ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis

Program: Economics and Business - Industrial Dynamics and Strategy

The relationship between agglomeration economies and firm profitability.

Name student: Mamasuev Nikita

Student ID number: 480100

Supervisor: Spyridon Stavropoulos

Second assessor: Martijn Burger

Date final version: 20.04.2018

Table of contents

Abstract

How does the profitability of manufacturing plant depend on whether it is located inside or outside the industry cluster? Using the measure of industrial concentration on NUTS 2 regional level in European Union, this thesis aims to investigate its relation with two profitability indicators for single manufacturing plants. One year cross sectional dataset with more than 410 thousands of firms observed is used for analysis. The results of tests of hypotheses point towards ambiguousness of effects of agglomeration economies on profitability of firms and suggest changes in approaches of further researches.

Chapter 1 - Introduction

Profit maximization is one of the main incentives behind any strategic decision of the firm. Some of those decisions can have a very long lasting effect on a variety of company's processes. Amongst those long term determinants of the company is its location. "To what extent does geographical positioning determine the success?"- is the question that is been asked throughout the history by the scientists of various specializations.

The locations that are chosen by the firms could be described by a variety of parameters as a place of operations brings a load of different externalities. Some of them are connected to the country of location that has its unique infrastructure and law enforcement and governance processes. Other ones of those parameters are described by the distances to, for example, transport hubs, or natural resources deposits. This thesis will mostly be focused on the distances between firms: their spatial proximity.

Choosing to locate close to the other firms or isolating your enterprise was always an open question to entrepreneurs. And as for the scientists, the idea of agglomeration economies first troubled Marshall (1890). Marshall investigated the different stages of production cycle in regional concentrations of trades that he described as an association of specialized firms, labor and trades. Until now, scientists keep returning to the idea of clusters in attempt to reinvent the approaches and definitions and to measure the effects they create.

A vast number of effects of agglomeration economies on firm performance were described by different researchers. Some of them make an emphasis on the benefits that clusters bring to the companies that compose those clusters. For example, Ciccone (2002) in his paper on agglomeration effects in Europe proves an existence of positive relation between density and productivity in Europe. Earlier, Ciccone & Hall (1996) concluded an increase in labor productivity in clusters in U.S. states. Audretsch & Feldman (1996) looked at another externality that occurs due to agglomeration and claimed that network effects tend to work in favor of companies that locate close to each other as the effectiveness of knowledge transmission depends on the geographical distance.

Other researchers, however, pay closer attention to the costs that those benefits bring. For example Glaeser and Mare (2001) as well as Combes (2012) bring evidence of drawbacks that clusters bring to the labor market; increasing the wage levels and making firms compete for the skilled workers.

There is an ongoing discussion in economic literature about the effects of agglomeration economies and firm performance, because, according to Jennen & Verwijmeren (2009) who studied the trade-offs of costs and benefits brought by clusters, in some cases benefits of agglomerations can outweigh the costs, and sometimes it's the other way around.

This thesis aims to investigate the connection of firm profitability and the production site location and to determine which combinations of firms and regions characteristics can result in firms benefiting from agglomerations.

On the following pages of that paper the theoretical framework that covers the literature relevant to the topic will be presented in the second chapter of that paper and three hypotheses will be proposed in coherence with the findings of cited literature. The data that was used to test those hypotheses will be described in the third chapter and the methodology of approach will be shown in the chapter number 4. Fifth chapter will cover the results of the used models and the conclusions that were made from those results will be presented. After all, additional robustness checks will be conducted in sixths chapter to answer questions regarding the quality of the research. Unanswered questions as well as flaws of the data and approach in general will be discussed in the limitations section of that paper that could be found in the chapter number 7. References to the cited literature as well as the appendix with the additional information can be found afterwards, in the very end of the paper.

Chapter 2 - Theoretical framework

2.1 First hypothesis framework

One of most important regional characteristics for a firm is how concentrated are firms from the same industry in this region. For example, Kukalis in his work looks at differences between performance of clustered and non-clustered firms (Kukalis, 2010). He claims that within an industry life cycle there are periods when clustered firms outperform isolated firms. However, those isolated laggards get ahead of the clustered firms in later stages of industry development. As the analysis in this paper is based on the one year sample, the change of the influence of the regional concentration could not be observed through the industry life cycle. But as the sample that is used for analysis in the research is composed of companies from developed manufacturing industries, according to Kukalis, the general negative effect of clusters on profitability should be expected.

But while some scholars use the dummy variable to identify whether the firm is in or out of the cluster (Kukalis, 2010), others decided to use a set of density measures. Jennen and Verwijmeren used a sample of single-establishment Dutch firms to find out that on average the costs of settling in an area with a dense spatial distribution of employment outweigh the benefits (Jennen & Verwijmeren, 2010). According to authors, the mechanism behind such effect is a trade-off between benefits of population and employment densities and the rise of costs of labor and real estate that the density of population brings. In their work they address the concept of spatial equilibrium, according to which high or low level of profitability in the region attracts or repulses the companies from that region until the equilibrium is reached. Jennen and Verwijmeren argue that such equilibrium cannot exist outside theoretical vacuum because firm's location choice is not driven solely by profit maximization incentives. Among other reasons for the location choice they bring up an example of business owner deciding to open a company close to his place of living. Taking into account all those additional causes that drive, for example, a choice of manufacturing plant location, it could be said that mentioned spatial equilibrium is expected to be pushed to the condition where companies still prefer to run business in a cluster despite the decrease of profitability. Again, location density was said to have a negative effect on firm's profitability.

Different conclusions were made by Ciccone and Hall (1994) in their article "Productivity and the Density of Economic Activity". In their aim to explain the large differences in labor productivity across U.S. states they found that a doubling of employment density increases average labor productivity by around 6 percent (Ciccone & Hall, 1994). They appear to be one of the first ones to address the problem of appearance of agglomeration externalities and the endogeneity issues that they bring. Ciccone and Hall argued that they could not include all the differences in the determinants of productivity that affected companies in researched states of USA and could not properly address named issues. However, this flaw of their work was later taken into account by Martin, Mayer and Mayneris in their paper named "Spatial concentration and plant-level productivity in France" (2011). In their case, the availability of panel data on firms and plants across eight years allows to account for endogeneity issues brought by existence of agglomeration externalities. As their main conclusion they still state that an increase in clustering will result in increase in plant productivity, because clustered plants benefit from the choice of the other companies to locate nearby. However, authors admit that clustering brings congestion that, in its turn, has an impact on utility of agents through the increase in, for example, pollution. Following that thought it is claimed that the presence of such trade-offs of positive and negative externalities is bound to result in non linearity of productivity and clustering relationship (Martin, Mayer & Mayneris, 2011).

In most cases, literature that focuses on the effects of agglomeration economies often describes them as either positive or negative. However, there are some factors, influence of which cannot be described so uniquely. The good example of such factor is reputation risk. While in today's world most of the relations between the companies are regulated with the strictly written contracts, it is not always the case for all business aspects and, especially, when talking about small firms. Ability of the company to fulfill its obligations, the way it treats its partners and employees and good relations with local government, press and competitors – this are all components of company's reputation. Good reputation could mean easier access to the qualified labor, increase in loyalty of currently employed people, also it attracts suppliers that could propose more lucrative and long term contracts. Positive company image may also result in pampering by the local government towards the company and, most importantly, attract the customers and increase their loyalty to the brand. Such condition is easier to achieve in a close business community, where all the firms interact with each other in one way or another. However, the drawbacks of undermining the reputation could hit company way harder if it is in a cluster because of two reasons. Firstly, the information about such fails will spread faster and further. Secondly, it is easier for both suppliers and customers to switch to a neighboring competitor. Effects of reputation risks in clusters are not researched and proven in the existing academic literature due to the fact that cases of reputation undermining are hard to track as well as the consequences they bring. However, the existence of reputation factors suggests that the effects of agglomeration economies are not as straight forward as it could appear from the cited literature.

Another argument that questions the existence of direct negative or positive effects of clusters on firm performance is the idea that there could be cases when combined with other factors some levels of spatial concentration could work in favor of firm's profitability. Martin, Mayer and Mayneris (2011) were not the only ones to suspect that the essence of connection between density of the industry in the region and the productivity of firms located in it was not linear. Multiple researches on the national scale proved that on the country level highly dense industrial structure could influence the firm's profitability in either way, depending on the level of concentration (Nickell, 1996; Gopinath, 2004). Similar findings were made by Stavropoulos and Skuras in their paper "Firm profitability and agglomeration economies: an elusive relationship". While using the location quotient as a measure of industry's concentration in a region, they proved that its effect on various measures of profitability was a quadric function. The analysis on that function proves that instead of overall negative effect of clusters on regional level, there are changes in firm profitability in regard to the level of concentration: at low levels of concentration the effect is negative, but with the increase in the value of location quotient it is followed by a significant positive shock which makes overall industrial density effect positive in a specific range of values of location quotient. What is also interesting is that according to authors, exceeding those values brings the analyzed function to the global minimum. At that point, apparently, the negative externalities that were created by congestion extremely diminish all the benefits provided by clusters (Stavropoulos& Skuras, 2016).

Overall, it is evident from the existing literature that the effect of regional specialization on firm profitability is not straight forward, but depends on exact values of cluster densities. Hence, in regard to that idea and the purpose of the thesis, the following hypothesis is tested:

1) Relation between regional specialization and company productivity exists and is not linear.

2.2 Second hypothesis framework

The effects of industrial concentration that are represented in the first hypothesis can significantly vary for companies with different characteristics. One of the things that influence company's ability to exploit the benefits of a cluster is size. Firstly, big manufacturers that are concentrated in one region reduce the market for local suppliers as they are often wired to international contracts and other nonlocal producers. That leads to decline in number of local independent suppliers which is a huge flaw for smaller companies in the same region (Enright, 1995; Porter, 1998). Another factor that favors big companies within clusters is ability to attract the most skilled workers. Qualified labor force is often attracted by stability of big companies, increased range of promotion possibilities and expectance of higher wages. Such behavior nullifies the effect of labor pooling economies for small firms in the cluster (Audretsch, 2001). While big companies could be better at exploiting cluster advantages, they could also be the first ones to suffer from the disadvantages. Negative externalities created by the cluster can have a much stronger impact on big companies compared to their smaller counterparts. As an example, local government can force a new law that will make the participants of the cluster reduce harmful exhausts of their plants. For small company that change in production process could be just a nuisance as it is easier to get rid of small amount of obsolete equipment and alter the production process. For less agile big companies this process of "going green" can be a way to suffer huge losses. The proposed example was precisely described by K. Hockerts and R. Wustenhagen in their paper "Greening Goliaths versus emerging Davids — Theorizing about the role of incumbents and new entrants in sustainable entrepreneurship"(Hockerts & Wustenhagen, 2010). The size of the company can have an impact both on the profitability of the company and on the factors that are taken into account by management when making a location decision. For example, it is said that location choice is frequently driven by management moving costs as well as personal preferences of the entrepreneurs (Jennen & Verwijmeren, 2010). That means that in case of small company, its only owner can make his decision solely on the fact that he wants his manufacturing plant to be close to his place of living. That, indeed, cannot have such weight in the decision made by Multinational Corporation. Also, the size of the company defines its acess to statistical information and its capabilities to analyze it in terms of both professional analysts and specialized software. Hence it could be concluded that two companies of different size located in that cluster could have been driven to that location by entirely different processes including the difference in the emphasis on profitability while making plant location choice. That leads to the thought about bigger companies being more aware of the pros and cons of clusters. Also, bigger companies have more resources to exploit the positive sides of clusters. Hence, size of the company appears to be an essential factor to control for and the following hypothesis is introduced:

2) Bigger[1] companies are more profitable and are better at exploiting cluster advantages.

[1] Size measured in sales.

2.3 Third hypothesis framework

It is evident that placing a manufacturing plant in a densely populated region can have a great impact on a wide variety of business processes.

In terms of productivity it is being known for a long time that, on average, the increase in population density results in an increase in labor productivity. It was researched, for example, by Sveikauskas in his work “Productivity in cities” in far 1975. In his paper Sveikauskas covered both, as he calls them, static and dynamic factors that are expected to have an effect on productivity in cities. “In the static sense a larger city permits more specialization and a greater division of labor and thus brings about an increase in productivity” (Sveikauskas, 1975). However, he claims that the new ideas and impressions that have high probability to occur in the city (the dynamic factors) are the ones to truly impact the technological progress and result in an upswing of productivity.

Those ideas were further developed in more modern researches. For example, Glaeser, Kallal, Sheinkman, Schleifer follow similar ideas, claiming that cities may drive the occurrence of knowledge spillovers between industries, which can result in an increase of profitability and growth rates (Glaeser, Kallal, Sheinkman, Schleifer, 1992). However, high level of spillovers within the industry can lower the incentives for research and development, which, consequently, affect the overall firm performance. Another mentioned bonus provided by regions with high population density is the wide access to the labor force. This could mean better specialists and lower wages, as the result of high competition on the labor market. But in case of the region with high level of specialization the labor poaching effect could occur. Such effect means the loss of key workers to other companies that in their turn have to offer higher wages to such specialists (Combes & Duranton, 2005). Together with Gobillon, Puga and Roux, Combes and Duranton wrote another interesting paper that offers various insights on the productivity of firms positioned in cities:”The productivity advantages of large cities: distinguishing agglomeration from firm selection”. While claiming that firms are, indeed, more productive in larger cities, authors provide two essentially different reasons for it: effects of firm selection and effects of agglomeration economies. Idea of firms selection in densely populated regions is built on the observation that cities often result in having more competitive markets that only allow the most efficient firms to remain operating. And as an impact of the agglomeration economies authors claim that “larger cities promote interactions that increase productivity” (Combes, 2012).

Moreover, more densely populated regions and, especially, cities, have a high level of local demand on production of the manufacturer, compared to more rural areas. On the other hand, attracting firms that produce homogenous products to the market will inevitably increase the competition between them. That could become the reason for lower prices and increased marketing costs both lowering firm profitability (Kukalis, 2010, 4).

In coherence with these arguments the following hypothesis is proposed:

3) Profitability increases in more urbanized areas. However, it decreases if the urbanized area is highly specialized

2.4 Theory on sectoral and regional heterogeneity

While this thesis focuses on manufacturing industry to observe the effects of clusters on production plant profitability, it is important to take into consideration the possibility of existence of deep firm- level heterogeneity even within seemingly homogenous manufacturing industry. The existence of heterogeneity across the regions of EU is much more obvious fact. However, it does not make it less essential to control for. There are some more or less widely used instruments that often included in aim to control for regional or firm level differences. Some of such instruments, specifically size and age of the company and degree of urbanization of the region were already mentioned in theoretical framework. And while regional and firm level characteristics appear to be distinct from each other for now, this section is dedicated to bringing up unobvious measures of heterogeneity that could affect the mechanisms of agglomeration that are being researched in that thesis. Surprisingly, according to the existing literature, similar approaches are used when identifying the core reasons for heterogeneity on regional and sectoral levels.

First ones to address such kinds of heterogeneity issues appear to be Frenken et al. (2007). In their work “Related variety, unrelated variety and regional economic growth” they distinguished between sectors where spatial proximity may support the process of knowledge dissipation (related variety) and sectors where knowledge was less likely to spill over (unrelated variety) in order to get insights on types and functioning of agglomeration externalities. Cortinovis and Van Oort (2015) in their research of economic growth of European regions followed the ideas of Frenken et al. (2007) and accounted for within sector related variety and between- sector unrelated variety using sectoral decomposition. Another study by Stavropoulos, van Oort and Burger (2018) on “Heterogeneous Relatedness and Firm Productivity” uses firm’s technological intensity and size to control for sectoral diversity. However, they also claim that regional hightech characterizations affect firm productivity by higher degree compared to the sectoral. All this evidence points towards the importance of mechanisms of knowledge spillovers and degree of knowledge intensity when talking about the differences between agglomeration processes across companies and regions.

Recent research made by Faggio, Silva and Strange (2017) was the closest one to the topic of that thesis and was focused on proving the differences in drivers of clustering across sectors of manufacturing industries specifically. Amongst the factors that cause the differences in mechanisms of agglomeration across industries they name entry and industry age, sector's technology orientation and workforce education. Also, according to their findings "high-technology sectors show stronger evidence of knowledge spillovers, while low-technology industries show stronger evidence of input sharing and labor pooling" (Faggio, 2017) which suggests knowledge intensity of the sector to be one of the main reasons behind the heterogeneity of companies in manufacturing industry. That heterogeneity issue will be addressed on the remaining pages of that thesis.

As all three hypotheses are introduced and all necessary remarks regarding further analysis are made, following parts of the thesis will be dedicated to the ways the tests of those hypotheses should be approached.

Chapter 3 - Data Description

In the following section, the data used to conduct empirical research required to test the hypotheses proposed in that thesis will be described.

While constructing the used dataset, a number of aims were pursued. First of all, the dataset had to be large. Increase of the size of the dataset increases the quality of statistical analysis in general, supports the strength of conclusions that are drawn from the analysis and decreases the probability of getting type II errors. Type II errors are especially bothering in case of any statistical research, including this thesis, as they infer the absence of something that is actually present. In case of this thesis, such false negative finding could lead, for example, to conclusion of absence of connection between firm profitability and industrial density while such relation could, in fact, be present. Secondly, the dataset should include wide variety of company level information to give an ability to draw measures of profitability from it as well as various company level descriptive statistics that could be further used to control for differences between the companies. Also, the dataset should include region level statistics that could be used to calculate the measures of industrial concentration/regional specialization that are essential for this paper. Constructing the dataset with described specifications was possible for the companies from manufacturing industry located on the territory of European Union.

Focusing on manufacturing industry brings noticeable benefits for the research of effect of regional specialization on profitability. The reason behind it is that manufacturing plants are more influenced by the effects brought by the spatial proximity than many other industries. For example, compared to the areas of financial or service economic activities, manufacturing is more dependent on the placement of the company. Plant location choices of the manufacturing plants are highly influenced by the presence of transportation hubs, wide access of to the labor force, possibilities to occupy extensive land and, ideally an easy way of delivering the final production to buyers and consumers. Such determinants of plant location choice in line with higher levels of economic size, educational attainment were named by Coughlin & Segev (2000) in their study on “Location Determinants of New Foreign Owned Manufacturing Plants”. And, for example, financial industry, on the other hand, is mostly constrained by the presence of qualified labor force and demand for its services. Also, manufacturing plants create arguably more noticeable externalities which make it easier to draw conclusions from analysis of companies from that industry. What is meant is that concentration of production sites overburdens the transportation system of the region and has a negative impact on environment due to inevitable exhaust of harmful substances appearance of which often accompanies manufacturing. Existence of such side effects of manufacturing industry clustering often draws attention of local communities, governments and policymakers to agglomerations. Hence, focus on the placement of manufacturing plants increases the value of that research as it could address the problems that often concern member states, regional authorities and media.

Firm-level information was obtained from Amadeus database, edition of year 2005. This is a commercial database collected by Bureau van Dijk. It includes the data on firms balance sheet and income statement positions as well as firm age, type of ownership, NACE codes for the European standard for industry classification and, most importantly, firm plant location. To avoid confusion and double counts, only firms with one production plants positioned in European Union were included. To have an opportunity to merge the described dataset with regional statistical information, plants locations were related to NUTS classification of the regions using the correspondence tables provided by Eurostat. NUTS 2 level of regional classification was used in the analysis, because NUTS 2 level represents geographical regions that have the highest level of autonomy amongst all EU member states (Stavropoulos& Skuras, 2016). And again, the value of that research increases as its conclusions could be applicable for the regional governments that are capable of making their own decisions regarding policies that affect the manufacturing clusters in their own regions.

The regional data was taken from the Structural Business Statistics (or SBS) published by Eurostat. SBS data is structured on the NACE activity classification and contains information about structure and performance of firms in the European Union. Using the data provided by the SBS the number of workers occupied in each industrial activity could be seen. That information, in its turn, was used to calculate the measures of industrial concentration that are essential for that thesis.

Also, information from the Eurostat Demographic survey was used to calculate the population density in the researched regions. This index of population density will be used as a proxy of level of urbanization, which, according to the literature, has a significant impact on all business activities which happen in the region and on effects of agglomeration economies as well.

The time period that will be covered by the research is only one year of 2005. The choice was made in favor of one-year cross-sectional sample, because making a panel dataset from a number of years turned out to be impossible with the data provided by the Amadeus database. Constructing the panel dataset over couple of years would mean a loss of a great portion of companies and result in obtained dataset being non representative for the manufacturing industry in the European Union. This is a legit decision in the framework of that thesis as it does not aim to look at any developments regarding the industrial clusters and their effects. Only the variation between companies, clusters and regions is of interest, meaning that representative one year cross-sectional sample will do just as good as a panel dataset. Specifically the year 2005 was chosen for the analysis because it should have no trace of upcoming financial crisis that appeared in 2008. It is essential to avoid the shocks that come with the financial crisis in the research because its impact on the companies can be highly dependent on the factors that could not be observed in scope of that thesis. For example, L. M. Gutu, A. I. Strachinaru, A. V. Strachinaru and V. Ilie in their work “The Macroeconomic Variables Impact on Industrial Production in the Context of Financial Crisis” research the impact of strong shocks in banking sector on industrial sector in Romania. It was proven that in state of financial crisis maintaining the corporation’s financial equilibrium becomes harder while debt burden gets heavier (L. M. Gutu, A. I. Strachinaru, A. V. Strachinaru, V. Ilie, 2015). First thing that suffers from such changes in the financial environment is firm’s profitability, which is the main variable of interest in that thesis.

In the end, the dataset that was used for analysis in this thesis includes 451,154 observations on different firms from manufacturing industry with single plant located on the territory of European Union in 2005. There are 15 countries of the EU included: Belgium, the Czech Republic, Finland, France, Germany, Greece, Hungary, Italy, the Netherlands, Poland, Portugal, Romania, Slovakia, Spain and Sweden. Table 1 in the appendix contains information on number of observations from each country as well as the percentage of the sample that belongs to each country. Highest proportion of observations belongs to Italy -22.2%. It is followed by France - 20.8% , and Spain – 17.2%. Despite seemingly under represented Germany with only 1.9% of observations belonging to it, it appears that the dataset covers the biggest economies of European Union. 200 different regions from those countries identified by NUTS 2 codes are present in the dataset. 11 industries from manufacturing sector according to NACE classification are represented:

1) Manufacture of food products and beverages;
2) Manufacture of textiles and textile products;
3) Manufacture of wood and wood products
4) Manufacture of pulp, paper and paper products; publishing and printing;
5) Manufacture of chemicals, chemical products and man-made fibers;
6) Manufacture of rubber and plastic products;
7) Manufacture of other non-metallic mineral products;
8) Manufacture of basic metal and fabricated metal products;
9) Manufacture of machinery and equipment;

10) Manufacture of electrical and optical equipment;
11) Not classified elsewhere manufacturing.

The sizes of firms included into the dataset differ significantly with sales varying from 0 to 7544500 thousands of Euros over the year, and with mean value at 7432 thousand Euros. The number of employees working at those firms in 2005 differs from 1 to 439400 with the mean at 50 workers. Enterprise value is also present in the data and could be used to measure the overall company value; it varies in range from -228881 to 59641000 thousand Euros, mean equaling 58050 thousands. Age of the firms in the dataset varies from 1 year to 628 years, mean at approximately 15 years. Return on total assets – is one of the main dependant variables and will be discussed later. In the data description section it will only be stated that the distribution of that variable in the dataset seems to be highly representative with return on total assets changing within full range from -100 to 100 and having its mean around 4 %. Also, the dataset includes broad information on the regions where manufacturing plants of those firms are located. For example, the GDP of the NUTS 2 level regions is present and varies from 1011 to 443134 millions of Euros with mean around 95440. Such information is essential for controlling for differences between the regions. Another important indicator: population density was already mentioned before. Number of inhabitants per square kilometer in the observed regions varies from 3 to 6253 and has a mean at 291.

Overall, the dataset contains a wide variety of statistical information on companies and regions where those companies are placed. Ones of them that were used to conduct econometrical research for that thesis will have the economical reasoning behind them presented in the methodology section of that paper.

Chapter 4 - Methodology

This part of the paper will be focused on describing the research approaches used in this thesis. First of all, the essential variables form the dataset will be highlighted. The economical reasoning behind them will be presented together with the expectances in behavior of those variables in coherence with the hypotheses. Overview of the used models will be given as well as the explanations why.

4.1 Dependant variables

The dependant variables are, indeed, the most substantial ones. As the main aim of this thesis is to investigate the relation between agglomeration economies and profitability, measures of profitability will be chosen as the main variables of interest. The used Amadeus database provides an easy access to two profitability measures: return on total assets and profit margin. Return on total assets is a frequently used measure of company's profitability as it shows the earnings of the enterprise before the taxation and subtraction of interest relative to total net assets. The following formula is used to calculate the return on total assets ratio:

$$ROTA = \frac{Net\ Income + Interest\ Expense + Taxes}{Total\ Net\ Assets}$$

The more the company earns relative to its assets, the more effective the use of those assets is and the more profitable is the firm. ROTA is expressed as a percentage and can vary from -100% to 100%. It reaches its maximum of 100% if every Euro invested in a company an Euro is earned by the company.

Second dependant variable is another popular measure of profitability: profit margin. Profit margin is a ratio that is calculated as net profit divided by sales of the company. Net profit equals total revenue minus all company's expenses. The formula for profit margin is as follows:

$$PRMA = \frac{Total\ Revenue - Operating\ Costs - Material\ Costs - Taxes}{Sales}$$

Profit margin is also expressed as a percentage and has the same variation interval as the Return on Total Assets ratio. It makes those two ratios very convenient to analyze together. Profit margin shows the proportion of total company's sales that actually ends up as earnings. A profit margin indicator equaling 1 means that for each one Euro of goods sold company increases its Total Revenue by 1 Euro.

Both Profit Margin and Return on Total Assets indicators are popular ways of measuring company's profitability, but they come with their own limitations. Comparison of two companies by those indices could only be made if the companies are assumed to have similar business models and revenue streams. Without that assumption the comparison loses all its value because the differences between revenue generating mechanisms could be too complex to be described with difference in profit from one Euro of sales. However, in case of that thesis the sample consists from companies that operate within one industry. And regarding manufacturing occupancy it is safe to assume that those companies have relatively similar business models to be comparable by introduced profitability measures. Both measures will be separately used in same models in attempt to find any differences in the estimations and inconsistencies in conclusions.

4.2 Explanatory and control variables

As this thesis aims to investigate the relation between introduced profitability measures and clustering of manufacturing industry, measure of geographical concentration of firms in the region should be used as a main explanatory variable. There are numerous ways to measure agglomeration economies, but with the access to industry level data on employment provided by the SBS the choice was made in favor of Location Quotient. With the used dataset the Location Quotient could be calculated for each sub sector of manufacturing industry in each NUTS 2 region. In that research the purpose of Location Quotient was to show how unique is each particular NUTS 2 region compared to the EU average in terms of industry concentration. European Union as a whole (and to be exact- 15 countries from EU that are represented in the sample) was chosen as a denominator in LQ fraction as it is assumed that each company represented in the sample, while making a decision on the manufacturing plant placement, was choosing between the locations around whole EU and not within specific region or cluster. It is implied that every country in the sample is a competitive player on the market of corporate investments. Following formula was used to calculate the Location Quotient:

$$LQ_{ir} = \frac{E_{ir}}{\sum_i E_{ir}} \Big/ \frac{\sum_r E_{ir}}{\sum_r \sum_i E_{ir}}$$

In that formula E_{ir} is the employment for manufacturing industry i in region r. Sum of employment across all manufacturing industries in region r is represented by $\sum_i E_{ir}$ in the formula above. The fraction in the denominator shows the overall employment in manufacturing industry m ($\sum_r E_{ir}$) divided by the overall employment in all industries across all regions ($\sum_r \sum_i E_{ir}$) – overall employment in manufacturing in EU. Exactly the same measure was used to describe the regional specialization by Glaeser (1992) and Stavropoulos (2016). It could be seen that LQ is calculated for specific industry in a specific region. Increase in LQ shows that that particular industry level of concentration in that region is higher than average level of concentration of manufacturing plants in EU.

As it was said earlier in this thesis, relation between the profitability indicators and LQ is expected to be non-trivial. That is why the analysis was conducted with the whole set of control variables:

1) First of all, the dataset gives an opportunity to treat the sample as a cross-sectional data with each observation corresponding to industry and region. That provides an opportunity to include regional dummies as well as industry dummies that will help accounting for structural and, otherwise, unobservable differences between different industries and regions.
2) Size of the company will be used as a control variable. As it was covered in the theoretical framework section of that thesis size of the company can affect the relation between profitability and LQ in a number of ways. First of all, bigger companies are expected to be generally more profitable as their resources, market power and analytical capabilities allow them to extract higher profit from their operations. But whether the trade-off between cluster advantages and disadvantages will be in favor of the company or not – is a tricky question answer to which also depends on the company's size. In that thesis, size of the company is measured by total sales of company's production in thousands of Euros. There is a number of ways to measure the firms size, however, sales in thousands of Euros seem like a valid choice in a framework of that research. Firm comparison by size is not biased by the differences in factor productivity, asset structure and business models of the firms; it appears to be a great way to control for differences in profitability that are connected to size within a manufacturing industry.
3) Age is another variable that is used to control for differences between the companies. It was expected to have a significant effect on profitability, as the firms that were operating in manufacturing for long time are bound to have well-established production processes as well as more lucrative long-term contracts with the suppliers.
4) Population density will be used as a measure of urbanization. Clusters that emerge in densely populated regions appear to have different characteristics compared to their counterparts in less populated zones. Also, the firms decision on whether to locate in the city or in a rural area highly depends on business processes established within the firm as well as type of cluster advantages the firm aims to exploit.

Descriptive statistics for introduced explanatory, dependent and control variables could be found in Table 2.

With all dependant and control variables introduced next section of the paper will aim to introduce the model most suitable for finding significant relations between variables of interest.

4.3 Model selection

The research for that thesis was conducted on single-year cross-sectional dataset. At the same time, main variables are profitability indexes that are continuous and may take values between -100% and 100%. Such data characteristics imply that the simple OLS model should be sufficient for getting reliable conclusions.

Another important insight regarding the model structure was obtained through the analysis of data. The illustration for it could be found on graph number 1 in the appendix section of that paper. Mentioned graph contains information on the relation of chosen profitability indexes and LQ. It was constructed by calculating the averages of profitability indicators over 20 quantiles of LQ separated based on equal number of observations in each quantile. The behavior of both ROTA and PRMA is quite similar in scope of that graph. Overall dependence trend of both indexes on LQ is negative. That observation is consistent with the big number of scientific findings about clustering having a negative effect on firm profitability. What is interesting is that the indexes take values both above and below the overall averages depending on the value of LQ. It is evident from the graph that the relations between main dependant variables and LQ are, indeed, non-linear. The number of changes in the direction of the trend suggests including up to 5 degrees of LQ in the model.

Summing up all the information on the dependant and control variables as well as the proposed type of the model, the formula for the basic model used to test the first hypothesis looks as follows:

$$ROTA\ (PRMA) = \beta_0 + \beta_1 * LQ + \beta_2 * LQ^2 + \beta_3 * LQ^3 + \beta_4 * LQ^4 + \beta_5 * LQ^5 + \beta_6 * age + \beta_7 * sales + \beta_8 * pop\_density + \beta_9 \dots \beta_{23} * country\_id + \beta_{24} \dots \beta_{34} * industry\_id$$

Where *ROTA* stays for Return on Total Assets, *PRMA* stays for Profit Margin, *LQ* and its degrees mean Location Quotient and its degrees, *age* means the age of the company in years, *sales* is the company's overall yearly sales in 10^10 Euros and *pop_density* is the number of inhabitants per square meter in the meant region. *Country_id* and *industry_id* refer to the country and industry dummies that were used to control for unobserved heterogeneity.

Chapter 5 - Results and discussion

In that section of the paper the outputs of the models that were used to test the proposed hypotheses will be presented. The interpretation to the obtained results will be given and discussed in scope of the framework of the paper.

5.1 Introducing the base model and testing the first hypothesis

First, the relation between profitability indexes of interest and LQ as seen from the graph 1 will be described. From the introduced graph it could be seen that first part of steep downward trend only continues until LQ reaches approximately 0.75, meaning that in regions with very low level of industrial concentration, higher LQ amongst them relates to less profitable firms. When the LQ reaches 1, meaning that the specialization of the particular region is on the average level of European Union, both ROTA and PRMA suddenly jump over the overall averages, identifying, that the average level of spatial proximity could be beneficial for the firm. However, that state does not last for long as the profitability seems to decrease in regions with LQ between 1 and 1.5. Regions with such characteristics are more specialized compared to the average level, but still have a small degree of industrial concentration. After 1.5 graphs for both ROTA and PRMA rocket up to what appears to be an overall maximum of represented profitability indexes. Maximum profitability appears to be obtained around LQ equaling 1.7 and is followed by the downward slope that remains for all higher values of LQ. Summing up, 5 sections of LQ are observed, each section identifying a change in the sign of relation between profitability and industrial concentration. Discussion of the reasons behind those changes will be presented further. Until then, the proof of statistical significance of those changes will be shown.

Table 3 contains the output of the base models that were used to research the relation between the Location Quotient and Return on Total Assets and Profit Margin. First thing to mention from the output is that Location Quotient variable is significant on 1 % significance level as well as all its degrees. This means that the level of industrial concentration in a region is connected with the profitability of the manufacturing plants located in that region. Moreover that connection is not linear but has an essence of fifth degree function: as each consequent degree of LQ changes it’s sign relative to the previous one, so does the trend of profitability. Such observations made from the output of the model are already enough to fully confirm the first hypothesis of this thesis about the relation between regional specialization and company productivity existing and being non-linear. That could be explained, again, by the point that was made in theoretical framework section of that thesis about the existence of various factors that are brought by agglomerations and the existence of trade-offs between those factors that could affect the company's profitability in any direction depending on the combination of different regional and company characteristics.

All other control variables in the base model of that thesis also turned out to be significant. The dummy variables for countries where the plants are located and the dummies for different manufacturing industries are also presented in the appendix. Even though those dummies are not essential to make conclusions regarding the proposed hypothesis, they bring important insights that will be discussed later. For now, it is important to state that all industry dummies turned out to be significant on 1 % significance level for both Profitability indexes. Also, the signs of coefficients of industry dummies are same for ROTA and PRMA models in all the cases. As "Manufacture of food products and beverages" is omitted from the model and used as a reference category, all other industries, except for "Manufacture of textiles and textile products", have positive influences on profitability indicators, meaning that those industry's activities are, on average, more profitable than the manufacture of food products and beverages. For the country dummies everything is not that clear: Not all of them turned out to be significant. Still, the majority of country dummies got coefficients that are significant on 1 % significant level, showing, one again, how important it was to control for the country to which the region and the company belong.

As for the age of the company variable, the results are quite ambiguous as well. Even though the effect of company age on both ROTA and PRMA was proven to be significant of 1% significance level, the sign of that coefficient varies for different profitability indexes and is negative for ROTA and positive for PRMA. There does not seem to be a logical explanation to it. But, including the fact that the size of the coefficient of age is very small relative to other variables in both models, it is safe to assume that age is not the strongest determinant of profitability and is overshadowed by the effects of agglomeration economies. Additional discussion on the role of age in regard of clustering and profitability could be found in "Robustness checks" section of this paper.

Situation with another control variable – size of the company, is not that complex. To remind, as the proxy of firm size, the yearly sales in 10 in 10^{th} degree Euros were used. In case of Return on Total Assets it could be concluded that an increase in total sales of the company by 10^10 Euros increases the ROTA by 3.84 percentage points, the effect is significant on 1 % significance level. In case of Profit Margin, same increase in sales will result in increase in PRMA by 3.45 percentage points, the effect is significant on 1 % significance level. That result suggests that overall companies with more sales are more profitable than those with less sales, or simply that bigger companies perform better than smaller ones. This, however, is not enough to test the second hypothesis of that thesis, as it is not yet clear how the improvement of performance driven by the size of the company is affected by the agglomeration economies. That is why, in order to test the second hypothesis, another model that includes the interaction of clustering measure and company size in it had to be proposed.

5.2 Testing the second hypothesis

The output of the model that gives insights regarding the second hypothesis of that thesis could be found in table 4. While significance levels and signs of coefficients of all main explanatory and control variables remained the same, the coefficient of the interaction between the sales and the first degree of location quotient provides questionable results. The interaction coefficient has a negative value in models with both profitability indexes as dependant variables, however in ROTA model it turns out to be insignificant, while in case of PRMA it is significant on 5 % significance level. So, according to first model, highlighting big companies in concentrated regions does not really add anything to existing conclusions about the relations between size and ROTA and LQ and ROTA. However, from PRMA model it could be concluded that profitability of bigger companies actually decreases in regions with high level of concentration. Such difference in estimation coefficients could occur due to the fact that Profit Margin index directly includes the company's sales in its formula, while the derivative of the ROTA index consists of Total Assets connection of which with the sales is questionable. So in case of interaction with the size, when total sales are used as a proxy of size, it is suggested to mostly focus on the results of the model where Profit margin is used as a profitability indicator.

The model output in table 4 only presents the results of the interaction of size proxy with the first degree of LQ. However, it is already proven that the relation between the profitability and clustering measure is not linear. And this thesis aims to investigate the effects of size for all values of LQ divided by differences in profitability indexes behavior. Hence, the models that included interactions of Location quotient and sales were also run. Their results will not be presented in that paper as they do not change for most of the control variables relative to the model where the interaction of sales with the first degree of LQ is used. However, it is important to notice that only the models that use the interaction of size and LQ for the first and the second degree of LQ have a significant interaction coefficient. That means that the behavior of relation between Profitability and agglomeration economies is not homogeneous across companies with different sizes. For both first and second degree of LQ the coefficient of interaction remains negative, meaning that the increase in concentration of companies in a region and increase of the size o those companies comes in hand with decrease in profitability of those companies, and that relation is linear.

That conclusion is contradictory to the second hypothesis that was proposed in that thesis. It was expected that due to their resourcefulness, bigger companies will be better at exploiting the advantages of the clusters. However, it appears that the trade-offs they are facing are not going in their favor after all. As it was mentioned in the theoretical framework of that thesis, big companies are stronger affected by the negative externalities that arise in the clusters due to the congestion. On average, that turns out to be the case of companies in the used sample as the manufacturing industries could be described as an activity especially responsible for appearance of negative externalities such as pollution. Also, manufacturing could be described as an industry with high spillover rates. And, in a cluster the ones inventing could be the bigger firms, and the ones benefitting from the spillovers of those inventions are the smaller ones.

There is, however, an idea that questions the truthfulness of this result. In the sample there could be some especially big companies that are the only ones responsible for the increase in the value of LQ in the region they locate. Such companies construct a cluster themselves and by themselves only, which leads to some non-orthodox consequences of clustering. That situation is called the "lock-in" problem and was addressed by Boshma (2005) as a source of negative influence on innovation. In such case, though, the company can obtain monopoly benefits in some aspects of its operations. For example, it could result in increase of bargaining power on labor market and on the market of local suppliers. Also, such company behavior decreases the probability of knowledge spillover outside of the company. However, that condition could backfire as those companies also close in from any spillovers from the outside. Question to which degree such rare but influential outliers have affected the output of the model will be answered in the "Robustness checks" section of that paper.

5.3 Testing the third hypothesis

Another goal of that paper was to highlight the relation between urbanization and profitability in clusters. As it is seen from the output of the base model, which results can be found in the table number 3, an increase in number of inhabitants per 1 square meter by one will, on average, result in an increase by 0.58 percentage points of ROTA index and in an increase by 0.71 percentage points of PRMA index. The effects are significant on 1 % significance level in both models. Number of inhabitants per square meter was used as a proxy of level of urbanization of the region. Results of the base model do support the statement that companies from more urbanized areas are more profitable. But the point of great interest is to see how profitability alters if the urbanized areas also contain an industry cluster. That question was approached the same way the second hypothesis was: an interaction of urbanization measure and clustering measure was introduced into the model. The output of that model can be found in an appendix of that paper in table 5.

Coefficient of interaction between Location Quotient and the number of inhabitants per square meter turned out to be negative and significant for both models. Regarding the coefficients of all other control variables, their signs and significance levels did not change compared to the base model that was proposed earlier in that paper. From the output of that model it could be concluded that while more urbanized areas are occupied by more profitable firms, firms that locate in the regions with high population density and high level of concentration of companies from the same industry turn out to be less profitable. That conclusion fully supports the third hypothesis that was proposed in the theoretical framework section. It appears that, indeed, the trade-offs of clusters do not go in favor of firms if those clusters are positioned in highly urbanized regions.

Again, because it was proven that the relation between the plant concentration and in a region and firm profitability is not linear, while looking at the interaction of LQ with any other variable it is important to look at the behavior of interaction at all 5 sections defined by the change in trends of profitability curve. The same way it was done while testing the second hypothesis, models with interactions of population density variable and all 5 degrees of LQ were run for both profitability indexes. Outputs of those check models will not be presented in the paper. However, there is a description of their results.

For all models, no matter what profitability index is used as a dependant variable and regardless the degree of LQ with which population density was interacted, the coefficient of the interaction is negative. In case of PRMA index, that negative interaction effect is significant on 1 % significance level in every model. Though for ROTA models it turned insignificant for urbanization measure interacting with fourth and fifth degree of LQ but, again, became significant on 1 % significance level for all three remaining degrees of LQ. With such results it is save to conclude, that there is a linear relation between profitability and urbanization in the clusters and that relation is negative. With that being said, third hypothesis of that thesis is fully proven and supported.

To sum up the results section of that paper; of three proposed hypothesis, first and second were fully supported by the used models. While testing the second hypothesis, unexpected results did occur, showing significant coefficients that contradict half of the second hypothesis.

While some robustness checks were already made to prove important conclusions made in the results section of that paper, some questions regarding the quality of conducted analysis remain unanswered. These questions will be addressed in the following “Robustness checks” section of that thesis.

Chapter 6 - Robustness checks

6.1 Controlling for regional heterogeneity

First question that arise in regard to conducted research is the sufficiency of control variables included in the model. Sure, some company level descriptive statistics were included as well as some regional level information. While the sample was treated as a cross sectional data, the included dummies accounted for unobserved country and industry parameters. However, there could be some company specific or region specific differences that were left unaccounted and biased the results.

Regarding the company specifics, no other parameters that could add explanatory power to the proposed models could be located in Amadeus database as well as any other publically accessed database. That drawback will be covered in the limitations section of the paper.

As it comes to the regional characteristics, Eurostat provides a wide variety of data that could be easily matched with the used dataset via NUTS 2 regional ids. In search for additional variables that could help controlling for regional differences, following information was added:

1) Regional area in square kilometers that could help accounting for regions with high LQ's, but land area so big it makes the plants of different firms pretty distinct in reality.
2) GDP per capita could control for macroeconomic differences in productivity and wealth between the regions. Both affect the profitability of the companies that locate in the region as the wages and the cost of assets increase in the region, while the customers buying power and productivity of workers increases as well.
3) Average annual wage could get more direct insights on differences in costs of workforce in a region. This regional parameter directly affects the profitability of the firm that locates in a region.
4) Distance to seaport appeared to be a meaningful variable during the analysis. Seaport accessibility could be an important factor that is often taken into account by manufacturers as it diminishes the costs of transportation of both materials needed for production to the manufacturing plant and final products from plant to the consumers.
5) Long term unemployment rate is an additional way of controlling for overall prosperity of the region as well as for accessibility of workforce.

6) Percentage of urban land could be used instead of population density as a measure of urbanization. This additional robustness check is used to see, if the model output changes with the change of the proxy.

These variables were included in the base model in order to observe any changes to previously observed relations. Population density variable was excluded from the base model and replaced with the percentage of land in use that is urban. The output of that model can be found in the Table 6 in the appendix.

First thing to notice is that compared to the base model the number of observations decreased significantly, as not all of the regions present in the original dataset got a match. Still, the number of observations remains sufficient to draw statistically significant conclusions, and the structure of the dataset almost remained. Most importantly, from comparison of the model outputs presented in tables 3 and 6 it could be concluded that the signs of main dependant and control variables remain unchanged. Same could be said about the significance levels of coefficients obtained for those variables, except for first degree of LQ in PRMA model, which significance level dropped from 1 % to 5 %. The coefficient for proxy of urbanization is positive and significant regardless the variable used. These facts serve as a proof of sufficiency of a conclusions made from previously analyzed models. Also it could be observed that all of the introduced regional characteristics do have a statistically significant relation with the profitability of the firm, showing again the importance of location choice.

6.2 Additional insights on country differences

Previous section of robustness checks was aiming to control for regional heterogeneity, however, this analysis still misses the higher scale country level differences.

Table 3 shows the output of the base model used to test the first hypothesis. Even though the country dummies were only used to control for country heterogeneity while testing the proposed hypotheses, the obtained coefficients for those dummies do bring some important insights. These coefficients should be interpreted relative to the companies situated in Belgium, because its dummy was used as a reference.

In case of model with ROTA as main dependant variable, all country dummies except for France, Netherlands, Slovakia and Sweden had effects on firm profitability significant on 1 %

significance level. And in case of PRMA model this list contains Czech Republic, France, Hungary, Slovakia and Sweden.

The obtained results, again, point towards the heterogeneity of countries in the sample. Changes in profitability of manufacturing plants relative to the level of Belgium appear to decrease for southern and western European countries. At the same time, country dummy coefficients do not appear to be correlating with some obvious indicators of country's wealth like GDP per capita. Instead, outliers like Romania that has, on average, unexpectedly high level of plant profitability, suggest that there could be a variety of unobserved country – specific factors that are wired to, for example, country's institutions or work ethics. The existence of such factors should be taken into account by further researches.

6.3 Additional checks for the effect of company age

Another variable that was promised to get more attention is company age. Its coefficient sign changes with the change of profitability indicator used in the model while being significant in both models. Its interaction with LQ was checked the same way it was done with proxies of size and urbanization. The coefficients of those interactions were proven to be significant, but their volume is as small as it is for age in the base model. Judging on this it is safe to conclude that age of the company is neutral to agglomeration economies.

6.4 Controlling for the influence of outliers

Last thing to do as a robustness check is investigate how the model outputs change if the data sample is cleared from the outliers. Both firms that are too small and too big were thrown out for sake of that check. Smallest firms could have potentially biased the results because of huge variance of profitability on small scale. As an extreme example, if company sold only one unit of final production, its profitability could change drastically compared to zero products sold. Extremely big companies could bias the results for number of reasons. First, they could be exploiting some unobserved advantages of the region, for example, have a very strong lobbying power in a regional government. Second, as it was said before, huge corporations could create the clusters themselves and be the only reason behind the high values of LQ in the region. Such cases start processes that are very different from the regular effects of agglomeration economies and should be reviewed separately.

After arranging the dataset by company's sales and deleting those observations where the amount of sales was not present, 2.5 % of observations were cut from each side of the sample. Then, the base model was run again on newly obtained sample. The output of that model could be found in table 7 in the appendix.

Thing to pay attention to is the fact that the significance levels as well as the signs of the coefficients of model from table 7 did not change relative to the base model output. The only questionable one is the coefficient of age, which should not be taken into account according to previous conclusions. The value of coefficient of sales increased drastically due to the major increase in average sales volume of firms in the sample. But in the end, all conclusions that were made in regard to dependant and control variables of the base model remain.

6.5 Controlling for sectoral heterogeneity

While sector-level dummies were already included in every model for sake of controlling for unobserved differences between researched manufacturing industries, attempts were made to account for industrial heterogeneity in a more accurate way. According to existing literature, numerous factors could complement to differences in mechanisms of agglomeration across sectors of manufacturing industry. For example, Faggio, Silva and Strange (2017) mention entry and industry age, sector's technology orientation, workforce education and knowledge intensity of the sector. "Limitations" section of that thesis will cover the unfortunate fact that gathering data on most of these characteristics turned out to be impossible. However, one of the most important drivers of heterogeneity between the industries, the knowledge intensity, could be included in the analysis as a robustness check of proposed models.

The sectors of the manufacturing industry that were used for analysis were defined using NACE REV 1.1 classification. That classification also implies aggregation of those sectors by technological intensity. Using the aggregations proposed by Eurostat, researched sectors were separated by the dummy variable that had a value 0 if related manufacturing industry was classified as the one having low technological intensity and 1 otherwise. Sectors with low technological intensity include manufacture of food products, beverages and tobacco, textiles and textile products, leather and leather products, wood and wood products, pulp, paper and paper products, publishing and printing and other non-classified manufacturing. All other sectors had a value of 1 of introduced dummy variable related to them.

Introduced dummy as well as its interaction with the Location Quotient were added to the base model. The output of the resulting model can be found in table 8 in the appendix. Adding these variables remained the coefficients and significance levels of main explanatory and control variables almost untouched relative to the base model output. At the same time the positive effect of knowledge intensity appears to have quite big effect (relative to the effect of LQ). This effect is significant on 1 % significance level for both ROTA and PRMA models. What is more interesting, the interaction of sector's knowledge intensity and LQ got a negative effect on estimated firm profitability. This effect is significant on 5 % and 1 % significant levels for ROTA and PRMA models respectively. According to that model, while manufacturing plants with low technological intensity appear to be less productive that their technologically intense counterparts, high technological intensity actually negatively relates with the performance of the plant located in a cluster. These findings are coherent with the existing academic literature as the knowledge intensive industries are claimed to be more affected by negative effects of knowledge spillovers (Faggio, 2017).

Firstly, it means that the conclusions regarding the hypotheses tested in the thesis remain unchanged even when adding more control variables. Secondly, this check shows the strong presence of heterogeneity within seemingly similar sectors of manufacturing industry. Additional controls for sector differences are recommended for further researches on mechanisms of agglomeration.

Chapter 7 - Limitations

One of the limitations that are faced by that research is the questionable approach to cluster definition. The data availability generally determined the research approach of that thesis and the geographical proximity between firms, or, to be exact, the fact of locating a manufacturing plant in the same region was used as way to measure agglomeration economies. How such economies should be measured – is an ongoing discussion in the academic literature. But in today's world getting more and more connected by the internet, geographical proximity is, obviously, losing its weight as a measure of agglomeration. For example, Rodrigues-Pose (2011) in his paper on the changing conception of distance in geography claimed that geographical proximity becomes an inferior determinant of location of economic activity.

Choosing the correct set of characteristics on which to define a cluster will not only lead to deeper understanding of essence of clusters and the reasons behind firms choices to locate in one region or another, but also could give additional insights on the positive and negative effects of clusters. In case of that research, unfortunately, only a relation could be proven, that could only be followed by a guess of set of reasons that led to that relation being negative or positive. For example, because of the lack of company-wide descriptive statistics, model used to test the third hypothesis has no way of distinguishing between the effects of agglomeration economies and effects of firm selection in cities. An emphasis on importance of such distinction was made by Combes (2012).

Generally, an attempt to figure out, how effects of agglomeration economies differ for firms with different characteristics, ended up on controlling for firm's age and size. These characteristics are the only promising ones from those that were obtained on practice, but theoretically speaking, many more unobserved firm parameters could alter the effects of clusters on profitability. Among such characteristics Stavropoulos and Skuras (2016) name absorptive capacity of knowledge spillovers, investments in R&D and development of human capital as well as participation in knowledge and innovation networks.

The inability to collect extended firm-level information also made it impossible to test for reputation risks. Their importance amongst other effects of agglomeration economies was discussed in the theoretical framework section of that paper. Collecting evidence on cases of reputation undermining and observing their effects in and outside the clusters could be a substantial topic for further research.

Another significant limitation that constrains that analysis is the fact that only the presence of the relation between agglomeration economies and firm profitability could be tested, but causal relation could not be proven and effects of clusters on firm profitability could not be proven. The reason behind it is the possibility of backwards causality between researched variables.

The data obtained from the Amadeus database is affected by a minor flaw because it has a bias on medium and large firms. Small companies are underrepresented in the dataset because for many of them the task of gathering accurate financial data appears to be impossible. Including small companies into the sample should definitely improve the accuracy of the analysis as such companies are highly affected by the externalities that are created by clusters. However, such addition is not expected to bring any drastic changes, because, according to the executed robustness checks, cutting the tails of the distribution does not change the conclusions indicating the dominance of medium sized companies over the model outputs.

Chapter 8 - Conclusions

By now, the theoretical framework of this thesis was presented and three hypotheses were proposed in coherence with the findings of cited literature. The data that was used to test those hypotheses as well as the methodology of approach were described. Results of the used models were presented and the conclusions were drawn. After all, additional robustness checks were conducted to answer questions regarding the quality of the research where possible. Unanswered questions as well as flaws of the data and approach in general were discussed in the limitations section of that paper.

Three hypotheses were proposed in a pursue of investigation of connection between firm profitability and the production site location. One of the hypotheses was aiming to prove the existence of relation in the first place and to describe it, and other two hypotheses were introduced to determine which combinations of firms and regions characteristics can result in firms benefiting from agglomerations. First hypothesis was fully concluded, statistically proving the presence of relation between the clustering proxy and two different profitability indicators. This relation was proven to have a shape of fifth degree function. Second hypothesis was partly concluded only partly as, while testing for company's size, its effect on profitability was proven to be positive, however, it was unexpected to find out bigger firms in clusters showing worse performance, compared to smaller firms in regions with high LQ's for respective industry. The third hypothesis was focused on effects of urbanization on trade-offs brought by agglomeration economies and was fully concluded. Companies in cities were proven to be more profitable while urbanization was diminishing the positive effects of clusters on profitability. The reasoning behind all these results was provided in the results section of the paper.

However, all these conclusions are only valuable for the agglomeration economies defined by spatial proximity. And whether it is a right way to measure the clusters is an open question. For example, Boschma (2005) argues that geographical proximity by itself cannot be used as an isolated factor, but other dimensions of proximity should be taken into account as well. Amongst those dimensions, apart from geographical proximity, Boschma names cognitive, organizational, social and institutional proximities. Unfortunately, in scope of this thesis it was impossible to control for additional dimensions of proximity. Hence, as a further research recommendation, adding more proximity indicators that could support or even overcome the impact of geographical proximity would greatly improve the truthfulness of the conclusions made in that paper.

Adding more dimensions of proximity also implies widening of and deepening of information gathered on firm level. While used models were only controlling for age and size of the company, all firm level indicators connected to measures of cognitive, organizational, social and institutional proximities could have additional ambiguous impact on effects of differently measured agglomeration economies. Expanding the dataset with such characteristics would also be recommended for upcoming researches.

Also, while the important issue of sectoral and regional heterogeneity was addressed in the scope of that thesis, not all of the potential differences that could affect the mechanisms of agglomeration were covered and controlled for. Different firms in different regions agglomerate for different reasons, understanding that is essential when researching the relation between agglomeration economies and firm performance.

Overall, the direction of this research is extremely valuable as the conclusions made in the process could be used by company's management to further rationalize the placement of manufacturing plants for profit maximization purposes. Also, obtained insights could be used by local governments that are trying to control and optimize agglomeration processes happening in their regions.

References

Audretsch, D., & Feldman, M. 1996. Knowledge spillovers and the geography of innovation and production, *American Economic Review,* 86: 630-640.

Audretsch, D.B., 2001. Research issues relating to structure, competition, and performance of small technology-based firms. *Small Business Economics* 16 (1), 37–51.

Boschma, R.A. (2005), Proximity and Innovation: A Critical Assessment, *Regional Studies* 39, pp. 61–74.

Ciccone, A. & R.E. Hall (1996), Productivity and the density of economic activity. *American* Economic Review 86, pp. 54–70.

Ciccone, A. (2002) Agglomeration effects in Europe, *European Economic Review*, 46, pp. 213–227.

Combes, P.-P. & G. Duranton (2006), Labour Pooling, Labour Poaching, and Spatial Clustering. *Regional Science and Urban Economics* 36, pp. 1–28.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga & S. Roux (2012), The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica* 80, pp. 2543–2594.

Cortinovis, N., & F. van Oort (2015). "Variety, economic growth and knowledge intensity of European regions: A spatial panel analysis". *Annals of Regional Science* 55: 7–32.

Coughlin, C. C., & Segev, E. (2000). Location determinants of new foreign-owned manufacturing plants. *Journal of Regional Science*, *40*(2), 323–351.

Enright, M.J., 1995. Organization and coordination in geographically concentrated industries. In: Lamoreau, N.R., Raff, D.M.G. (Eds.), Coordination and Information: Historical Perspectives on the Organization of Enterprise. *University of Chicago Press*, pp. 103–146.

Faggio, G., O. Silva & W.C. Strange (2017), “Heterogeneous agglomeration”. *Review of Economics and Statistics* 99: 80-94.

Frenken K, van Oort F, Verburg T (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies* 41:685–697

Glaeser E., H. Kallal, J. Sheinkman & A. Schleifer (1992), Growth in Cities. *Journal of Political Economy* 100, pp. 1126–1152.

Glaeser, E. L. and Maré, D. C. (2001) Cities and skills, *Journal of Labor Economics*, 19, pp. 316–342.

Gopinath, M., Pick, D., Li, Y., 2004. An empirical analysis of productivity growth and industrial concentration in US manufacturing. *Applied Economics* 36 (1), 1–17.

Gutu L. M., A. I. Strachinaru, A. V. Strachinaru, V. Ilie (2015), The Macroeconomic Variables Impact on Industrial Production in the Context of Financial Crisis, *Procedia Economics and Finance* 32, pp. 1258-1267.

Hockerts K., R. Wujstenhagen (2010), Greening Goliaths versus emerging Davids - Theorizing about the role of incumbents and new entrants in sustainable entrepreneurship, *Journal of Business Venturing*, Volume 25, Issue 5, pp. 481-492

Jennen, M. & P. Verwijmeren (2010), Agglomeration Effects and Financial Performance. *Urban Studies* 47, pp. 2683–2703.

Kukalis, S. (2010), Agglomeration Economies and Firm Performance: The Case of Industry Clusters, *Journal of Management* 36, pp. 453–481.

Marshall, A. 1890. *Principles of economics*. London: Macmillan.

Martin, P., T. Mayer & F. Mayneris (2011), Spatial Concentration and Plant-level Productivity in France. *Journal of Urban Economics* 69, pp. 182–195.

Nickell, S.J., 1996. Competition and corporate performance. *Journal of Political Economy*

104 (4), 724–746.

Porter, M.E., 1998. Clusters and the new economics of competition. *Harvard Business Review* 77–90 November/December.

Rodriguez-Pose, A. (2011), Economists as Geographers and Geographers as Something Else: On the Changing Conception of Distance in Geography and Economics, *Journal of Economic Geography* 11, pp. 347–356.

Stavropoulos, S. and Skuras, D. (2016), Firm Profitability and Agglomeration Economies: An Elusive Relationship. *Tijdschrift voor Economische en Sociale Geografie*, 107: 66–80.

Stavropoulos S., F.G. van Oort & M.J. Burger (2018), “Heterogeneous Relatedness and Firm Productivity”

Sveikauskas, L. (1975), Productivity of Cities. *Quarterly Journal of Economics* 89, pp. 393–413.

Appendix

Table 1

Country representation summary.

Country	No.	%
BELGIUM	21.974	4,9%
CZECH REPUBLIC	7.850	1,7%
FINLAND	9.640	2,1%
FRANCE	93.890	20,8%
GERMANY	8.794	1,9%
GREECE	6.502	1,4%
HUNGARY	24.274	5,4%
ITALY	100.362	22,2%
NETHERLANDS	1.727	0,4%
POLAND	8.760	1,9%
PORTUGAL	10.855	2,4%
ROMANIA	55.639	12,3%
SLOVAKIA	1.241	0,3%
SPAIN	77.579	17,2%
SWEDEN	22.067	4,9%
Total	451.154	100,0%

Table 2

Descriptive statistics of main variables.

	mean	sd	count
Return on Total Assets	3.699	19.624	451154
Profit Margin	2.611	16.483	414131
LQ	1.246	.695	451154
Age	14.910	14.273	451130
Sales 10^10 EUR	.0007	.019	414235
People per m^2	.288	.525	451154

Table 3

Base models output.

VARIABLES	(1) Return on total assets	(2) Profit margin
LQ	-8.39***	-5.72***
	(1.13)	(0.96)
LQ^2	7.94***	5.39***
	(1.12)	(0.96)
LQ^3	-3.23***	-2.14***
	(0.47)	(0.41)
LQ^4	0.57***	0.37***
	(0.09)	(0.07)
LQ^5	-0.04***	-0.02***
	(0.01)	(0.00)
Age	-0.01***	0.02***
	(0.00)	(0.00)
Sales 10^10 EUR	3.84**	3.45**
	(1.93)	(1.55)
People per m^2	0.58***	0.71***
	(0.10)	(0.09)
Czech Republic_dummy	1.38***	0.06
	(0.35)	(0.29)
Finland_dummy	4.89***	2.86***
	(0.35)	(0.31)
France_dummy	0.16	-0.26
	(0.26)	(0.25)
Germany_dummy	2.71***	-0.54**
	(0.31)	(0.26)
Greece_dummy	-1.61***	-1.46***
	(0.29)	(0.29)
Hungary_dummy	-0.99***	0.17
	(0.31)	(0.29)
Italy_dummy	-2.60***	-2.08***
	(0.26)	(0.24)
Netherlands_dummy	0.41	-1.21**
	(0.79)	(0.60)
Poland_dummy	3.81***	0.62**
	(0.33)	(0.28)
Portugal_dummy	-3.88***	-3.61***
	(0.29)	(0.28)
Romania_dummy	4.49***	2.67***
	(0.30)	(0.27)
Slovakia_dummy	0.74	-0.38
	(0.51)	(0.40)
Spain_dummy	-3.36***	-2.83***
	(0.26)	(0.25)
Sweden_dummy	-0.46	-0.41
	(0.29)	(0.27)

Textiles	-1.05***	-0.57***
	(0.14)	(0.12)
Wood	1.10***	1.01***
	(0.15)	(0.13)
Paper and Printing	1.89***	2.49***
	(0.13)	(0.11)
Chemicals and Fibers	2.77***	2.53***
	(0.18)	(0.16)
Rubber and Plastic	2.20***	1.92***
	(0.15)	(0.13)
Non-Metallic minerals	1.94***	1.86***
	(0.14)	(0.13)
Metal	3.65***	3.31***
	(0.10)	(0.08)
Machinery	4.09***	3.45***
	(0.12)	(0.10)
Electrical and Optical	4.45***	3.70***
	(0.14)	(0.11)
Not classified	0.79***	0.75***
	(0.13)	(0.11)
Constant	5.78***	3.03***
	(0.48)	(0.42)
Observations	414,211	407,643
R-squared	0.026	0.019

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4

Model with the interaction of size and LQ included.

VARIABLES	(1) Return on total assets	(2) Profit margin
LQ	-8.40***	-5.73***
	(1.13)	(0.96)
Sales 10^10 EUR	6.55**	5.70***
	(2.79)	(2.06)
LQ-SALES Interaction	-1.04	-0.86**
	(0.64)	(0.44)
LQ^2	7.95***	5.40***
	(1.12)	(0.96)
LQ^3	-3.23***	-2.14***
	(0.47)	(0.41)
LQ^4	0.57***	0.37***
	(0.09)	(0.07)
LQ^5	-0.04***	-0.02***
	(0.01)	(0.00)
Age	-0.01***	0.02***
	(0.00)	(0.00)
People per m^2	0.58***	0.70***
	(0.10)	(0.09)
Constant	5.78***	3.03***
	(0.48)	(0.42)
Observations	414,211	407,643
R-squared	0.026	0.019
Country dummies	YES	YES
Industry dummies	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5

Model with the interaction of urbanization proxy and LQ included.

VARIABLES	(1) Return on total assets	(2) Profit margin
LQ	-8.10***	-5.41***
	(1.14)	(0.97)
People per m^2	1.14***	1.31***
	(0.16)	(0.14)
LQ-URBANISATION Interaction	-0.40***	-0.43***
	(0.09)	(0.08)
LQ^2	7.68***	5.11***
	(1.13)	(0.97)
LQ^3	-3.07***	-1.97***
	(0.48)	(0.41)
LQ^4	0.53***	0.33***
	(0.09)	(0.08)
LQ^5	-0.03***	-0.02***
	(0.01)	(0.00)
Age	-0.01***	0.02***
	(0.00)	(0.00)
Sales 10^10 EUR	3.82**	3.43**
	(1.86)	(1.48)
Constant	5.56***	2.80***
	(0.48)	(0.42)
Observations	414,211	407,643
R-squared	0.026	0.019
Country dummies	YES	YES
Industry dummies	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6

Extended base model with regional statistics added as robustness checks.

VARIABLES	(1) Return on total assets	(2) Profit margin
LQ	-3.09***	-2.38**
	(1.19)	(1.00)
LQ^2	3.07***	2.64***
	(1.18)	(1.01)
LQ^3	-1.34***	-1.18***
	(0.50)	(0.43)
LQ^4	0.25***	0.22***
	(0.09)	(0.08)
LQ^5	-0.02***	-0.01***
	(0.01)	(0.00)
Age	-0.01***	0.02***
	(0.00)	(0.00)
Sales 10^10 EUR	3.45**	3.27**
	(1.71)	(1.40)
Area in 10^5 km^2	0.33*	-0.03
	(0.20)	(0.18)
GDP in 1000 EUR per capita	0.08***	0.11***
	(0.02)	(0.02)
Average annual wage in 1000 EUR	-0.13***	-0.14***
	(0.03)	(0.02)
Distance to Seaport (1000 km)	-1.82***	-0.96***
	(0.36)	(0.29)
Long term unemployment rate (%)	-0.16***	-0.09***
	(0.03)	(0.03)
Percentage land use that is urban	0.17***	0.12**
	(0.05)	(0.05)
Constant	8.14***	5.34***
	(0.74)	(0.63)
Observations	341,749	337,523
R-squared	0.026	0.019
Country dummies	YES	YES
Industry dummies	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7

Base model run on the sample that excludes top and bottom 2.5 % of the distribution.

VARIABLES	(1) Return on total assets	(2) Profit margin
LQ	-9.50***	-6.54***
	(1.11)	(0.92)
LQ^2	9.00***	6.14***
	(1.11)	(0.92)
LQ^3	-3.65***	-2.43***
	(0.47)	(0.39)
LQ^4	0.64***	0.42***
	(0.08)	(0.07)
LQ^5	-0.04***	-0.03***
	(0.01)	(0.00)
Age	-0.04***	-0.00
	(0.00)	(0.00)
Sales 10^10 EUR	1,789.11***	1,290.15***
	(49.21)	(40.02)
People per m^2	0.67***	0.72***
	(0.10)	(0.09)
Constant	4.45***	2.24***
	(0.50)	(0.44)
Observations	393,498	389,456
R-squared	0.036	0.026
Country dummies	YES	YES
Industry dummies	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8

Controlling for industry heterogeneity using knowledge intensity dummy.

VARIABLES	(1) Return on total assets	(2) Profit margin
LQ	-8.19***	-5.50***
	(1.14)	(0.97)
Knowledge Intensity	4.70***	3.97***
	(0.18)	(0.15)
LQ- KnowledgeIntensity Interact	-0.22**	-0.24***
	(0.10)	(0.09)
LQ^2	7.88***	5.33***
	(1.12)	(0.96)
LQ^3	-3.22***	-2.13***
	(0.47)	(0.41)
LQ^4	0.57***	0.37***
	(0.09)	(0.07)
LQ^5	-0.04***	-0.02***
	(0.01)	(0.00)
Age	-0.01***	0.02***
	(0.00)	(0.00)
Sales 10^10 EUR	3.92**	3.53**
	(1.95)	(1.58)
People per m^2	0.58***	0.70***
	(0.10)	(0.09)
Constant	5.63***	2.87***
	(0.49)	(0.43)
Observations	414,211	407,643
R-squared	0.026	0.019
Country dummies	YES	YES
Industry dummies	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Graph 1

Illustration of relation between LQ and profitability indicators.