THE INFLUENCE OF SKILL RELATEDNESS ON COAGGLOMERATION

ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Author: A.E.H. Hendriks
Student number: 323369
Thesis supervisor: dr. F.M.H. Neffke
Date: 08-2011

Contents

Summary

Firms seek locations which gain the most possible advantages for their business. Firms which are working in the same industry seek for the same advantages. They tend to end in the same location, this creates agglomeration. Literature suggests that the most important advantages gained in an agglomeration are: the transportation costs of goods, skills and knowledge. Labor costs are becoming more important within agglomeration, because the transportation costs of persons is rising, while the transportation costs of goods is falling.

This article reveals the relationship between skill relatedness and coagglomeration. It shows that coagglomeration depends on skill relatedness. The reasons why agglomerations and coagglomerations arise is discussed in this article. It summarizes the advantages which are gained by firms who enter an agglomeration.

These advantages of an agglomeration also arise between firms of different industries. This is why coagglomerations exist. One of the factors why agglomerations exist is: skill relatedness. Firms use the same labor, because they are in the same business. If they are located near to each other they gain advantages. This happens with firms who have the exact same business, but it also happens when they don't. Firms in different industries locate near to each other because they use the same labor.

This paper uses a method to calculate if there are coagglomerations between industries (Ellison, Glaeser, & Kerr, 2010). This method returns an index which reveals if industries coagglomerate more than other industries. This paper discusses the method used and shows that there are some positive and negative side effects to this method.

Data of skill relatedness is used (Neffke & Henning, 2010), which is based of Swedish inter industry labor flows. A short review on the skill relatedness method is present in this article. The skill relatedness data is compared to the coagglomeration index based on Dutch employment data. It reveals that coagglomeration is partly based on skill relatedness and that the more industries are skill related the more they tend to coagglomerate.

Skill relatedness does not fully explain why firms coagglomerate, but this article reveals that skill relatedness is one of the coagglomeration reasons. It reveals that skill relatedness has the most influence on coagglomeration between service industry combinations.

1. Introduction

A shipbuilding company won't sell a ship out in the desert. And if it would, it can't easily be transported to the sea. Shipbuilding companies naturally choose a location close to water. In modern times not only the location of manufacturing firms are important. Service industries and headquarters of large companies also ask themselves what is the best possible location for their offices. The example of a ship building company shows one of the issues which are important in the location choice. This issue is about the transportation costs of goods. Literature suggests that there are more issues which need to be taken into account when a firm chooses its location. One of these issues is the labor force. It is important that firms to stay close to the labor force used by the firm. A company seeks a location close to the most important labor needed for their activities. If multiple companies use the same labor, they could start offices or factories in the same region. This is a result of the fact that they try to seek the best labor for their company. When multiple industries move to the same region they coagglomerate. In this article the relationship between coagglomeration and the use of skills is exposed.

Industries which use the same labor force are skill related. Neffke and Henning (2010) created a skill relatedness index, this index in calculated for all different pair industry combinations. The main question of this paper is: "Does the level of skill relatedness influence coagglomeration?". Literature suggests that skill relatedness causes coagglomeration within industries, this paper proves that this also happens between multiple industries. And it will conclude that skill relatedness has a positive influence on coagglomeration.

This paper consists of seven sections. The first section is a theoretical back ground on the reason why industries agglomerate. It shows the advantages which are gained by firms in an agglomeration. The second part elaborates coagglomeration. It explains what coagglomeration is, and how it appears. The third section elaborates the skill relatedness index, developed by Neffke and Henning. The data is described in the fourth part. In this chapter two methods are revealed which are used to calculate a coagglomeration index. This chapter also explains which data is used, and the most important statistics of the variables used are given in this section. The next part of the article will reveal the results. This is a summary of the most important findings of this article. Also the methods used to examine the data are elaborated. Fifth, the conclusions from the literature and data research are stated, this part is only a brief overview of the main findings of the article. The last part starts a discussion, this will show that further research is needed, to give better results. It also discusses some weaknesses of this article, just like some assumptions, which may not reveal the reality. The discussion mostly contains possibilities for further research.

2. Agglomeration of industries

In the past few decades, agglomerations of industries are an important topic for research. Areas like Silicon Valley for the IT industry, and London as a main area for the European financial industry, are agglomeration of a single industry. Authors have tried to explain why industries agglomerate, others focused on the consequences of agglomeration.

This chapter provides in insight on the basic reasons of agglomerations of industries. Agglomerations arise because advantages of knowledge, goods or labor skills which are gained at lower costs within the agglomeration. It will also show that natural advantages generate an incentive for firms to agglomerate.

Literature explains that agglomeration is caused by knowledge spillovers and natural resource (Boschma & Frenken, 2009). Spillovers occur more easily between industries which use the same technologies. These industries create spin offs, which are new firms in the same industry. This creates an agglomeration effect. The combination of these effects will lead to clusters on geographical areas with the most natural resources important for the particular industry, if there are no other factors of influence on the location of companies. Factors like customers, end users and path dependency are not added to this theory, so reality will probably show a different view.

Spin offs, mentioned by Boschma and Frenken (2009) create more competiveness and turbulence. Turbulence in the market leads to positive results in productivity, because resources are used in an improved way. This holds at least for the service sector (Bosma, Stam, & Schutjens, 2006). They noted that the labor productivity rises with the entry and exit of new entrepreneurs. Spin offs add value to the firms in the region.

Agglomeration provides advantages on transportation costs of skills, materials and machinery according to Marshall (1920), where the latter is known as technology or knowledge now a day. Advantages of skills are created by a shared labor pool of the agglomeration. Marshall notices that agglomerations enable the opportunity of knowledge spillovers. Advantages of materials arise because of low transportation costs of goods within local areas. Marshall even stated that the laborers move to agglomerations, because wages rise when labor is getting scarce. Other models under scribe, the finding that labor will move to an area where the real wages are the highest (Krugman, 1991). Thus a labor pool grows because more companies join the agglomeration which attracts laborers who move to the agglomeration side.

Not all industries tend to agglomerate. The retail industry does not agglomerate, looking at a regional level. Carree and Dejardin (2007) concluded that entry of retail firms is less in areas which contain a lot firms in the same retail branch at the moment. This indicates that the retail branch does not create advantages of agglomeration with industries of the same type. This may differ on scale. For example, if you take a look to neighborhoods in a city, you can find agglomerations, think of the shopping centers.

Maskell and Malmberg (1999) stated that it is important for firms to agglomerate with competition. This is the most important way to renew their knowledge and to perform well. The problem of gaining knowledge is that there are no certain results of an investment, but the location will influence some factor. First of all, proximity matters. This is most important for transferring tacit knowledge. Agglomeration provides the opportunity to transfer knowledge. And history matters, they explain this by stating that: ways how knowledge was gained in the past will probable used in the future. Knowledge will be gained by certain routines and processes.

The remaining of this chapter explains the sources of agglomeration. It will discuss knowledge spillovers, transportation costs of goods, labor pools and natural advantages. The first tree reasons where mentioned by Marshall while the latter is also used as a factor of agglomeration in a lot of literature.

2.1 Industries can gain natural advantages

The example in the introduction of this paper provides a good idea of natural advantages. A shipbuilding company will only start a firm close to water. Water is a natural advantage for this company. Natural advantages, in the context of agglomeration, are advantages of a certain area compared to other areas. These advantages may not be an advantage for all industries, but it will attract some industries. The advantages are based on natural resources. They can't be created by mankind. For example a ski resort is build more easily in the Alps then in the middle of the Netherlands. In the Alps there is snow in the winter, and the slopes of the hills are great for skiing and snowboarding. In the Netherlands on the other hand, an indoor ski track is needed, because there are almost none hills or mountains and there is seldom snow naturally.

The location of industries in influenced by natural advantages (Ellison & Glaeser, 1999), about 20 percent of the concentrations can be derived from natural advantages. The concentrations which are subject of research are agglomerations of a single industry type. Natural advantages are advantages for firms in the same industry.

2.2 Knowledge disseminate easier

The previous section shown that agglomeration is based on natural advantages, but there are also "unnatural advantages", created by mankind. One of these factors is the creation and distribution of knowledge. This section will elaborate the advantages of knowledge in an agglomeration. First explaining ways of how knowledge moves in time and space.

Patent citations where used as measurement tool of knowledge spillovers by Jaffe, Trajtenberg, & Henderson (1993). In this case researchers used patents to find out if they were cited more than average in the region of the patent subscriber. Evidence indicates that there is a relationship between the original patent location and the amount of times it is cited in the region. Even correcting, for industry focused regions, by using patents of the same industry, won't change the results. Thus the conclusion can be made that: knowledge contained by patents does not disseminate outside the region fast.

There is scientific evidence for the fact that knowledge spillovers appear more in local areas. This does not seem to be logical, because a patent exists of codified knowledge. In other words everybody should be able to reuse this knowledge, inside or outside the region. The results of Jaffe et al. (1993) suggest that the patents contain tacit knowledge. Tacit knowledge is hard to transfer between persons, because nobody can explain it. They either don't know exactly what they are doing or they don't have the words to explain what they are doing. Learning by doing is the best way to gain tacit knowledge according to the traditional theory of Polanyi (1966). If patents would contain some kind of tacit knowledge, then there is a reason why citations are not normally divided among other regions. This means that tacit knowledge is gained while inventing patents and this knowledge is only disseminated in the same region where the invention is created.

But if knowledge spreads only within locations, how does the knowledge spread outside of this region. First of all Jaffe et al. found evidence that the localization effect evaporates over time. They found that the effect of localization was the strongest on new patents. Eventually the patent will be cited relatively more and more outside it's originally region. They only used two cohorts to test, thus this evidence is weak. One of the reasons why it takes time for patents to be cited in other reasons could be, because it takes time for inventors in other regions to notice the patent. And use the information in their invention. In the region of the original patent there are friends and other persons who are related somehow to the inventor. They are probable able to reuse this knowledge in other inventions.

Agrawal, lain, & McHale (2006) proved that inventors are more cited in regions where they worked or lived before. They said that inventors, who move, still spread their new knowledge to the previous location. Regions, where an inventor has lived or worked, cite patents more than twice as much than other regions. Thus movers create a second reason how knowledge is able to travel. Social relationships are involved in the process of knowledge exchange. This is another reason why knowledge does not travel fast. It is more likely that social relationships exist between persons who live of work in the same region then persons who don't. Both papers suggest that knowledge spillovers appear when people are related to each other, in physical distance, or in their social network. Agglomerations create more relationships between inventors and workers. This leads to faster knowledge spillovers. The socials ties are important for dissemination of knowledge.

Now it is clear that it is hard to gain new knowledge. Gertler (2003) describes how tacit knowledge can be acquired, kept and Spread within organizations. He argues the fact that tacit knowledge can only be spread within regions. He state that tacit knowledge can disseminate through international organizations, if it is done the right way. The most important factor to spread this knowledge is by interaction of agents face-to-face. For example knowledge enablers can go to different locations in the organization. If they work long enough on this location, they have the chance to disseminate knowledge within this location.

According to Gertler it is possible to disseminate knowledge outside a region. But he also points out that the creation of new knowledge is only possible within a region. This means that new knowledge will only be created within an agglomerated region. But the agglomeration is not so important to disseminate knowledge.

This section shows that knowledge can spread within a local area faster than outside of this area. This creates an incentive for industries to agglomerate. Agglomerations are able to develop more inventions than average firms. There is a difference in codified and tacit knowledge, where codified knowledge should travel more easily, because it is understandable for everyone. New knowledge is important for companies, they agglomerate to gain new knowledge. By agglomerating employees create social ties with employees of related firms. By these social ties new knowledge can be gained.

2.3 Transportation costs drop

Marshall did not only mention the transportation costs of knowledge as an important factor for agglomeration. He mentioned transportation costs of goods as well. Industries, of which the price of the final goods is substantially based on raw materials or intermediates, depend on transportation costs of goods. Competition in these markets creates the incentive for companies to search for lower costs. Let's assume that these industries don't use specific labor skills or knowledge to create a competitive advantage. Then they try to reduce the cost of their product by reducing transportation cost.

Transportation costs depend on the geographical position of the suppliers of the firm. Firms try to position their factory close to the suppliers. But they also need to take in account that there will be transportation costs of their products to the customers. So if customers are all close to each other and the suppliers are spread all over the world, it could be the best option to locate near to the customers instead of the suppliers.

Krugman (1991) created a model containing 2 regions, where there was only manufacturing labor (workers) and peasants. He showed that transportation cost of goods influenced the division of workers within the regions, and consequently the division of manufacturing firms. Low and high transportation cost create a total agglomeration in one region, this creates cities, but in between there is a large area where workers are divided between the two regions. He took not only transportation cost into account, but also price elasticity of the manufactured goods and the share of income spend on manufactured goods. This model suggests that labor and firms agglomerate when the transportation costs are low or high. Note that this only holds for manufacturing industries

This model Krugman used is simplistic, and makes a lot of assumptions. It does not show how the real world looks like, but it shows the importance of transportation costs in the creation of agglomerations. The main contribution to the literature contains the conclusion that there is a realistic influence of the transportation cost to the division of labor. This creates an incentive for labor pools to arise. The model could imply that the drop of transportation costs which is ongoing last century creates more opportunities for agglomeration. The model is abstract, so reactions of the real world could differ with the findings of Krugman.

Advantages of transportation costs have a long history. In the United States before 1900, large cities became main hubs for transportation. Because of the great advantages in transportation, these cities became attractive for manufacturing plants. These plants moved to the big cities to gain the advantage of transportation (Glaeser & Kohlhase, 2003). But with the decline of the relative transportation costs compared to the total price of goods, the question remains, will these costs be as important as they have been in the past. Glaeser & Kohlhase argue that this won't be the case. They even extend this vision and assume that in the future transportation costs of goods will be neglect able.

If this happens then there is no raison for industries to move to a location close to their suppliers, so manufacturing industries leave big cities. Face-to-face contact is important for service industries. These industries tend to move into cities. Being close to the labor force will be much more important in these industries. Glaeser & Kohlhase under scribe this view. They even calculated that the transportation costs of persons have been rising past decades. And because wages are rising and congestion increases, they assume that the transportation cost of persons rises in the future. This all together will make transportation cost of skills more important than the transportation costs of goods.

2.4 Agglomerations use a labor pool

Transportation costs of goods are important for manufacturing industries, but they are becoming less important. Service industries don't depend on transportation costs of goods. For these industries the proximity of labor skills and knowledge is much more important.

When industries agglomerate they make use of the same labor force. This labor force creates positive externalities for all firms which use this labor. Laborers can gain skills at one firm and later move to another. The skills which are acquired by the laborer can be used in the new firm.

Peck (1992) shows four different advantages of a labor pool: first, labor turnover rates are positively correlated with the market size. Big labor pools increase the amount of job switches. Second, employees are attracted to agglomerations, because there are more job opportunities. The agglomeration attracts new labor and this again attracts more companies. Third, the cost for finding and searching new employees will fall for companies. Labor is always near to the company's geographical location. Last, the labor force creates a new social environment. In this environment norms and values will become a common good.

These are the main advantages of a labor pool within an agglomeration. In literature these advantages are called the externalities. Next to these externalities, which indicate why labor pools arise, flexibility is a main advantage of companies to move their firms to a labor pool. This advantage is connected to the low search cost and the high amount of labor turnover rates.

Companies need to become more and more flexible. This means that the companies should be able to addapt changes in the environment fast. New technologies are important to create relative advanages over competitors. Skills can be gained easely by acquiering new employement. Within a laborpool it is attractive to hire temporary employees. They learn in other companies and use their skill temporary to learn the employees who work permanent in the company. Skills are important in all industries but most important in service industries. Frims need to try to acqire new skills, because they sell skills instead of goods.

2.5 Conclusion

Firms are not alone, relationships arise between firms in the same industry. Firms in the same industries use eachothers knowledge, skills and/or goods. But they pay transportation costs for all these factors. It is important for firms to reduce these costs, to gain advantages. firms locate near to firms in the same industry, this creates agglomerations. Knowledge is important for industries, but it is hard to measure knowledge gains. Transportation costs of goods have had a major influence of agglomerations past centuries. But it seems that these costs are declining. Thus they are losing relevance. Skills becoming more and more important. New agglomerations arise because companies share the same skills. Creating a labor pool creates advantages. In the future these labor skills remain to be a substantial factor for industries.

This chapter elaborated the factors of agglomeration. The main point of this chapter is that agglomeration creates advantages. Agglomeartions arise because of related industries, the factors of agglomeration provide different types of industry relatedness. Thus the advantages can differ between industries. Labor skills are becoming more imporatant at the moment.

3. Coagglomeration

The previous chapter showed reasons why industries agglomerate. It concluded that agglomeration appears because of advantages which are gained by firms in the agglomeration. In this chapter coagglomeration will be elaborated. What is coagglomeration? What are consequences of this coagglomeration? And how coagglomeration can be calculated? This chapter finally explains a coagglomeration index, which is used in the data examination of this paper.

Coagglomeration is an agglomeration of multiple industries. Firms in the same industry agglomerate because they gain scale advantages. But a coagglomeration consists of multiple industry types. So what are the advantages which can be gained in a coagglomeration? According to recent literature these advantages are not so different from the advantages of an agglomeration.

Ellison, Glaeser, & Kerr (2010) have extended the theory of Marshall and searched for the reason of coagglomeration, they found that reasons would differ between different industry combinations. For example: if industries depend on transportation cost of goods, they expect that the industry coagglomerate with suppliers and consumers. For all industries they calculated covariates, which indicate the importance of goods, skills and technology for these industries. They found that all covariates, including the skill relatedness covariate, had a significant effect on the type of coagglomeration which appears. The covariate of goods has the most influence on coagglomeration between companies which use the same goods. Input output tables were used to prove this relationship. One of the reasons that industries are more likely to measure the benefits of coagglomerating with suppliers and consumers could be that transportation costs of goods are fairly measurable. Knowledge spillovers are much harder to measure and covariates could be biased. Input output tables are more robust then the covariate of knowledge spillovers. Those are measured by patent citations. Patent citations can be incomplete or patents can cite other patents which didn't have a real contribution to the patent. (Jaffe, Trajtenberg, & Henderson, 1993) So more spillovers are indicated then the amount of patents which really provided the spillover of knowledge.

This means that skills are not unique in one industry, instead certain industries use the same kind of labor. This is an indication that industries are related to each other. Marshall (1920) already noted skills as one of the reasons for agglomeration. But as Ellison Glaeser and Kerr (2010) discover, it is also a reason for coagglomeration. Coagglomerations exist because of industries are using the same labor force.

Using one big labor force of multiple industries will increase the opportunities for innovation (Boschma & Frenken, 2011). Knowledge spillovers occur more between agents with a limited cognitive distance, agents need to understand each other. But the cognitive distance shouldn't be too large either. This blocks further innovation, because agents don't understand each other. Thus creating a labor pool of skill related labor will get the best effect if this is done by multiple industries, which use similar labor but not the same (Nooteboom, 2000).

Neffke, Henning and Boschma (2009), proved that firms which enter a region are mostly related to the current industries. Firms who leave the region are probable unrelated to the relative important industries of the region. They also indicated that new growth paths depend on the history of a region. They suggest that just like an agglomeration of one industry, a coagglomeration attracts related firms, and a coagglomeration grows, meaning that more related firms join the agglomeration.

3.1 Relatedness of industries

Geographical coagglomeration exists according the theory, this is proven by Ellison, Glaeser and Kerr (2010), but how do they measure this? Earlier literature used two kinds of codes to make a distinction between multiple industries, "Nomenclature statistique des Activités économiques dans la Communauté Européenne"(NACE) and "Standard Industrial Classification"(SIC). Every firm in Europe has a NACE code. These codes are constructed hierarchically. So from top down, all categories contain fewer industries and are more specific. These kinds of codes can be used to define a relatedness index. But this does not always provide good information on the relation between industries, if you compare this with relatedness based on skills, goods or ideas. You would expect that categories with the same parents are more related than categories with the same grandparents. But it occurs that related industries haven't got the same root level, which would indicate that they are not related at all. For example manufacturing and repairing motor vehicles. You would expect that a car mechanic is able to build a car, and vice versa. The same skills and goods are used in industries which differ on root level.

Within recent literature new ways to measure relatedness are created. Neffke and Henning (2008) showed relatedness by using the amount of production facilities which produced multiple products. They compared this with a predicted number of co-occurrences. This method shows other relations, which won't be revealed using NACE codes. With this method of relatedness, more evidence is found that related industries coagglomerate. A plant uses resources optimal, products manufactured in one plant use some of the same resources at least.

3.2 coagglomeration index

Another contribution is made by Ellison and Glaeser (1997). They created a method to calculate concentrations and coagglomerations of industries. They divide industries using the NACE codes. But they do not use the NACE code to explain the relation between industries. One of the main eye openers of this article is that clusters do not only arise within one industry, but they contain multiple industries. It is there for important to gain information about which industries are related and cluster together.

The main reason which Ellison and Glaeser use as explanation for coagglomeration is natural resources. Just like single industry agglomeration, where one industry gain advantages of natural resource. Multiple industries gain advantage for of the same resources. For example, both bananas and coffee need a tropical climate to grow. So firms which produce these goods probable locate in the same regions. Ellison and Glaeser suggest that if all firms in one industry choose the location to maximize their profits. They will end up in clusters near to natural resources. Ellison and Glaeser suggest that an industry fully agglomerates in one region. When there are regions which contain the same amount of natural resources, then it is uncertain where the cluster will arise. The location of the first company is important for the location of the cluster, others will follow. This pattern is called a dartboard. Clusters are spread randomly on the map, if there are no natural advantages.

Industries won't coagglomerate totally with another, but some companies of an industry coagglomerate, other companies coagglomerate with other industries. This paper clearly stated that the evidence of coagglomeration is only weak, there is no strong relationship between one industry and another. A reason for this could be that: firms are not always aware of related industries. Companies may not be familiar with all other industries, and may not know that some industries are related. If they don't know this they won't use these industries to gain new knowledge or other advantages. These situations prevent clusters form arising. Other reasons can be found in an uncertainty of knowing what will be the best cluster.

Ellison and Glaeser created a method to calculate a coagglomeration index, to show which industries coagglomerate. The method estimates a variable which indicates the power of natural resources and spillovers in the region for the type of industry. A value of 1 means that all industries depend on the natural resources and spillover effects, while a value of 0 implies that there is no effect at all. To calculate this value, data on all industries are needed. Next to this an overview of amount of plants, amount of jobs in the industry and plant sizes are needed. One of the problems using this model is that not all data is available. It is hard to find data on all plant sizes for example.

Another weakness is that this paper only shows coagglomeration of industries, it does not explain why some industries coagglomerate. The paper suggests two main reasons for coagglomeration: natural resources and spillovers. But the index calculated does not make a separation between those factors. It measures the consequence, or the so called coagglomeration, but the paper is not clear on the reason why certain industries coagglomerate and if the industries are related to each other. The paper lacks information about history of regions. History is an important aspect for current industries, because it enables or disables the opportunity for other industries to develop in a region. The current industry portfolio of a region has a significant influence on the future industry portfolio of the same region (2007). The last weakness of this article is that the evidence of coagglomeration is weak. There is no proof that industries coagglomerate because of the spillovers or natural resources. In the real world a lot of different aspects may influence the location decision of a company.

In 2010 Ellison, Glaeser and Kerr redesigned the coagglomeration index. They used it to tell if coagglomerations are influences by the same reasons as agglomeration. And found proof that this is the case. Now that this is known, the effect of these influences should be further investigated.

This article further develops this and tries to find a relationship between skill relatedness and coagglomeration. Later in this paper the method created by Ellison, Glaeser and Kerr is used to calculate coagglomeration indexes. It provides insight in which industries coagglomerations together. To overcome the weaknesses that this paper does not provide a reason of coagglomeration, this index is used as a consequence of relatedness, using skill relatedness as input. Skill relatedness is measured and explains coagglomeration.

3.3 Conclusion

This chapter gives insight in coagglomerations. These agglomerations of multiple industries arise because advantages gained. The same advantages which are gained by agglomerations are gained by coagglomerations. Natural advantages, goods, skills and knowledge are an important factor for the creation of coagglomerations. There are multiple methods to reveal the relatedness between industries. Already some of these methods are used to prove that the advantages of coagglomeration are the same as agglomeration advantages. Coagglomerations can be measured using different tools and methods. But these methods do only provide explanations why coagglomerations exist. It does not show the influence of one of these factors on coagglomeration.

4. Skill relatedness

The first chapter showed that agglomeration appears because of advantages in transportation costs. The second showed that coagglomerations arise because of the same reasons, but there is no method to quantify the influence of these reasons. Coagglomeration can be explained using one of these advantages for example skill relatedness. Firms which are skill related use the same kind of labor, they create an advantage if they take labor out of the same labor pool. Transportation of persons is becoming more expensive (Glaeser & Kohlhase, 2003), so skill relatedness should get more influence on coagglomerations. This chapter explains what skill relatedness is and how it can be quantified. It explains a methodology used to calculate skill relatedness. It provides some advantages and disadvantages of this method. In later chapters this index is used to explain agglomerations.

Skill relatedness means that companies are related because they use the same kind of laborers. The activities of the companies create jobs which fit the same labor force. This does not mean that two companies can exchange all their employees, but some employees which are important for the core business of the companies can be exchanged. For example a welder can work in the metal industry, but the same employee could easily find a job in the ship building industry, because welding is an important part in shipbuilding. This example implies that companies in the metal industry are skill related to shipbuilding companies. Another example is a nurse who can work in a hospital, but also in a retirement home.

When companies hire new employees which don't have new knowledge compared to the existing employees, then the productivity growth is negative. Instead, when a company hires new employees with new related knowledge and/or skills, then the productivity of the company is positive (Boschma, Eriksson, & Lindgren, 2009). Companies benefit hiring employees from related industries, and they are better off when they don't hire employees from companies with the exact same business. Coagglomeration creates labor pools of skill related workers.

Other studies pointed out which companies use the local skills most. Entrepreneurs use the local skill resources more often than expanding companies. This is because entrepreneurs seek the best region for their business. They move their business to the labor pool, while larger businesses use more of the labor available in the company. Large businesses also invent more new technologies, while entrepreneurs use the technologies already available in the region. (Neffke & Henning, 2010)

Labor can and will be exchanged within a group of companies and this has a positive effect on firms. The effect is best when related industries exchange personel, so the assumtion, that cross industry laborflows reveal the relatedness of industries, can be made. Neffke and Henning (2010) created a method to turn this realtionship into figures. They used swedish data of industry laborfows to calculate an index of skill relatedness of all industries. This index is a comparisson of the predicted cross industry laborflow between industries and the real labor flow. The predicted cross industry labor flows is based on a model which takes in account that some industry combinations do not have inter industry laborflows at all. An index above one reveals relatedness of skills between the industries. A index below one implies that there is a negative skill relatedness effect between the industries.

In their method Neffke and Henning left out general labor skills like management, because management is not bounded to a certain industry. All industries need them, so management turnover is no indication of relatedness. An other factor which is corrected for in the data is labor which move to new plants. This could be a concequence of the agglomeration, and does not reveal the skilled relatedness.

This method has also some negative sides. Cross industry labor flows do not always reveal relatedness. Workers could search for a job within a region, because they do not want to move to another region. If this has happened too much, it will bias the data. And it shows that there is a lot of labor spillover within a region, compared to outside the region. This would imply that related industries coagglomerate, while this is a consequence of industries being in the same region. Note: Workers tend to search for jobs mostly related to their old jobs. They want to reuse their current skills as much as possible, because they will earn higher wages (Neal, 1995). Cross industry labor flows in a region develop between most related industries available.

The size of an industry in Sweden could influence the relatedness factors. Small industries could easily be a positive or negative outlier. Other problems may arise if some industries are going through difficult times. They could fire a lot off employees. The employees try to find related jobs and not always succeed, because there are too much laborers with the same skills on the market. Then these laborers will find less related jobs, which would bias the data.

In this paper the index which is created by Neffke and Henning is used to compare with the coagglomeration indexes. This comparison provides evidence if relatedness of skill creates coagglomerations.

Conclusion

This chapter showed that relatedness can be revealed using cross industry labor flows. Laborers tend to move to jobs which are related to their current jobs, because they will earn the most. An index crated by Neffke and Henning gives important insights in skill relatedness of industries. This index compares the expected labor flows with the real ones. The index and the method used for this have cons and pros. But it is used to compare with agglomeration figures in the next chapters.

5. Data and methods

This chapter shows how to make a connection between theory and practice. First of all a hypothesis is added which is tested later in this paper, further the data which is examined is described. The main methods used to calculate coagglomeration are described. This chapter helps to understand how the results of this paper originate.

While the previous chapters gave an overview on relatedness and coagglomeration, the next two chapters will add new scientific evidence for the relationship between skill relatedness and coagglomeration. The hypothesis which can be concluded from the theoretically background is: *Skill relatedness positively influences coagglomeration*.

To prove this Dutch data of industry employment is used to calculate coagglomeration indexes. Also data containing skill relatedness index is used, to see if skill relatedness influences coagglomeration. The Dutch data contains 40 regions, which together cover the whole surface of the Netherlands. For all these regions the data contains 184 industries of different 3 digit NACE codes. The amount of jobs in 1996 and 2002 of all industry region combinations are present in the data. So in total the data consist of 184*40 observations with employment rates of 2002 and 1996. Next to this data which is used to calculate a coagglomeration index, more data is used containing all industry combinations of the 184 industries. All industries are combined with all other industry types, all these industry combinations contain a skill relatedness value. This leads to 184*183 observations with the skill relatedness index. This data is calculated by Neffke and Henning (2010) based on Swedish employment data.

Not all 3 digit NACE code categories are taken into account. Most agricultural industries are left out, because agricultural industries can't settle their plants everywhere they want. They use the space which is not used by other industries or by housing, this makes it difficult for these industries to agglomerate. The industries left out of the research would bias the results, because of this reason.

To calculate the coagglomeration index, using the method created by Ellison and Glaeser (1997), figures of industries in multiple regions are necessary. For each region all plants, with their industries, and the industry size in amount of jobs are required. In the original article Ellison and Glaser already suggest that it is hard to acquire data on plant size distributions. This is a main indicator of the concentration index. The data used in this article didn't provide plant level figures. In 2010 Ellison Glaeser and Kerr created a new method to calculate the coagglomeration index, this method didn't need exact plant information. Because this data was not available, the method of 2010 is used in this article. The coagglomeration index should not differ from the method developed by Ellison and Glaeser in 1997.

The method used controls for differences in industry and region size. Tough some facts of the distribution of the data could be useful to clarify the results. The aggregated data of regions show that the data is right skewed, with a mean regional employment rate of 149000 jobs in 1996 and 179000 jobs in 2002. The data is even more right skewed when it is aggregated on industry. Most industries consist only of a small share of the total employment, where a fraction of the 184 industries are responsible for a large amount of jobs. The median of the industry figures show an employment rate of 11000 employees in 1996 the maximum value was 437000. The appendix contains more figures of the distribution of the data. The data contains some industries which are relative small. The coagglomeration index of these industry combinations could be biased. Table 1 contains an overview of all descriptive statistics.

Table1. Overview of industry and region employment sizes.

	N	Mean	Median	Std. Dev.	Min	Max
Total region 1996	40	178786.2	101340	155851.1	17110	684854
Total region 2002	40	149402.8	120957.5	126113.9	17497	538961
Total industry 1996	184	32478.88	32478.88	59159.86	0	436731
Total industry 2002	184	32478.88	32478.88	59159.86	0	436731

The largest industries of the Netherlands consists mostly service industries. In the top 5 of largest industries only the "Building of complete constructions or parts thereof; civil engineering" industry is non service. The service sector in the Netherlands is relative large compared to the manufacturing industries. Literature has shown most agglomeration effects and relatedness effects on these manufacturing industries. If the service industry is less sensitive for the effects mentioned in the previous chapters, the results could differ from what is expected. With the method used in this paper it is easy to separate service from manufacturing industries. This separation proves if there is a difference in effect between service and manufacturing industries.
The skill relatedness variable is also right skewed. More than half of the data contains 0 values; this happens when no cross industry labor flows has taken place in the period of measurement by Neffke and Henning. The maximum value is 252.9, so there is a lot of difference between low and high values. Unlike the industry size variable, the size of the skill relatedness index does matter. Because the main purpose of this paper is to find out the relationship between the size of the skill relatedness index and coagglomeration index,

5.1 Coagglomeration index

This section elaborates how this index can be calculated. The original method is designed by Ellison and Glaeser (1997). This is a complex method, which is equivalent to the method used in the research. This original method was the first definition and is refined in 2010. This method is elaborated to enlarge the view on coagglomeration. γ^c is the coagglomeration index. i indexes the region and j the industry. Let w_j be the employment share of the jth industry. $\tilde{\gamma}_j$ is the concentration index of industry j. Coagglomeration is calculated using the following method:

$$\gamma^c = \frac{[\frac{G}{(1-\sum_i x_i^2)}] - H - \sum_{j=1}^{r} \tilde{\gamma}_j w_j^2 (1-H_j)}{1-\sum_{j=1}^{r} w_j^2}$$

Where $G = \sum_{i=1}^{I}(s_i - x_i)^2$ and $s_i = \sum_{j=1}^{J} w_j s_{ij}$. s_i indicates the total employment of industry j in all sub areas together. H_j is the Herfindahl index of industry j and $H = \sum_i w_j^2 H_j$. This coagglomeration index controls for distribution of the industry across the country and for the size of the industry.

First sum of all the relative industry size is calculated compared to the total region size. This is divided by the relative size of all regions. If the size of the regions is equally divided among all regions, this leads to a smaller coefficient. The coagglomeration coefficient rises if the employment share of the industries differs.

The method used in this paper is less complex. This method is also developed by Ellison and Glaeser together with Kerr (2010). They argue that coagglomeration is measurable without the data of the plant distribution. The only variables they use are s_i and x_i. This data is available for the Netherlands. And the results should be equivalent to the complex method. The formula to calculate the coagglomeration index is the following:

$$\gamma^c = \frac{\sum_{i=1}^{I}(s_{m1} - x_m)(s_{m2} - x_m)}{1-\sum_{i=1}^{I} x^2}$$

This method uses the relative size of the industries, in all regions. When industries are equally spread among the country, higher coefficients will be the result. If one region is relative large in total employment size compared to other regions, then the coefficient will be higher.

5.2 Conclusion

This chapter gave insight in the data used to examine the effects of skill relatedness on coagglomeration. The main variables used are employment numbers of multiple industries in 40 Dutch regions. Statistics of the most important variables are added in this chapter. The chapter explained how to calculate the coagglomeration index. The coagglomeration index can be calculated using two methods. Both methods are discussed in this chapter, because one method was used to create the other.

6. Results

Until this point the relationship between coagglomeration and skills is only described theoretical. In this section elaborates the results of tests, used to prove the hypothesis. The results show that there is a significant relationship between skill relatedness and coagglomeration. In the appendix all test results can be found which are not added to this chapter, the results of these tests did not provide new useful information.

The coagglomeration index can be calculated in using two methods. A method described in 1997, using the Herfindahl index, and a less complex method developed by Ellison Glaeser and Kerr in 2010. The second method is used, because the available data gave the opportunity to use this method. And not all data needed for the first method was available.

Both methods lead to the same coagglomeration indexes, so it doesn't make any difference for the final results which method is used. The correlation between the coagglomeration index and the skill relatedness index is 0.12 for both cohorts. Spearman rang correlation coefficient leads to almost the same result. A value of 0.13 is more or less the same as the Pearson correlation value. The difference between the two correlation coefficients is that the spearman uses the order of the observations, while the Pearson correlation uses the absolute value. The results of the spearman test reveal that skill relatedness and coagglomeration are dependent, concluding that there is significant evidence for a relation of skill relatedness with coagglomeration.

Table 2 provides an overview of the descriptive statistics of the coagglomeration indexes of 1996 and 2002($\boldsymbol{\gamma^c}$ 1996, $\boldsymbol{\gamma^c}$ 2002), the skill relatedness (SR) and the transformed skill relatedness, which will be elaborated later this chapter. All these variables are also transformed to a standard deviation scale, the statistics of these values are also added to the table.

Table2. Descriptive statistics

Variable	N	Median	Mean	Std. Dev.	Min	Max
γ^c 1996	33306	.0004685	-.00038	.0184668	-.0756567	.4132963
γ^c 2002	33672	.0007865	-.0001197	.0193904	-.090303	.4229152
SR	33672	106.269	0	4.936.654	0	2.528.636
SRT	33672	-.5787397	-1	.5780755	-1	.9921218
SD γ^c 1996	33306	-.0207447	-.0599632	.8536082	-3.539.544	1.906.176
SD γ^c 2002	33672	-.0198921	-.0594007	.845358	-39.911	1.838.356
SD SR	33672	-.0031029	-.2041511	.9339552	-.2041511	4.763.459
SD SRT	33672	.0173344	-.7128851	1.002.045	-.7128851	2.740.291

To reveal and quantify this relationship a regression analysis is executed for both cohorts. The coagglomeration is the dependent variable and SR is the independent. Both regressions return a significant positive coefficient. See the left panel table 3 and 4 for the exact figures. The results imply that the more skill related industries are the more they coagglomerate. This regression analysis reveals that the hypothesis is true. Data gives significant results on a positive relation of SR on coagglomeration. Figure 1 and 2 show the regression analysis on a scatter plot. The scatter plot is not clear because the data is right skewed.

Figure 1. Coagglomeration 1996 size by SR

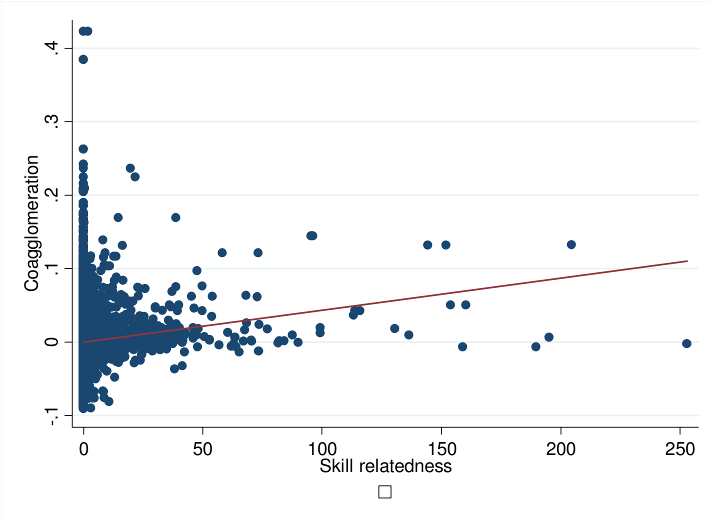


Figure 2. Coagglomeration 2002 size by SR

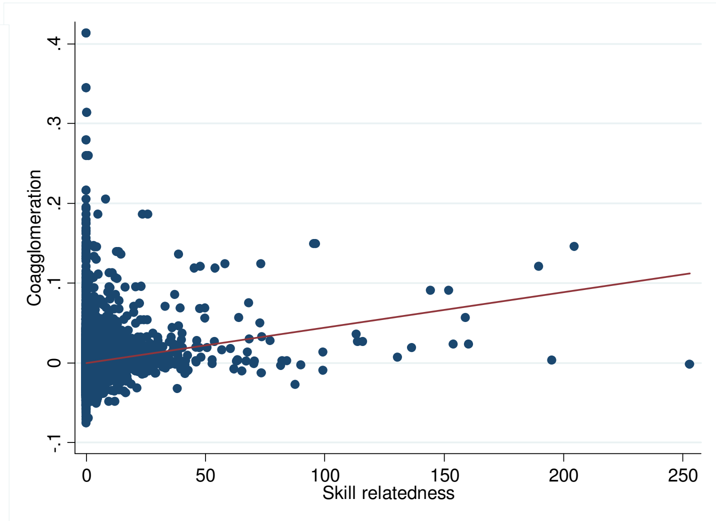


Table 3. Regression analysis of coagglomeration 1996

Coagglomeration 1996	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR(T)	4.42 e−04	0.000	2.98 e−03	0.000
Constant	−4.06 e−06	0.968	2.18 e−03	0.000
R Squared	0.014		8.7 e−03	
N observations	33306		33306	

Table 4. Regression analysis of coagglomeration 2002

Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR(T)	4.34 e−04	0.000	3.27 e−03	0.000
Constant	3.25 e−04	0.002	2.68 e−03	0.000
R Squared	0.012		9.5 e−03	
N observations	33672		33672	

6.1 Transforming skill relatedness

The previous chapter mentioned that the SR data is very right skewed, and contains a lot of values which are 0. The coagglomeration index has a theoretical maximum value of 1, which creates very small coefficients. Without the 0 values of SR, half of the values lie between 0 and 1 and the other half above 1. The values between 0 and 1 mean that there is negative skill relatedness, while values above 1 suggest positive SR between industries. Because the data is divided this way, a transformation could be made, which divide the data between -1 and 1.

This transformation makes the differences between the largest and smallest SR values less, because all values of SR are positive or 0. An easy method to create an index which contains values between -1 and 1 is. To calculate a new variable, named SRT which contains the value (SR-1)/(SR+1). The result is a variable with the min of -1 and a max of 0.992 using the available data. Using this new variable in a regression reveals the relationship more clearly, because the range of the independent variable is much smaller than in the previous regression and the minus sign is meaningful. All values between -1 and 0 reveal negative skill relatedness while the values between 0 and 1 reveal positive skill relatedness. Thus negative values reveal negative skill relatedness where positive values reveal positive relatedness. Next to this the steps are more equally divided, meaning the scale of SRT is more linear between negative and positive effect.

Figure 3. Coagglomeration 2002 size by SRT

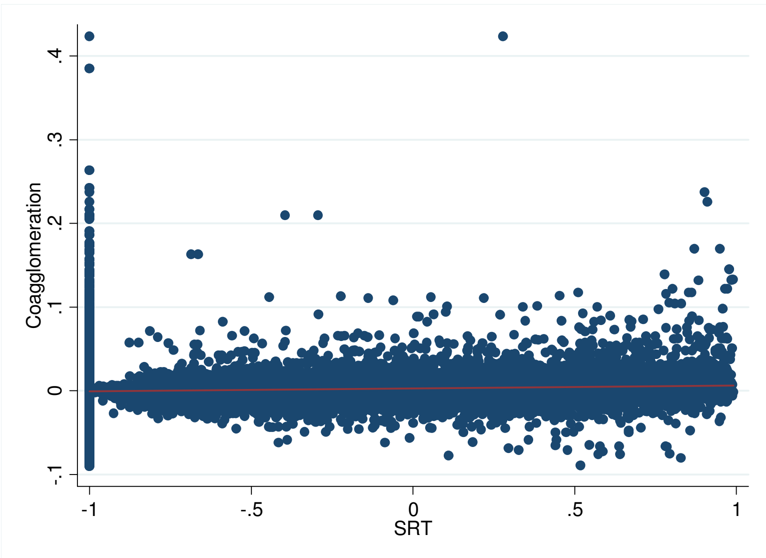


Table 1 contains descriptive statistics of SRT. A regression, of the coagglomeration indexes with the newly calculated transformed variable, leads to larger significant coefficients. This is what is expected, and makes the statement that coagglomeration dependents on skill relatedness more robust. The right panel of Table 3 and 4 show the results of the regression. Figure 3 shows a scatter plot of SRT versus coagglomeration. The red line shows the regression line, which hardly comes above the 0 line.

While it looks like these results are very clarifying, the note should be made that all the previous regression analysis result in a low r squared. The model does not explain coagglomeration very well. This result is not unexpected. Literature shows multiple advantages can be gained by location choice. This paper tries to explain this only using SR, but more factors are important to explain all coagglomeration effects. Another reason already mentioned that not all companies are aware of which industries are (skill) related. Unaware companies won't move to the (skill) related industries. Thus not all effects are used by firms.

6.2 Quantify the regression analysis

The regression analysis until so far show the influence of SR on coagglomeration, but it is difficult to estimate the influence of SR exactly, because indexes are used. To give more insight in the size of the relationship between skill relatedness and coagglomeration standard deviation scales are used. How are these scales calculated? First of all the mean and the standard deviation of a variable is calculated. Then the following transformation is executed (value-mean)/standard deviation. The result is a variable, which indicates the division of the variable. Only 15.87 % of the values are larger than 1 of smaller than -1 and 2.28 % is larger than 2 or smaller than -2. With this method it is possible to give a better indication of the impact of skill relatedness. Variables on the standard deviation scale have SD in front of their name, so SD SR is the standard deviation variable of the skill relatedness variable. And SD SRT is the standard deviation scale of the SRT variable.

Table 5. Regression analysis of standard deviation coagglomeration 1996

SD Coagglomeration 1996	SD Skill relatedness index		SD Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SD SR(T)	.108	0.000	.079	0.000
Constant	-.021	0.000	-.022	0.000
R Squared	0.014		8.7 e-03	
N observations	33306		33306	

Table 6. Regression analysis of standard deviation coagglomeration 2002

SD Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SD SR(T)	0.100	0.000	.082	0.000
Constant	-.020	0.000	-.021	0.000
R Squared	0.012		9.5 e-03	
N observations	33672		33672	

Table 5 and 6 show a Regression analysis of the standard deviation attributes. Both dependent and independent variables are transformed on the standard deviation scale. This shows that if the SR value is one standard deviation above average, then the coagglomeration would be 0.1 times standard deviation above the average coagglomeration index. This reveals that the impact of SR on coagglomeration is only small, because only 16 percent of all observations contain an SR value of 1 or more.

6.3 Differences between manufacturing and service industry

So far most literature concentrated on manufacturing industries. Skills may not be the most important factor of manufacturing industries. Service industries depend on skills and knowledge, they do not sell goods. It is interesting, if there are differences between the coagglomeration of manufacturing industries only, service industries only and combinations of service and manufacturing industries, because service industries are depend less on goods and more of skills.

Table 7 contains a regression analysis of the manufacturing industry only. So the data is filtered that only results of combinations between manufacturing industries are taken into account. This regression uses also the standard deviation variables. It seems that the coagglomeration of these industries is less dependent on SR than when the all data is used. This would indicate that Manufacturing industries are less dependent of SR. The same results are found using a combination of service and manufacturing industries. The effect is only marginally larger than the effect in the manufacturing industries.

Table 7. Regression analysis of standard deviation coagglomeration 1996 only manufacturing industries

SD Coagglomeration 1996	SD Skill relatedness index		SD Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SD SR(T)	.075	0.000	.055	0.000
Constant	−2.16 e−4	0.985	−4.31 e−4	0.970
R Squared	5.8 e−03		3.1 e−03	
N observations	7832		7832	

Table 8. Regression analysis of standard deviation coagglomeration 1996 combinations of service and manufacturing

SD Coagglomeration 1996	SD Skill relatedness index		SD Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SD SR(T)	.089	0.000	.055	0.000
Constant	−5.79 e−05	0.994	−2.10 e−04	0.979
R Squared	7.9 e−03		3.0 e−03	
N observations	15842		15842	

Table 9. Regression analysis of standard deviation coagglomeration 1996 only service industries

SD Coagglomeration 1996	SD Skill relatedness index		SD Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SD SR(T)	.161	0.000	.139	0.000
Constant	1.96e-10	1.000	1.04e-09	1.000
R Squared	0.026		0.019	
N observations	7832		7832	

A real difference is shown when the influence of SR on coagglomeration of service industries is regressed. This reveals a significant difference with the regression of all industries. Instead of the manufacturing industries, the SR has more influence on coagglomeration of service industries. Table 9 shows the results of regression values of the 1996 cohort, but the same results appear if the 2002 cohort is used.

Not only the effect of the service industries is the same in both cohorts, but also the manufacturing and the combination between manufacturing and service industries reveal the same results. The separation between manufacturing industries and service industries is done by NACE codes.

6.4 Conclusion

Test between skill relatedness and coagglomeration show that skill relatedness influences coagglomeration positive. This relation is shown by regression analyses. Regressions are executed with absolute and transforming the data. The transformations create a better view of the size of the influence of skill relatedness. It is clear that the total influence of skill relatedness is only minor, but there are more relatedness factors influencing coagglomeration. This is all in favor of the hypothesis. The results show that skill relatedness positive influences coagglomeration.

The effect of skill relatedness is larger between service industries than between other combinations. This suggests that skills are more important in the service industry than it is in other industries.

7. Discussion

This paper describes the possitive effect of skill relatedness on coagglomeration. There are still some weaknesses and onresolved issues which are elaborated in this chaper.

First of all lets discuss how relevant this topic will be in the future. At this moment firms create opportunities for employees to work at home. The question remains will the transportation cost of persons stay as important as it is right now? The results of this paper may be less relevant in the future. But at this moment working at home is not well enough integrated and the relationship between SR and coagglomeration is still visible. In the future more research can prove if home working has a negative effect on coagglomeration. If transporation costs of skills becomes less important, there will be a trend to agglomerations based on knowledge and technology. This could be a topic for furtere research

The method used in this arcticle to calculate a coagglomeration index could bias the results. Further research could see if this is the best method to define coagglomerations. The same holds for the skill relatedness index. Like allready mentioned the skill relatedness index reveals a lot of 0 values, which could be due to a lot of small industries. This indicate that industries are having a negative skill relatedness, while this does not need to be the case. SR is based on Swedisch data. Some industries are very small in Sweden, this could influence the skill relatedness variable. So a comparison should be made with other data or other methods indexing SR

Neffke and henning(2010) wrote that entrepeneurs tend to move to agglomerations, because they use the skills of the agglomeration. Large firms use more skills which are already in the company. If this also holds for coagglomerations, then it is important to know how industries are divided. Further research could use firm level data, this could reveal more information on the effect of skill relatedness on coagglomeration.

This article revels that the effect of skill relatedness is the larger between service industries than average. Further research could elaborate this effect. The reason is not found within this article, further research is nessecary. Maybe the use of goods as main product for manufactoring firms is right, but this still need to be proven.

This article uses Dutch data on coagglomeration and Swedisch data on Skill relatedness. The SR values could differ in the Netherlands. To prove if all asumptions are valued, a skill relatedness variable of dutch industries could be used. The method of Neffke and Henning (2010) is not executed with Dutch data yet.

8. Conclusion

Coagglomerations or agglomerations of multiple industries use the same advantages, or at least they use the advantages gained by using the same skills. This article revealed that industries which use the same labor locate near to each other. This means that firms seek locations, to drop transportation costs of labor and gain advantages. Not only transportation costs but also seeking and other labor related costs drop for the firm. Firms do not only look for related firms in their own industry, but they use other industries which use the same labor skills.

Coagglomeration can be quantified using data of employment of industries and regions. Until so far the influence of skill relatedness on coagglomeration was not revealed. But skill relatedness has a significant influence on coagglomeration. Using a standard deviation scale the influence of skill relatedness is about 0.1 on coagglomeration. This result is minor, but this article does not take into account other effects.

The influence of skill relatedness on coagglomeration differs between industry types. The effect of a combination of two service industries is twice as large then the effect of two manufacturing industries. An explanation for this phenomenon is that service industries depend more on skills than they do on goods. And they could save more money on coagglomerating with skill related firms than if they coagglomerate with firms who use the same goods.

The main question of this article: "Does the level of skill relatedness influence coagglomeration?", can be answered positively. The level of skill relatedness does influence coagglomeration. The hypothesis can also be assumed as true, so skill relatedness for influences coagglomeration positively

This result can be used to explain why certain firms are located near others. And why other firms don't move to the same area. Findings reveal that the effect is larger between services industries than between manufacturing industries. This implies that skills are more important for service industries than for manufacturing industries.

9. Bibliography

Agrawal, a., lain, C., & McHale, J. (2006). Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography* , 571-591.

Boschma, R., & Frenken, K. (2009). Technological relatedness and regional branching. In M. F. H. Bathelt, *Dynamic Geographies of.* Taylor and Francis.

Boschma, R., & Frenken, K. (2011). The emerging empirics of evloutionary economic geography. *Journal of Economic Geography* .

Boschma, R., Eriksson, R., & Lindgren, U. (2009). How does labor mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economis Geography* , 169-190.

Bosma, N., Stam, E., & Schutjens, V. (2006). *Creative Destruction and Regional.* Zoetermeer: SCALES.

Carree, M., & Dejardin, M. (2007). Entry thresholds and actual entry and exit in local markets. *Small Business Economics* , 203-212.

Ellison, G., & Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* , 889-927.

Ellison, G., & Glaeser, E. L. (1999). Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *The American Economic Review,* , 311-316.

Ellison, G., Glaeser, E. L., & Kerr, W. (2010). What Causes Industry Agglomeration? Evidence from coagglomeration. *American Economic Review* , 1195–1213.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or The undefinable tacitness of being(there). *Journal of Economic Geography* , 75-99.

Glaeser, E. L., & Kohlhase, J. E. (2003). Cities, regions and the decline of transport costs . *Papers in Regional Science* , 197-228.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* , 577-598.

Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy* , 483-499.

Marshall, A. (1920). *Principles of Economics: An introductory volume.*

Maskell, P., & Malmberg, A. (1999). Localised learning and industrial competitiveness. *Camebridge Journal of Economics* , 167-185.

Neal, D. (1995). Industry-SpecificH uman Capital:. *Journal of Labor Economics* , 653-677.

Neffke, F., & Henning, M. S. (2008). Revealed Relatedness: Mapping Industry Space. *Papers in Evolutionary Economic Geography* .

Neffke, F., & Henning, M. (2010). Seeds of regional structural change. The role of entrepreneurs and expanding firms in shaping local path dependencies.

Neffke, F., & Henning, M. (2010). Skill-relatedness and firm diversification. *Evolutionary Economics Group* .

Neffke, F., Henning, M., & Boschma, M. R. (2009). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Papers in Evolutionary Economic Geography* .

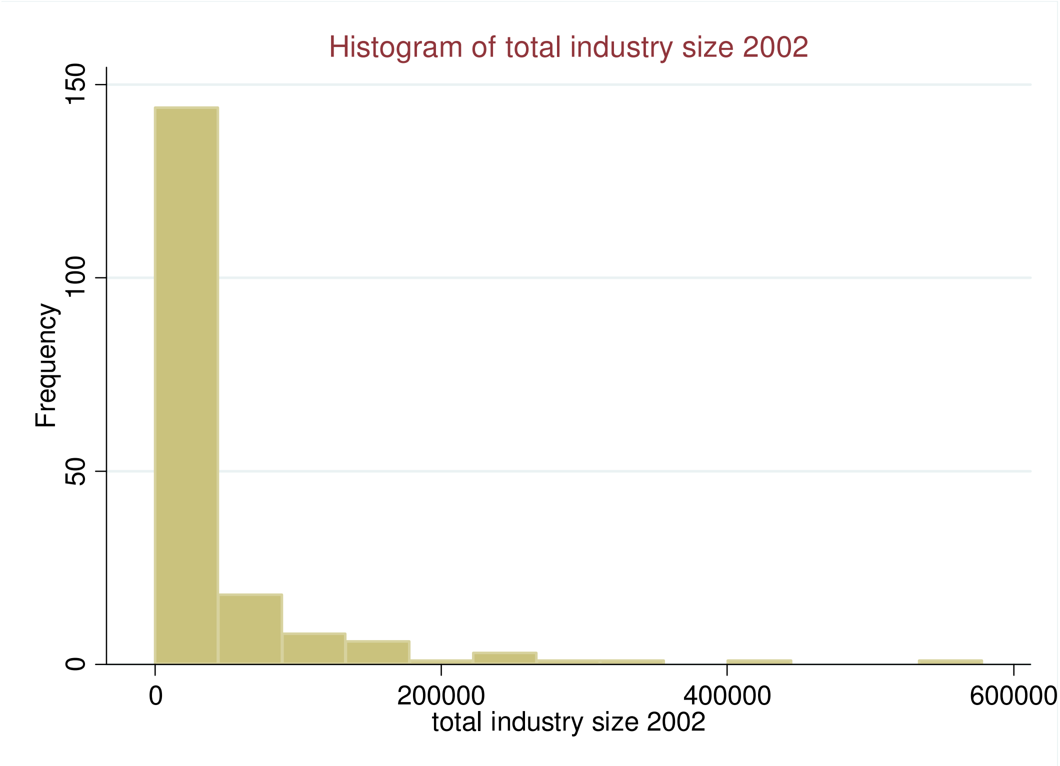
Nooteboom, B. (2000). *Learning and Innovation in Organizations and Economies.* Oxford: Ofcord university press.

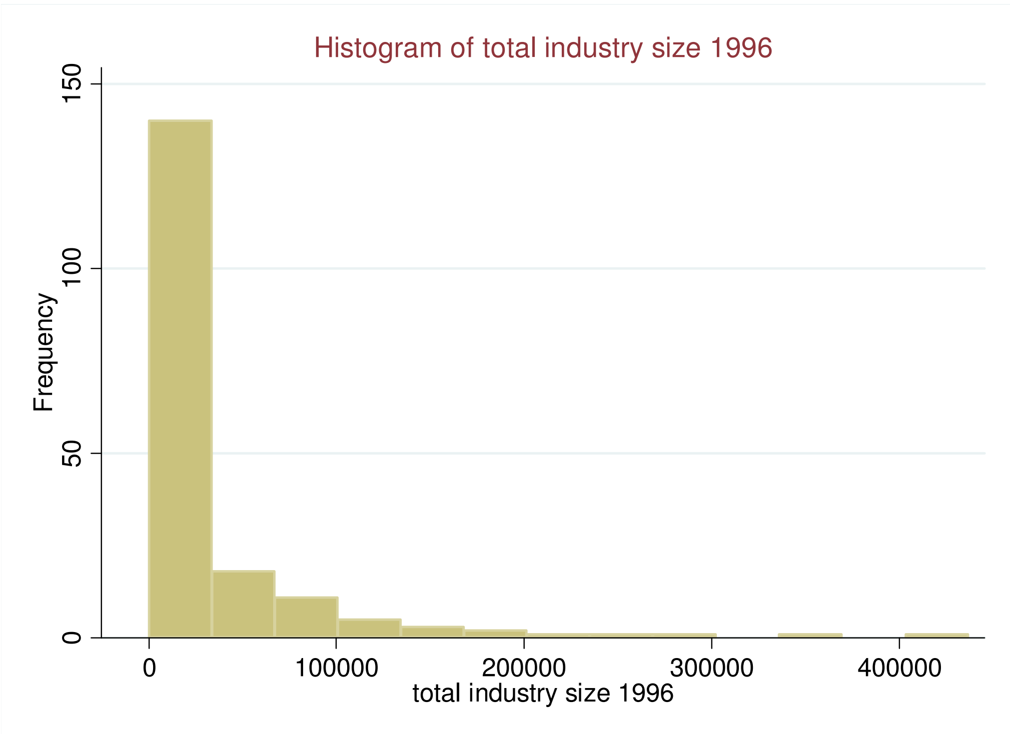
Peck, J. (1992). Labor and Agglomeration: Control and Flexibility in Local Labor Markets. *Economic Geography* , 325-347.

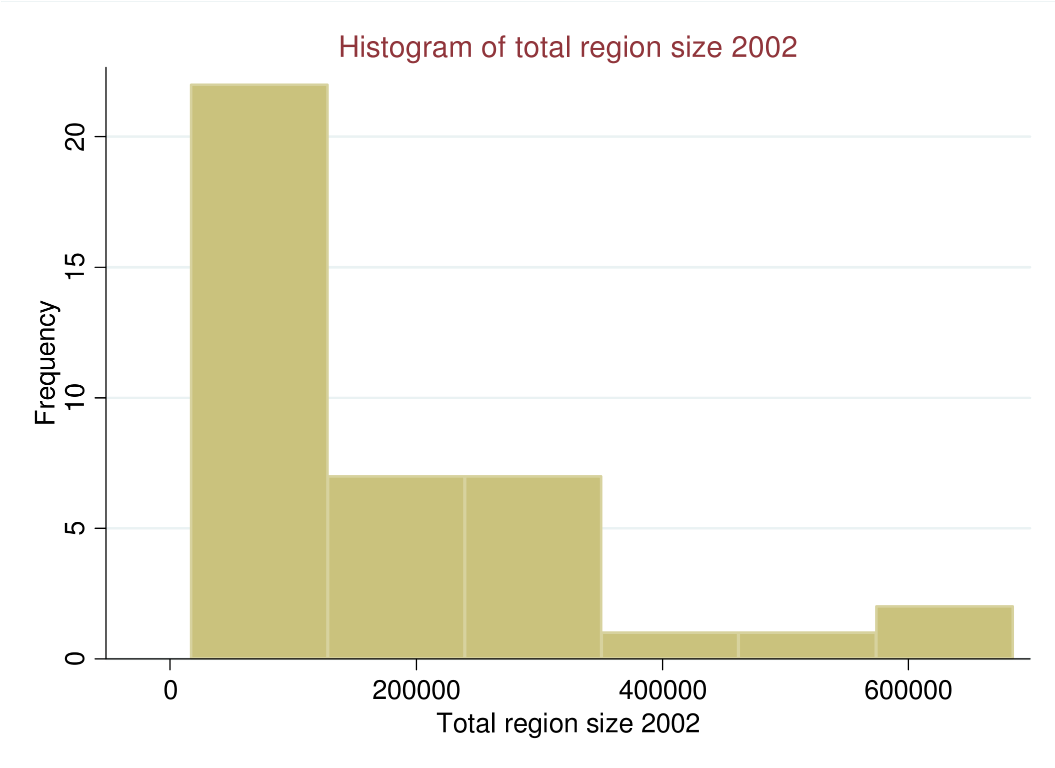
Polanyi, M. (1966). *Tacit dimension.* Garden: Doubleday & Company, Inc.

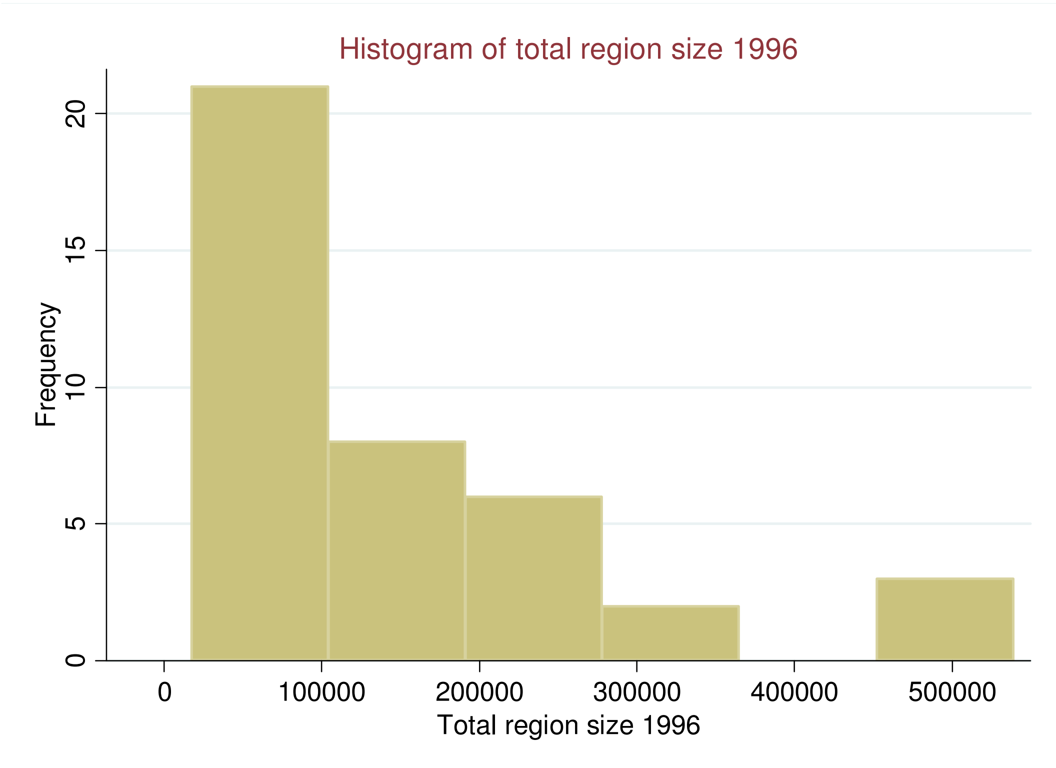
Appendix

Appendix A: distributions of employment






Histogram of total region size 2002
Frequency
0
5
10
15
20
0
200000
400000
600000
Total region size 2002


Histogram of total region size 1996
Frequency
0
5
10
15
20
0
100000
200000
300000
400000
500000
Total region size 1996

Appendix B: correlation coefficients

	Coagglomeration 1996	Coagglomeration 2002	Skill relatedness	SRT
Coagglomeration 1996	1.000			
Coagglomeration 2002	0.7538	1.000		
Skill relatedness	0.1187	0.1212	1.000	
SRT	0.0933	0.1074	0.4394	1.000

Appendix c: regressions using a part of the industries

Regression analysis of coagglomeration 1996 manufacturing industries only

Coagglomeration 1996	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR	2.73 e-04	0.000	2.14 e-03	0.000
Constant	7.22 e-03	0.000	8.87 e-03	0.000
R Squared	5.8 e-03		3.1 e-03	
N observations	7832		7832	

Regression analysis of coagglomeration 2002 manufacturing industries only

Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR	3.02 e-04	0.000	2.04 e-03	0.000
Constant	8.92 e-03	0.000	1.06 e-02	0.000
R Squared	6.9 e-03		2.8 e-03	
N observations	8010		8010	

Regression analysis of standard deviation coagglomeration 2002 manufacturing industries only

SD Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SD SR	.083	0.000	.053	0.000
Constant	4.95e-10	1.000	4.43e-10	1.000
R Squared	6.9 e-03		2.8 e-03	
N observations	8010		8010	

Regression analysis of coagglomeration 1996 only manufacturing/service industry combinations

Coagglomeration 1996	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR	4.73 e-04	0.000	1.67 e-03	0.000
Constant	-4.43 e-03	0.000	-3.03 e-03	0.000
R Squared	7.9 e-03		3.0 e-03	
N observations	15842		15842	

Regression analysis of coagglomeration 2002 only manufacturing/service industry combinations

Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR	2.16 e-04	0.000	2.35 e-03	0.000
Constant	-4.60 e-03	0.000	-2.90 e-03	0.000
R Squared	1.4 e-03		5.2 e-03	
N observations	16020		16020	

Regression analysis of standard deviation coagglomeration 2002 only manufacturing/service industry combinations

SD Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SD SR	.038	0.000	.072	0.000
Constant	6.21e-11	1.000	1.22e-09	1.000
R Squared	1.4 e-03		5.2 e-03	
N observations	16020		16020	

Regression analysis of coagglomeration 1996 service industries only

Coagglomeration 1996	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR	3.97 e-04	0.000	3.13 e-03	0.000
Constant	1.79 e-03	0.968	3.568 e-03	0.000
R Squared	0.026		0.019	
N observations	7832		7832	

Table2. Regression analysis of coagglomeration 2002 service industries only

Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR	3.86 e-04	0.000	3.39 e-03	0.000
Constant	1.83 e-03	0.002	3.68 e-03	0.000
R Squared	0.021		0.019	
N observations	7832		7832	

Table2. Regression analysis of standard deviation service industries only

SD Coagglomeration 2002	Skill relatedness index		Transformation skill relatedness	
	coefficient	*P*	*coefficient*	*P*
SR	.144	0.000	.138	0.000
Constant	-2.21e-10	1.000	6.68e-10	1.000
R Squared	0.021		0.019	
N observations	7832		7832	