sparked the creation of factories and railways, leading to mass production and mass transportation. (Brynjolfsson & McAfee, 2016). The invention of steam engine and the following innovations, as in applications of the invention enabled people to replace muscle and manual labor with mechanical power. (Brynjolfsson & McAfee, 2016; Deane, P., 2010).

The Industrial revolution brought enormous wealth and production skyrocketed. As a result, people from rural areas started emigrating to cities in search of factory work, doubling the emigration rate from rural areas over the 19th century. (Jr., 2002; Williamson, 1988). This transformation came with negative side effects, as a wave of unemployment followed. The reason for this was employee displacement, as machines had reduced the need for certain types of labor. (Feinstein, 1998). For these reasons, the reception for new inventions was not all positive, as can be observed with the Luddite movement during the British industrial revolution. These groups composed of textile artisans sought to destroy the machines, as they felt their employment being threatened by the industrial machines. (Autor, 2015).

Since the introduction of the steam engine, there have been successive periods of progress, with technologies such as electricity and telephones in the second industrial revolution. (Mokyr & Strotz, 1998). These inventions have paved the way for electronical devices and long-distance communication, eventually leading us to what is known as the digital era. Digital era encompasses the transformation from the industrial revolution into a world dominated by informational technology. In the digital era, innovation comes largely from digital technologies, that shape the society. (Shepherd, 2004).

What we are now experiencing is the second machine age. Where previous waves of automation have affected physical labor, new inventions are expected to advance mental power, automating mental work that has been untouched by previous waves of automation. (Brynjolfsson & McAfee, 2016). This advancement is happening as a result of a technology called Artificial Intelligence. Artificial Intelligence has been booming in the past decade, and it is expected to create economic growth on a global scale, particularly through a subset of Artificial Intelligence known as Machine Learning. (McKinsey Global Institute, 2018; Brynjolfsson et al., 2018). Companies are always striving to gain a competitive edge over competing businesses. By adopting new technology, companies can be the first to enter a new market, gaining an advantage on their competition. (McKinsey Global Institute, 2018). The unknown factor here is how will this economic transformation affect Finland, and if there are varying changes within the country between municipalities.

Before the arrival of industry, the population of Finland used to be evenly distributed. The arrival of industry and railroad paved way for industrializing cities, which created a population flow from rural areas to urban areas. This led to a phenomenon called the rural flight or urbanization, as population started to transition to larger cities to look for work. Globally the trend has been the same as in 1950, only 30 percent of population was living in urban areas. In 2050, it is expected to rise to 68 percent. Population decline in rural areas has been a concern in Finland, as companies and services move away from rural areas to more competitive cities for financial reasons. (Kotaniemi et al., 2011; United Nations Publications, 2019; Tervo et al., 2018).

1.1. Background of the research

Machine Learning is a technology that is used to automate tasks from various occupations, and industries and thus generate economic value. (Brynjolfsson et al., 2018). It is used to create accurate approximations of algorithms, as there are problems where the input and the output is known, but not the algorithm. (Alpaydin, 2014). Machine Learning has become prevalent as there are increasing efforts to analyze, make decisions, and profit from large-scale data. (Jordan & Mitchell, 2015).

Competitiveness is dependent on the efficiency of the output that an area produces. There is a strive to increase their competitiveness, as it increases the standard of living, and the economic wellbeing of the area. (Case et al., 2013; Porter, 2000; Begg, 1999). Economic performance is typically measured by the Gross Domestic Product of an area, which measures the added value of production. Alternatively, salaries of an area are also used in certain situations, as it is a better estimate of economic wellbeing in an area, since the GDP does not measure if the produced wealth stays in the area (Huovari et al., 2001). The relationship between growth in GDP and growth in productivity is a significant one, making productivity growth a key determinant in economic prosperity on a regional scale. (Gardiner et al., 2004).

The consensus on Machine Learning is that it will have a vast impact on the surrounding economy, but the estimates vary. It has been expected to automate anywhere from 8 to 35 percent of work currently done by people. (Arntz et al., 2016; Frey & Osborne, 2013). Countries are expected to benefit from the adoption of Machine Learning, with some areas of the world benefitting more than others. There is cause for alarm of increased inequality between countries if the gaps in economy are increased. (McKinsey Global Institute, 2018).

While it is often companies who are searching for new ways to gain the competitive advantage, it also needs to be stated that technological shifts affect regions as well, whether it is continents, countries, or municipalities. The issue of competitiveness has spread to national, regional, and local level, becoming a major issue for the policymakers. Governing bodies are searching for indicators of competitive performance and adjusting policies in hopes of improving the regions competitiveness. These comparisons are mainly conducted by creating competitiveness indexes on which different regions are ranked. (Kitson et al., 2004).

Regional competitiveness has been studied on an EU level. A yearly competitiveness research is published, which measures 268 regions within EU on their ability to provide an attractive and sustainable area for companies and people to work and live in. To measure the competitiveness, an index is built, containing a total of 84 variables. These variables are divided into eleven different pillars, which form the aspects of competitiveness. (Annoni, P., & Dijkstra, L., 2019). In Finland, regional

competitiveness has been measured by surveying business leaders in Finland to enquire them on what makes a region attractive for business purposes. (Saario, 2016). Regional competitiveness has also been studied through the lens of building a competitiveness index for Finnish regions. The competitiveness index was built from four categories, which formed the aspect of competitiveness in a Finnish region. These four pillars contained a total of 16 variables, from which a competitiveness index score was calculated for 85 regions of Finland. (Huovari et al., 2002).

The scale of automation caused by Machine Learning has been researched by calculating each occupation a percentage chance of getting automated. The occupations that cross a certain percentage threshold are then labeled high-risk occupations. This method was used on US occupational data to create a percentage for each occupation, and the scores used on employment data to see the amount of employed people being at high-risk of getting automated. These scores were then mapped on Finnish and Norwegian employment data to calculate the percentage of population that has a high-risk of being automated in the future by Machine Learning and mobile robotics. (Frey & Osborne, 2013; Pajarinen et al. 2015). Alternative approach was used to calculate a similar score for every occupational task. (Brynjolfsson et al., 2018; Arntz et al., 2016).

1.2. Research problem and research questions

The main objective of this research is to find out how Machine Learning affects competitiveness of municipalities in Finland. In order to achieve this, a measurement for both the impact of Machine Learning and regional competitiveness is formed. The main research question for the thesis is the following:

1. What are the economic and competitiveness impacts of Machine Learning on municipalities of Finland?

To answer the question, it was split into three sub-objectives of which the first one aims to search for the impact that Machine Learning has on competitiveness of municipalities in Finland. The second one is to figure out the economic ramifications

that competitiveness has on municipalities in Finland. The third sub-objective inspects how SML affects economy of Finnish municipalities. More specifically, how does Machine Learning affect the Gross Domestic Product and salaries of Finnish municipalities. The research questions are thus further examined with these three sub-questions:

1.1. How does Machine Learning affect the competitiveness of municipalities in Finland?
1.2. How does competitiveness affect the economic development Finnish municipalities?
1.3. How does Machine Learning affect the economic development of Finnish municipalities?

1.3. Structure of the research

This thesis has been split into 7 chapters, of which the first one is the introduction of the topic and research. Next chapter is Machine learning and its impacts, which covers the relevant topics related to Machine Learning, and how it affects the surrounding economy. Third chapter covers the other important component of the research, which is competitiveness. How competitiveness is measured, what it is, and how does it relate to Machine Learning. Fourth chapter covers what data was used, how it was collected and how it will be used in analysis. Next chapter covers the analysis, the empirical part of the research. The results of the research are presented in chapter six, and the final chapter covers the discussion and thoughts on further research into the topic.

2. MACHINE LEARNING RELATED CONCEPTS AND THEIR IMPACTS

This chapter will cover the concepts presented in the study. It will cover machine learning, its related concepts, and the economic implications that machine learning will have on business and society.

2.1. Automation

Processes which are achieved without the assistance of human labor is called automation. People can participate in an automated process, but the process must operate under self-direction. (Groover, 2019, pp. W-97).

Agriculture used to employ roughly 40 percent of the workforce in the US back in 1900. Hundred years later, that figure had fallen to 2 percent. (Autor, 2015). Workplace technology is used to replace labor, technological advancements have transformed industries. In agriculture, manual labor was largely replaced with tractors, and powered tools have affected employment in construction. Transformations are not exclusive to manual labor; implementation of computers has also affected office work. Automation that is achieved with computer-controlled equipment is often referred to as computerization or digitalization. (Frey & Osborne, 2013). Computers have automatized payroll tasks and other work that had previously been done by humans. (Autor, 2015).

Technological advancement drives long-term productivity growth of economies through automation. (Warr & Ayres, 2006). From an economic perspective, there are conflicting results on the effect that automation has overall. Negative effects stem from job losses caused by automation. (Acemoglu & Restrepo, 2017, p. 36). In the past decades, the increase in IT has helped automation drive productivity and disrupt labor markets through the automation of human labor. (Autor, 2015). This has led to the fear of large-scale job destruction and insufficient number of job creation, causing mass unemployment. (Pajarinen et al., 2015). Consensus among economists suggests that while job losses are a negative side effect of automation, it is offset by the job growth from the resulting productivity increase. (Brynjolfsson & McAfee, 2016). Technologies

that enable automation raise productivity, allowing for the output to stay constant with less input, reducing cost of production (Case et al., 2013, pp. 253). However, historically automation often has a net positive effect on employment, as the technologies causing automation to have created demand for new type of employment. Lower costs may introduce additional demand, spiking the demand for labor. Automation induced labor productivity may also result in increased labor income either through wages, employment, or both. (Arntz et al., 2016).

2.2. Digitalization

Digitalization describes the process of adopting and using digital technologies in individual, societal, and organizational contexts. (Legner et al., 2017). By equipping machines with digital systems, digitalization enables machines to operate and communicate with other machines. (Lerch & Gotsch, 2015).

Digitalization has emerged over time in waves. First wave started with the computers replacing paper in business, which led to higher automation. Internet started the second wave, as with it came a global communication infrastructure, which changed the way companies create value, and created opportunities for new types of business, like e-commerce. The third wave came in the 2010s, it brought social, mobile, analytics and cloud computing technologies. These transformations allow for new types of business models to emerge, as well as enhance existing products and services. (Legner et al., 2017, p. 305).

Results show that digitalization may have strong positive effect on economic growth, employment growth, and drive forward productivity of labor. (Evangelista et al., 2014). Digitalization can have a large impact on productivity even on the smallest job components, making routine aspects of occupations easier and more efficient. (Kraut et al., 1989).

2.3. Artificial Intelligence

Artificial Intelligence (AI) is a field that analyzes the intelligently acting computational programs. Intelligently acting programs can act according to their goals, are capable of changing goals and environments and learn from experience while working the limitations set to it. The goal is to get an understanding of what makes behavior intelligent in a system. This can be achieved by analyzing the programs, conducting experiments on building intelligent programs, and building these programs to perform intelligent tasks. (Poole & Mackworth, 2017, pp. 4–9).

AI is defined as a branch of Computer Science. (Luger, 2008, pp. 1–2), which has itself spawned multiple subfields, such as Machine Learning and robotics. (McCorduck, 2004, pp. 417). AI is expected to transform work through automation, freeing up employees to become more productive. (Davenport & Ronanki, 2018).

According to PricewaterhouseCoopers report, implementation of AI will increase the global GDP by roughly 15 trillion by the 2030. This equates to a 14 percent increase (PricewaterhouseCoopers, 2017). Report by McKinsey presents a similar estimate as AI is expected to increase global economy by 13 trillion by 2030, with 70 percent of companies implementing AI into their business. While we do not have specific numbers on how Machine Learning will affect the economy on its own, Machine Learning is prevalent in majority of AI technologies currently in use. Machine Learning algorithms are often being used to speed up processes and improve matching (McKinsey Global Institute, 2018).

The economy boost is expected to come with increased productivity from automation of business processes, and from AI technologies assisting labor force. Productivity gains come from automation of tasks and allowing employees to focus on higher-value work. Another aspect is the consumer demand, which will create a need for products and services that use AI applications. The benefit of AI is customer insight, that allows companies to better cater to their customers need on an individual level. (PricewaterhouseCoopers, 2017).

While AI has the potential of delivering economic growth, it may not be distributed equally across regions. (McKinsey Global Institute, 2018). The effect that AI will have on economy differs between regions in the world. The expectation is that it will affect China the most with a potential 26 percent increase by 2030, while Northern Europe's economy is expected to gain up to a 10 percent over the next decade. (PricewaterhouseCoopers, 2017).

2.4. Machine Learning

Computer problems are typically solved using mathematical algorithms, which takes numbers as inputs and turns them into desirable outputs for the user. There are also tasks, for which the input and output is known, but there is no algorithm. This lack of knowledge is tackled by making predictive systems by analyzing large quantities of data. These predictive systems are used to make accurate approximations based on past performance. This is known as Machine Learning. (Alpaydin, 2014, pp. 2–3) Machine learning (ML) is a subfield within AI that is expected to automate tasks from various occupations, and industries in the upcoming decades and generate economic value. (Brynjolfsson et al., 2018).

The essence of ML is that it allows the program to adapt to their results and allows it to modify its parameters to improve accuracy. (Marsland, 2014, pp. 25-26). ML algorithms search through a large quantity of data and look for the choice, searching for the best outcome based on a performance metric. (Jordan & Mitchell, 2015).

We have seen a surge in mobile and networked systems in the recent decade. To handle the large quantities of data that come with this change, ML has become more prevalent as people try to analyze, make decisions, and benefit from large-scale data. Prominence of data in organizations has also increased demand for new ML algorithms. (Jordan & Mitchell, 2015).

The recent explosion in computing has made it possible to use ML on increasingly complex tasks and the explosion of the amount of data has allowed the application on

ML on a larger range of domains. (Louridas & Ebert, 2016). ML models are now being used to automate tasks that previous waves of automation have not been able to reach. (Brynjolfsson et al., 2018). ML is prominent in society today, it can be used to turn speech into text, in recommendation systems, and many other things. (LeCun et al., 2015). ML has impacted a variety of industries which are centered around data, from consumer services and e-commerce to the control of logistics chains and fault diagnosis in complex systems. (Jordan & Mitchell, 2015). It is described as a general-purpose technology (Brynjolfsson et al., 2018), meaning that it can possibly affect the surrounding economic system and lead to changes in social factors like working hours and constraints on family life. (Helpman, 1998, pp. 3–4).

ML is a young field with a lot of underexplored research opportunities, even though it has had practical and commercial success. (Jordan & Mitchell, 2015). Businesses have generally been using ML to automate business processes, engage with customers and employees, and gain insight from data analysis. However, companies have faced challenges introducing initiatives related to ML and AI. The biggest obstacles being the integration of projects into existing systems and processes, the cost of knowledge and technology, the lack of understanding of how technology works, difficulty to get knowledgeable people, and immaturity of the technologies. (Davenport & Ronanki, 2018).

3. COMPETITIVENESS OF REGIONS

This chapter covers the aspects that form competitiveness. It starts by defining the meaning of competitiveness, then exploring the ways it can be measured. Then it takes a closer look on what components drive competitiveness. Lastly, its connection to the topic of ML is explored

3.1. Definitions of competitiveness

When the output of an economy increases, it is called economic growth. Economic growth is achieved by increasing the number of resources, or by producing more with existing resources. An increase in resources can mean either an increase in capital stock or a larger labor force. Increase in production, or productivity can be improved with an innovation or a use of a new technology or production technique. Alternatively, the use of new equipment and machinery can be used to increase productivity. (Case et al., 2013, pp. 44-45).

The goal of economic development is to reach long-term development in a nation's standard of living. Standard of living is measuring the productivity of a nation's economy, it is measured by the value of the goods and services produced per unit of human, physical, and capital resources. Prosperity of an economy depends on productivity, how factors are used and upgraded in a particular location. Location then influences productivity and its growth, affecting the competitive advantage. Competitiveness is thus defined by productivity. (Porter, 2000). Begg (1999), defines competitiveness as the ability to sustain change in factors that raise the productivity growth in the long term. The derived seven aspects that form competitiveness between regions of a country are standard of living, employment rate, productivity, company characteristics, business environment, capacity for innovation and learning, and top-down sectoral trends and macro influences. These interconnected factors form the urban competitiveness of a region. (Begg, 1999).

The most common method for measuring economic performance of a region has been Gross Domestic Product (GDP), which measures the added value of production. (Huovari et al., 2001). The change in GDP from one period to another can be

measured as growth in the economy. (Lepenies, P., 2016). The problem with using GDP as an indicator is that not all the capital stays within the region, as some of it transfers over to other regions and countries. However, GDP is not always the best way to measure economic performance in a country, especially when measuring the economic wellbeing of the population. For these reasons, the regional salaries are also used to measure economic performance. While there are issues related to salaries like transfer of wealth that distort the accuracy of the region's economic performance, this narrows down the differences between regions, but does not change the performance order. (Huovari et al., 2001). There is a significant relationship between growth in GDP and growth in productivity, making productivity growth a key determinant in economic prosperity regionally. (Gardiner et al., 2004).

3.2. Measuring competitiveness

There are multiple studies done on regional competitiveness in Finland and the EU region. In Saario, K (2016), Finnish business owners and executives are interviewed about the strengths and weaknesses of Finnish municipalities that they observe from a business standpoint. The responses were broken down based on the industry that the business was in, which are the service, construction, industrial, and the commerce industry. Among that, the answers were also detailed for 19 different municipalities in Finland. The survey contains questions about regions in terms of quality of life, transportation and logistics, workforce availability and suitability, and growth and internationalization. Based on the results, the study concludes that for a region to increase their competitiveness, the most important aspects were to provide suitable workforce, and traffic connections. For companies with abroad operations, the availability of air travel was critical.

Huovari et al. (2002), considers an index for 85 municipalities in Finland to measure competitiveness. The index is formed from four parts, which are Reachability, Innovativeness, Centralization and Social Equity. Social Equity covers the skills and knowledge of the region as much of the success of any given municipality rests on how skilled the people working there are, and how informed they are as decision-makers. Five indicators were selected to represent this concept, which are: The amount of working age people in the region, employment participation percentage of

working age people, the number of students, number of engineering students, and the number of people holding a tertiary degree. The second part is innovation in which innovation is seen as a factor in increasing productivity. The indicators are: Research and Development costs, Patents, Innovative places of work, Value added by cutting edge technology. The third part is centralization, where region is measured if it has a strong local market which can be either diverse or specialized in something. Begg (1999) states that centralization can offer cooperation between companies in the region and more subcontractors. The indicators selected for centralization were: Population density, Percentage of employees working in a centralized industry, Percentage of employees working in a supporting industry, Percentage of population working in the largest industry. Reachability is the last part of the competitiveness index. It is based on geographical location as well as the quality of infrastructure. This part measures how well a region grants access to a market. The indicators chosen for this were: Distance of markets, Distance of Airports, Foreign export and import connections.

European Union conducts yearly regional competitiveness research on NUTS2 regions. NUTS is a hierarchical system designed to divide the economic territories of the EU for regional statistics, to run socio-economic analyses on the regions, and for framing of the EU policies. In the study, regional competitiveness is described as: “Regional competitiveness is the ability of a region to offer an attractive and sustainable environment for firms and residents to live and work in”. The Regional Competitiveness Index (RCI) is originally composed of 84 indicators, 10 of which were removed due to not being statistically significant. The rest were used to measure the competitiveness level of the 268 regions across the 28 member states of the EU. These indicators are grouped into eleven different pillars, which are Innovation, Institutions, Business Sophistication, Technological readiness, Market size, Labor market efficiency, Higher education and lifelong learning, Basic education, Health, Infrastructure, and Macroeconomic stability. The purpose of the index is to measure the productivity and efficiency of enterprises, but also cover long-term potential and societal well-being of the region. It differs from traditional approaches of regional economic performance as it does not focus on only the business competitiveness and environment. (Annoni, P., & Dijkstra, L., 2019).

3.3. Drivers of competitiveness

Productivity in a location is not dependent on the industry, but on how the companies compete within the industry. (Porter, 2000). Innovation is one aspect in competitiveness, as it allows companies to develop new products and processes. (Begg, 1999). Industry does not bring in prosperity if the companies are not productive. Companies can increase their productivity in any industry by differentiating from their competition, whether it by offering new products and services, by employing new processes, or using a new advanced technology. (Porter, 2000). Efforts in innovation can further create new industry segments. The level of innovativeness in a region is dependent on factors, such as the ability to harness investment in the region, availability of support for research efforts, pressure from customers, and surrounding network with which experiences and benefits can be shared with. (Begg, 1999).

Clusters are geographic concentrations of specific companies and industries. Clusters can be found in both advanced, as well as developing areas on different geographical levels, such as national, or regional levels. Concentrations create opportunity for closer linkages with suppliers and other institutions, which can benefit the rate of improvement and efficiency. Thus, location can affect the productivity and productivity growth, creating competitive advantage. (Porter, 2000). This centralization is defined as sectoral trends, as the mix of industries within a region are shaped by the regions historical development. (Begg, 1999). On a country level, advantages should be recognized and exploited, making specialization inevitable. (Krugman, 1996). However, if the demand for certain specialization diminishes over time, it will have damaging effects to the economy, meaning that over-specialization can pose a serious threat to a region (Cuadrado-Roura and Rubalcaba-Bermejo, 1998).

The business environment refers to the factors that are outside the control reach of the companies within the region. This includes social equity factors, such as availability and skills of the labor force, availability of schools, and training facilities. To attract companies to a specific region, the business environment needs to be suited to their needs. Business environment is molded by regional policy with the intent that it can

serve the private sector by affecting the costs to do business in the region, as well as ease of doing business in the region. (Begg, 1999).

Access to an extensive market offers benefits in lower cost which allows for increase in productivity. It also provides an opening for information access, relationships within an industry increase companies' chances to innovate. (Porter, 2000). Begg (1999) exemplifies this with the fact that a location is under disadvantage if the delivery of products takes a longer time, or there is no quick access to the needed business services.

Ciampi (1996, p. 144) states, that increase in competitiveness does not come at the expense of someone else, meaning that competition is not a zero-sum game. Increase in competitiveness in one area can have positive effects on others, as their method is integrated elsewhere. (Ciampi, 1996). However, in cases where productivity growth is higher than GDP growth, gains in one area comes at the expense of another, as the growth comes from market share redistribution. (Begg, 1999). There are opposing views in how differences in competitiveness are expected to change over time. Neoclassical Growth theory specifies that innovation spills across borders, evening out the productivity differences between regions over time. Endogenous Growth theory is based on the idea that convergence is a slow process, and innovative regions attract skilled labor from other regions, which leads to persisting or widening differences. (Gardiner et al., 2004).

3.4. ML and competitiveness

Frey and Osborne (2013) inspect automatization from the angle of how computerization affects employment in the United States. For each occupation, a probability of automation by computerization was calculated. This probability is derived from nine different variables, on which the occupation is rated. The variables are finger dexterity, manual dexterity, cramped workspace, originality, fine arts, social perceptiveness, negotiation, persuasion, assisting and caring for others. Pajarinen et al. (2015) study the impact that computerization will have on the employment of Finland and Norway using the approach from Frey and Osborne (2013). Arntz et al.

(2016) criticize Frey and Osborne (2013) for assuming that workers in the same occupation share the same tasks, when there is actually a lot of differentiation and thus risk of automation may vary even within an occupation. (Arntz et al., 2016).

In Frey and Osborne (2013), three different Gaussian process classifiers were tested on sample data. The exponentiated quadratic model was chosen, as it outperformed the linear model corresponding to logistic regression and the rational quadratic. The model was then used to predict all the occupations from the O*NET database with the crafted variables. These probabilities are then used to place occupations in low, medium, and high-risk categories. Occupations with a probability of less than 0.3 are considered low-risk, and probability of over 0.7 is considered high-risk. They conclude that while the new technology is increasingly able to perform non-routine cognitive tasks, that is associated with high-wage and high-education occupations, the model predicts that computerization will affect low-income and low-education work the most. The transportation, logistics, office and administrative support, and service industries are predicted to be most affected by computerization. Another observation is that following the wave of automation will be a technological plateau. The challenge for the next wave of computerization will be to overcome the constraints related to creativity and social intelligence, such as negotiation and fine arts. (Frey & Osborne, 2013). With this method, Pajarinen et al. (2015) calculate the probability of computerization which is implemented on Finnish and Norwegian occupational data to produce the results, placing 35 percent of Finnish, and 33 percent of Norwegian employment into high-risk category for computerization. The paper also finds that high-wage, high-education, public sector, and service industry jobs are generally safer from computerization than low-wage, low-education, private sector, and manufacturing jobs, and that in Norway women tend to be more secure from computerization than men. Conclusion is that while computerization does increase welfare globally, it is unknown how gains from computerization are distributed as lucrative and popular digital platforms are owned by foreign entities. (Pajarinen et al., 2015). Arntz et al. (2016), criticize the Pajarinen et al. (2015) study for the implementation of Frey and Osborne (2013) occupation-based method for calculating automation, as they make an assumption that occupations are identical between Finland, Norway, and the US.

Arntz et al. (2016) implements a task-based approach instead of an occupation-based one used in Frey and Osborne (2013). Automation probability is calculated for tasks, and the occupational automation risk is calculated based on which tasks the occupation contains. Occupations with more automatable tasks have a larger risk of getting automated than the occupations with less automatable tasks. The report sides with Frey and Osborne (2013) in that low-education jobs are under a greater risk of automatization, but in the US only 9 percent of occupations are estimated to be in the high-risk of automatization, which is an automatability of over 70 percent. The percentages were calculated for all OECD countries, the difference to Pajarinen et al. (2015) is that Finland received a high-risk automation score of 7 percent instead of 35, and Norway received 10 percent instead of 33. (Arntz et al., 2016).

The main crux of this research revolves around Brynjolfsson et al. (2018), which observes how ML affects the labor force through a task-based method, similar to the Arntz et al., (2016) approach. In the study, a Suitability for Machine Learning (SML) score was calculated for each occupation and occupational task based on how well a job task is suited for ML. The higher the SML score, more likely it is that it can be automated with ML. The study concludes that very few occupations could be automated completely, alternatively there were few instances of occupations that were completely safe from ML, as almost all occupations contained tasks that can be automated and tasks that cannot be automated with ML. Because of this, a shift in the labor market is assumed, as jobs will be reorganized because of ML. Conclusion is that it is work that is automated, not occupations or people.

3.5. Model and hypotheses

Using existing theory established in the chapter, a competitiveness index was created to measure the competitiveness of each municipality in Finland. For ML, the SML score calculation method from the Brynjolfsson et al., (2018) was used to measure the chances of occupations to get automated with ML. The higher the SML score is, the more likely something is to get automated, thus expectation is that SML affects competitiveness negatively. SML is also expected to negatively impact GDP and salary.

Figure 1 shows the process of building the model for the research. Each part of the process is separately documented and presented in the thesis.

Figure 1. Model building process.

Based on this, the hypotheses for this study are the following:

H1: SML score has a negative impact on the competitiveness of a municipality.
H2: SML score has a negative impact on GDP.
H3: SML score has a negative impact on salary.

4. RESEARCH METHODS

This chapter will cover the collection on data, and the descriptive statistics of the created dataset. Additionally, the data preparation methods are explained, and data evaluation described.

4.1. Data collection

The data for the competitiveness index created for this study contains 17 variables for all Finnish municipality regions for years 2000-2019. The dataset replicates the Huovari et al. (2002) competitiveness index for variables, with one added exception as knowledge workers variable is added to the dataset from the Annoni et al. (2019) competitiveness index for NUTS2 regions in the EU. The competitiveness index is then compared against the yearly mean salary per capita, and GDP per capita data of Finnish municipalities. For this section, the parts of competitiveness that can be quantified were selected to create the competitiveness index.

All data for the competitiveness index dataset was collected from secondary sources. Patent data was requested from Patent Registry Office of Finland, industrial foreign connections data was received from Customs Finland, Market reachability was crafted using OpenStreetMap distance data, measuring each municipality's distance to other municipalities in Finland. All other data was collected from Statistics Finland's public database, or directly requested if not available. Some of the air traffic data was missing from Statistics Finland and was requested from Finavia, and the Lappeenranta Airport.

The SML score data was used to measure the potential of ML affecting employment in each Finnish region. It was measured using the Brynjolfsson et al. (2018) as a basis. The paper contained a list of occupations in the United States from O*NET database on which an SML score was calculated for each occupation. A complete list of Direct Working Tasks (DWA) from the U.S Department of Labor's O*NET database was listed, and questionnaire was conducted with 23 questions containing a grading scale of 1 to 5. With this questionnaire, the participants of the survey answered for the DWA's they were familiar with. After calculating the SML scores for each occupational task, the SML score were aggregated for each of the 964 jobs in the United States

found in the O*NET database based on the occupational tasks they contain to see how much can be automated in each occupation. To have SML scores for Finnish occupational data, a list of occupations in Finland was collected from the ministry of Economic Affairs and Employment of Finland, and the US occupational data was mapped on to the Finnish occupational list. The corresponding SML score from the US occupational data was then used in the Finnish occupational data. SML scores for each municipality was calculated by summing up the amount of each occupation in a region and taking the weighted mean of the SML scores.

4.2. Data preparation

There are a total of 17 variables in competitiveness index, as seen in *table 1*. The data is collected from 310 regions for all but 5 variables. One variable data is from 70 regions, and the rest are from 19 regions. The populations for each of the 70 and 19 regions were calculated, then the values were distributed to the 310 municipalities based on which region they belonged to, and what percentage of the region's population did the municipality have. All variables in dataset have been normalized to output values between 0 and 1, except for variables where the values are percentages, as the range is already between 0 and 1.

Variable	Source	Coverage	Role
Working age population	Statistics Finland	2007-2018, 310 regions	Social Equity
Participation in the workforce	Statistics Finland	2000-2018, 310 regions	Social Equity
Students	Statistics Finland	2000-2018, 310 regions	Social Equity
Engineering students	Statistics Finland	2000-2018, 310 regions	Social Equity
Tertiary degrees	Statistics Finland	2000-2018, 310 regions	Social Equity
Share of R&D investment	Statistics Finland	2000-2018, 70 regions	Innovativeness
Patents	Patent Registry office of Finland	2000-2018, 310 regions	Innovativeness
Innovative places to work	Statistics Finland	2007-2018, 310 regions	Innovativeness
Share of high technology in economic output	Statistics Finland	2013-2018, 19 regions	Innovativeness
Knowledge work	Statistics Finland	2010-2018, 310 regions	Innovativeness
Centralization of population	Statistics Finland	2000-2018, 310 regions	Centralization
Share of centralized industries in employment	Statistics Finland	2007-2018, 310 regions	Centralization
Share of business service industries in employment	Statistics Finland	2007-2018, 310 regions	Centralization
Share of biggest industry in employment	Statistics Finland	2007-2018, 310 regions	Centralization
Market reachability	OpenStreetMap	2000-2018, 310 regions	Reachability
Reachability of air traffic	Statistics Finland, Finavia, Lappeenranta airport	2000-2018, 19 airports	Reachability
Industrial foreign connections	Customs Finland	2015-2018, 19 regions	Reachability

Table 1. Variables used in competitiveness index.

R&D investment, Patents, Innovative places to work, share of high technology in economic output, share of business service industries in employment, and industrial foreign connections variables were calculated by the formula in *Equation 1*.

$$Result = \frac{municipality_{\ i}/whole\_country_{\ i}}{municipality_{\ p}/whole\_country_{\ p}} \qquad \begin{array}{l} i = variable \\ p = population \end{array} \qquad \textbf{(1)}$$

Largest employment sector and participation in the workforce were calculated by dividing the share of the variable in the municipality by the share of the variable in the whole country. In market reachability, the population within 200-kilometer radius is calculated for each municipality. The population is also multiplied with a value between 0 and 1, based on the distance from the inspected municipality. The closer the municipality is, the higher the coefficient in multiplication. For air traffic, a similar formula was used. Instead of nearby population, the number of passengers travelling

by air was used in the calculation. The formula for market reachability and air traffic reachability is found in *Equation 2*.

$$result_i = \frac{\sum_{j=1} max\left(0; municipality_j * (1 - \frac{distance_{ij}}{200})\right)}{max\ (municipality_i)} \qquad \textbf{(2)}$$

Centralization of population was calculated using the population and the area size of municipalities. Equation 3 shows the formula used for the calculations.

$$result_i = ln\left(\frac{population_i/area_i}{population_{country}/area_{country}} + e - 1\right) \qquad \textbf{(3)}$$

The competitiveness index is calculated by dividing the variables into 4 different groups, which are Reachability, Innovativeness, Centralization and Social Equity. (Huovari et al., 2001). Each variable in a group is given an equal weight, and the score of each group is summed together to form the final score for the competitiveness index.

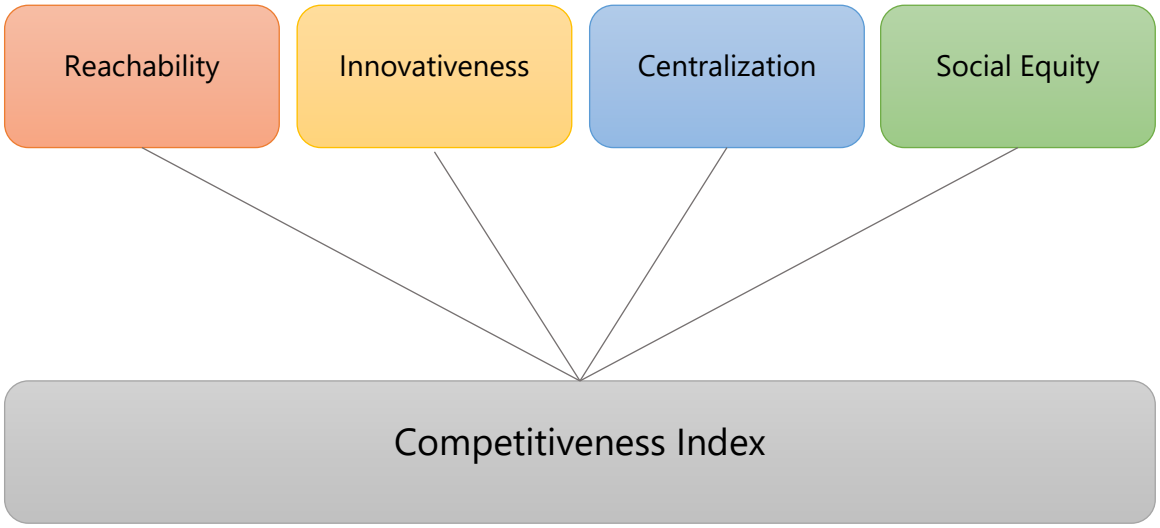


Figure 2. Four dimensions of competitiveness

After calculating the SML scores, they were standardized, as the range for the scores was low, ranging between 3.4 and 3.5. These standardized results were then normalized, with each value having a score between 0 and 1.

The data collected from secondary sources contains varying number of municipalities based on the year the data is from. The dataset was formed based on the 2019 municipality divide, meaning that municipalities that no longer exist were merged to the municipality they belonged to in 2019. The total amount of municipalities was 310. The data was collected from 2000-2019 if available. The GDP and salary data for Finnish municipalities was collected from Statistics Finland. The data used for analysis ended up being from years 2010-2018, because of this, variables Industrial foreign connections, and share of high technology in economic output had to be cut from further analysis.

4.3. Analysis methods

This part explains the quantitative methods that were used in this study for analysis. The analysis starts with exploratory analysis to understand the nature of the data. Data exploration findings are covered in the descriptive statistics for the SML scores, the competitiveness index, the dimensions, and their relationship with the GDP and Salary variables. For the dimensions, we want to do a dimension reduction of our data using factor analysis. After cutting two of the initial 17 variables, there are 15 variables in the index, and Factor analysis allows us to trim down the number of variables used. Factor analysis results can then be used for further analysis, which will be panel data analysis.

The competitiveness index in this study is a formative by nature. This means that the index Is formed from the variables associated with it, but It does not affect the variables its composed from. In order to use the index in analysis, it was first tested for reliability and validity. For this, explorative factor analysis was used, a principal component analysis (PCA) method.

After conducting the factor analysis on the competitiveness index, cross-correlation analysis is performed. The point of the analysis is to capture any effects that occur

between variables over a certain period of time. The final analysis is the panel data analysis, which will be conducted using the OLS model, the fixed effects model, and the random effects model.

4.3.1. Principal Component Analysis

PCA is a common technique used for dimensionality reduction. (Brooks, 2014, pp. 170). The aim of PCA is to estimate the correlation matrix between variables in the matrix, using characteristic vectors and characteristic roots, these are also called eigenvectors and eigenvalues. (Kline, 1994, pp. 30-31). PCA method is useful in cases where the explanatory variables are closely related, it takes the explanatory variables in the model and turns them into same amount of new uncorrelated variables. The components in PCA are calculated so that they are in a descending order based on importance, each variable having a percentage of variance that it is explaining. Some components may only account for a fraction of the variance and are therefore discarded. (Brooks, 2014, pp. 170-171).

Principal components are orthogonal, uncorrelated linear combinations of actual scores. (Kline, 1994, pp. 39). These principal components are calculated so that in every principal component, each of the explanatory variables is multiplied with the coefficient of the explanatory variable in said principal component. This is shown in equation 4. (Brooks, 2014, pp. 170-171).

$$
\begin{aligned}
P_1 &= \alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \ldots + \alpha_{1k}x_k \\
P_2 &= \alpha_{21}x_1 + \alpha_{22}x_2 + \alpha_{23}x_3 + \ldots + \alpha_{2k}x_k \\
&\ldots \qquad \ldots \qquad \ldots \\
P_k &= \alpha_{k1}x_1 + \alpha_{k2}x_2 + \alpha_{k3}x_3 + \ldots + \alpha_{kk}x_k
\end{aligned}
\qquad (4)
$$

P = principal component
α = coefficient of principal component i,
explanatory variable j.
x = explanatory variable

The coefficients in the equation 4 are also called factor loadings. The sum of squared coefficients for each component is required to be one. This requirement is expressed with sigma notation in *equation 5*. (Brooks, 2014, pp. 170-171).

(5)

$$\sum_{j=1}^{k} a_{ij}^2 = 1\,, \qquad i = 1,2,\dots,k$$

Factor loading is a combination of the variables with the factor which best explain the variance. (Kline, 1994, pp. 36). The factor loading value is achieved by multiplying the matrix of observations for original variables on transposed matrix of observations for original variables. Each eigenvalue is then divided by the sum of all eigenvalues. The resulting ratio is the percentage of total variance that the principal component explains. *Equation 6* shows us the formula for calculating the ratio. (Brooks, 2014, pp. 171).

$$\varphi_i = \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i}, \lambda = \text{eigenvalue on variable i} \qquad \textbf{(6)}$$

Before factor analysis results can be implemented, rotation is performed on the results. One of the methods this can be achieved is with Varimax rotation. The idea behind Varimax method is to maximize the sum of variances of the squared loadings of the factor matrix columns. Varimax aims for simple structure while keeping factor axes orthogonal. (Kline, 1994, pp. 64-68). In rotation of factors, simple structure has the following requirements:

1. All rotated matrix rows have at least one zero in them.
2. The minimum number of zero loadings should be the number of factors in the rotation for each factor.
3. For every pair of factors there should be both variables with zero loadings, and significant loadings on other.
4. Large proportion should be zero in matrices with large number of factors.
5. Every pair of factors should only have a few significant loading variables on both factors.

4.3.2. Panel data analysis

In this study, cross-correlation is run prior to panel data analysis to help establish the model parameters. Cross-correlation is a function, which describes the relationship

between two different signals at different time scales. (Jun et al., 2006). The correlation is the measure of the similarity between the signals when a lag in time is applied. This is also referred to as the sliding dot product or the sliding inner-product. Cross-correlation calculation for two series x(i) and y(i) is explained in equation 7. The cross-correlation r is defined with a delay of d, mx and my representing the mean values of the corresponding series. (Bourke, P., 1996).

$$\mathrm{r} = \frac{\sum_i [\,(x(i) - mx) * (y(i-d) - my)]}{\sqrt{\sum_i (x(i) - mx)^2}\,\sqrt{\sum_i (y(i-d) - my)^2}}, \qquad \mathrm{i} = 0,1,2 \ldots \mathrm{N} - 1 \qquad \textbf{(7)}$$

Furthermore, calculating the formula for all delays will results in a cross-correlation series that has twice the length of the original series. This formula can be seen in equation 8. (Bourke, P., 1996).

$$\mathrm{r(d)} = \frac{\sum_i [\,(x(i) - mx) * (y(i-d) - my)]}{\sqrt{\sum_i (x(i) - mx)^2}\,\sqrt{\sum_i (y(i-d) - my)^2}}, \qquad \mathrm{d} = 0,1,2 \ldots \mathrm{N} - 1 \qquad \textbf{(8)}$$

Panel data, also known as longitudinal data, combines the characteristics for both the cross-sectional data, and time-series data. Panel keeps the same entities and measure their differences over time. There are different ways of dealing with panel data, the simplest way being through pooled regression. This method would only require all of the cross-sectional and time-series data to be listed in a single column. Pooled regression could be conducted using the Ordinary Least Squares (OLS) method, the equation for a simple panel data regression be seen in equation 9. (Brooks, 2014)

$$y_{it} = \alpha + \beta x_{it} + u_{it} \quad , \quad \begin{array}{l} \text{a = intercept} \\ \text{B = slope coefficient} \\ \text{u = error term} \end{array} \qquad \textbf{(9)}$$

However, the use of pooled regression in panel data analysis has limited use, as there are restrictions to keep in mind when choosing the model. The model assumes that the average values of each variables and their relationships are constant over time

and cross-sectional units. If the data does not meet these restrictions, panel data analysis is best to conduct using an alternative method (Brooks, 2014).

The fixed effects model takes the error term of the OLS model, and splits it into two components, one of which is v, calculated for each unit in each point of time. The other error term, u in the fixed effects model is to capture the variables cross-sectional effects, but not the changes that occur over time. (Brooks, 2014). The basic equation for the fixed effects model can be seen in *equation 10*. The β acts as the vector of coefficients, α is the intercept, u and v are the previously described error terms. (Allison, 2009).

$$y_{it} = u_i + \alpha_i + \beta x_{it} + v_{it} \qquad \begin{matrix} i = 1,2,3\ldots N \\ t = 1,2,3\ldots T \end{matrix} \qquad \textbf{(10)}$$

The alternative to the fixed effects model is the random effects model. The random effects model differs in that there is a random variable ϵ_i, which measures the deviation from the common intercept α. ϵ_i stays the same over time but changes cross-sectionally. The relationships between the explanatory and explained variables stay the same cross-sectionally and between times. The difference between the two models, is that the intercepts in the cross-sectional units are expected to come from the common intercept α and a random variable ϵ_i, that stays constant between time periods, but varies cross-sectionally. It measures the deviation between the common intercept, and the random deviation of each entity's intercept. The equation of the random effects model is shown in *equation 11*. (Brooks, 2014).

$$y_{it} = \alpha + \beta x_{it} + \omega_{it}, \qquad \omega_{it} = \epsilon_i + v_{it} \qquad \textbf{(11)}$$

The choice between the fixed effects and the random effects model is done by conducting a Hausman test. The idea of the test is to compare the coefficient estimates from the fixed effects model and the random effects model. Hausman tests if the estimators in both models are consistent. (Hill, 2011).

After making the decision between the fixed effects model, and the random effects model, another test is done to see, whether it is possible to use the OLS model instead of the fixed effects or the random effects model. For the fixed effects model, the F-test is conducted on the fixed-effects model. If the null hypothesis of the F-test is rejected, the appropriate model should be the fixed effects model. Otherwise, the OLS should be the chosen model, as it is the simpler model. The idea behind the F-test is to see if the constants are similar between the entities in the dataset. The OLS model has a global constant, and thus it can be used only if there are no differences between entities. For the random effects model, a Breusch-Pagan test is conducted because OLS does not consider the presence of heterogeneity in the data. If the null hypothesis is rejected, the random effects model is the appropriate model, otherwise OLS should be used. (Hill, 2011).

5. RESULTS

This part will cover the analysis of the data. The section will start with descriptive analysis, where some basic relationships and information is established. The next part is factor analysis, where the existing competitiveness index goes through a dimensionality reduction. The resulting index and other variables are then inspected in the cross-correlation analysis part, where lagged variables of the existing variables are compared. Last part of the analysis is the panel data, which inspects the changes in the data over a period of time.

5.1. Descriptive analysis

The SML scores were calculated for each of the 310 municipalities. The standardized score results are seen on a heatmap of Finland in *figure 3*a. From the figure, we can see that SML scores tend to be lower in rural areas, as well as in the eastern and northern parts of Finland. Highest scores can be observed around the Finnish coast and the capital region. The range for the standardized scores was between -6.1 and 4.9, with the lowest and highest values being in the Åland islands, as Kökar had the lowest SML score, and Vårdö had the highest. The descriptive values for SML and standardized SML scores can be seen in *table 4*.

In *figure 3b*, the heatmap is based on the competitiveness index of Finnish municipalities. The largest values center around large cities of Finland, while rural areas have the lowest index scores. The capital region seems to form a large semicircle of high competitiveness values.

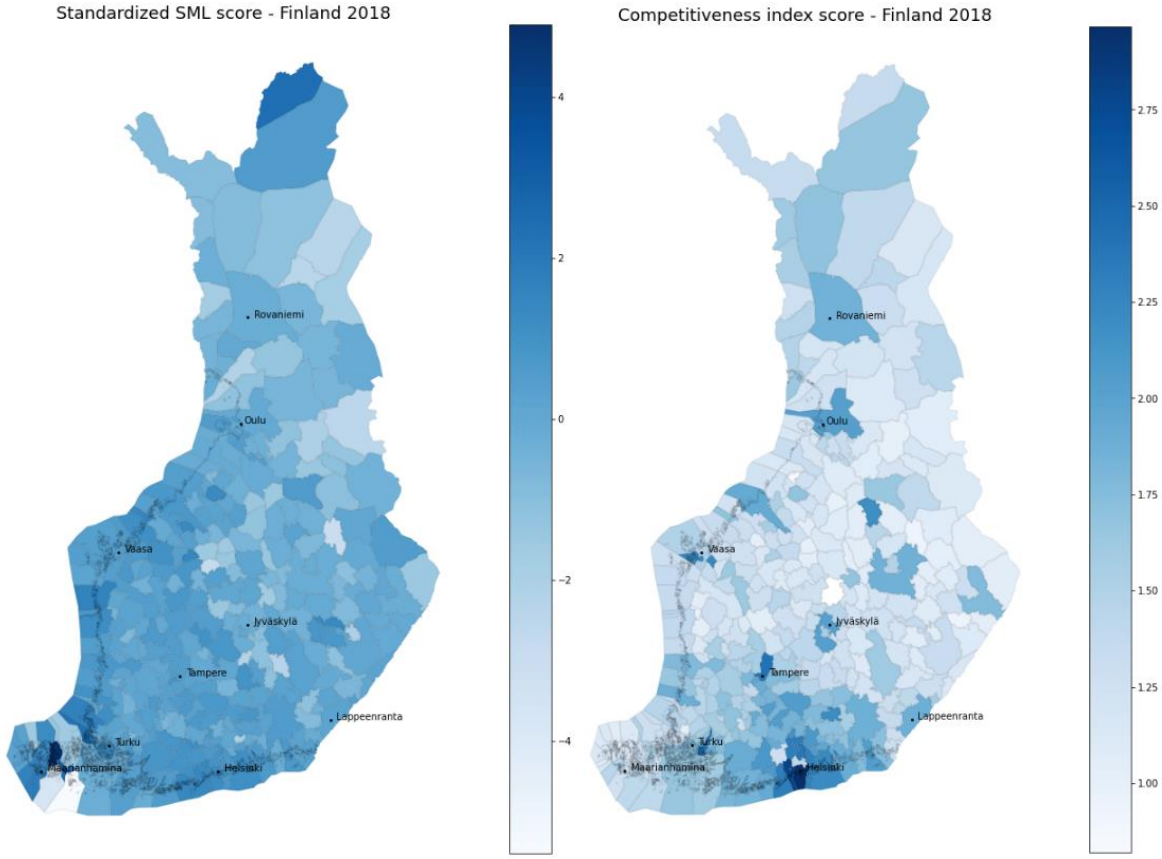


Figure 3 a&b. Standardized SML, and competitiveness index scores of municipalities in 2018.

After calculating the scores for each of the variables and the dimensions, we give each dimension an equal weight to get the final competitiveness score. For each municipality, the index score was calculated by taking the sum of the averages for each dimension. The results can be seen in *table 2a & 2b*. The capital region of Helsinki, Espoo, and Vantaa received the largest scores, top 10 also contains municipalities that are close to the capital region, such as Kauniainen, Kerava, Järvenpää, Tuusula, and Hyvinkää. The other two municipalities in the top 10 are two large cities in the southwestern Finland: Tampere, and Turku. The bottom 10 features municipalities from central Finland and the Åland islands.

TOP 10		BOTTOM 10	
Municipality	Competitive Index	Municipality	Competitive Index
Helsinki	2.71	Merijärvi	0.70
Espoo	2.58	Rääkkylä	0.81
Vantaa	2.51	Rautavaara	0.82
Kauniainen	2.43	Karijoki	0.84
Järvenpää	2.27	Kärsämäki	0.86
Kerava	2.26	Vesanto	0.87
Tuusula	2.20	Puolanka	0.89
Turku	2.20	Keitele	0.89
Tampere	2.17	Kiuruvesi	0.90
Nurmijärvi	2.15	Kinnula	0.90

Table 2 a&b. Top 10 and bottom 10 municipalities on the competitiveness index in 2018.

In *table 3*, we can see descriptive statistics for all variables of the competitiveness index. All variables have a range between 0 and 1. There are five variables in both social equity, and innovativeness, four variables in centralization, and three in reachability. The distributions of all the variables for the year 2018 can be seen in *appendix 1.1-1.15*. There are multiple variables with very low mean values. With variables engineering students, patents, research & development, and air traffic, this is because there are outlier values with very high values, pushing the majority of the variables near zero. For participation rate, the distribution is pushed very close to the maximum value. This is because the values are clumped on a very small range, as the participation rate is 74 percent at it is lowest. Market size variable shows a large spike in the lower end of the distribution. This is likely, because most of the municipalities do not have large cities nearby and thus the market size value is low.

	Mean	Std. Dev.	Min	Max	Category
Working age population	0.38	0.09	0.18	1	Social Equity
Participation in the workforce	0.87	0.04	0.74	1	Social Equity
Students	0.38	0.17	0.03	1	Social Equity
Engineering students	0.05	0.14	0.00	1	Social Equity
Tertiary degrees	0.40	0.11	0.18	1	Social Equity
Share of R&D investment	0.04	0.09	0.00	1	Innovativeness
Patents	0.06	0.10	0.00	1	Innovativeness
Innovative places to work	0.36	0.12	0.00	1	Innovativeness
Knowledge work	0.42	0.12	0.18	1	Innovativeness
Centralization of population	0.22	0.14	0.10	1	Centralization
Share of centralized industries in employment	0.51	0.11	0.23	1	Centralization
Share of business service industries in employment	0.43	0.15	0.00	1	Centralization
Share of biggest industry in employment	0.35	0.11	0.11	1	Centralization
Market reachability	0.37	0.28	0.00	1	Reachability
Reachability of air traffic	0.15	0.24	0.00	1	Reachability

Table 3. Descriptive statistics of index variables.

In *table 4,* we can see the descriptive statistics of the competitiveness index, GDP, and salary, standardized SML scores, as well as the SML scores. The standard deviation of SML is 0.01, meaning that the values are centered very close to the mean. This can be observed in *figure 3a*, where the heatmap is relatively static in its color. Salary is often presented in a monthly form, but ultimately it was not chosen for this research as it is easier to compare yearly salaries with GDP per capita numbers because of the similar scale.

	Mean	Std. Dev.	Min	Max
Competitiveness index	1.23	0.39	0.63	2.78
SML	3.45	0.01	3.40	3.53
Standardized SML	0.00	1.00	-5.60	9.57
GDP per capita	32191.87	8028.13	16957.30	77681.00
Yearly salary per capita	25490.60	4061.57	17666.00	67790.00

Table 4. Descriptive statistics of competitiveness index, SML, Standardized SML, GDP, and Salary.

The variables of the competitiveness index were tested against the SML scores, GDP, and salaries of the regions to see the relationships. Results can be seen in t*able 5*. All variables have a positive relationship with GDP and Salary. Variables with a significant correlation with GDP are Tertiary degrees, Centralization of population, share of biggest industry in employment, Knowledge work, and Reachability of air traffic, which all have strong positive correlations. For salary, participation in the workforce, tertiary degrees, knowledge work, centralization of population, share of centralized industries in employment, and reachability of air traffic had strong positive correlations. Working age population, students, engineering students, patents, and Research & Development investments all have insignificant correlation for both GDP and salary, having a positive correlation of less than .30. The rest of the variables have a weak positive correlation with GDP and salary. The index itself has a strong positive relationship with GDP as the correlation is .58 and having positive correlation of .65 with salary. The SML score has a positive correlation with both the GDP and salary. The correlation with GDP is .18, and .25 with salary.

Variable	Category	GDP	Salary
Working age population	Social Equity	.19	.04
Participation in the workforce	Social Equity	.42	.63
Students	Social Equity	.24	.23
Engineering students	Social Equity	.17	.17
Tertiary degrees	Social Equity	.53	.78
Share of R&D investment	Innovativeness	.12	.15
Patents	Innovativeness	.05	.08
Innovative places to work	Innovativeness	.30	.21
Knowledge work	Innovativeness	.55	.80
Centralization of population	Centralization	.55	.64
Share of centralized industries in employment	Centralization	.43	.59
Share of business service industries in employment	Centralization	.28	.19
Share of biggest industry in employment	Centralization	.37	.56
Market reachability	Reachability	.40	.45
Reachability of air traffic	Reachability	.50	.54
Index		**.58**	**.65**
SML		**.18**	**.25**

Table 5. Correlation matrix of index variables with GDP and salaries, 2010-2018.

Competitiveness index and SML scores were plotted to observe the relationship between the variables. The relationship between the two is a positive correlation of 0.545. The resulting plot is seen in *figure* 4, where each region represents a point on the map, and the red line representing the relationship. The municipalities that have a high SML score tend to score high on the competitiveness index. The *figure* uses non-standardized values of SML.

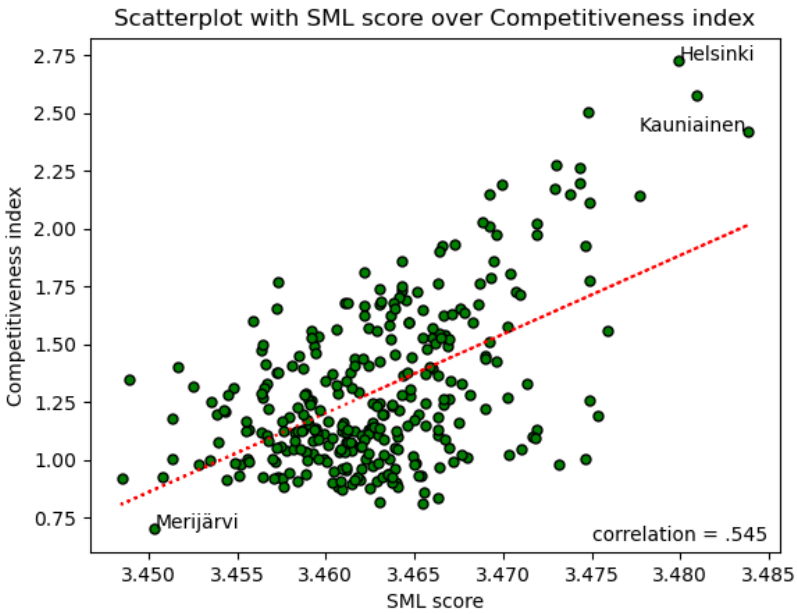


Figure 4. Scatterplot of SML score over Competitiveness index for 2018 values.

In *table 6*, we can see the correlation matrix of the competitiveness indexes explanatory groups. From the table, we can see that all relationships between the dimensions are positive, and that Social equity, Innovativeness, and Centralization all have strong positive correlations. Reachability has a moderate positive correlation with Centralization, and a low positive correlation with Social equity, and Innovativeness.

	Social Equity	**Innovativeness**	**Centralization**	**Reachability**
Social Equity	1	.777	.651	.288
Innovativeness	.777	1	.717	.319
Centralization	.651	.717	1	.555
Reachability	.288	.319	.555	1

Table 6. Correlation matrix of the Competitiveness index dimensions.

The dimensions Reachability, Innovativeness, Centralization, and Social Equity were compared with the Index, SML score, GDP per capita, and Salary per capita. The correlations are seen in *appendix 1.16*. All dimensions seem to have a strong positive correlation with the Index, Centralization and Social Equity having the largest correlation at .845 and .822. Centralization had the highest correlation with all the variables, while Reachability scored the lowest correlation against the index and SML, and Innovativeness scored the lowest against GDP and salary per capita.

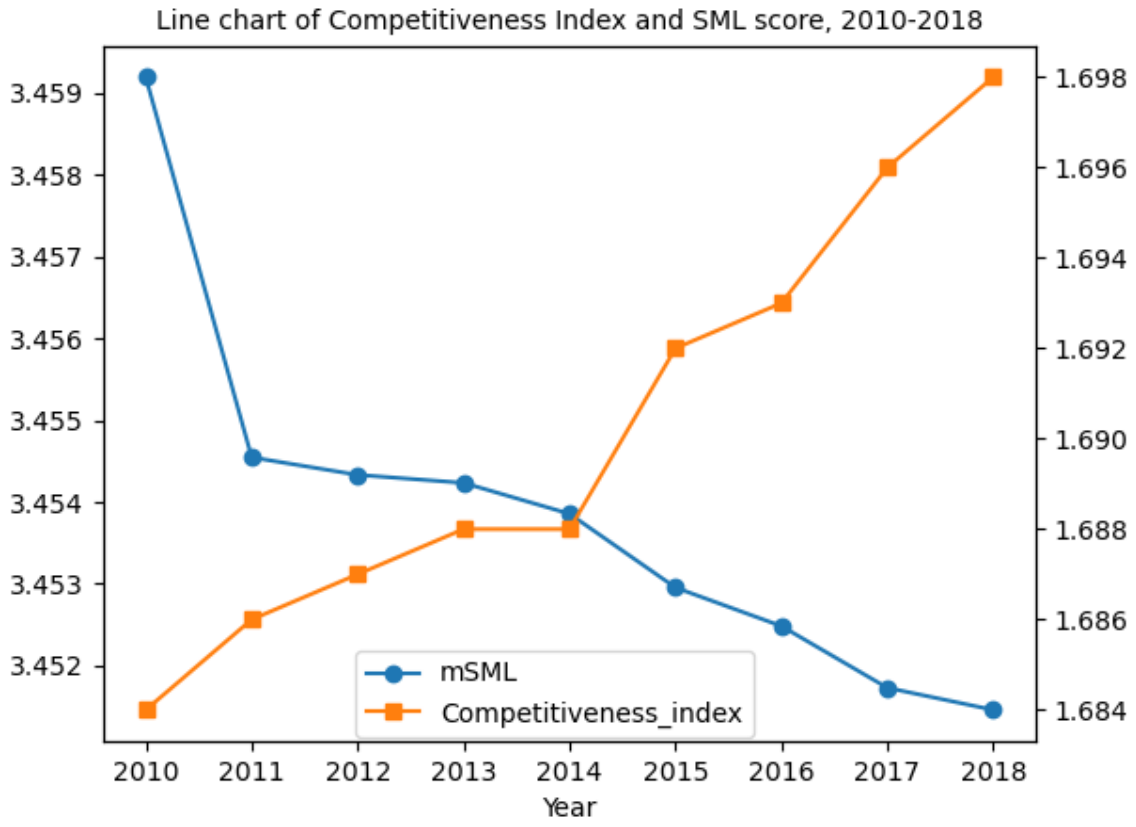


Figure 5. Line chart of SML score and Competitiveness index. SML score y-axis on the left, competitiveness index y-axis on the right.

Figure 5 shows us the movement of both the competitiveness index and the SML score over the years 2010-2018. The variables seem to be moving in opposite directions, implying some sort of negative correlation.

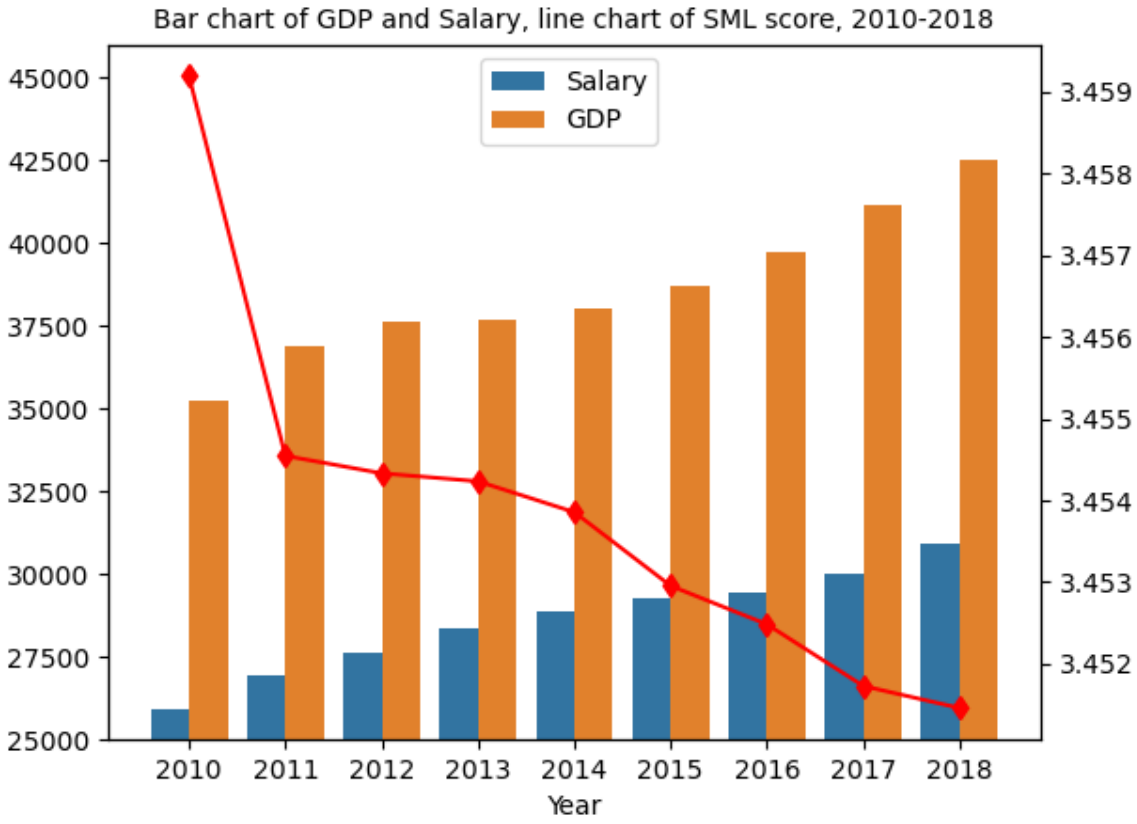


Figure 6. Bar chart of GDP and Salary, with SML score as a line chart.

In *figure* 6, GDP and salary are plotted together with the SML score for years 2010-2018. GDP and salary are represented by different colors of bars, SML score is seen as the red line. Both the GDP and salary show incremental growth each year, while the SML score moves downwards. The numbers are the mean values for the entire country of Finland.

5.2. Measure development for competitiveness

In this section, PCA was conducted to reduce the dimensionality of the index. We start off by calculating the correlations of all the variables in the competitiveness index to see what relationships they have. (Appendix 2.1). We start by removing the variables with no seeming correlation with other variables in the index. The correlation threshold is .30, which the Patent variable does not satisfy.

Initial factor analysis as shown in *table 7,* is performed on the remaining variables with the principal component method. The factors with an eigenvalue of over one was retained. The analysis had a total of 2790 observations. The analysis showed us that there a total of four factors with the eigenvalue being over one. These four factors had a total explanatory power of 77 percent, with the first one having 24 percent, the second one with 21 percent, third and fourth both having 16 percent explanatory power of the variance. Factor rotation was also done using the Varimax rotation method and applying the Kaiser normalization method, removing all loadings that were under .35. The uniqueness values are under the threshold of .50 for all variables, except Research and Development, which is at .53. This variable is thus removed from further analysis. Another variable that is removed is Innovative places to work, as it has a very high correlation of .96, and near identical loadings with the Share of business service in employment. The reason Innovative places to work is removed out of the two variables is that the variable is an approximation of what it is trying to measure. The Kaiser-Meyer-Olkin (KMO), that is also called measure for sampling adequacy showed that all of the variables had a KMO score of over the .60 threshold, passing the test with an overall score of .75.

Variable	Rotated loadings				Uniqueness	MSA
	Factor 1.1	Factor 1.2	Factor 1.3	Factor 1.4		
Working age population		.69			.51	.64
Participation rate	.72				.34	.83
Students		.81			.32	.73
Engineering students		.85			.24	.80
Tertiary degrees	.64	.49			.22	.80
R&D		.67			.53	.85
Innovative places to work				.93	.05	.64
Knowledge work	.69	.44	.36		.18	.76
Centralization of population	..43	.54	.46		.28	.88
% of Centralized industry	.86				.18	.72
% of Business services in industry				.93	.04	.65
Largest industry %	.90				.17	.70
Market size			.93		.09	.70
Air traffic			.89		.11	.75
Eigenvalue	3.3	3.1	2.3	2.1		
Cum.%	.23	.46	.62	.77		.75

Table 7. Loadings, uniqueness, and KMO values of the initial factor analysis for the competitiveness index.

After the removal of the two variables in the initial analysis, a refined analysis was run to see how the results improve. The refined analysis seen in *table 8* resulted in three factors having an eigenvalue of over one, which is one less than the initial analysis. Eigenvalues are visualized in a scree plot in *appendix 2.2*. The explanatory power of the refined analysis dropped slightly to .72, meaning that the loadings declined a bit. Tertiary degrees, and Centralization of population, and Knowledge work have significant loadings with more than one factor, with very small differences. For these reasons, both variables are eliminated from further analysis. Business services

uniqueness level crosses the .50 threshold with a value of .59 and is thus eliminated from further analysis. Uniqueness levels for the other variables are at an acceptable level, and the KMO values show good results, with an overall KMO value being .76.

Variable	Rotated loadings			Uniqueness	MSA
	Factor 1.1	Factor 1.2	Factor 1.3		
Working age population	.73			.47	.60
Participation rate		.75		.38	.83
Students	.82			.32	.70
Engineering students	.85			.27	.80
Tertiary degrees	.50	.65		.22	.79
Knowledge work	.44	.71		.19	.75
Centralization of population	.60	.44	.45	.26	.89
% of Centralized industry		.84		.25	.74
% of Business services in industry	.52		.37	.59	.89
Largest industry %		.88		.21	.75
Market size			.92	.11	.70
Air traffic			.91	.10	.74
Eigenvalue	3.3	3.0	2.3		
Cum.%	.28	.53	.72		.77

Table 8. Loadings, uniqueness, and KMO values of the refined factor analysis for the competitiveness index.

The final factor analysis is run after removing the additional three variables. The final factor analysis still has the three factors with an eigenvalue of over one, as seen in *table 9*. This is also visualized in the scree plot in *appendix 2.3*. Explanatory power of the factor analysis is now .78, with each variable only belonging to one factor. Uniqueness values are all at acceptable levels being under .50, and the KMO overall value is down from before, but still over the .60 threshold. Cronbach alpha is .70,

exceeding the .60 threshold for reliability in exploratory factor analysis. External validity was confirmed by comparing the performance of 2010 results to the 2018 results. (*Appendix 2.4*). The validity test shows good performance with no significant differences between the samples, except for variable Market size, which belongs to the first factor in 2010. Additional inspection showed that it is an outlier, as it is in the third factor for all other years in the data. Thus it can be concluded, that the factors can be used for further analysis.

Variable	Rotated loadings			Uniqueness	MSA
	Factor 1.1	Factor 1.2	Factor 1.3		
Working age population	.77			.40	.72
Participation rate		.70		.44	.81
Students	.85			.25	.57
Engineering students	.87			.24	.62
% of Centralized industry		.89		.16	.62
Largest industry %		.93		.11	.62
Market size			.93	.09	.62
Air traffic			.91	.11	.63
Eigenvalue	2.3	2.1	1.8		
Cum.%	.28	.55	.78		.64
Cronbach α	.70				
Variable-test r	.30-.80				
Variable-rest r	.19-.63				

Table 9. Loadings, uniqueness, and KMO values of the final factor analysis for the competitiveness index.

The first of the final three factors contains the variables Working age population, Students, and Engineering students, this factor will be Social Equity. The second variable contains the Share of centralized industry, Share of largest industry, and Participation rate. This factor is centralization. Last factor contains the market size, and the air traffic, the factor is named reachability. The result of the PCA is that variables related to the innovativeness category were not statistically reliable and were

thus removed from further analysis. Variables from Social equity and centralization were also removed as part of the factor analysis.

Summary statistics of the factors and the histograms can be found under *appendix 2.5,* distributions of factors can be seen in *appendix 2.6-2.8.* From the results, we can interpret that mean is fairly low in all three factors, the highest mean being in Centralization, then Social equity, and the smallest mean is in Reachability factor. Reachability has a long tail skewed to the right. The final result is that innovativeness factor was removed in its entirety. Social equity was reduced by two variables, one of those was participation rate, which moved into the centralization factor. Centralization gained one variable into the factor, but lost two, leaving it with three variables total. Reachability kept the two variables it originally had. The new competitiveness index score is a mean of the centralization, social equity and reachability scores having a range between 0 and 1, instead of the sum of categorical values used in the old competitiveness index. This value is multiplied by 100, so the range of the new competitiveness index is between 0 and 100.

The results from factor analysis are taken into cross correlation analysis, to see how the new competitiveness analysis built from the factors reacts with GDP and salary variables. The original competitiveness index is also used in this analysis, as the idea is to see which index has a better explanatory power over the other variables in the analysis.

After finishing the PCA, a correlation matrix was formed with the variable's competitiveness index, social equity, innovativeness, centralization, reachability, salary, GDP, new social equity, new centralization, new reachability, new competitiveness index, and SML. The descriptive statistics of the variables used in cross-correlation analysis can be seen in *appendix 3.1*. The results for the correlation matrix can be seen in *table 10.*

	1	2	3	4	5	6	7	8	9	10	11	12
1. Social equity (ORIG)	1											
2. Innovativeness (ORIG)	.36	1										
3. Centralization (ORIG)	.60	.46	1									
4. Reachability (ORIG)	.24	.18	.52	1								
5. Competitiveness index	.57	.49	.80	.88	1							
6. Social equity (NEW)	.95	.34	.44	.08	.41	1						
7. Centralization (NEW)	.36	.23	.80	.43	.61	.15	1					
8. Reachability (NEW)	.23	.17	.51	.97	.86	.08	.43	1				
9. Competitiveness (NEW)	.57	.30	.73	.89	.95	.41	.63	.91	1			
10. SML	.26	.04	.19	.28	.29	.16	.23	.27	.32	1		
11. Salary	.45	.16	.63	.51	.62	.19	.66	.50	.61	.25	1	
12. GDP	.41	.15	.56	.45	.56	.25	.48	.45	.55	.18	.64	1

Table 10. Correlation matrix for cross correlation analysis variables, years 2010-2018.

From *table 10* we can see that competitiveness index is nearly identical with factor competitiveness index because the latter is a stripped-down version of the original. The correlation of GDP and Salary with both competitiveness indexes are roughly the same, suggesting that the new competitiveness index does have roughly the same explanatory power as the original did. Of the three categories withing the new index, social equity has a marginal positive correlation with salary and GDP, centralization has a high positive correlation with salary, and a moderate correlation with GDP, reachability has moderate positive correlation with both the salary and GDP. SML seems to have a low positive correlation with all the other variables. There seems to be a spike in the 2010 results for SML, as noted in descriptive analysis, *figure 5&6*. Running the correlation matrix again without 2010 numbers increases the correlation for SML with every variable. SML has a correlation of over .30 with both competitiveness indexes and salary. These results suggest that the newly created competitiveness index can be used in further analysis, as the dimensionality reduction has not caused reduction in accuracy.

5.3. Panel regression analysis

The data that is used in this study spans across multiple years, thus it is worthwhile to see if there are relationships that occur over time. For this, cross-correlation analysis is used. By adding time lags to our data, we can see what kind of a relationship SML

score, competitiveness index score, and the categories within the index have on GDP and salary, and if there is a prolonged effect.

The analysis starts with the creation of lagged variables for the SML score, Social equity, Centralization, Reachability, and the Competitiveness index score. Since GDP and salary are the dependent variables, they are not used to create lagged variables. These lagged variables were created from one to eight lags for each variable, as there are 9 years covered in the data.

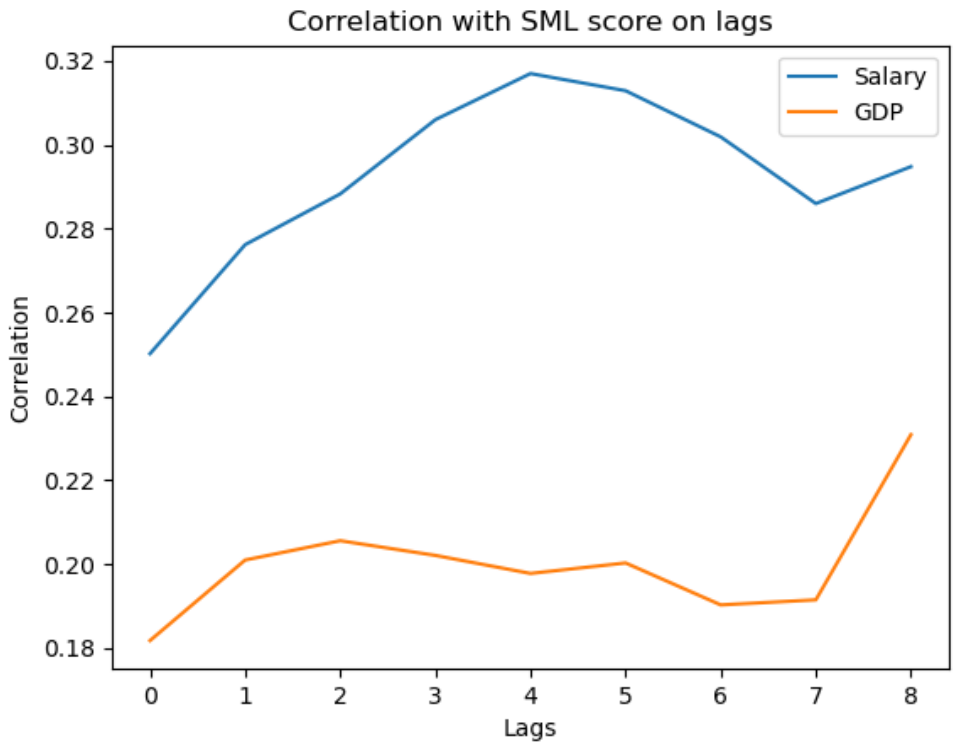


Figure 7. Correlation of salary and GDP with SML score and its lags. Statistically significant at the 1% level (p=.000).

SML score was the first variable from which the lagged variables were created. As seen in *figure 7,* the correlation for GDP does peak at the eighth lag, but the correlation is low for all lags, as it stays under .30. For the salary variable, the correlation seems to peak at the fourth lag, suggesting that the relationship is at its strongest with a four-year delay between the SML score and salary. The correlations with the competitiveness index categories can be seen in *figure* 8. Again, there seem to be clear spikes on the eighth lag. The reason could be the spike in the SML score, which

can be seen in *figure 5* and *figure 6* as there SML score in year 2010 is higher than in the later years. SML score has a negative relationship with centralization and social equity, but a positive correlation with reachability. Overall, lags do not seem to have a noticeable difference on the competitiveness index categories. Same can be seen in the appendix 3.2, the correlation table between SML score lags and competitiveness index lags does not show significant changes in correlation.

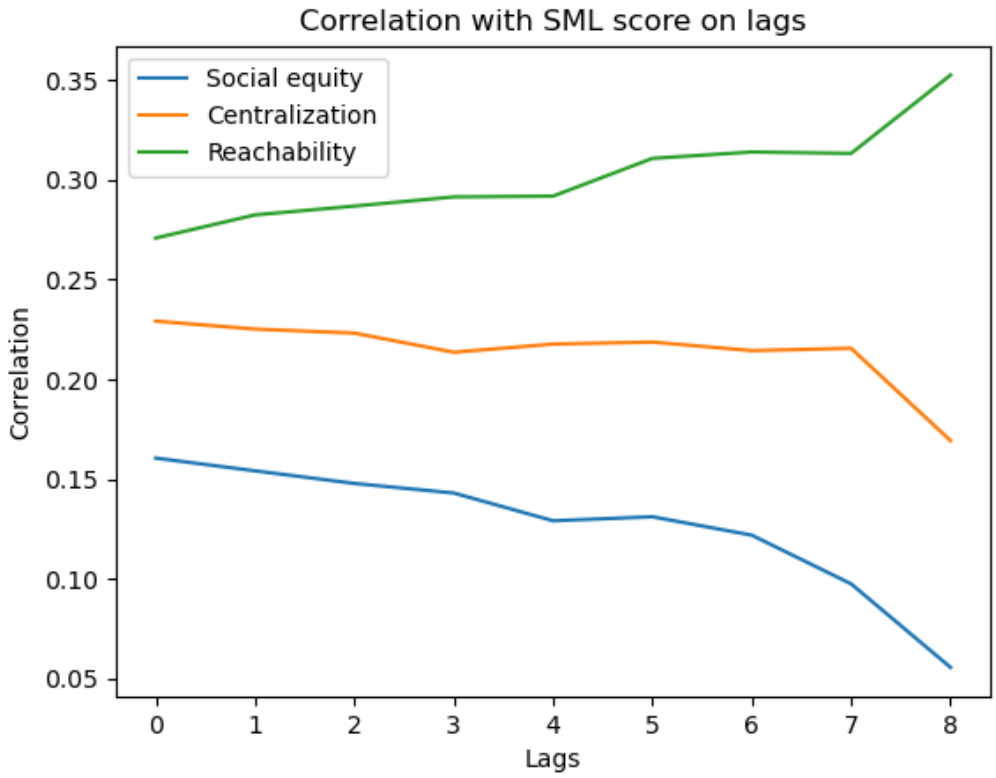


Figure 8. Correlation of social equity, centralization, and reachability with SML score and its lags. Statistically significant at the 1% level (p=.000).

Competitiveness index lag correlations with salary and GDP are visualized in *figure 9*. The figure shows us moderate decline for both the GDP and salary with each lag. Cross-correlations of competitiveness index categories with the salary and GDP can also be seen in *appendix 3.3-3.5*. The categorical figures explain the dip on the eighth lag of the competitiveness index, as there is a similar drop off for the reachability variable. Social equity also shows a dip on the eighth lag for GDP.

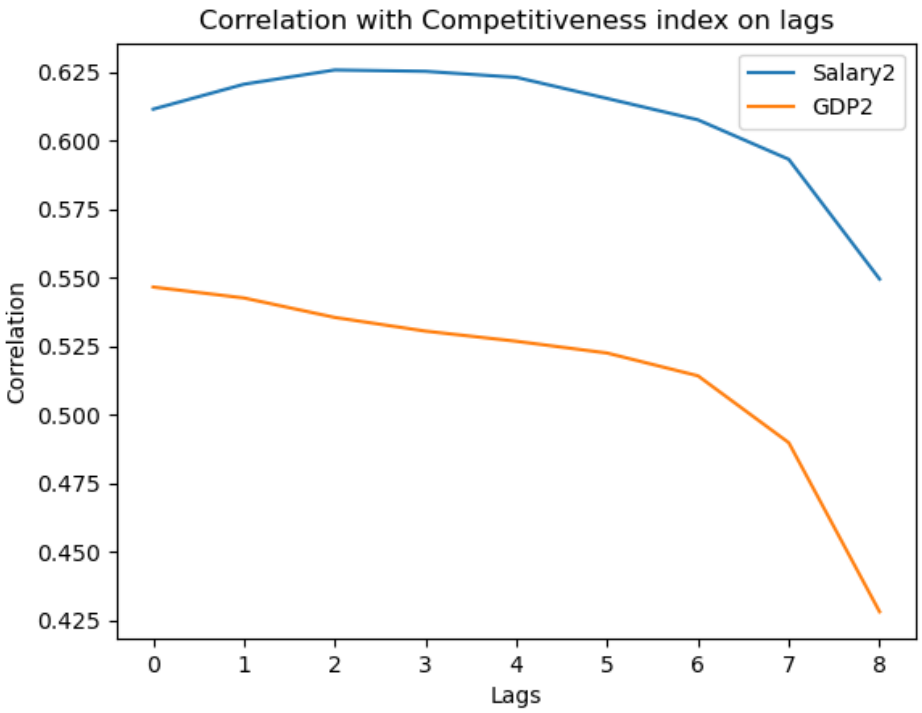


Figure 9. Correlation of salary and GDP with competitiveness index score lags. Statistically significant at the 1% level (p=.000).

Overall, the cross-correlation analysis shows that the use of lagged variables for SML score may be warranted for salary, as there is a spike at the fourth lag. For competitiveness index, there is a spike in the eighth lag, but it appears to be more of an outlier than an underlying trend in the data. The cross-correlation analysis did not show any meaningful improvement between the relationship of SML score and competitiveness index score using lagged variables.

With the competitiveness index built in the factor analysis, and the correlation information that was revealed in the cross-correlation analysis, a regression model can be created to estimate the impact that SML has on competitiveness index, GDP, and salary. The aim is to find out whether SML has a direct impact on GDP and salary or does it have an indirect impact through the competitiveness index. This analysis will be conducted as a panel data analysis, as we are dealing with a dataset that has both the cross-sectional elements with the SML and competitiveness index variables for each municipality, as well as time-series elements, since the data has a span of multiple years.

In *table 11*, we can see the descriptive statistics for all the variables that are used in the panel data analysis. There are four different analyses that will be run using these four variables. The analysis consists of multiple parts, first of which will have SML as the explanatory variable for competitiveness index, as the idea is to see how SML affect the competitiveness of a municipality. The second analysis will have SML and competitiveness index as the explanatory variables for GDP and salary. The analysis will show the effect that SML and competitiveness index have on the GDP and salary variables. The last part will measure the effect of SML and competitiveness index on GDP and salary separately. No other control variables were included in the analysis, as they are already contained within the competitiveness index.

Variable	Mean	Std. Dev.	Min	Max
Competitiveness index	35.88	10.59	21.36	75.87
SML	3.45	0.01	3.40	3.53
GDP per capita	32191.87	8028.13	16957.30	77681.00
Yearly salary per capita	25490.60	4061.57	17666.00	67790.00

Table 11. Descriptive statistics of Panel Data analysis variable.

5.3.1. SML on competitiveness index

The first analysis was conducted with SML scores and competitiveness index variables. Fifth lag of the variable seemed to give the best results, as can be seen in *appendix 3.6.* T*able 12* shows us the results of the panel data regression analysis. The table contains results for multiple lags of SML to see if they differ from the cross-correlation analysis results. For the regression, a OLS model, fixed effects model and the random effect model was built. There are a total of 2790 observations from 310 municipalities, and four different lags were tested. The chosen lag can be seen in bolded font.

Method	SML	SML -2	SML -5	SML -6
Fixed effects coefficients	16.24* (9.06)	21.66*** (3.06)	**18.86***** **(3.36)**	7.13*** (3.64)
Observations	**2790**			
Municipalities	**310**			
R-squared	0.101	0.105	**0.114**	0.113
Probability > F	0.000	0.000	**0.000**	0.000
Breusch-Pagan test p-value	**0.000**			
Hausman test p-value	**0.000**			

*Table 12. SML effect on competitiveness. Panel regression analysis with the Fixed effects model. (*** p < .01, ** p < .05, * p < .1), standard errors in brackets.*

The Breusch-Pagan test resulted in a value of .00, null hypothesis is therefore rejected. We can conclude that there is significant difference between the municipalities. The Hausman test for the fixed effects and random effects resulted in a p-value of .00. Since the resulting threshold is below .05, the null hypothesis of the Hausman test is rejected, meaning that the difference in the coefficients is systematic. Rejecting the null hypothesis of the Hausman test means, that random effects and fixed effects are not identical, thus the random effects model is not appropriate. After running the fixed effects model the F-test was conducted, which tests if there is a same constant for all units. The p-value of the F-test in the fixed effects model is .000, which is below the threshold of .05, so zero hypothesis is rejected, meaning that constants are different for all units. Since the zero hypothesis is rejected, the fixed effects model is the appropriate model for this analysis. The chosen fixed effects model explains 11.4 percent of the competitiveness indexes variance, it has a small effect on the total value of the index. The coefficients are positive for all lags, a point increase in the chosen SML lag results in an 18.86 increase in competitiveness index. This result is surprising as it contradicts the notion that automation increases competitiveness. The effect of SML on competitiveness index could be increased by adding variables to the competitiveness index that measure innovativeness aspect of competitiveness. Innovativeness was a part of the original competitiveness index, but the variables belonging to the category were all excluded during factor analysis.

5.3.2. SML and competitiveness index on GDP and salary

In the part of the analysis, the aim is to find out how competitiveness index and SML scores affect the GDP and salary. This will be carried out with two different analyses, where one has salary as the dependent variable, and the other having GDP as the dependent variable. Lagged variables will be utilized based on the cross-correlation analysis done in the previous part. The results can be seen in *table 13,* the chosen lags are bolded.

Method	Competitiveness index SML	Competitiveness index SML -2	Competitiveness index -1 SML -2
Competitiveness index coefficients (FE)	-95.17*** (25.64)	**-639.22***** **(93.75)**	-765.27*** (93.44)
SML coefficients (FE)	-280150.90*** (11576.05)	**-157661.6***** **(12521.32)**	-161147*** (12354.25)
Observations	**2790**		
Municipalities	**310**		
R-squared	0.121	**0.303**	0.303
Probability > F	0.000	**0.000**	0.000
Breusch-Pagan test p-value	**0.000**		
Hausman test p-value	**0.000**		

*Table 13. SML and competitiveness effect on GDP. Panel regression analysis with the fixed effects model. (*** p < .01, ** p < .05, * p < .1), standard errors in brackets.*

Selection of the model between OLS, fixed effects, and random effects was chosen by conducting a series of tests, first of which is the Bresuch-Pagan test. The resulting p-value was .00. Since that is below the threshold of .05, it means the null hypothesis for the Breusch-pagan test is rejected and there is difference between municipalities. The next test conducted was the Hausman test. Similarly, the Hausman test returned a p-value .00, which means that the null hypothesis was rejected as it was below the threshold. Random effects coefficients are not identical

to the fixed effects coefficients and thus the random effects model is ruled out as inappropriate for further analysis. To choose between the fixed effects model and the OLS model, the F-test was conducted on the fixed effects model. The result was .00, meaning that the null hypothesis is rejected in this test. Since constants are the same for all units, the OLS model is ruled out as inappropriate, and the fixed effects model was chosen. The fixed effects model results for GDP can be seen in *table 13.* The chosen model is bolded, containing SML score with a lag of two, and competitiveness index with a lag of zero. R-squared for the lag one of competitiveness index is the same, thus the simpler model was chosen. The table contains three columns of variables with a different set of lags. SML and competitiveness index have a negative impact on GDP in all three sets. The chosen set explains 30.3 percent of the variance in GDP. A point increase in competitiveness decreases GDP by 639.22, and a point increase in SML decreases salary by 157661.6. Because of the tight range of the SML variable, we can divide salary by 100 to get the impact of a 0.01 increase in SML, which is 1576.62.

The final analysis for the salary variable is done in a similar fashion to the previous analysis on GDP. SML score and competitiveness index scores are used in the analysis to see their effect on the salaries. Similarly, for the selection of the model, a Breusch-pagan test was conducted. The test resulted in a value of .00, meaning that the null hypothesis was rejected as the value is below the threshold of .05. Thus, it can be stated that there are differences between municipalities. The next test was the Hausman test, which also got a value of .00, meaning that the null hypothesis is rejected, and the random effects are not identical to the fixed effects. Random effects model was ruled out as not appropriate for the analysis. Additionally, an F-test was done on the fixed effects model, which resulted in a .00 value, rejecting the null hypothesis. The chosen model is the fixed effects model, as it was in the previous analyses.

Three different sets of lags were used create the initial models, the chosen lag for SML was four, and a lag of zero was chosen for the competitiveness index score. All sets had negative impact on salary for both SML and the competitiveness index. The chosen set of lags explained 43 percent of the variance in salary, a point increase in

competitiveness decreased salaries by 238.88, and a 0.01 increase in SML decreased salary by 443.97.

Method	Competitiveness index SML	**Competitiveness index** **SML -4**	Competitiveness index -2 SML -4
Competitiveness index coefficients (FE)	-31.43** (12.97)	**-238.88***** **(45.95)**	-245.92*** (42.60)
SML coefficients (FE)	-189219.90*** (5855.66)	**-44396.92***** **(5189.45)**	-45402.90*** (5139.24)
Observations	**2790**		
Municipalities	**310**		
R-squared	0.123	**0.430**	0.424
Probability > F	0.000	**0.000**	0.000
Breusch-Pagan test p-value	**0.000**		
Hausman test p-value	**0.000**		

*Table 14. SML and Competitiveness index effect on salary. Panel regression analysis with the fixed effects model. (*** p < .01, ** p < .05, * p < .1), standard errors in brackets.*

5.3.3. Competitiveness index on GDP and salary

To see how competitiveness index affects GDP and salary without SML, separate regression models were created for it. *Table 15* shows the results for competitiveness index against GDP Like in the previous analyses, the null hypotheses of Breusch-Pagan test, Hausman test, and the F-test were all rejected. Thus, the fixed effects model was chosen.

Method	Competitiveness index	Competitiveness index – lag 1	Competitiveness index – lag 2
Fixed effects coefficient (Standard error)	**-117.49***** **(28.48)**	-103.58*** (26.05)	-39.12*** (25.29)
Observations	**2790**		
Municipalities	**310**		
R-squared	**0.30**	0.29	0.29
Probability > F	**0.000**	0.000	0.000
Breusch-Pagan test p-value	**0.000**		
Hausman test p-value	**0.000**		

*Table 15. Competitiveness index effect on GDP. Panel regression analysis with the fixed effects model. (*** p < .01, ** p < .05, * p < .1), standard errors in brackets.*

The competitiveness index was then tested against salary variable, of which the results are seen in *table 16.* Four lags of competitiveness index are seen in the table, of which the first lag was chosen. Lags two and three produced more accurate results, but both had a t-test p-value of over 0.10. A point increase in competitiveness index decreases salary by 38.41, all lags had a negative effect on salaries. The null hypotheses of Breusch-Pagan test, Hausman test, and the F-test were rejected, and fixed effects model was chosen.

Method	Competitiveness index	Competitiveness index – lag 1	Competitiveness index – lag 2	Competitiveness index – lag 3
Fixed effects coefficient (Standard error)	-46.51*** (15.45)	**-38.41***** **(12.90)**	-18.34 (11.20)	-5.14 (9.72)
Observations	**2790**			
Municipalities	**310**			
R-squared	0.374	**0.385**	0.392	0.391
Probability > F	0.000	**0.000**	0.000	0.000
Breusch-Pagan test p-value	**0.000**			
Hausman test p-value	**0.000**			

*Table 16. Competitiveness index effect on salary. Panel regression analysis with the fixed effects model. (*** p < .01, ** p < .05, * p < .1), standard errors in brackets.*

To gain additional understanding of why competitiveness index has a negative impact on GDP and salary variables, the same regression model was built using the factors established in *table 9.* The resulting model for GDP and salary can be seen in *appendix 3.7.* We can see that centralization has a positive coefficient on GDP and salary, while reachability has a slightly negative coefficient. However, social equity has a strongly negative coefficient to GDP and salary. Fixed effect model measures the differences within panels, and the time range in the research is 2010-2018. This could explain the contradiction with existing theory on competitiveness. Social equity factor is composed of working age population, students, and engineering students. The OLS model for GDP and salary can be seen in *appendix 3.8,* which suggests that there is possibly an opposite long-term effect, as the OLS coefficients are positive.

5.3.4. SML on GDP and salary

In order to see what impact the SML score has on the GDP and salary, two additional analyses were conducted with SML as the only variable in the regression models against GDP and salary. The results for GDP can be seen in *table 17.* There are four columns for four different time lags, the chosen one being bolded. As in the previous analyses the null hypothesis were rejected for the Breusch-Pagan test, the Hausman test, and the fixed effect F-test. Thus, the fixed effect model was chosen as the appropriate model. All of the lags had a negative impact on GDP. The chosen SML lag explained 4.2 percent of GDP's total variation. All of the SML lags had a negative impact on GDP. A 0.01 increase in SML decreased GDP by 1715.09.

Method	SML	SML -1	**SML -2**	SML -3
Fixed effects coefficient	-281696.70*** (11598.34)	-202785.80*** (11881.92)	**-171509.30***** **(12505.77)**	-170801.60*** (13674.48)
Observations	**2790**			
Municipalities	**310**			
R-squared	0.033	0.040	**0.042**	0.041
Probability > F	0.000	0.000	**0.000**	0.000
Breusch-Pagan test p-value	**0.000**			
Hausman test p-value	**0.000**			

*Table 17. SML effect on GDP. Panel regression analysis with the fixed effects model. (*** p < .01, ** p < .05, * p < .1), standard errors in brackets.*

The analysis for salary was conducted in a similar manner. Four columns with different lags for SML. The results can be seen in *table 18*, the chosen model is bolded. For the selection of the model, Breusch-Pagan test, the Hausman test, and the F-test for the fixed effects model was conducted. All returned a value of .00, and the null hypotheses were rejected as before. Fixed effects model was thus selected here as well. The chosen SML lag explains 10.1 percent of the salary variance, and all lags have a negative impact on salary. SML increase of 0.01 decreases salary by 489.48.

Method	SML	SML -3	**SML -4**	SML -5
Fixed effects coefficient	-189730.50*** (5857.62)	-82857.56*** (4891.34)	**-48947.83***** **(5168.54)**	-62719.86*** (5456.85)
Observations	**2790**			
Municipalities	**310**			
R-squared	0.063	0.094	**0.101**	0.098
Probability > F	0.000	0.000	**0.000**	0.000
Breusch-Pagan test p-value	**0.000**			
Hausman test p-value	**0.000**			

*Table 18. SML effect on salary. Panel regression analysis with the fixed effects model. (*** p < .01, ** p < .05, * p < .1), standard errors in brackets.*

The end result based on the panel data analysis, is that SML has a 11.4 percent direct variance effect on competitiveness index, 4.2 percent effect on GDP, and 10.1 percent effect on salaries. As for the indirect effect, the coefficient of SML on GDP was -171509.3, and SML with competitiveness index had -157661.6. The coefficient on salary for SML was -48947.83 and -44396.92 for both SML and competitiveness index. There is not a lot of differences between the two, indicating that the indirect effect is very small.

There is also an indirect effect, as SML score together with the competitiveness index score have a 30 percent effect on GDP variance, and a 43 percent effect on salary variance.

6. CONCLUSIONS AND DISCUSSION

In this chapter of the study, the results from the previous chapter are added together to see how they answer the research questions and the objective of the research, and the contribution to existing theory of the topic is explored. There is also discussion on the reliability and validity of the research, the limitations, and ideas for further research on the topic.

6.1. Summary and main findings

The objective of this research was to find how ML affects competitiveness regionally in Finland by inspecting all 310 municipalities of Finland. To achieve this, the effect of ML was inspected against a formative index created to measure the potential competitiveness of a municipality, as well as the GDP, and salary of a municipality. To measure ML, a SML score was calculated for every municipality.

GDP was chosen as it measures the added value of production, therefore following the actual competitiveness of a region. Salary was chosen as a supportive variable for GDP to measure competitiveness, as wealth produced in a region does not always stay within the region. Example of such a case is when an international company operates in a region and moves their wealth abroad. The formative index was created by selecting variables to measure four different categories of competitiveness, which were social equity, innovativeness, centralization, and reachability. This measure does not necessarily give the current competitiveness of a municipality, but rather a potential competitiveness that the municipality could reach. The final number of variables for the competitiveness index was 17, the variables used for the index were from the Huovari et al., (2001) index, which contained 16 different variables for the four categories. An additional variable was created to measure knowledge work, and also added to the index (Huovari et al., 2001; Arntz et al., 2016). Data was collected of the 17 variables for the time range of 2000-2018 and of 310 Finnish municipalities, of which the years 2010-2018 were selected for analysis. GDP and salary values are mean per capita values of 310 Finnish municipalities, collected for years 2010-2018.

Two competitiveness index variables were dropped because of lacking data for all the years, 15 variables remain.

The measure for ML was created using the Brynjolfsson et al., (2018) method of calculating a likelihood score of getting automated for each occupational task. After calculating the SML score for all of the tasks, the score was then calculated for every occupation by aggregating the score of the tasks that belong to the occupation and taking the mean. The SML score for a municipality is calculated by getting the employment data for the Finnish municipalities and applying the SML scores for the occupations found in the data. The municipality SML score is calculated by taking the weighted mean of the occupations in the municipality based on how many people are working the occupation. The SML scores were calculated for all 310 municipalities for the years 2010-2018.

The initial analysis on the dataset shows that there is a positive moderate relationship between the calculated competitiveness index and the SML scores in a municipality. Correlations between the categories showed that social equity, centralization, and innovativeness categories had high positive relationships, whereas reachability had a moderate to low positive relationship with the rest of the competitiveness index categories. The correlation of competitiveness index was moderately positive with GDP and salary. SML score had a positive relationship with both, but the correlation was weak. SML score also had a positive moderate correlation with competitiveness index score.

Dimensionality reduction was performed on the competitiveness index in order to remove variables with marginal informational value to the index. Factor analysis model PCA was performed on the 15 variables in the index, and a total of 7 variables were removed for low correlation with other variables, near identical correlations, having loadings with multiple factors, and having high uniqueness values in the PCA model. There were four categories featured in the index before the dimensionality reduction. All of the innovativeness factor variables were removed in the analysis, leaving the index with three factors, which were the social equity, centralization, and reachability. The reduced competitiveness index scored very similar correlation values with the

SML score, GDP, and salary variables. The new competitiveness had a range of 0 to 1 which was then multiplied by 100.

The reduced competitiveness index was then used to run a cross-correlation analysis, where time lags are applied to variables to see if there are delayed effects on other variables. SML score was plotted against the GDP and salary variables to see how it correlates with the variables. The fourth time lag seemed to give the most accurate results for salary, suggesting that the SML score likely impacts salaries with a four-year delay. For GDP, there was no single time lag, but lags one to five seemed to have very similar accuracy scores for SML. There was a spike at lag eight, but it is ignored as there were underlying differences in the data for the employment data in year 2010, probably causing an outlier value in SML. The cross-correlation plot of SML on competitiveness index reveals, that the accuracy stays roughly the same with every lag, the only difference between the lags is that the categories seem to be drifting apart from each other. Competitiveness index plotted on salary and GDP showed, that the lag on GDP should be zero, as the correlation decreases with every additional lag. For salary, there is a small spike at time lag 2, which shows the largest correlation. These lags were then taken into consideration for the panel data analysis in the regression models.

The panel data analysis was conducted in four parts. In the first part, panel regression models were built to see how big of an impact the SML score has on the competitiveness index. In each analysis, a OLS model, fixed effects model and a random effects model was built. In the first part of the analysis, the effect of SML on competitiveness index was expected. The chosen model was the fixed effects model, and SML had a time lag of 5. The regression model explained 11.4 percent of competitiveness index variance. In the second part of the panel data analysis, panel regression models were built to see, how both the SML score, and the competitiveness index score affect GDP, and another set of models constructed to see the effect on salaries. For the GDP, different time lags were tested for both variables, and the second lag was chosen for the SML score, and no lags for competitiveness index as it was the simplest model and gave the best accuracy. All of the tested lags impacted GDP negatively. The model explained 30.3 percent of the variance of GDP. For salary, the best performing model was with a lag of 4 for SML, and a lag of zero for

competitiveness index. Like in the case of GDP, increases in competitiveness index had a negative salary on all tested lags. The R-squared of the model was 43 percent. Validity tests for both models suggested the fixed effects model as the appropriate model for the GDP and salary analysis. To get a better understanding of the negative coefficients, the factors of competitiveness index were separately tested against GDP and salary. The result was that the high negative coefficients for social equity factor pushed the competitiveness index to have a negative effect on GDP and salary. Social equity factor was composed of variables students, engineering students, and working age population. For the last part, the same analysis was performed as in the previous part, this time with just the SML score. In the first analysis, different lags were tested to see the impact SML score has on GDP. The R-squared of 4 percent indicated that SML score has a small impact of 4 percent on the variance of GDP. The validity tests suggested the fixed effects model for the analysis. For the salary, SML score with a lag of four affected 10 percent of the salary variance. The validity and reliability tests again suggested the use of fixed effects model. The coefficient of SML was negative in the model for GDP and salary variables, meaning that SML has a negative impact on GDP and salary.

The objective of this research was to find out how ML affects competitiveness of a region. There was a total of three hypotheses, which were formed based on the research questions. This research aimed to understand the impact of ML by inspecting its impact on a formative index for competitiveness, and by inspecting its impact on GDP and salary data, which were chosen as the economic measures of competitiveness. There was a total of three research questions and each one is addressed individually, and the conclusions explained.

The first sub-question: **How does Machine Learning affect the competitiveness of municipalities in Finland?**

In the research, this question was set up to be answered by the formative competitiveness index created to measure the potential of competitiveness in a municipality. The effect was examined in the panel data analysis part of the research, where the SML score impact on competitiveness index score was studied. The hypothesis for this question was that SML score would have a negative effect on

competitiveness index score, which is opposite to what is seen in the panel data analysis results. The coefficient of SML in the chosen model, the fixed effects model, is a positive value. The null hypothesis of the first hypothesis of the research cannot be rejected based on these findings, as there are no suggestions of SML affecting competitiveness negatively.

The second sub-question was: **How does competitiveness affect the economic development Finnish municipalities?**

The panel regression model showed that the competitiveness index seemed to have a 30 percent impact on the GDP variance, and a 43 percent impact on salary variance. Negative coefficient indicated that competitiveness has a negative impact on GDP and salary, which is contrary to the existing theory. The major reason was the social equity factor, which was composed of working age population, students, and engineering students. Students are generally considered a low-income, low-consumption group. Since we are looking at GDP per capita, and mean salary and not total values, an influx of low-income and low-consumption people would likely bring down both of these measures. Fixed effects model does look within panels, and as the time range was from 2010-2018, it does seem plausible that social equity would react negatively with GDP per capita, and mean yearly salary. Competitiveness would thus have a negative short-term economic effect, and possible a positive long-term effect as was seen in the OLS model.

The third sub-question was: **How does Machine Learning affect the economic development of Finnish municipalities?**

The regression model showed the SML score having a clear negative effect on the GDP of a municipality. The explanatory power of the model was only 4 percent, meaning that SML score has a small negative impact on GDP. The second hypothesis of the research can be accepted, as the model seems to pick up SML score having a negative effect on GDP.

The panel regression model showed that the SML score has a clear negative effect on the salary of a municipality. The model explained SML score having a 10 percent

impact on the variance of salary. The third hypothesis of the research can be accepted, as SML had a significant negative impact on the salary in the chosen fixed effects model.

Higher SML score means that the employment in a municipality is focused on occupations in which more work can be automated by ML in the future. Automation has a positive impact on GDP through the growth of productivity (Huovari et al., 2001; Lepenies, 2016), which would explain why increase in highly automatable work affects GDP negatively. This would explain the negative relationship between SML score and salaries, if routine work tends to have lower salary compared to non-routine work.

Main question: **What are the economic and competitiveness impacts of Machine Learning on municipalities of Finland?**

Having discussed the sub-questions, the main question can be addressed. The concluding remarks are that SML score appears to affect competitiveness positively, but it does appear to have direct negative effects on the economic performance measures, which are the GDP and salary. In other terms, automation by ML seems to decrease competitiveness short-term, but ML automation does seem to have a positive impact on the economy.

6.2. Connection to existing literature

This research contributes to the existing literature on ML. The research does support previous research done on social impacts of ML, as this research does support that reducing the amount of work that is highly automatable by ML increases GDP. It is probable that the cause is an increase in productivity. This means that there is an incentive to invest into ML, as it can provide an economic boost. (Brynjolfsson et al., 2018; Pajarinen et al., 2015; Frey and Osborne, 2013). The economic impact of AI and ML was expected to be 14 percent of the global economy, which would be in line with the results of this study, as the impact of AI was not studied, and there might be differences across the globe (McKinsey Global Institute, 2018). The ML score used in

the research was directly responsible for 4 percent of the changes in GDP, as well as having an indirect impact through the competitiveness index. What is perhaps new to the research concerning the implications of AI and ML is that the reduction of highly automatable jobs seems to impact competitiveness negatively short-term.

While this research is outside of the realm on how ML affects employment, it does bear mention that the results of this research do not seem to show support for large wave of automatization, as mentioned in Pajarinen et al. (2015), or Frey and Osborne (2013). Rather, the results seem to line up with Arntz et al. (2016), where ML seemed to have small positive impact on automation. Though it is worth noting, that the SML score used in this thesis focuses solely on ML, whereas studies like Pajarinen et al. (2015), and Frey and Osborne (2013) also focused on other technologies, like mobile robotics.

This thesis featured the formative competitiveness index introduced by Huovari et al. (2001) to measure the level of potential competitiveness in the Finnish regions. The result is that there seem to be issues with the reliability of this index, as many of the variables featured in the original index were excluded in the factor analysis conducted in this thesis using the PCA method.

6.3. Limitations and further research

In order to think about the limitation of this study, there are two underlying concepts which the limitation cover. First being the quality of the data, and the second one covering the objective of the research.

While most of the variables used in the data were collected from all 310 municipalities of Finland, there were a total of five variables which had to be approximated, because the data covered only 19 or 70 regions. Availability of data was a problem that may have hindered the results of the research, like the absence of railroad traffic data. Additionally, some variables had to be removed from analysis for not covering the needed timeframe of 2010-2018. The quality of the data also suffered from the

inaccuracies that arose mostly from Åland islands or other areas with small population figures.

For the limitations of the objective, the scope of the research was limited to Finland, thus the effects of neighboring countries or nearby cities on competitiveness of municipalities was not measured. Innovativeness was found to be a critical component of measuring competitiveness, but all of the variables measuring innovativeness in the competitiveness index built for this thesis were excluded in the factor analysis part of the research. The final competitiveness index had a total of eight variables, of which two were associated with students, possibly skewing the index score.

The objective of the research was how ML will affect the competitiveness of the municipalities in Finland, and whether it has an impact on the salary or the GDP of the region. ML measure created for this study measures the amount of work that could be automated using ML tools, there is no knowledge on how much work is actually replaced with these tools. An assumption was made for the SML score that the occupations are similar between the United States, and Finland. This method was criticized in Arntz et al., (2016), and may have hindered the results.

The results of this thesis can be used to further research how the competitiveness and economic wellbeing of Finnish regions is affected by different technologies or policies. There could be value in seeing if the effect on competitiveness does change dramatically if other technologies are added onto ML, as there are also other innovations contributing to automation. Results of this research can also act as a comparison point for conducting a similar type of research for regions of other countries.

REFERENCES

Acemoglu, D., & Restrepo, P. (2017). Robots and Jobs: Evidence from US Labor Markets. Robots and Jobs: Evidence from US Labor Markets, 35–36. https://doi.org/10.3386/w23285

Allison, P. D. (2009). Fixed effects regression models. SAGE publications.

Alpaydin, E. (2014). Introduction to Machine Learning (third edition). The MIT Press.

Annoni, P., & Dijkstra, L. (2019). EU regional competitiveness index RCI 2019. Luxembourg: European Commission, Joint Research Centre, [Accessed 28.12.2020] Accessible at: https://ec.europa.eu/regional_policy/sources/docgener/work/2019_03_rci2019.pdf

Arntz, M., T. Gregory, and U. Zierahn (2016), "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis", OECD Social, Employment and Migration Working Papers, No. 189, OECD Publishing, Paris, https://doi.org/10.1787/5jlz9h56dvq7-en.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. Journal of Economic Perspectives, 29(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Begg, I. (1999). Cities and Competitiveness. Urban Studies, 36(5–6), 795–809. 10.1080/0042098993222

Bourke, P. (1996). Cross correlation. Cross Correlation”, Auto Correlation—2D Pattern Identification.

Brooks, C. (2014). Introductory Econometrics for Finance (3rd ed.). Cambridge University Press.

Brynjolfsson, E., & McAfee, A. (2016). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies (Reprint ed.). W. W. Norton & Company.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock. 2018. "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" AEA Papers and Proceedings, 108: 43-47.

Case, K. E., Fair, R. C., & Oster, S. E. (2013). Principles of Microeconomics (11th ed.). Pearson.

Cuadrado-Roura, J. R., & Rubalcaba-Bermejo, L. (1998). Specialization and Competition amongst European Cities: A New Approach through Fair and Exhibition Activities. Regional Studies, 32(2), 133–147. https://doi.org/10.1080/00343409850123026.

Davenport, T. H., & Ronanki, R. (2018, January). Artificial Intelligence for the Real World (Vol. 96 Issue 1). Harvard Business Review. https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

Deane, P. (2010). The First Industrial Revolution (2nd ed.). Cambridge University Press.

Derrick, Timothy R. and Thomas, Joshua M., "Time Series Analysis: The Cross-Correlation Function" (2004). Kinesiology Publications. 46.

Evangelista, R., Guerrieri, P., & Meliciani, V. (2014). The economic impact of digital technologies in Europe. Economics of Innovation and New Technology, 23(8), 802–824. https://doi.org/10.1080/10438599.2014.918438

Feinstein, C. H. (1998). Pessimism Perpetuated: Real Wages and the Standard of Living in Britain during and after the Industrial Revolution. The Journal of Economic History, 58(3), 625–658. https://doi.org/10.1017/s0022050700021100

Frey, Carl & Osborne, Michael. (2013). The Future of Employment: How Susceptible Are Jobs to Computerisation? Oxford Martin. 114. 10.1016/j.techfore.2016.08.019.

Gardiner, Ben & Martin, Ron & Peter, Tyler. (2004). Competitiveness, Productivity and Economic Growth across the European Regions. Regional Competitiveness. 10.4324/9780203607046.

Groover, M. P. (2019). Fundamentals of Modern Manufacturing: Materials, Processes, and Systems. John Wiley & Sons Inc.

Helpman, E. (1998). General Purpose Technologies and Economic Growth. Amsterdam University Press.

Hill, C. R. (2011). Principles of Econometrics (4th ed.). Wiley.

Huovari, Janne & Kangasharju, Aki & Alanen, Aku. (2002). Constructing an Index for Regional Competitiveness. 10.1007/978-3-540-24823-1_7.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jun, W. C., Oh, G., & Kim, S. (2006). Understanding volatility correlation behavior with a magnitude cross-correlation function. Physical Review E, 73(6). https://doi.org/10.1103/physreve.73.066128

Jr., R. L. E. (2002). Lectures on Economic Growth (1st ed.). Harvard University Press.

Kitson, M., Martin, R., & Tyler, P. (2004). Regional Competitiveness: An Elusive yet Key Concept?. Regional Studies, 38, 991-999.

Kline, P. (1994). An Easy Guide to Factor Analysis. Routledge.

Kotavaara, O., Antikainen, H., & Rusanen, J. (2011). Urbanization and Transportation in Finland, 1880–1970. The Journal of Interdisciplinary History, 42(1), 89–109. https://doi.org/10.1162/jinh_a_00206

Kraut, R. E., Dumais, S. T., & Koch, S. (1989). Computerization, productivity, and quality of work-life. Communications of the ACM, 32(2), 220–238. https://doi.org/10.1145/63342.63347

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., Mädche, A., Urbach, N., & Ahlemann, F. (2017). Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community. Business & Information Systems Engineering, 59(4), 301–308. https://doi.org/10.1007/s12599-017-0484-2

Lepenies, P. (2016). The Power of a Single Number: A Political History of GDP. Columbia University Press.

Lerch, C., & Gotsch, M. (2015). Digitalized Product-Service Systems in Manufacturing Firms: A Case Study Analysis. Research-Technology Management, 58(5), 45–52. https://doi.org/10.5437/08956308x5805357

Louridas, P., & Ebert, C. (2016). Machine Learning. IEEE Software, 33(5), 110–115. https://doi.org/10.1109/ms.2016.114

Luger, G. (2008). Artificial Intelligence: Structures and Strategies for Complex Problem Solving (6th ed.). Pearson.

Marsland, S. (2014). Machine Learning: An Algorithmic Perspective (2nd ed.). Chapman and Hall/CRC.

McKinsey Global Institute. (2018, September). Notes from the AI frontier: Modeling the impact of AI on the world economy.

https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx.

McCorduck, P. (2004). Machines Who Think (2nd ed.). Routledge.

Mokyr, J., & Strotz, R. H. (1998). The second industrial revolution, 1870-1914. Storia dell'economia Mondiale, 21945.

Pajarinen, Mika, Rouvinen, Petri & Ekeland, Anders (22.4.2015).“Computerization Threatens One-Third of Finnish and Norwegian Employment”. ETLA Brief No 34. http://pub.etla.fi/ETLA-Muistio-Brief-34.pdf

Poole, D. L., & Mackworth, A. K. (2017). Artificial Intelligence: Foundations of Computational Agents (2nd ed.). Cambridge University Press.

Porter, M. E. (2000). Location, Competition, and Economic Development: Local Clusters in a Global Economy. Economic Development Quarterly, 14(1), 15–34. https://doi.org/10.1177/089124240001400105

PricewaterhouseCoopers. (2017). Sizing the prize What's the real value of AI for your business and how can you capitalise? https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Saario, K. (2016) Alueiden kilpailukyky 2016: Yritysten näkökulma.

Shepherd, J. (2004). Why the Digital Era?. Social and Economic Transformation in the Digital Era, 1.

Tervo, H., Helminen, V., Rehunen, A., & Tohmo, T. (2018). Onko urbanisaatio maaseudun turma? Kaupunkien väestönkasvun vaikutukset erityyppisen maaseudun väestökehitykseen Suomessa ajanjaksolla 1990–2015.

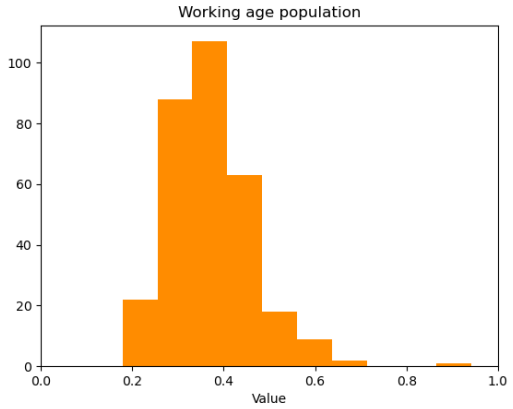
United Nations Publications. (2019). World Urbanization Prospects: The 2018 Revision (Revised ed.). United Nations.

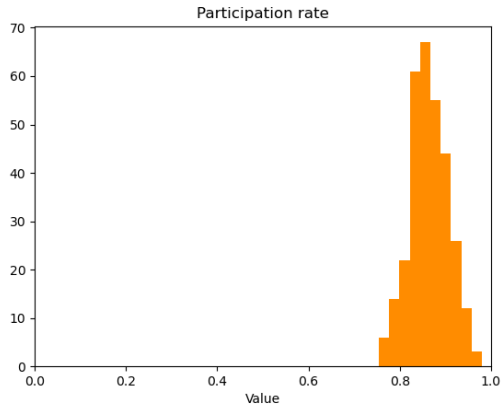
Warr, B., & Ayres, R. (2006). Working Paper Economic Growth , Technological Progress and Energy Use in the US over the Last Century : Identifying Common Trends and Structural Change in Macroeconomic Time Series.

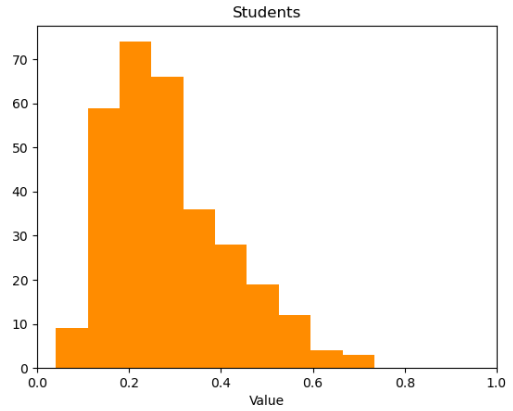
Williamson, J. G. (1988). Chapter 11 Migration and urbanization. Handbook of Development Economics, 425–465. https://doi.org/10.1016/s1573-4471(88)01014-9

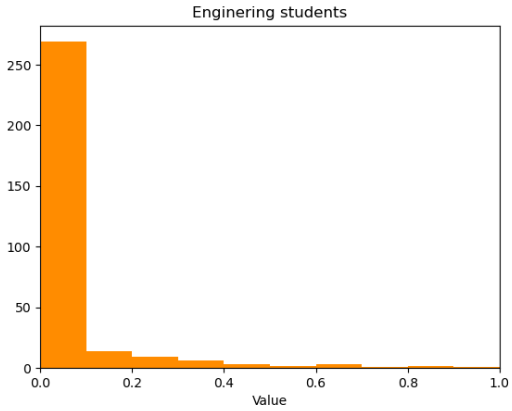
APPENDICES

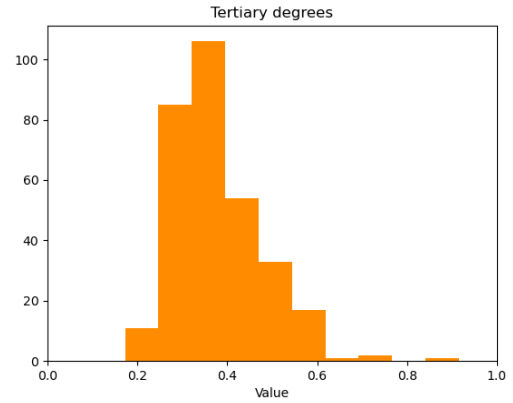
Appendix 1. Data & Methodology

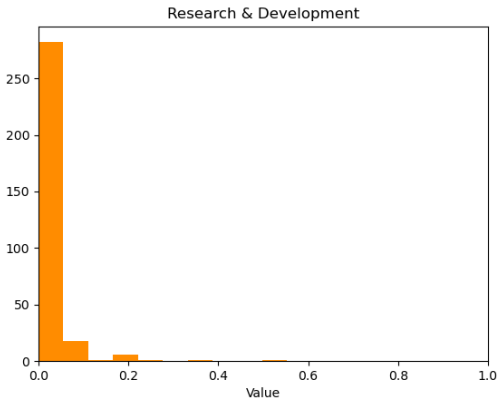


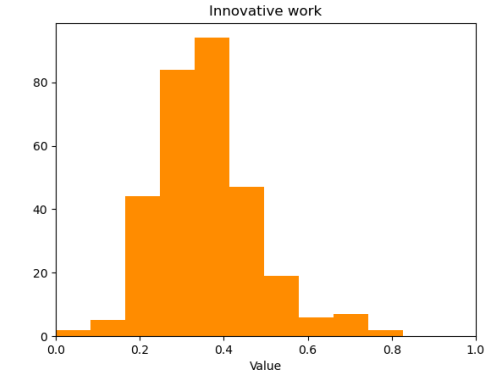




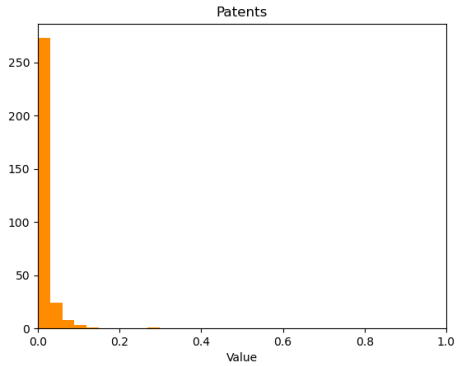


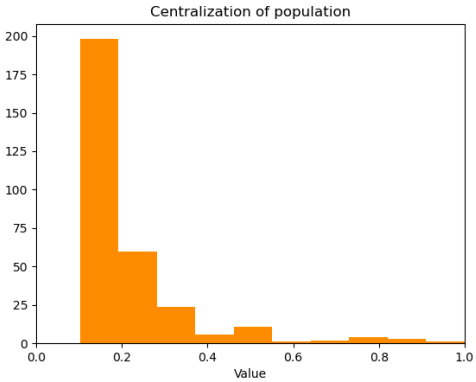


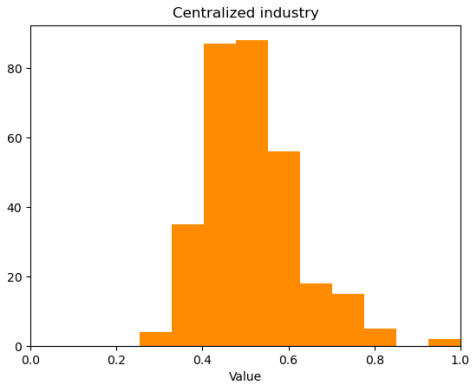


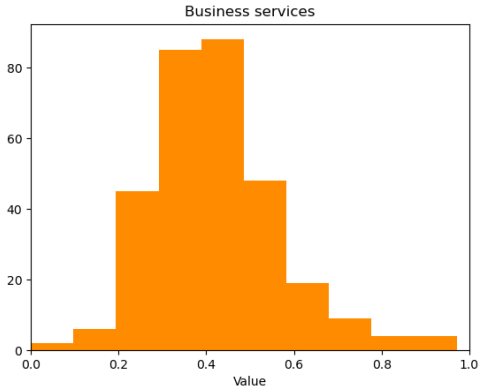

Innovative work
Value

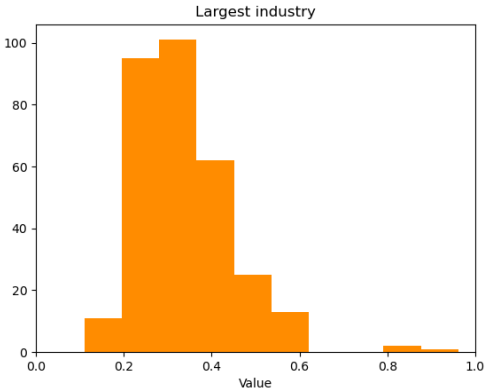

Knowledge work
Value

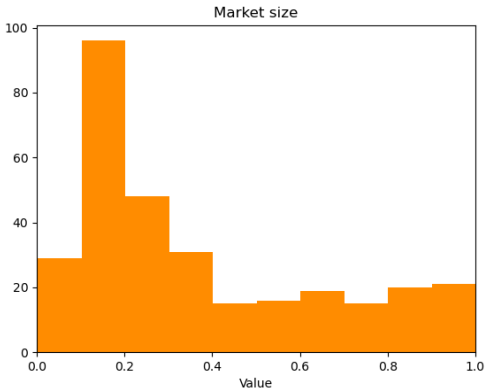

Patents
Value

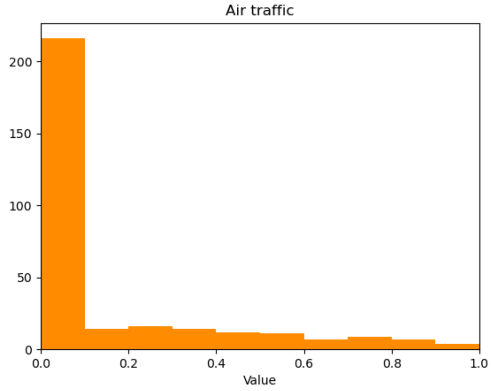

Centralization of population
Value


Centralized industry
Value


Business services
Value


Largest industry
Value


Market size
Value



Appendix 1.1-1.15. Distributions of competitiveness index variables.

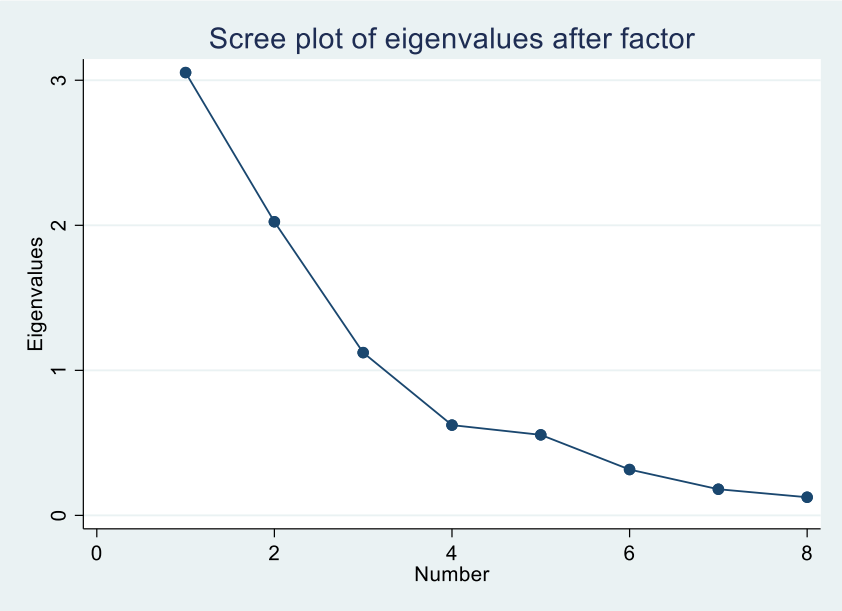
	Index	GDP	Salary
Social Equity	0.789	0.443	0.536
Innovativeness	0.839	0.447	0.536
Centralization	0.906	0.581	0.669
Reachability	0.731	0.539	0.561

Appendix 1.16. Correlation matrix with four dimensions against the index, GDP, and salary variables.

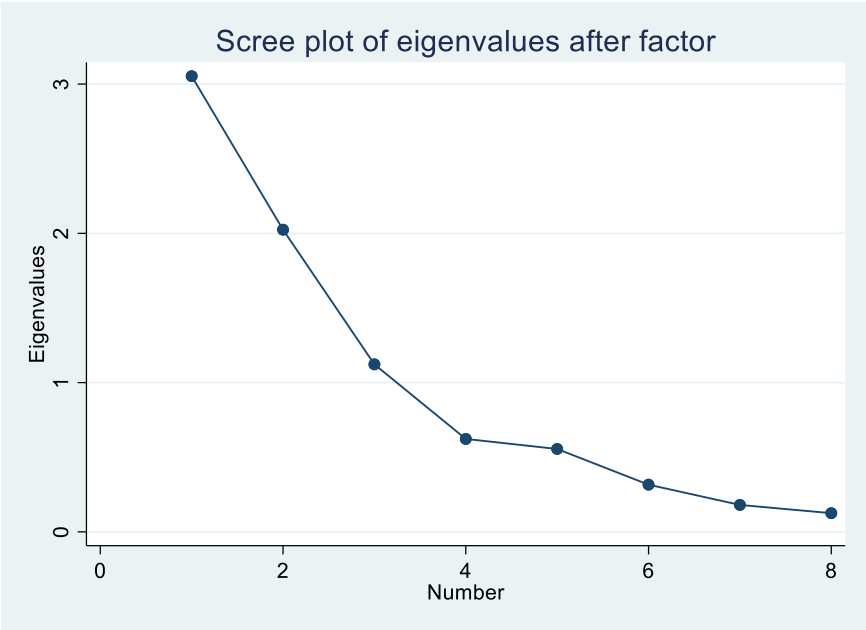
Appendix 2. Measure development for competitiveness.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Working age pop	1														
2. Participation rate	-.02	1													
3. Students	.48***	.05***	1												
4. Engineering stude	.48**	-.06***	.65***	1											
5. Tertiary degrees	.23***	.45***	.43***	.47***	1										
6. R&D	.31***	.04*	.41***	.56***	.36***	1									
7. Patents	.1***	-.03*	.08***	.09***	.11***	.07***	1								
8. Innovative places	.25***	-.02	.32***	.35***	.38***	.26***	.04**	1							
9. Knowledge work	.14***	.50***	.44***	.41***	.93***	.35***	.11***	.38***	1						
10. Centralization of	.41***	.30***	.48***	.46***	.71***	.31***	.13***	.39***	.74***	1					
11. % of Centralized	.11***	.47***	.09***	.20***	.55***	.15***	-.00	.35***	.61***	.51***	1				
12. % of Business se	.25***	-.05**	.32***	.36***	.37***	.27***	.05**	.96***	.38***	.40***	.33***	1			
13. % of Largest ind	.06***	.54***	.17***	.17***	.53***	.12***	.00	.24***	.59***	.41***	.85***	.25***	1		
14. Market size	-.04*	.38***	.14***	.06***	.43***	.11***	.03	.22***	.44***	.48***	.36***	.25***	.32***	1	
15. Air traffic	-.02	.33***	.12***	.06***	.40***	.10***	.04*	.27***	.48***	.47***	.43***	.29***	.37***	.81***	1

*Appendix 2.1. Spearman correlation of Social and informational equity (*p<.1, ** p < .05, *** p < .01).*



Appendix 2.2. Scree plot of refined factor model.



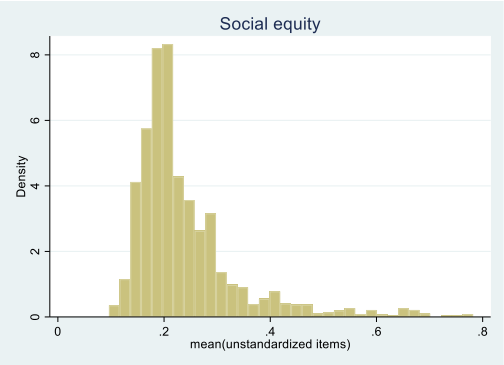
Appendix 2.3. Scree plot of the final factor model.

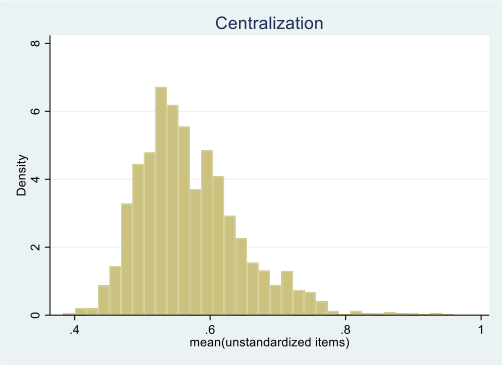
	2010					2018				
Variable	Factor 1.1	Factor 1.2	Factor 1.3	Uniqueness	MSA	Factor 1.1	Factor 1.2	Factor 1.3	Uniqueness	MSA
Working age population		0.79		0.37	0.73		0.76		0.40	0.73
Participation rate	0.77			0.40	0.77	0.77			0.38	0.78
Students		0.87		0.24	0.56		0.83		0.27	0.55
Engineering students		0.87		0.23	0.62		0.86		0.26	0.62
% of Centralized industry	0.87			0.22	0.57	0.86			0.18	0.60
Largest industry %	0.89			0.20	0.59	0.92			0.11	0.61
Market size	0.62			0.58	0.75			0.96	0.05	0.59
Air traffic			0.98	0.05	0.58			0.93	0.06	0.59
Eigenvalue	2.7	2	1			2.3	2.1	1.9		
Cum.%	0.33	0.59	0.71			0.29	0.55	0.79		
KMO					0.63					0.62

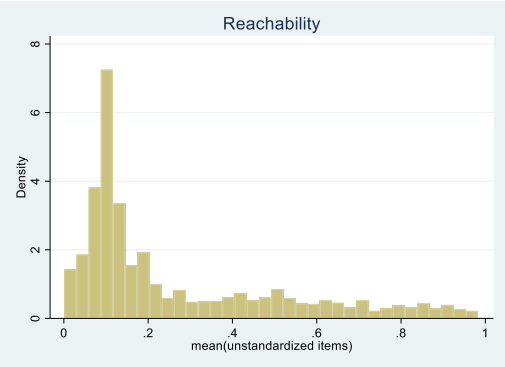
Appendix 2.4. External validity check for all three factors, years 2010 and 2018 statistically significant at the 1% level (p=.000).

Variable	Mean	Std. Dev.	Min	Max
Social equity	0.24	0.10	0.10	0.78
Centralization	0.57	0.07	0.38	0.96
Reachability	0.27	0.25	0.00	0.98

Appendix 2.5. Summary statistics of factors; Social equity, Centralization, and Reachability.







Appendix 2.6-2.8. Histogram of Factor analysis factors; Social equity, Centralization, and Reachability.

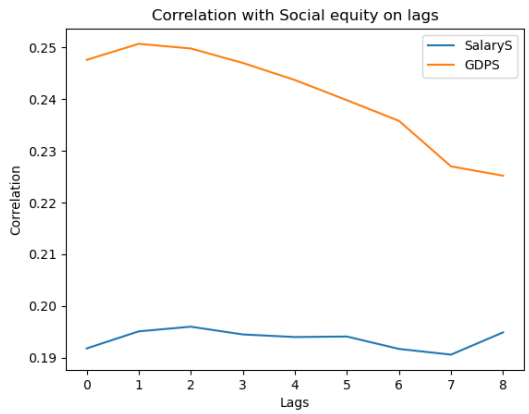
Appendix 3. Panel regression analysis.

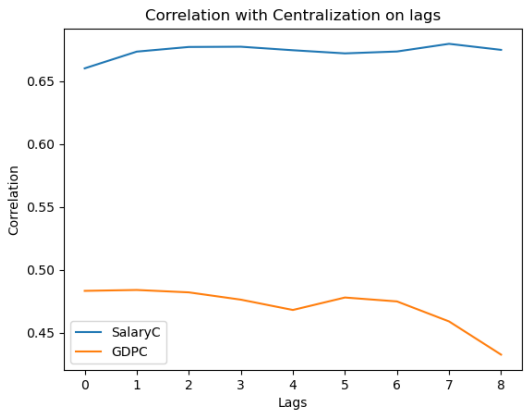
	Obs.	Mean	Std. Dev.	Min	Max
New competitiveness index	2790	0.36	0.11	0.21	0.76
Standardized SML	2790	0.00	1.00	-5.60	9.57
GDP	2790	32191.87	8029.57	16957.30	77681.00
Salary	2790	25490.60	4062.30	17666.00	67790.00

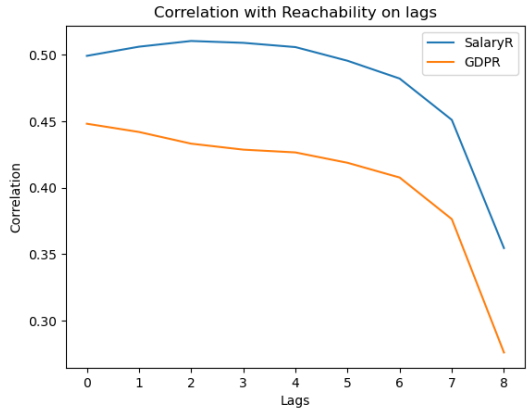
Appendix 3.1. Descriptive statistics of the variables used in cross-correlation analysis.

	Competitiveness index	index-1	index-2	index-3	index-4	index-5	index-6	index-7	index-8
SML	0.32	0.33	0.33	0.32	0.32	0.33	0.32	0.33	0.30
SML-1	0.32	0.32	0.33	0.33	0.32	0.32	0.34	0.32	0.32
SML-2	0.32	0.32	0.32	0.33	0.32	0.31	0.32	0.33	0.29
SML-3	0.32	0.33	0.32	0.32	0.33	0.33	0.31	0.32	0.36
SML-4	0.32	0.32	0.32	0.32	0.31	0.33	0.31	0.29	0.27
SML-5	0.34	0.34	0.33	0.34	0.34	0.33	0.35	0.35	0.33
SML-6	0.34	0.34	0.34	0.33	0.34	0.34	0.33	0.36	0.36
SML-7	0.33	0.33	0.33	0.33	0.32	0.33	0.33	0.31	0.33
SML-8	0.34	0.35	0.35	0.34	0.34	0.34	0.32	0.35	0.33

Appendix 3.2. Cross-correlation matrix for SML score and competitiveness index lags. Statistically significant at the 1% level (p=.000).







Appendix 3.3-3.5. Cross-correlation figures with GDP and salary against social equity, centralization, and reachability.

	SML	SML -1	SML -2	SML -3	SML -4	SML -5	SML -6	SML -7	SML -8
Competitiveness index	0.319	0.323	0.325	0.325	0.322	0.338	0.336	0.329	0.336

Appendix 3.6. Competitiveness index correlations with SML at different lags.

Method	GDP FE	Salary FE
Social equity	-210.89*** (47.66)	-194.97*** (25.57)
Centralization	79.29* (43.03)	99.68*** (23.09)
Reachability	-35.76*** (9.87)	-11.59** (5.29)
Observations	2790	
Municipalities	310	
R-squared	0.08	0.00
Probability > F	0.000	0.000
Breusch-Pagan test p-value	0.000	0.000
Hausman test p-value	0.000	0.000

Appendix 3.7. Competitiveness index factor fixed effects regression.

Method	GDP OLS	Salary OLS
Competitiveness index	414.70*** (12.03)	234.69*** (5.75)
Observations	2790	
Municipalities	310	
Root MSE	6724.4	3214.6
R-squared	0.30	0.37

Appendix 3.8. Competitiveness index OLS model.