Firm innovation and productivity

A regional analysis

Sandra Bladh

Supervisor: Johanna Palmberg

Södertörns högskola | Economics Institution

Masters thesis 30hp

Economics | May 2017

SÖDERTÖRNS HÖGSKOLA | STOCKHOLM
sh.se

ABSTRACT

This thesis studies the effect of innovation activities and productivity by using the CDM-model and extend the existing knowledge by using the CIS-dataset in combination with official statistics performing a such detailed regional analysis that have not been done before. By using the different labour market codes interacted with the industry codes I can capture informative deviations between different industries in different regions. The results show a significant variation between the different regions and industries, and that the urban and metropolitan areas are more innovative and more productive than the rural areas. However, the financial sector and health sectors showed a steady innovation input activity across most regions while the metropolitan areas showed to invest less in innovation inputs in the real estate sector compered to rural and urban areas.

ACKNOWLEDGEMENTS

First of all I like to thank by beloved husband Carl for all supporting words and love during this spring. You are my rock!

I like to thank Andreas Poldahl at Statistics Sweden for all help and support with the data and estimation process, this would never have worked without you. I am truly grateful!

And of course a big thank you to Johanna Plamberg my supervisor for all helpful comments and for pointing me into the right direction.

CONTENTS

1 INTRODUCTION

"Everything that can be invented has already been invented."

- Charles Holland Duell (1889)

The famous quote by Duell have later been established to be falsely quoted, but what role does new inventions and innovation activities play in the society today. Joseph Schumpeter expresses innovation as the driving force of the economic development (Schumpeter, 1942) and investments in research and development (R&D) are essential for gaining its contributions.

Around the world there are many on-going projects and strategies for encouraging R&D investments and entrepreneurship. Sweden was in 2016 ranked by WIPO's Global Innovation Index as number two[1] of the most innovating countries in the world (Wipo, 2016). The EU-2020 strategy is one example of an important project with the objective of creating growth and new jobs by investing in R&D. A project that as well contributes to the environmental research and poverty reduction. The project aims to invest three per cent of the European GDP in R&D (European Commission, 2016).

Schumpeter appoints the phenomena of creative destruction, the cycle were old outdated ideas are replaced by new, and innovation to be two forces that goes hand in hand (Schumpeter, 1942). Start-up firms' have shown to create multiple new jobs. In the US, 40 million jobs were created by start-up firms in a period of 25 years from 1980 to 2005 (Braunerhjelm, et al., 2011) and a study in Sweden shows that 240 000 new jobs were created by start-up firms between 2000-2009 (Heyman, et al., 2013). Schumpeter's theory claims creative destruction to be the main force behind firm failure, and the cycle were the new firms are established the core of the economic growth is found (Schumpeter, 1942). Studies have confirmed the Schumpeterian hypothesis by showing that for every 10 firm that goes to bankruptcy, about 10 new are introduced on the market (Davis, et al., 1993).

Innovation have been widely discussed and examined for year and one model that is the leading edge for the studies is the CDM-model. The CDM-model was introduced by Crépon, Duguet and Mairesse in 1998 has and have since then been used in hundreds of innovation studies in more than 40 countries. The model is today the most appropriate model when analysing micro data based on the Oslo Manual[2] (Lööf, et al., 2017). The

[1] Switzerland was appointed as number one.

[2] A guideline for collecting and interpreting the CIS-data

model is based on the assumption that it is innovation output and not input that increases the firms' productivity and is using proxies for those variables that are unobserved, for example the demand conditions and technical opportunities (Crépon, et al., 1998). When estimating the productivity, the augmented knowledge based production function introduced by Griliches in 1990 is used in a systematic and clear manner (Lööf, et al., 2017). The model was originally used on cross-sectional data but it is easily adapted and modified to fit all sorts of data and studies have be conducted using a variation of different measuring methods (Lööf, et al., 2017).

This study applies the CDM-model when studying innovation activities and productivity growth in different labour markets in Sweden and the model is modified to fit panel data. The analysis combines the variables obtained from the Community Innovation Survey (CIS) with microdata from the Structural Business Statistics (SBS) and regional labour statistics based on administrative sources (RAMS). This give a wide range of detailed firm variables and the possibility to draw conclusions on an aggregated level about how the age, size, group and industry belong and position of the firms are affecting the innovation output and productivity. The data covers ten year of observations between 2004 and 2014 and more than 14 000 firms are observed.

Tons of earlier research has been done on the innovation subject. Lööf, Mairesse and Mohenen examined the CDM-model by a bibliometric study of 12 paper from 25 researcher between 1990 and 2012 to measure the impact it had on the scientific literature (Lööf, et al., 2017). Five of the 12 papers used the CDM-model and the result showed no conflict between the 12 different authors. A study comparing German and Swedish firms from 2003 showed that group belonging increased the probability of participating in innovation activities (Lööf, et al., 2003). A study conducted in 2015 on Swedish CIS-data form 2008 to 2012 using the CDM-model showed significant evidence of heterogeneity across technology and knowledge sectors but that the influence of R&D investment is in line with earlier research (Baum, et al., 2015).

This study contributes to earlier research by using micro level data in combination with official statistics performing a detailed regional analysis in a manner that have not been done before. Gaining information about each firm's location and main sector will provide detailed information which will enable conclusions about how the location of different firm and sectors will affect the innovation output and productivity. For simplicity the locations will be divided into rural, urban and metropolitan areas.

The CDM-model will be used to answer the following questions: *Are firms as innovative in metropolitan, urban and rural areas in Sweden? Are there any regional or industrial differences in investment? Are there any differences in the productivity between different areas and industries?*

The outline of this paper is as follows: first a brief overview over my gained results. The second section of the paper start with earlier research on the topic and will then handle the theoretical parts of innovations, regions and the CDM-model. Section 3 describes the dataset, the empirical model and the treatment of the data. Section 4 presents the empirical results and section 5 concludes the paper.

The hypothesis that start-up firms are more innovative than the established have in recent studies been discussed and the result is deviating between different studies. This study shows that start up firm are more likely to be innovative and that it is beneficial to be in a urban area with a modern amount of competition. The human capital is an important factor in both productivity and innovation output as well as the population density in showing to increase innovation output and productivity significantly.

2 Background

"I see no advantage in these new clocks. They run no faster than the ones made 100 years ago."

— Henry Ford

2.1 Earlier research

Table 1 presents an overview of some of the earlier research on innovation, divided into if CDM-model has been used, or not. The first six studies have been conducted using the CDM-model while the other two have been using other methods. I have included both researches with Swedish CIS-data as well as some researches on foreign data. Three of the researches has done a regional analysis, but only one of them have been using the CDM-model.

Table 1 - Earlier research

Earlier research using the CDM-model		
Author	**Data, country, measure**	**Result**
Lööf, Peters & Janz 2003	▪ CIS: 1998-2000 ▪ Germany and Sweden ▪ CDM-model	The national market is more important for German firms. Group belonging decrease the probability of innovation in Sweden. The intensity of both innovation input and innovation output decreases with firm size in Germany. The R&D subsidiary system in Germany is more oriented towards larger firms than its Swedish equivalent and that the average size of innovative firms are higher in Germany.
Criscuolo, 2009	▪ CIS: year 2000 ▪ 18 countries ▪ CDM-model	The results show similar and consistent patterns within the different countries. There are some notable exceptions, especially the relationship between innovation policy and investments in innovation. In Europe the correlation between sales from product innovation and productivity is higher for larger enterprises, and for Brazil, Canada and New Zealand the correlation is higher among SMEs[3]. As expected, in most countries the productivity effect of product innovation is larger in the manufacturing sector than in the services sector. Exceptions are Germany and New Zealand where the innovation-productivity link seems to be stronger in the services sector sample.
Crépon, Duget & Mairesse 1998	▪ SESSI innovation survey ▪ Year 1990 ▪ French manufacturing firms ▪ First CDM-model	Probability of engaging in R&D increase with firm size. The innovation output rises with research effort. Firm productivity correlates positively with innovation output.
Lööf & Heshmati, 2006	▪ CIS year: 1996-1998 ▪ Sweden ▪ CDM-model	Employment increases with innovation output only for services. There is a close association between the level of profit and innovation for services as well as for manufacturing firms. The growth rate of productivity increases only with innovations new to the market when manufacturing firms are considered. The positive relationship between innovation and employment growth and innovation and productivity growth for service firms is independent of the degree of novelty of the innovations.
Baum, Lööf, Nabavi & Stephan, 2015	▪ CIS year: 2006-2012 ▪ Sweden ▪ CDM-model	Measures of the infuence of R&D investment on innovation sales and innovation sales on labor productivity generally in line

[3] Small-Medium Enterprises

		with the original CDM values. Significant evidence of heterogeneity across technology and knowledge sectors in their magnitudes. The impact of other explanatory factors on the key variables also exhibits considerable differences across sectors, with significant effects in some sectors and not others. These results cast doubt on earlier research which does not allow for this heterogeneity
Goya, Vaýa & Suriñach, 2013	▪ Technological Innovation Panel ▪ Spain ▪ 2004-2010 ▪ CDM-model	The firm's decision whether to engage in R&D activities is influenced by other firms dicision. Innovation carried out by other firms (intra- and inter-industry externalities) have a positive impact on firm's productivity.
Earlier regional reasearch on innovation activities		
Braunerhjelm, Borgman, 2004	▪ Regional data, 1975-1999 ▪ Sweden ▪ 143 industries ▪ 70 labour market regions	This study examines the relationship between concentration and regional growth by using the Ellison–Glaeser indexes and Gini location. The econometric results imply a 2–6% higher growth in regionally concentrated industries. The effect is more pronounced for knowledge-intensive manufacturing, network industries and industries intensively using raw material. It is also found that regional entrepreneurship and regional absorption capacity are important explanations of regional growth, whereas the impact of the skill-level and economies of scale is more mixed.
Lööf, Johansson, 2014	▪ CIS, 2002-2006 ▪ Sweden, metropolitan analysis	Productivity premium associated with persistent R&D is close to 8 per cent in non-metro locations and about 14 per cent in the largest city. Firms without any R&D engagement does not benefit at all from the external milieu in metro areas. No productivity premium is associated with occasional R&D effort regardless of the firm's location.

3 Theoretical framework

3.1 Innovations

Joseph Schumpeter emphasizes the importance of the production factors: labour, capital and raw material and their contribution to the economic development (Schumpeter, 1911/1934). The productions factors need to be combined in new or more efficient ways in order to contribute to the development. Schumpeter establishes that the one that invents the new combination is "the inventor" and the one that brings it to the market is "the entrepreneur" (Schumpeter, 1911/1934). The act of bringing the new invention to the market, is the driving force of the economic development and economic growth (Schumpeter, 1911/1934).

The entrepreneur is not just someone that sets up a new business, but a person or a group of persons that are able to transform a new idea or invention into something new and successful; a process that will generate a totally new product or a new market for an existing product (Mazzucato, 2013). Firms are considered being innovators if they have implemented a new innovation and the degree of novelty is of significance; whether the innovation is new to the firm, new to the market or new to the world. Entrepreneurs that implement products that are new to the market or the world are being considered as drivers of the process of innovation (OECD, 2009).

The economist Frank Knight defined in 1921 two important dilemmas that the entrepreneur is facing – risk and uncertainty. The risk is something that the entrepreneur possible can protect himself against, like a building fire or theft, while the uncertainty is much harder to retaliate against. The possibility that the new inventions will be something ground breaking is one example of uncertainty (Johansson, et al., 2014).

New inventions are not always material items, but might as well be new thought and theories or new social institutions and organisations (Kaiserfeld, 2005). An idea is to be called an invention if it is a new and unique thought. The philosopher Jon Elster defines innovation as "the production of new technical knowledge" and inventions as "the generation of some scientific idea, a theory of concept that may lead to an innovation when applied to a process of production" (Kaiserfeld, 2005). Kaiserfeld states that there are inventions that are not applied to any process of production, fire for example was discovered long before science existed.

Schumpeter defines five different types of innovations that are presented in Table 3, divided into technological and non-technological innovations.

Table 2 - Schumpeters five innovation types

Innovation type	
Technological innovations:	
1. Product	Product innovations gives new products to the market, that is a product that the consumers are not familiar with, or a

	product of new quality
2. Process	A new method is presented, that might be new production techniques that changes the production possibilities or new ways of using the existing raw material. Or something that changes other method like marketing or payments.
Non technological innovations:	
3. Business model	A new market opens up, that is a market that have not been available for that manufacturing industry before, that might be a new market or an already existing market.
4. Source of supply	Innovations that lead to the availability of new sources of raw materials or semi-finished products, that might be already existing or brand new sources.
5. Mergers & divestments	New organizations are developed, that might be new mo-nopoly settings or the resolution of monopoly. Schumpeter argue that the private property rights are basic for the pros-perity of the western countries (Johansson, et al., 2014)
	(Schumpeter, 1911/1934)

Earlier studies have showed that the new industrial sector earlier mainly contributed to product innovations, while it later has been a change to a higher rate of process innovations instead (Kaiserfeld, 2005). It is showed that the producers tend to develop different types of innovations, mostly because of the asymmetric information, due to that the consumer and the developer has different knowledge about the good or service. The asymmetric information between the consumer and the producer appears between the information that the consumer hold about its' need and "context-of-use" and the information that the producer that specializes in the specific demanded good holds. Since the information is "sticky" and not easterly exchanged between the producer and consumer, the consumer has a much more detailed picture of its preferences. Concurrently the manufacturers have a better model of the solution approach in which they specialize than the user has (von Hippel, 2005).

One important aspect of innovation that Schumpeter also stresses is the phenomena of creative destruction, that is the natural path that the development of new ideas causes the old idea to be outdated, something that can be observed as the closing of some firms while others persist. Schumpeter claims the creative destruction as the core of the economic growth, mainly because of the creation of new occupations and the maintaining of employment flow (Schumpeter, 1942). Schah, Davis and Haltiwanger showed that start-up firms created a greater share of the job base outside the manufacturing sector and that every year about ten per cent of the jobs are being destructed but the same share is every year created by start-up firms, creating a sustainable work flow (Davis, et al., 1993).

3.2 Productivity growth

When discussing economic development, productivity is one of the key indicators and it is defined as the ratio between output and inputs. The different production inputs like labour, raw material and capital need to be as efficiently used as possible in order to receive the fullest productivity. In order to succeed, innovation processes in which new and more efficient combinations are being invented are a key element in achieving the fullest productivity growth (Fujita, 2008).

The concept of creative destruction and the production function are two of the main subjects when discussing productivity. The process of creative destruction where the production structure continuously seeks more upgraded technology, processes and output mixes by excluding unproductive segments (Caballero & Hammour , 2000). And it has been shown that the job reallocation from less productive businesses contributes heavily to the productivity growth, in linking to knowledge intensity. Firms located in clusters are often highlighted when discussing productivity because of the close relationship between the different production stages in a delimited area, often referred to as knowledge and technological externalities or spillovers (Swann & Baptista, 1998).

The production function measures the highest level of output that the firm can obtain by its given inputs. Equation 1 shows the production function which describes the output of a firm given the inputs of physical labour and capital (Solow, 1957). A equals the proportion of the output that is not explained by the inputs, the total factor productivity, TFP, which also can be denoted the level of efficiency that the inputs are utilized in the production. TFP and innovation are closely related and studies have shown that an increase of R&D tend to increase the TFP growth (Karafillis & Papanagiotou, 2010)

The production function

$$Y = A(K, L) \qquad \textit{Eq. 1}$$

where : Y=output K= capital L= labour
A= the level of efficiency

Equation 2 is defined as the knowledge based Production function and is used in the final step of the CDM-model (Crépon, et al., 1998). The knowledge production function shows how investments in different knowledge based activities, for example R&D, increases the knowledge. An increase in knowledge will increase the innovation output which will increase the productivity (Griliches, 1998).

The Knowledge Production Function

$$Q = AX^{\beta}K^{\gamma} \qquad \textit{Eq. 2}$$

Where

Q= the output

X= index of conventional inputs

including physical capital

K= the "stock of knowledge" (or R&D)

A=is the level of disembodied technology

β and γ are the parameters of interest

(Griliches, 1998)

3.3 Regional labour market, Metropolitan, urban and rural areas

When analysing regional economics two forces are due to be investigated, agglomeration which means moving toward a centre and the force of dispersion, when moving away from a centre. All societies are faced with the same dilemma, individuals must get together in order to benefit from the advantages of the division of labour (Fujita, 1996). Since location is of such importance for the economic development the economic activity, resources and economic agents will not be evenly distributed across the country (Palmberg & Backman, 2015). The regions can be divided into metropolitan, urban and rural areas by a hierarchical structure of locations. Each of the category is associated with a specific level of services, demand, resources and different growth patterns (Palmberg & Backman, 2015). The metropolitan areas in Sweden are: Stockholm, Gothenburg and Malmö. Studies have shown that a majority of the university educated students move towards metropolitan areas during and after their academic studies, making the metropolitan areas a net receiver of human capital, while the urban and rural areas are exporters of human capital (Andersson, et al., 2015). It has also been shown that the students that move towards a metropolitan areas after graduation generally have higher grades and higher educated parents than the individuals who decide to stay in the urban and rural areas (Andersson, et al., 2015). The reasons to why the students decide to move have been widely discussed, but some important factors might be that the metropolitan areas have better labour market perspectives with dense labour market and a diversity in work sectors, professions and employers. The diversity of educated staff in the metropolitan areas gives the firms an advantage in finding specially educated personnel (Andersson, et al., 2015). The increased concentration of labour increases the competition, forcing the companies to invest in innovation in order to stay as sharp on the market. Increased competition will also eliminate obstacles of establishment, developed infrastructure and improved quality (Braunerhjelm, et al., 2011).

3.4 The CDM-model

The CDM-model was introduced by Crépon, Duget & Mairesse in August 1998 in the article "Research, innovation and productivity: an econometric analysis at the firm level". It is a framework for linking the relationship between productivity, innovations and research at the firm level. It was the first model that showed the fact that the innovation inputs determine the innovation outputs which affects the productivity (OECD, 2009). The model summarizes the process from the firm's decision to invest in research to the impact of innovation on the firm production activities (Crépon, et al., 1998). The model introduces three new features into the analysis that will be presented in table 3.

Table 3 - New features introduced in the CDM-model

New features	Description
Innovation output increases the productivity	The firms invest in innovation in order to develop new processes that will increase the productivity and economical performances
New data on innovation output	The data used in the report is more detailed allowing for new innovation variables.
More efficient econometric methods	The model uses econometric methods that correct for usual biases like selectivity bias and endogeneity problems.

The CDM framework introduces a structural model that explains productivity by innovation output and the innovation output by research investment, and it suggests a method of correcting for the selectivity and the endogeneity inherent in the model (Lööf, et al., 2017). The CDM-model consists of four different equations. Two of the equations are for research, one equation is an innovation function and the last one is a productivity function.

When investigating the firm's research behaviour, the first equation will answer if the firms are engaged in any research activities or not and the second equation will check for the intensity of that research investments. Originally the third equation was an innovation function that studied the number of patents and innovative sales, presenting if the firms have introduced any new products on the market. This thesis uses a dummy for the innovation output that takes the value one if the firms has introduced a new or improved good, service or process. The final equation studies the productivity by using the Cobb Douglas production function. The Cobb Douglas production function includes physical capital, employment, skill composition and innovation output, where the innovation output is measured by patents per employee or by the latent share of innovative sales (Crépon, et al., 1998).

4 METHOD

"Where all think alike there is little danger of innovation."

- Edward Abbey

4.1 DESCRIPTIVE STATISTICS – INNOVATION INDICATORS

This thesis uses mainly variables from the CIS-database in combination with variables from The Statistical Business Register (SBR) database and the Regional labour statistics based on administrative sources (RAMS). The "Oslo Manual" is a guideline for measuring, collecting and interpreting innovation data. The first version of the manual was released in 1992 and the latest version in 2005. The OSLO Manual is an analytical framework for the study of innovation with its focus on technical product and process innovations in manufacturing. The Oslo Manual is the reference for the European CIS-data (OECD, 2005).

SBR contains information about all firms, government offices and organisations in Sweden. The database provides information about the firm's location, industry codes, number of employees (Statistics Sweden , 2017). RAMS is commonly used in research. It is a yearly conducted data collection and consists of every person that is a registered resident in Sweden the 31 of December that year. The data contains information about all the firms in Sweden linking individuals to both enterprises and establishments by person-, organization- and establishment number and provides information about the personnel structure in the firms (Statistics Sweden, 2017).

In order for the CIS-survey to provide sufficient information on innovation the different questions are including both technical and non-technical innovation. When studying technical innovation, the indicators focus on individual elements of product and process innovations. The process innovation includes the improvements that the firm has done on its' intramural processes such as new technologies or other internal developments leading to an increase of new knowledge creation. The product innovations are when the new goods or services are established on the firm's markets (Criscuolo, 2009).

The CIS-data includes firms with a minimum of 10 employees in all the regions of Sweden chosen from a sampling frame. The variable of the number of employees and the firm's turnover is collected from the Statistical Business Register (SBR) and the industry codes (SNI) is collected from the Business Registers (BR) (Statistics Sweden, 2014).

The data covers information about the amount of capital spent on innovation activities, as well as information about co-operations with universities and other research facilities. Microdata-based indicators reflect the behaviour of individual firms and firms' heterogeneity, and by giving detailed information about the sizes of the firms it gives possibilities to draw conclusions of correlations between innovations decisions and the heterogeneity between firms. The firms also differ in what type of innovation activities they perform, whether it is product, process, organisation or marketing innovations. The designing of efficient innovation policies with the target of increasing the innovativeness along some firms needs an understanding of why some firms are innovative while some are not. An increased knowledge of the firms is crucial in the work of policy formation, if the policies do not take the heterogeneity along the firms into account the policies tends to miss the main target (OECD, 2009). Microdata gives many advantages since it will provide the researcher with information at micro level, such as firm size, firm location, industry and the education level of the employed. By getting the information about each firm's innovative profile it can be aggregated to country or regional level giving rise to much more complexed research methods that will identify similarities and differences in certain characteristics or certain groups of firms and allow for estimating functional relationships between sub-groups of firms (OECD, 2009).

The data used in this study is extracted from the CIS database from Statistics Sweden combined with the SBS and RAMS. In order to get the information of the firms' location the

Statistical Business Register dataset have been matched to the CIS dataset for each year using the firms company registration number as the matching key. The analysis covers the CIS dataset from 2004, 2006, 2008, 2010, 2012 and 2014 and contains 12864 different firms. Figure 1 presents how many observations per year the data set contains, the division slightly uneven with 26% of the observations in 2014 and 11% of the observations are for year 2004 and 2006. Figure 2 presents how many of the firms that have answered the survey one or multiple years. 6027 firm have only answered the survey one year, while 462 have answered the survey all six years.

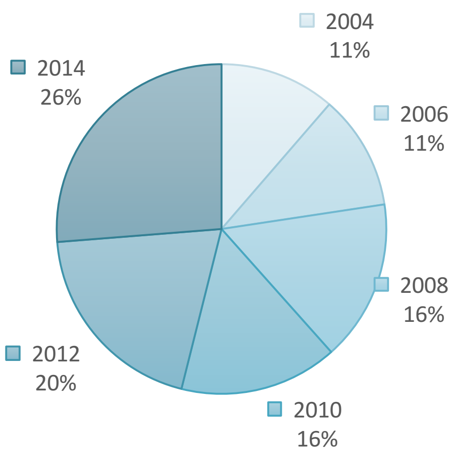


Figure 1 - The proportion of observations per year

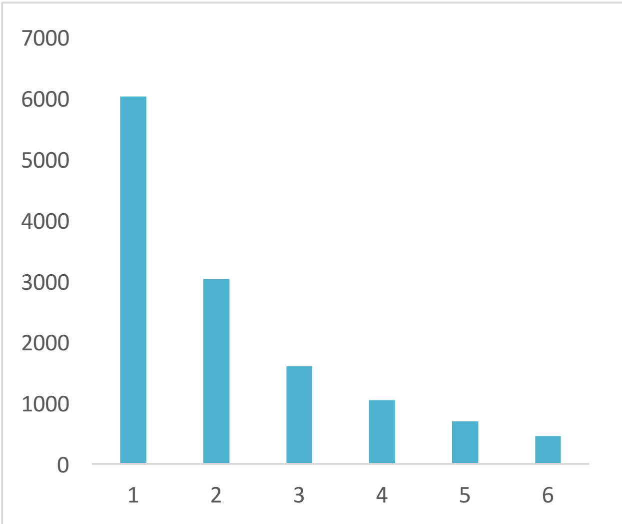


Figure 2 - Number of firms with one or multiple years of observations

The panel data set provides many important advantages compared to cross sectional data. Observations of different firms over time accounts for heterogeneity, provides more informative data with less collinearity and better detection of different measure effects (Gujarati, 2012). The dataset is an unbalanced short panel, it is a short panel since the number of firms are more than the number of years studied, with 12 864 different firms for six different points of time. And it is an unbalanced set of observations because the number of time observations are not the same for each firm (Gujarati, 2012).

The dataset measures different types of innovation, good, service and process innovation. Among the firms that are included the innovativeness is varying slightly throughout the years. Figure 3 and 4 shows how the innovation output is showing a steady percentage between the years, showing that about 40% of the firms has introduced a new or significantly improved good, service or process each year.

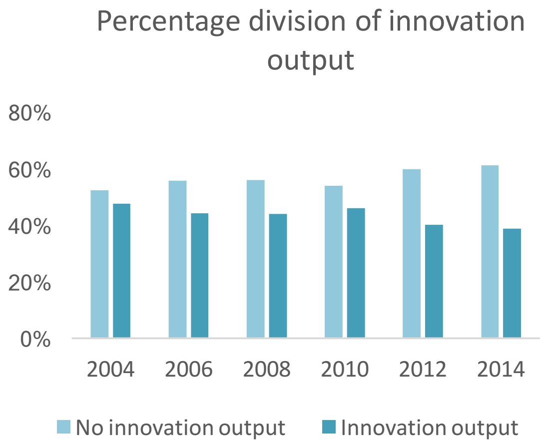


Figure 3 - Percentage of firms with innovation output, comparison when belonging and not belonging to a Group

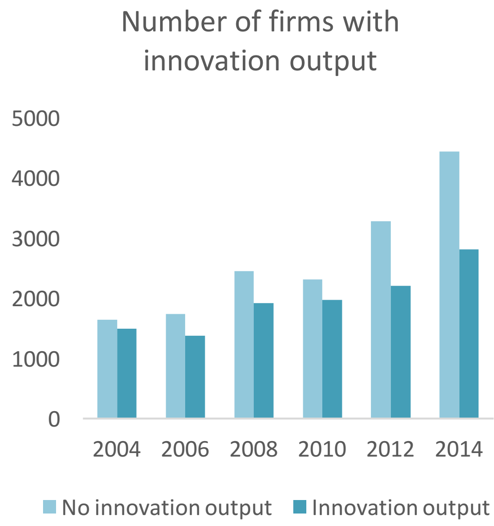
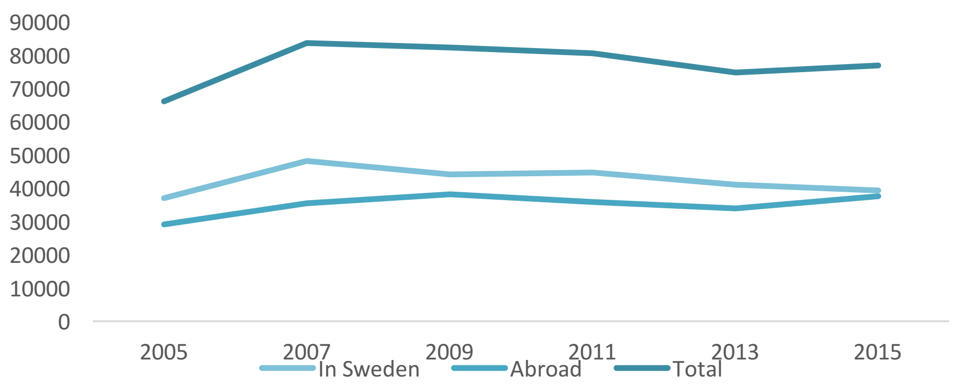


Figure 4 – Number of firms with innovation output comparison between firms belonging and not belonging to a Group

The R&D expenditures has shown to increase among firms that are part of an international group throughout the recent years. In 2005 the R&D expenditure that the Swedish enterprises spent was about 66 billions while in 2015 the expenditure had increased to 76 billions. Figure 5shows the R&D expenditure development (Tillväxtanalys, 2015).



	2005	2007	2009	2011	2013	2015
Total in the world	66 106	85 523	82 230	80 459	74 754	76 813
In Sweden	36 988	48 133	44 166	44 629	40 928	39 225
Abroad	29 118	35 391	38 064	35 831	33 825	37 588

Figure 5 - The expenditure in millon of SEK that swedish Groups spend on (Tillväxtanalys,

R&D abroad and in Sweden between 2005-2015.	2015)

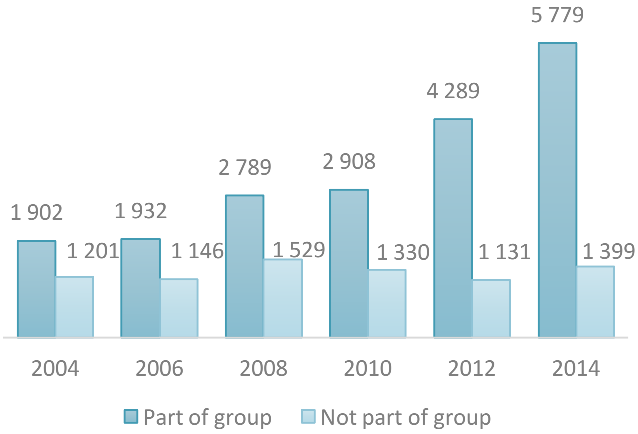


Figure 6 - Number of firms in the dataset that are part of a Group

In 2013, 2 831 of the Swedish firms belonged to a group with subsidiary abroad (Tillväxtanalys, 2013). Figure 6 present the number of firms in the dataset that are belonging to a group. As seen even when the number of observations per year increase, the number of firms that are not part of a group maintain constant, while the fraction that are part of a group increased significantly. The advantage for a firm to belong to a group is that they have a larger availability too financial medium, that can be through group loans from abroad without any safety (Statistics Sweden, 2013).

About 70 per cent (95 billion SEK) of the Swedish R&D expenditure is in the corporate sector, and the second largest is the university sector, while the R&D expenditures in the public sector was about three per cent (Tillväxtanalys, 2015). Today the public sector is not included in the CIS-selection, but the topic is discussed and around the world the public sector is increasing the R&D expenditures. The Swedish Growth analysis released a report where they established that Sweden needs to develop strategic collaborations in order to increase the innovation in the public sectors in order to keep up with the increasing innovative rate in the world. The study showed that for example India have found ways to decrease the cost with about 10 per cent compered to USA for some health treatments (Tillväxtanalys, 2016). Figure 7 and 8 below shows a comparison between the innovation output, the number of new or significantly improved service, good or processes that are introduced by firms that are part of a group compared to those that are not. Both the percentage and the absolute number show a higher innovation output among those firms that are belonging to a group.

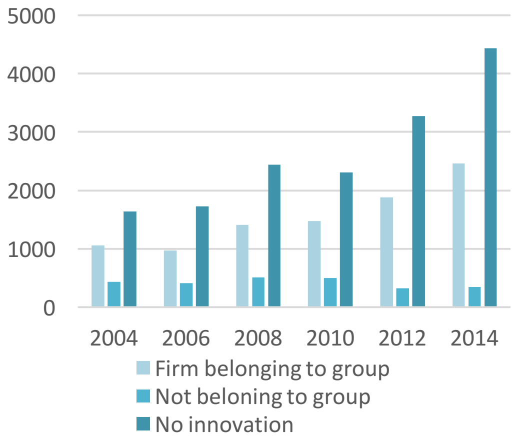


Figure 7 – Number of firms with innovation output, comparison between firms belonging to a group and not.

Figure 8 – Percentage of firms with innovation output, comparison between firms belonging to a group and not.

As discussed before, Schumpeter expresses start-up firms as one primary resource for gaining economic development. In this study the start-up firms are defined as businesses that are two years or younger. Figure 9 present number of firms in the dataset that are defined as start-up firms distributed over the years and figure 10 presents their innovation output compered to established firms. Even though the vast majority of the dataset are established firms, the innovativeness among those firms that are between 0 (started the same year as the survey is conducted) and 2 years in per cent is just about the same. For both the established and the start-up firms the majority (about 50-60%) have not introduced any new or significantly improved product, service or process.

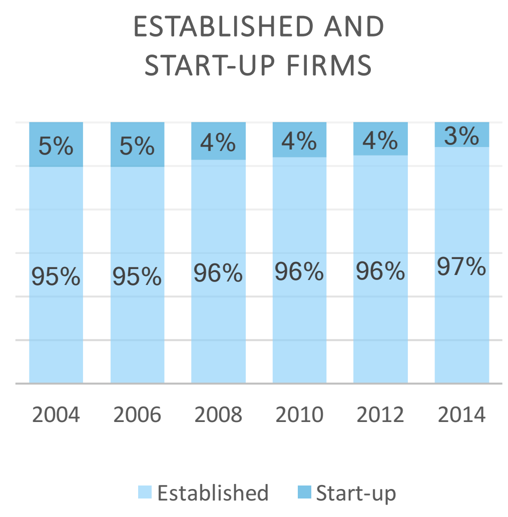


Figure 9 – The distribution of the dataset between established and start-up firms over the years, the vast majority of the firms are older than 2 years.

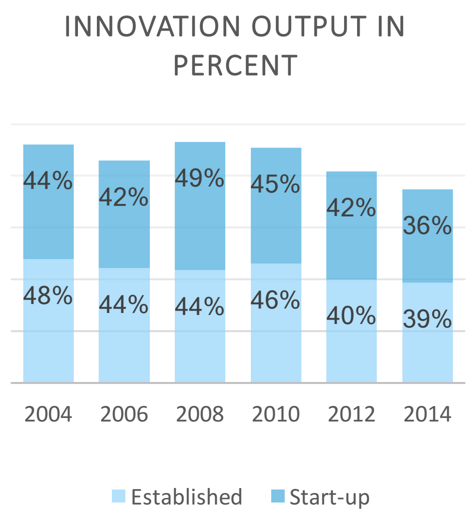


Figure 10 - The percentage of start-up firms and established that have introduced a new or significantly improved product. About 50-60% of both have no innovation output.

Figure 10 - Percentage of firms that are active on the international market

4.2 Descriptive statistics – Regional and Industrial Indicators

When studying the different areas of Sweden, the municipality code that is included in the Statistical Business Register (SBR) (Statistics Sweden , 2017) will be used and matched to the CIS-survey by the company registration number. The different municipality codes of the enterprises are divided into 70 labour market (LM) regions. The regional labour markets are areas of several municipalities that are in some sense independent of other areas concerning supply and demand of labour forces. The division is based on statistics on commuting; every labour market region is a local centre from which less than 20 per cent of the acquisition workers commutes from that region to another. The number of workers commuting from that local centre to another specific municipality is less than 7,5 per cent (Statistics Sweden , 2017).

Figure 11 presents an overview of the size of the different 290 municipalities in Sweden. The size of the bubble represents the population density, were the three metropolitan areas are three largest bubbles. The urban areas are the medium size bubbles and as can be seen, the urban areas are often close to the metropolitan areas. the division done by Swedish Agency for Economic and Regional Growth has been used when dividing the LM-regions into metropolitan, urban and rural areas (Tillväxtverket, 2017).

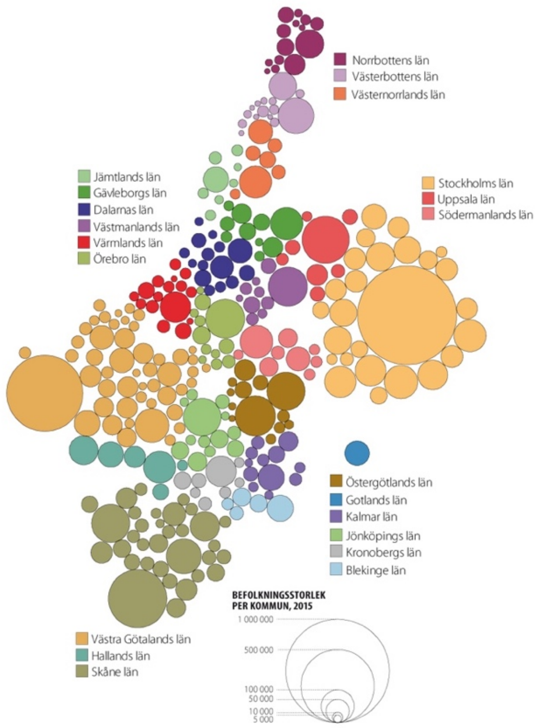


Figure 11 – The size of the bubble represents the population density of the municipality. The three largest bubbles are the three metropolitan areas of Sweden: Stockholm, Malmö and Gothenburg. The smallest bubble represents the rural areas

Table 1 in appendix gives detailed information of how the observations are divided between the different LM-regions. The number of firms that is located in metropolitan, urban and rural areas are plotted in Figure 12. The three major regions, the metropolitans, contains a sum of 51% of the total observations, 37% of the observations are the urban areas and the remaining 12% are from the rural areas. Because the firms are mainly located in the metropolitan and urban areas, the sample represent the distribution of the firms in Sweden well. Though in order to maintain significant results on the rural areas as well, a more even distribution between the different areas is needed. Rural areas, such as Åsele, Vilhelmina, Pajala, Jokkmokk and Överkalix has less than 10 observations over the whole time period, this unfortunately makes it impossible to draw significant results from those LM-regions. Since 30 observations, according to the rule of thumb in statistics, is the number of observations needed to assume normal distribution[4], 14 of my LM-regions will give untrustworthy estimates[5] (Wackerly, et al., 2008). Figure 13 and 14 presents the innovation output in the different areas.

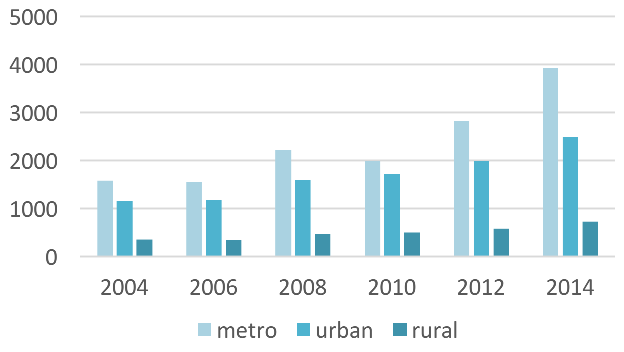


Figure 12- Number of firms that is located in metropolitan, urban and rural areas. The number of firms located in metro areas are significantly more than in urban and rural areas.

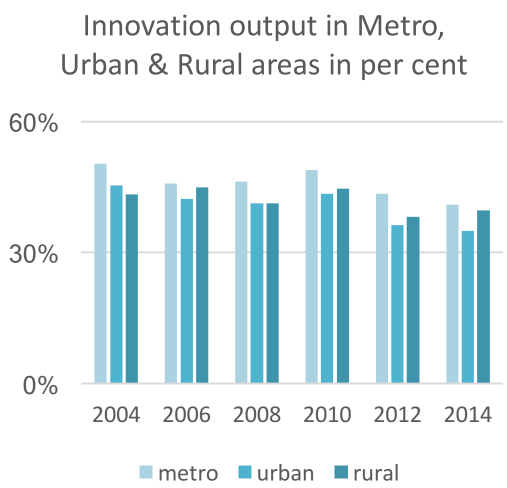


Figure 13 – Percentage of the firms in metropolitan, urban and rural areas that have introduced a new or significantly improved process, good or service.

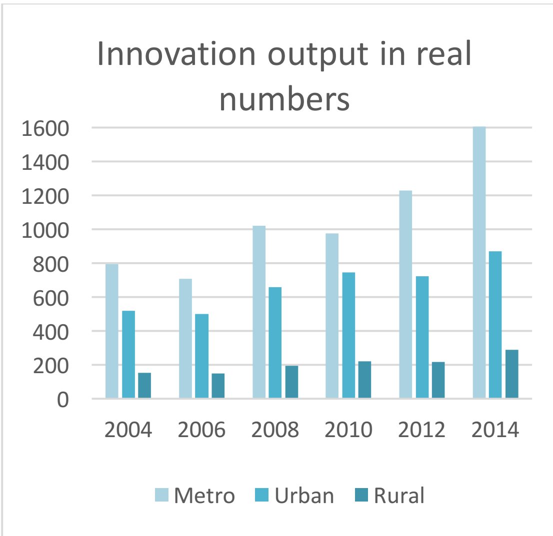


Figure 14- Number of the firms in metropolitan, urban and rural areas that have introduced a new or significantly improved process, good or service.

[4] According to the Central Limit Theorem

[5] For more information of which, please see Appendix table 1.

The industrial codes are as well as the municipality codes maintained from the SBR. The industry-codes are exceedingly informative. The firms can be divided into a five-digit code giving information down to the detail group that the firm belongs to The first two digits give information about the main industry of the firm. Since the study only is interested in the innovation in the different branches the industry codes have been translated into the GICS codes. GICS is short for “Global Industry Classification Standard”, that is an overall division of the branches into 11 sectors. It was developed in 1999 and have gone through some changes since (MSCI, 2017). The advantage of using the overall classification before the industry codes was since the industry codes are interacted with the different LM-regions in order to maintain overall information how the different branches differ between the regions. GICS 1 is the “energy” sector, including the sub-industries of oil and gas drilling, gas and oil production and marketing and other services including oil and gas. GICS 2 is the “material sector” chemicals, construction materials, metals, mining, paper and forest products are some examples of the industries included. GICS 3 is the “industrial” sector, that includes airlines, marine, transportation infrastructure, commercial services, trading companies, building products and electrical equipment among many others. GICS 4 is the “consumer discretionary”, that is consumer products that is purchased occasionally. The sector is including automobiles, household durables, leisure equipment, textiles, apparel, luxury goods, hotels, restaurants etcetera. The fifth GICS is “consumer staples” that includes more daily goods and services, such as food, beverages, tobacco and personal products. The sixth GICS is “health care” that includes health care equipment and supplies, health care providers, health care technology, pharmaceuticals and biotechnology. The financial sector, GICS 7, includes banks, insurance, consumer finance, capital market and real estate investment trusts. GICS 8 is the “information technology” sector, including internet and software services, IT, computers, electronics equipment and office electronics. GICS 9 includes telecommunication services. GICS 10 includes utilities such as electric, gas and water utilities. And the final GICS, 11, includes the “real estate” sector, that is architectural services and real estate agents (Standard & Poor’s, 2006).

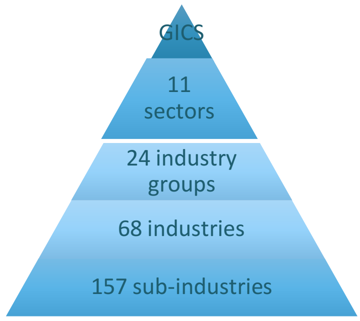


Figure 15- The GICS hierarchy

4.3 The variables

The dependent variable in equation one, R&D, is the participation in R&D a dummy variable that adopt the value one if there are positive values in firm's investments in any sort of innovation activity for example intramural R&D, extramural R&D, acquisition of machinery or expenditure in some other external knowledge. The variable adopts the value zero if there is no R&D expenditure.

Variable innovation is a dummy variable for if the firm managed to introduce a new or significantly improved good or service on the market.

The variable Investment Prediction represent the predicted values of the total expenditure on innovation obtained from equation two. The variable is lagged one year in order to see if positive predicted values on innovation expenditures increase the probability of achieving innovation. Since the data is for every two years, lagging the variable with one step, means that the model will account for the time needed for innovation input to turn into innovation output (OECD, 2009).

Size10 is dummy for if the firm has 10-100 employees. Size100 is a dummy for if the firm has more than 100 employees but less than 1000. Size1000 is dummy for if the firm has more than 1000 employees. The size of the firm is of significance when examining innovation. Pavitt showed in a study in 1987 that firms with less than 1000 employees tend to be more innovative than firms with more employees, but that the relationship is U-shaped, meaning that at some point the correlation between innovation and size of firm will become positive again. The innovativeness in the large firms tend to be more technical driven in the question of product and not process innovation (Pavitt, et al., 1987).

Continuous R&D means that the firm continuously invests in research and development. A dummy that has been transformed from the CIS variable "rdeng" present what type of investment in innovation the firm generally does. The dummy variable obtains the value one if the firm has answered that they continuously invest in R&D. Lagging the variable one step makes it possible to investigating how the continuous investments in innovations will affect the innovation outcomes and still account for the lag between the input and output.

Human capital is a measure of the ratio between the number of employees with higher education and the total number of employees. Knowledge play a crucial role in the production and is the primary source of value. All the human productivity is knowledge based, and the machinery is simply an embodiment of knowledge (Grant, 1996). Studies have shown human capital to have a crucial role in a firm's innovativeness since the knowledge is imbedded in the workers. Higher levels of human capital are expected to have a positive impact on the probability of innovation since that will result in a higher education rate and an increase of the knowledge (Grant, 1996).

The variable "start-up" represent firms that are two years or younger. This is to test if start-up firms are more innovative then already established firms. There have been two sides to this hypothesis, those that claim that already established firms are more innovative because of their opportunity to use existing firm knowledge while others claim new firms to be more innovative because of the efficiency of not needing to filter new knowledge through organizational routines and already fixed and ill-suited structures and the fact that their innovative efforts do not cannibalize their existing products (Katila, 2005). Joseph Schumpeter argued for the positive relationship between start-up firms and innovation. That argument is however not established in the modern literature, instead many recent studies have shown that there is not any positive relationship (Andersson, et al., 2013).

Group variable is a dummy that assumes the value one if the firms is part of a group. That indicates both access to finance as well as intra-group knowledge spill-overs (Goya, et al., 2013). Sweden have thanks to not participating in the two world wars been able to build big world leading groups that invest great amount of money in R&D (Andersson, et al., 2013). Data from Statistics Sweden have in later years shown a positive trend in R&D investments, mainly abroad were the firms uses the foreign R&D departments to evolve new technologies leading to higher efficiency (Andersson, et al., 2013).

International competition is a dummy variable indicating that the firm is operating on an international market when assuming the value one and only on the national market if zero. The increased global competition gives incentives to not fall behind the rest of the world, this is particularly for those firms that are established on the foreign market (Andersson, et al., 2013).

The variable Innovation Prediction is maintained from the Probit regression for equation three. It shows the predicted probability per firm of introducing an innovation to the market. The variable tells if there is any relation between the firm's productivity and the innovation output.

The LM-codes are included in the regressions as explanatory variables. The variable assumes the value one when the firm is located in that region. The variable will indicate if the regions have different impact on innovation activities or productivity.

Table 4 present an overview of the variables:

Table 4- The Variables

Variable	Description
CIS-variables	
Dependent variables	
R&D	1 if the firms have any R&D expenditure 0 otherwise
Innovation	1 if the firm introduced a new or significantly improved good or service 0 otherwise
Innovation investment per employee	This variable shows how much the firm have invested in innovation activities divided by the number of employees.
Productivity	The firms turnover divided by the number of employees
Independent variables	
Group	1 if the firm is part of a group 0 otherwise
Investment prediction	A prediction of the investments in R&D maintained from equation 2, 1 year lag
International competition	1 if the firm is active on the international market 0 otherwise
Innovation prediction	A prediction of the probability of innovation activity maintained from equation 3
Continuous R&D	1 if the firm continuously invests in innovation activities 0 otherwise
Cooperation	1 if the firm cooperates with other firms on innovation activities 0 otherwise
International cooperation	1 if cooperating with consultants, commercial labs or R&D institutes abroad 0 otherwise
FDB-Variables	
Size10	1 if the firm has 10-99 employees 0 otherwise
Size100	1 if the firms has 100-999 0 otherwise
Size1000	1 if the firm has more than 1000 employees 0 otherwise
LM-codes	1 is the firms is located in that LM-region. 0 otherwise
Startup firm	1 if the firm is 2 years or younger 0 otherwise
GICS*LA	Interaction between LA regions and industries
GICS	The main activity of the firm
RAMS-Variables	
Human capital	Share of employees with a university or college degree
Higher studies	Number of employees with higher studies
Lack education	1 if no employees with higher education (used as instrument for human capital) 0 otherwise

4.4 Data diagnostics

Breusch and Pagan Lagrangian multiplier test for random effects and the Hausman specification test was performed in order to check that the model fits the data. The result of the Breush and Pegan test showed that the data does not fit a pooled regression but a random effects estimation. In the choosing between the random and the fixed effects models, the Hausman test showed that the data contains fixed effects.

The correlation matrix shows some correlations between a few of the variables[6], I performed a VIF test[7] to control that the correlations does not mean problems as multicollinearity. Sometimes low VIF values can mean problems as well, but a golden rule is that values under seven is accepted, the test came out clear and as long as the estimations seems accurate the correlations should not induce any bias.

4.5 The method

When preparing the dataset for the study different dataset had to be merged using the firms' company registration number as the matching variable. To maintain the variables on the postal codes, the firms number of employees, the industry codes and the year that the firms was established every year had to be matched separately to the Structural Business Statistics (SBS). The variables of human capital and number of employees with higher degree of education was maintained by matching the dataset to the regional labour statistics based on administrative sources (RAMS). In order to exclude the firms' that have not answered the innovation survey the companies were sorted by one variable included in the CIS while those companies that have no data on that variable were removed. The dataset was then appended so that one large data set was maintained, containing all the variables for every year with up to six years of observations on some firms. The data set was constantly checked so that not any years or observations were lost when merging and appending the dataset. A lot of transformation of the data was needed to be done in order to match the different variables and to maintain the LM-codes from the postal codes as well as the firms' ages and the human capital. All the processing of the data was done in SAS 9.4 while the estimation of the regressions was done in STATA 14.1. The dataset was controlled and any duplicates removed by STATA.

[6] The correlation matrix is presented in Table 4 in Appendix
[7] Presented in Table 5 in Appendix

The CDM-model contains four different equations in three different steps and is commonly considered for introducing simultaneity and sample biases (Lööf, et al., 2003). The estimation method used for the first two equations is the Tobit model were the dependent variable for equation one is unobserved and instead a dummy indicating the effect is used (Lööf, et al., 2003). The Heckman two stage estimation model that estimates the two equations in one step is frequently used for this procedure, unfortunately that model does not suit panel data (Briggs, 2004). Instead the estimations were performed using the Wooldridge method from 1995 were time dummies and inverse Mills ratios are included in order to correct for selection bias and simultaneity (Wooldridge, 1995).

The first equation investigates how the different explanatory variables determine the firm's decision of engaging in R&D. The explanatory variables included are human capital, that is the percentage of the firm's employees that have a higher education. The group dummy variable indicating if the firm is part of a group. Revenue, the age and size of the firm and if the main market for the firm's product are the local market or the market abroad. By also including the different LM-codes and interaction dummies between the LM-codes and GICS industry codes there is an ability to compare the different regions and industries influence on the innovation decision. The first equation is estimated by a random effect (RE) panel data Probit model with bootstrapped standard errors. The RE model is not the most appropriate model for this case since it assumes no correlation between the repressor's and the error term, an assumption that is not held in this case. According to the Hausman specification test the appropriate model is the fixed effects (FE) model which however is not operative together with the Heckman model. Instead the simultaneity bias is corrected for by using the method introduced by Mundlak in 1978. The method includes mean vectors of the time correlated regressors as control variables. The procedure provides a new method for gaining best linear unbiased estimators (BLUE) with data that have correlations between the individual effect and the within-individual effect (Mundlak, 1978). By including the within mean vectors for each firm separately it is possible to maintain more steady and significant estimators with lower standard errors.

The second equation investigates the innovation investment intensity per employee. Since the sample is not random when only looking at the firms with innovation activities, that might give rise to selection bias. The Heckman two stage estimation model that uses inverse Mills ratios (IMR) integrated in model is used in order to correct for the selection bias. One IMR is also created for every time period by running the first probit regression separately for each year, and by predicting the outcome for every year and transform that into the probability function density and the cumulative distribution function and then divide them by each other. When adding them to the pooled regression, the second equation, an indication of the selection bias will appear. The IMR estimators was significant showing the importance of taking the selection problem into consideration. When comparing the estimators from the pooled regression and the Heckman regression they were almost identical (Vartanian, 2013).

The second step and third equation investigates the innovation output dependent on different explanatory variables. The dependent variable is a dummy assuming the value one if the firm introduced a new or significantly improved service or good. The variable is lagged two years and is a prediction over positive innovation investments (from equation one). It is used as the indicator of innovative activities. The equation is estimated by a panel data RE Probit model with the within means per firm included in order to correct for the auto-correlation and bootstrapped corrected standard errors.

When estimating the equations bootstrapped standard errors have been used. The method is a procedure for estimating standard error and have been shown to work well with large sample sizes as well as with non-normal data (Chan, 2009). Since the assumption that the error terms are independently and identically distributed is not always held, tools are needed to correct for the occurred errors and give satisfactory results. In panel data this is often occurs as serial or auto correlations. The bootstrapped standard error are drawn from clusters defined by the id, this will give results similar to those from robust standard errors, though slightly smaller (Guan, 2003).

4.6 THE CDM-EQUATIONS

The CDM-model is a three step procedure with four equations. The regional analysis is conducted by including different functional dummy variables for the labour market regions and interactions between the LM-regions and the industry codes. The “I” in the equations equals the different firms and the “t” is a time-index for every two years from 2004 to2014.

The first equation explains whether a firm is engaged in innovative activities or not by using a Tobit model were the dependent variable in not actually observed but instead a proxy[8] for innovation participation is used. The dependent variable “R&D participation”

[8] More information about the procedure is under the method section

were 1=yes, that is the innovative firms, and 0=no. Equation three will be referred to as " The research equations".

$$R\&D\ Part = \begin{cases} 1\ if\ R\&D^*_{it} = \beta_0 X_{oit}\ \geq 0 \\ 0\ if\ \ R\&D^*_{it} = \beta_0 X_{oit} \leq 0 \end{cases} \quad \text{Eq.3}$$

Were the dependent variables are:

$x_1 = size10$
$x_2 = Human\ capital$
$x_3 = International\ competition$
$x_4 = Start\ up\ firm$
$x_5 = Log\ turnover$
$x_6 = International\ cooperation$
$x_7 = Group$
$x_{8-n} = Differnt\ LA\ regions$

The fourth equation investigates the research investment intensity. The research intensity is measured as the logarithm of the ratio between research expenditure divided by the number of employees at the firm, RDI/emp* is the unobserved latent innovation activity effort estimated as a determine of innovation expenditures.

$$\frac{RDI}{emp_{1it}} = \frac{RDI^*}{emp_{1it}} = \beta_{1it} X_{1it} + \mu\gamma_{it} + \alpha + \varepsilon_{1it}\ \ if\ y_{0it} = 1 \quad \text{Eq.4}$$

$x_1 = size10$
$x_2 = Human\ capital$
$x_3 = International\ competition$
$x_4 = Start\ up\ firm$
$x_5 = Group$
$x_6 = International\ cooperation$
$x_{7-n} = Differnt\ LA\ regions$
$\gamma = Time\ intergrated\ Mills\ inverted\ ratios$
$\alpha = Time\ consistent\ estimator\ correcting\ for\ the\ firm\ heterogeneity$[9]

The fifth equation in the CDM-model studies the innovation output, the equation is referred to as "Innovation equation". It is a Probit equation giving the probability for a firm to engage in innovation activities.

$$P(Innovation)_{it} = \gamma \frac{RDI^*}{emp_{1it-2}} + \beta_{it} X_{it} + \varepsilon_{1it} \quad if \quad y_{0it} = 1 \quad \text{Eq. 5}$$

$x_1 = Group_{t-2}$
$x_2 = Human\ capital$
$x_3 = International\ competition$
$x_4 = Start\ up\ firm$
$x_5 = Size10$

[9] The Wooldridge method

$x_6 = International\ cooperation$
$x_7 = Continous\ expenditurs\ in\ R\&D_{t-2}$
$x_{8-n} = Differnt\ LA\ regions$
$\frac{\gamma RDI}{emp} = Predicted\ value\ from\ equation\ 2\ of\ innovation\ expenditurs$

The sixth equation calculates the productivity according to the Cobb Douglas function, it will be referred to as "the productivity equation". Labour productivity is estimated as the ratio between the logarithm of the firm's turnover divided by the number of employees.

$$P(Labour\ productivity)_{it} = yI + \beta_{3it}X_{1it} + \mu\gamma_{it} + \varepsilon_{1it} \quad if \quad y_{0it} = 1 \qquad \text{Eq. 6}$$

$yI = Is\ the\ predicted\ value\ of\ innovation\ output\ from\ eq\ 3$
$x_2 = Lack\ education$
$x_3 = International\ competition$
$x_4 = Size\ 1000$
$x_5 = Size10$
$x_6 = Group$
$x_6 = International\ cooperation$
$x_7 = Established\ buisness$
$x_{8-n} = Differnt\ LA\ regions$
$\gamma = Time\ intergrated\ Mills\ inverted\ ratios$

5 RESULTS

"We cannot solve a problem by using the same kind of thinking we used when we created them".

- A. Einstein

5.1 FIRST STEP – THE RESEARCH EQUATIONS

In equation 1, it is explained whether a firm practises innovation activities or not. The dependent variable is:

$y_i = 1$ if firm i is engaged in innovation activities

$y_i = 0$ if firm i is not engaged in innovation activities

Equation two is a censored regression that investigates the innovation investment intensity in firms with some sort of innovation activities. That is the equation were $y_i = 1$

The dependent variables are the logarithm of total expenditure of own R&D research, expenditure in buying external R&D research, acquisition of machinery, equipment or software and acquisition of another external knowledge.

Since the observations will not be randomly chosen when only investigating those firms that have any innovation activities selection selectivity and simultaneity biases might arise the model proposed by Wooldridge in 1995 are used that accounts for the selection and simultaneity bias. The marginal effects were supposed to be presented in the table, unfortunately the regression contained to many variables for STATA to perform the command.

Table 5 – The results of the research equations

(Dep. Var)	Engagement in R&D activities (RE Probit)		R&D intensity (Consistent estimator)	
Time period 2004-2014	Coefficient	Bootstrap stn.err	Coefficient	Bootstrap stn.err
Firm size 10-100	-0,4025***	0,0311	0,4387***	0,0854
Firm size >1000	0,2755***	0,0882	-0,1262	0,1250
Human capital	0,0046	0,0533	0.0199***	0,0017
Group	0,0092***	0,0007	0.1342**	0,0691
Start up firm	0,1209***	0,0258	0,4607***	0,1193
International competition	0,6498***	0,0235	0,5184***	0,1190
Int. Cooperation	1,9239***	0,0869	0,9702***	0,2026
Log Turnover	-	-	0,0199	0,0329
Cooperation	-	-	0,3257***	0,0463
Within means	Yes***	-	Yes***	-
Mills Ratios	-	-	Yes***	-
LM- regions	Yes	-	Yes[10]	-
GICS	Yes	-	Yes	-
GICS*LM	-	-	Yes[11]	-
Materials	0,2845***	0,0451	0,8328***	0,2164
Industrials	-0,5710***	0,0488	-0,5092**	0,2368
Consumer Discretionary	0,0469	0,0535	0,2665	0,2474
Consumer Staples	-0,2305***	0,0686	1,5027***	0,2729
Health care	0,1749**	0,0617	0,4079**	0,1944
Financials	-0,1603***	0,0507	0,6010**	0,2363
Information technology	-0,3518***	0,1067	2,0385***	0,5545
Utilities	-0,4569***	0,0640	-0,0835**	0,2459

***The variable is significant at 1% level **Significant at 5% *Significant at 10%
Rho=0,4316

In line with earlier research the results indicate that human capital has a positive effect on both the probability for a firm to invest in R&D as well as the amount invested, an increased number of personnel with higher degree increases the amount spent on R&D. Small firms, with a maximum of 100 employees tend to be less innovative than larger firms. A study conducted by Cohen and Kleppen in 1996 showed that larger firms have an

[10] Only those regions which have shown significant estimates will be included in Table 5, all the estimates can be seen in Appendix Table 5.
[11] Only the significant estimates are presented in this table, Appendix Table 5 includes all estimates

advantage because of the larger output over which they can apply the results. Their results show that R&D expenditures advantage is in expenses for processes relative to product R&D (Cohen & Klepper, 1996). The amount spent on R&D per employee shows a strong positive relationship in the small firms, this is not so surprisingly since the amount spent is divided on a smaller number of people. Larger firms with more than 1000 employees are more likely to invest in innovation activities. When controlling this with the total expenditure spent on R&D without dividing by the number of employees, the smaller firms spent less on R&D. The start-up firm in this thesis confirms Schumpeter's hypothesis of the creative destruction when indicating that the firms are both more likely to invest in innovation as well as showing an increased amount invested in R&D. Being part of a Group shows a small positive impact on the probability of innovation activities. Lööf et al. gained the results that being part of a group decrease the probability of innovation activities in Sweden compared to Germany (Lööf & Heshmati, 2006). The number of firms that are part of a group and included in the CIS have increased a lot the last years together with a higher responding rate, that might be an explanation together with the fact that the groups innovativeness might change over the years. As presented earlier in Figure 5, the last ten years the R&D expenditure in Group has increased with 10 billions of SEK. Firms that are active on the foreign market tend to be more innovative compared to those only active on the national market. Increased competition means that the firm need to introduce more differentiated products, so when being active on the international market the firms need to invest more in R&D in order to keep up with the competitors (Klepper, 1996).

Unfortunately, most of the regions gave insignificant results, meaning that it is not impossible that the estimates in fact are equal to zero. This is because of the lack of enough observations in some regions. Though, those estimates giving significant results show some variations between the different regions and industries.

All the metropolitan areas show positive probability of investing in research activities. A distinguish feature of the probability of investment in research in both the rural and urban areas, are that the large regions tend to show a positive probability of investing in innovation, while the smaller regions such as[12] Kiruna[u], Gällivare[u], Överkalix[r], Strömsund[r] and Gotland[r*] show to have a negative effect on the probability of research activities. Unfortunately only Gotland have significant estimator. When it comes to investment in research the health sector along with the material sector tend to be more likely to invest in innova-

[12] Raised to U indicates urban area, raised to R indicates rural area

tion activity. While among those who invest in innovation, industrial sector along with utilities tend to invest much less than the other. When investigating the intensity invest in innovations, the material industry tend to be more innovative in urban areas than in the metropolitan and rural areas. The material industry shows a clear pattern of higher investment in urban areas in the southern Sweden except from Umeå that tend to invest more. The industry sector did not contain enough observations for all locations, but seemed overall to invest much less than the other sectors. The rural area Värnamo invest most in research activities while Jönköping and Borås that is close by tend to invest much less. The consumer discretionary sector also shows large differences between the different regions. The metropolitan areas Stockholm and Gothenburg differn a lot when it comes to consumer discretionary, Stockholm show a positive estimate while Gotenburg tend to invest much less than most of the other regions, even so that it shows a negative estimate and the neighbouring region Borås tend to follow. The northern regions of Sweden invest much less in innovations in this sector then the southern parts. The health care sector only provided information from a few regions, the metropolitan areas Malmö-Lund and Stockholm is much more innovative compared to Borås which gives a large negative estimate on this sector as well. In the financial sector the metropolitan areas Stockholm and Gothenburg is investing much less in innovation activities than the urban and rural areas. In the real estate sector, Stockholm is the only region showing negative estimators, urban areas tend to invest more.

The Mills Ratios gave all significant results, indicating the importance of taking the simultaneity bias into consideration.

5.2 Second Step - The innovation equations

The second step studies the innovation output. The dependent variable Innovation indicates the innovation output of the firms and the equation is a Probit equation explaining how the different independent variables are affecting the probability of innovation output. The numbers reported is the marginal effects, the advantage of using the marginal effect over the coefficients are that the marginal effects show the size of the effect and not just in which direction the effect goes (Gujarati, 2012).

Table 6- The results of the innovation equations

(Dep. Var)	Innovation (Re Probit)	
Time period 2004-2014	**Marginal effects**	**Bootstrap stn.err**
Lag PRDI	0,0385**	0,0158

Lag group	0,0018	0,0135
Human Capital	0,0015***	0,0015
Lag Size 10-100	-0,1170***	0,0146
Lag Continuous R&D	0,1183***	0,0126
International competition	0,0853***	0,0133
International cooperation	0,3366***	0,0272
Time dummies	Yes	-
Means	Yes*	-
LM-regions	Yes*	-
Industry dummies	No	-
Interaction industry & LM	No	-
***The variable is significant at 1% level **Significant at 5% *Significant at 10%		
Rho=0,4316 Observations: 11 385		

By lagging the variables two year it is possible to measure how inputs is affecting the output in the same estimation. Because the panel is unbalanced, a lot of observations will be unable to include because there is no observations two years back.

The smaller firms 10-100 employees tend to be less likely to provide innovation output than the larger firms, if the firm has under 100 employees the probability of innovation output decreases by 12%. Something that is interesting is that the variable lag continuous R&D show that if the firm continuously invest in R&D two years back, the probability of innovation output increases by almost 12 percent. International competition and international cooperation both increases the probability of innovation output, cooperation as much as 34%, this is in line with earlier research as discussed in section 5.1. The regression also shows that the prediction of the R&D intensity of from equation 4 mean that an increased investment in innovation two years back will increase the probability of innovation output with about 4%, meaning that there is a casualty between the amount invested in innovation activities and innovation output.

The whole table can be seen in appendix table 8. The regional analysis shows that the metropolitan areas Stockholm, Gothenburg and Malmö are less likely to introduce a new or significantly improved good compared to the urban areas. The probability of innovation output increases in the urban areas and decreases in rural and metropolitan areas. Except from the rural areas in the north were the probability of innovation activities increases.

Among the industries the real estate sector along with the consumer discretionary and health sector is less likely to provide innovation output.

5.3 THIRD STEP – THE PRODUCTIVITY EQUATIONS

The productivity equations measure the firms labour productivity, that is the turnover divided by the number of employees, it is estimated by a random effect Probit model.

Table 7 – The productivity regression

(Dep. Var)	Labour productivity (Re Probit)	
Time period 2004-2014	**Coefficient**	**Bootstrap stn.err**
Prediction innovation output	0,0186***	0,0075
Size 10-100	0,2354***	0,0177
Size <1000	-0,0508***	0,0059
Group	-0,0509***	0,0067
Lack education	-0,0928***	0,0104
Established	-0,0031	0,0073
Means	Yes	
Time dummies	Yes	
***The variable is significant at 1% level **Significant at 5% *Significant at 10%		
Rho=0,4316 Observations: 17635		

The predicted innovation output shows that the predicted innovation output has a positive effect in the firms productivity. If the firms are expected to be innovative, the firm's probability of being productive increases. Small firms tend to be more productive then large firms and being part of a group decreases the probability of being productive. If the firm don't have any personnel with higher education, the firm is less productive as well. The fact that established firms tend to be less likely to be productive is something that confirms the Schumpeter hypothesis of creative destruction.

Among the different regions in Sweden, the regions that are associated with clusters are more likely to be productive than the most other, that is Stockholm, Gothenburg, Trollhättan this is something that also is in line with the study of productivity in Swedish clusters (Borgman & Braunerhjelm, 2004). The rural areas are less likely to be productive, especially those located in the northern part of Sweden. The regression shows that firms located in urban areas, it tends to be more likely to be productive.

6 CONCLUSION

"Innovation has nothing to do with how many R&D dollars you have. When Apple came up with the Mac, IBM was spending at least 100 times more on R & D. It's not about money. It's about the people you have, how you're led, and how much you get it."

- Steve Jobs

This study has analysed the impact that different regions in sweden has on the posibillity of investing in innovation, probability of innovation output and the probability of firms being productive, and how innovation and productivity is connected. The follow questions have been aimed to answer: *Are firms as innovative in metropolitan, urban and rural areas in Sweden? Are there any regional or industrial differences in investment? Are there any differences in the productivity between different areas and industries?* And the study have found some interesting connections. First of all, agglomeration were people move towards a centre, this is the urban and metropolitan areas, as well as clusters. This seem to increase the productivity significantly while the probability to innovation output seem to decrease in the metropolitans. The areas that are most likely to present a innovative good are the urban area, something that might be associated with that a moderate level of competition is increasing the need to provide new goods in order to stay on top. In the metropolitan areas the might be a higher level of spill over effects, making it easier to the firm to benefit from another firm's innovativeness. The higher level of competition in the metropolitan areas might as well make it harder to be the first on the market with a new invention. Even if there are some clear deviations between the different areas of Sweden, especially does the northern parts deviate. This is something that also is associated with the population density, since those part are much more sparsely populated the possibility to spill over effects decrease, making the northern rural areas less innovative and less productive. Location is of importance, but the size or the firm, the level of human capital and the firms age is of most importance. In all stages these variables have shown to have a large effect on innovation and productivity and the deviation between the different industries are of importance. The are some deviation between the different industries in the regions, but the largest difference is between the innovativeness in the different industries.

Unfortunately many of the region estimates gave insignificant results and in order to gain more significant estimators, more observations are needed from the rural areas. For future research a more detailed analysis of the different industries in the different regions would be of great interest, perhaps a other estimation method is needed to provide sufficient information.

As a concluding remark I like to say that the study seems to both agree and disagree with Steve Jobs innovation quote. The amount spent on innovation have shown significant results of increasing the innovation output which increases the productivity. But he is right in the fact that the human capital is of significant importance. So innovation is not just the amount spent, but as well being surrounded by higher knowledge and of course, different spill over effects from firms nearby.

REFERENCES

Andersson, M. et al., 2013. *Det Innovativa Sverige Sverige som kunskapsnation i en internationell kontext.* Stockholm: ESBRI och Brunzell Design.

Andersson, M. et al., 2015. *Det innovativa Sverige 2 Innovation och attraktion i stad och på landsbygd.* 2 ed. Stockholm: ESBRI och Brunzell Design.

Baum, C. F., Lööf, H. & Nabavi, P., 2015. A new approach to estimation of the R&D–innovation–productivity relationship. *Economics Of Innovation And New Technology*, 26(1-2), pp. 121-133.

Borgman, B. & Braunerhjelm, P., 2004. Geographical Concentration, Entrepreneurship and Regional Growth: Evidence from Regional Data in Sweden, 1975-99. *Regional Studies*, November, 38(8).

Braunerhjelm, P. & Borgman, B., 2004. *Geographical Concentration, Entrepreneurship and Regional Growth: Evidence from Regional Data in Sweden, 1975–99,* Stockholm: Center for Business and Policy Studies (SNS).

Braunerhjelm, P. et al., 2011. *Ett innovationspolitiskt ramverk - Ett steg vidare ,* Stockholm: Entreprenörskapsforum.

Briggs, D., 2004. Causal Inference and the Heckman model. *Journal of Educational and Behavioral Statistics*, Vinter, 29(4), pp. 397-420.

Caballero, R. J. & Hammour , M., 2000. *CREATIVE DESTRUCTION IN DEVELOPMENT: INSTITUTIONS, CRISES, AND RESTRUCTURING1.* Washington, DELTA.

Chan, W., 2009. Bootstrap Standard Error and Confidence Intervals for the Difference Between Two Squared Multiple Correlation Coefficients. *Educational and Psychological Measurement*, August, 69(4), pp. 566-584.

Cohen, W. & Klepper, S., 1996. Firm Size And The Nature Of Innovation Within Industries: The Case Of Process And Product R&D. *The Review of Economics and Statistics*, 78(2), pp. 232-243.

Cook, G. A., Pandit, N. R., Lööf, H. & Johansson, B., 2013. Clustering, MNEs, and Innovation: Who Benefits and How?. Vol. 20(No. 2), p. 203–22.

Crépon, B., Duget, E. & Mairesse, J., 1998. Research, innovation and productivity- an econometric analysis at the firm level. *National Bureau of Economic Research ,* Volume 6696.

Criscuolo, C., 2009. Innovation and Productivity: Estimating the Core Model 1 Across 18 Countries. In: *Innovation In Firms: A Microeconomic Perspective.* London: OECD.

Davis, S. J., Schuh, S. & Haltiwanger, J., 1993. Small business and job creation: dissecting the myth and reasessing the facts. *NBER working paper series* , October, Volume 4492.

European Commission, 2016. *Europa 2020.* [Online]
Available at: http://ec.europa.eu/europe2020/index_sv.htm
[Accessed 08 02 2017].

Eurostat, 2012. *Eurostat statistics explained.* [Online]
Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Urban-rural_typology
[Accessed 01 03 2017].

Eurostat, 2016. *Glossary:Nomenclature of territorial units for statistics (NUTS).* [Online] Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Nomenclature_of_territorial_units_for_statistics_(NUTS) [Accessed 01 03 2017].

Foster, L., Haltiwanger, J. & Krizan, C., 2001. Aggregate Productivity Growth Lessons from Microeconomic Evidence. In: U. o. C. Press, ed. *New Developments in Productivity Analysis.* s.l.:The National Bureau of Economic Research, pp. p. 303 - 372.

Fujita, M., 1996. *Economics of Agglomeration.* Pennsylvania: Kyoto University and University of Pennsylvania.

Fujita, S., 2008. Creative Destruction and Aggregate Productivity Growth. *Business Review*, pp. 12-20.

Goya, E., Vayá, E. & Suriñach, J., 2013. Do spillovers matter? CDM model estimates for Spain using panel data. *Search Working Paper*, 4(28).

Grant, R. M., 1996. Toward A Knowledge-Based Theory Of The Firm. *Strategic Management Journal*, Winter Special Issue, Volume 17, pp. 109-122.

Griliches, Z., 1998. R&D and Productivity: The Unfinished Business. In: *R&D and Productivity: The Econometric Evidence.* Chicago: University of Chicago Press, pp. p. 269 - 283.

Guan, W., 2003. From the help desk: Bootstrapped standard errors. *The Stata Journal* , Volume 1, pp. 71-80.

Gujarati, D., 2012. *Econometrics by Example.* New York: Palgrave Macmillan.

Heyman, F., Norbäck, P.-J. & Persson, L., 2013. *Var skapas jobben? En ESO-rapport om dynamiken i svenskt näringsliv 1990 till 2009,* Stockholm : Finansdepartementet .

Johansson, D. et al., 2014. *Företagsskattekommittén och Entreprenörskapet – En kommentar till företagsskattekommitténs arbete med utgångspunkt i effekterna på entreprenörskapet i sverige,* Stockholm: Entreprenörskapsforum.

Kaiserfeld, T., 2005. *A review of Theories of Invention and Innovation,* Stockholm: CESIS Royal Institute of Technology .

Karafillis, C. & Papanagiotou, E., 2010. Innovation and total factor productivity in organic farming. *Applied Economics*, 43(23), pp. 3075-3087.

Katila, R., 2005. When Does Lack Of Resources Make New Firms Innovative?. *Academy of Management Journal*, 48(5), pp. 814-829.

Klepper, S., 1996. Entry, Exit, Growth, and Innovation over the Product Life Cycle. *The American Economic Review*, 86(3), pp. 562-583.

Lööf, H. & Heshmati, A., 2006. On the relationship between innovation and performance: A sensitivity analysis. *Economics of Innovation and New Technology*, 15(4), pp. 317-344.

Lööf, H., Mairesse, J. & Mohnenc, P., 2017. CDM 20 years after. *Economics of innovation and new technology,* 26(1-2), pp. 1-5.

Lööf, H., Peters , B. & Janz, N., 2003. Firm Level Innovation And Productivity- Is There A Common Story Across Countries?. *Centre for European Economic Research*.

Mazzucato, M., 2013. *The Entrepreneurial State: Debunking Public vs. Private Myths in Risk and Innovation.* London: Anthem Press.

MSCI, 2017. *MSCI.* [Online]
Available at: https://www.msci.com/gics
[Accessed 16 05 2017].

Mundlak, Y., 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica*, January, 46(1), pp. 69-85.

OECD, 2005. *Oslo Manual - Guidelines For Collecting And Interpreting Innovation Data.* 3 ed. Paris: European Commission.

OECD, 2009. *Innovation in firms: A microeconomic perspective.* Paris: OECS.

Palmberg, J. & Backman, M., 2015. Contextualizing small family firms: How does the urban–rural context affect firm employment growth?. *Journal of Family Business Strategy,* Volume 6, pp. 247-258.

Pavitt, K., Robson, M. & Townsend, J., 1987. The size distribution of innovative firms in the UK: 1945-1983. *The Journal of Industrial Economics*, March, XXXV(3), pp. 297-316.

Schumpeter, J. A., 1911/1934. *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle.* Cambridge: Harvard University Press.

Schumpeter, J. A., 1942. *Capitalism, Socialism, and Democracy.* New York: Harper & Bros.

Solow, R. M., 1957. Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, August, 39(3), pp. 312-320.

Standard & Poor's, 2006. *Global Industry Classification Standard (GICS®).* [Online]
Available at: http://www.unm.edu/~maj/Security%20Analysis/GICS.pdf
[Accessed May 2017].

Statistics Sweden , 2017. *Metoden att skapa lokala arbetsmarknader (LA).* [Online]
Available at:
http://www.sverigeisiffror.scb.se/contentassets/c2d754bcaf964bcca33ac7cc2510c765/metoden-att-skapa-lokala-arbetsmarknader.pdf
[Accessed 14 04 2017].

Statistics Sweden , 2017. *Statistical Business Register.* [Online]
Available at: http://www.scb.se/nv0101-en
[Accessed 27 03 2017].

Statistics Sweden, 2012. *Innovationsverksamhet i Sverige (CIS).* [Online]
Available at:
http://www.scb.se/contentassets/9e6a00ac2fc7421cabab329528166232/uf0315_bs_2014_as_151216.pdf
[Accessed 01 03 2017].

Statistics Sweden, 2013. *Företagens finansieringsmöjligheter.* [Online]
Available at: http://www.scb.se/sv_/Hitta-statistik/Artiklar/Foretagens-finansieringsmojligheter/
[Accessed Maj 2017].

Statistics Sweden, 2014. *Innovationsversamhet i Sverige,* Stockholm: Statistics Sweden.

Statistics Sweden, 2017. *Här bor sveriges befolkning.* [Online]
Available at: http://www.scb.se/Grupp/Artiklar/Valfard-2017-1/Har-bor-sveriges-befolkning/Har-bor-Sveriges-befolkning-stor.jpg
[Accessed 29 03 2017].

Statistics Sweden, 2017. *SCB Registerbaserad arbetsmarknadsstatistik.* [Online]
Available at: http://www.scb.se/sv_/Vara-tjanster/Bestalla-mikrodata/Vilka-mikrodata-

finns/Registerbaserad-arbetsmarknadsstatistik-RAMS-/
[Accessed may 2017].

Statistics Sweden, 2017. *Statistics Sweden Innovationsverksamhet i svenska företag.* [Online]
Available at: http://www.scb.se/sv_/Vara-tjanster/Bestalla-mikrodata/Vilka-mikrodata-finns/Innovationsverksamhet-i-svenska-foretag/
[Accessed 08 02 2017].

Swann, P. & Baptista, R., 1998. Do firms in clusters innovate more?. *Research Policy*, Volume 27, pp. 525-540.

Tillväxtanalys, 2013. *Svenska koncerner med dotterbolag i utlandet 2013,* s.l.: Tillväxtanalys, SCB.

Tillväxtanalys, 2015. *Forskning och utveckling i internationella företag 2015,* Östersund: Tillväxtanalys, SCB.

Tillväxtanalys, 2016. *Hänt i världens Våren 2016 Innovation i offentligt sektor ,* Östersund: Tillväxtanalys .

Tillväxtverket, 2017. *Regiontyper.* [Online]
Available at: https://tillvaxtverket.se/statistik/regional-utveckling/regionala-indelningar/regiontyper.html
[Accessed 05 2017].

Wackerly, D. D., Mendelhall, W. & Scheaffer, R. L., 2008. *Mathematical Statistics with Applications.* Florida: Brooks/Cole.

Vartanian, T. P., 2013. *Generating and Using the Inverse of the Mills Ratio in Heckman Selection Models Vartanian: SW 683.* [Online]
Available at:
http://www.brynmawr.edu/socialwork/GSSW/Vartanian/Handouts/Inverse%20Mills.pdf
[Accessed 05 2017].

Wipo, 2016. *The Global Innovation Index 2016 - Winning with Global Innovation.* [Online]
Available at: http://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2016.pdf
[Accessed 23 03 2017].

von Hippel, E., 2005. *Democratizing Innovation.* Cambridge: Massachusetts Institute of Technology .

Wooldridge, J., 1995. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* , Volume 68, pp. 115-132.

APPENDIX

Appendix Table 1 – The distribution of the observations between the different LM-regions

LA	Region	Freq.	%	LA	Region	Freq	%
1401	Stockholm-Solna	7867	29,0%	1439	Vansbro	27	0,10%
1403	Eskilstuna	330	1,20%	1440	Malung-Sälen	36	0,13%
1404	Linköping	668	2,42%	1441	Mora	83	0,30%
1405	Norrköping	402	1,46%	1442	Falun-Borlänge	377	1,37%
1406	Jönköping	857	3,11%	1443	Avesta-Hedemora	94	0,34%
1407	Värnamo	527	1,91%	1444	Ludvika	88	0,32%
1408	Vetlanda	166	0,60%	1445	Ljusdal	71	0,26%
1409	Älmhult	124	0,45%	1446	Gävle	302	1,10%
1410	Växjö	237	0,86%	1447	Söderhamn	67	0,24%
1411	Ljungby	296	1,07%	1448	Bollnäs-Ovanåker	118	0,43%
1412	Kalmar	301	1,09%	1449	Hudiksvall	139	0,50%
1413	Oskarshamn	125	0,45%	1450	Sundsvall	409	1,48%
1414	Västervik	123	0,45%	1451	Kramfors	85	0,31%
1415	Vimmerby	124	0,45%	1452	Sollefteå	33	0,12%
1416	Gotland	108	0,39%	1453	Örnsköldsvik	235	0,85%
1417	Karlskrona	192	0,70%	1454	Strömsund	33	0,12%
1418	Karlshamn-Olofström	96	0,35%	1455	Härjedalen	41	0,15%
1419	Malmö-Lund	2928	10,62%	1456	Östersund	317	1,15%
1421	Halmstad	379	1,38%	1457	Storuman	26	0,09%
1422	Bengtsfors-Dals-Ed	28	0,10%	1458	Dorotea	14	0,05%
1423	Göteborg	3289	11,93%	1459	Vilhelmina	8	0,03%
1424	Strömstad	60	0,22%	1460	Åsele	5	0,02%
1425	Trollhättan-Vänersborg	404	1,47%	1461	Umeå	386	1,40%
1426	Borås	682	2,47%	1462	Lycksele	58	0,21%
1427	Lidköping-Götene	283	1,03%	1463	Skellefteå	8	0,03%
1428	Skövde	465	1,69%	1464	Arvidsjaur	274	0,99%
1429	Torsby	89	0,32%	1465	Arjeplog	11	0,04%
1430	Årjäng	46	0,17%	1466	Jokkmokk	8	0,03%
1431	Karlstad	498	1,81%	1467	Överkalix	7	0,03%
1432	Filipstad	39	0,14%	1468	Övertorneå	10	0,04%
1433	Hagfors	28	0,10%	1469	Pajala	6	0,02%
1434	Arvika-Eda	72	0,26%	1470	Gällivare	13	0,05%
1435	Hällefors	15	0,05%	1471	Luleå	505	1,83%
1436	Örebro	649	2,35%	1472	Haparanda	26	0,09%
1437	Västerås	613	2,22%	1473	Kiruna	81	0,29%
1438	Fagersta	65	0,24%				

Appendix Table 2 – The division between rural urban and metro

Urban		Rural		Metropolitan	
LM	Municipality	LM	Municipality	LM	Municipality
1403	Eskilstuna	1407	Värnamo	1401	Stockholm-Solna
1404	Linköping	1411	Ljungby	1419	Malmö-Lund
1405	Norrköping	1415	Vimmerby	1423	Göteborg
1406	Jönköping	1416	Gotland		
1408	Vetlanda	1422	Bengtsfors-Dals-Ed		
1409	Älmhult	1424	Strömstad		
1410	Växjö	1427	Lidköping-Götene		
1412	Kalmar	1429	Torsby		
1413	Oskarshamn	1430	Årjäng		
1414	Västervik	1439	Vansbro		
1417	Karlskrona	1440	Malung-Sälen		
1418	Karlshamn-Olofström	1441	Mora		

1421	Halmstad	1445	Ljusdal
1425	Trollhättan-Vänersborg	1447	Söderhamn
1426	Borås	1448	Bollnäs-Ovanåker
1428	Skövde	1449	Hudiksvall
1431	Karlstad	1451	Kramfors
1432	Filipstad	1452	Sollefteå
1433	Hagfors	1454	Strömsund
1434	Arvika-Eda	1455	Härjedalen
1435	Hällefors	1456	Östersund
1436	Örebro	1457	Storuman
1437	Västerås	1458	Dorotea
1438	Fagersta	1459	Vilhelmina
1442	Falun-Borlänge	1460	Åsele
1443	Avesta-Hedemora	1464	Arvidsjaur
1444	Ludvika	1465	Arjeplog
1446	Gävle	1466	Jokkmokk
1450	Sundsvall	1467	Överkalix
1453	Örnsköldsvik		
1461	Umeå		
1462	Lycksele		
1463	Skellefteå		
1468	Övertorneå		
1469	Pajala		
1470	Gällivare		
1471	Luleå		
1472	Haparanda		
1473	Kiruna		

Appendix Table 3 – The GICS division

GICS 1 = Energy	GICS 4 = Consumer Discretionary	GICS 7 = Financials	GICS 10 = Utilities
GICS 2 = Material	GICS 5 = Consumer Staples	GICS 8 = Information Technology	GICS 11 =Real Estate
GICS 3 = Industry	GICS 6= Health Care	GICS 9 = Telecommunication Services	

Appendeix Table 4 – Descriptive statistics

Variable		Mean	Std,Dev	Min	Max	Observations	
Group (0/1)	overall	0,717	0,450	0	1	N=	27335
	between		0,434	0	1	n=	12864
	within		0,196	-0,116	1,550	T-bar=	2,125
Humancapital	overall	16,713	19,735	0	100	N=	27335
	between		20,324	0	100	n=	12864
	within		3,494	-20,243	73,237	T-bar=	2,125
Metro (0/1)	overall	0,515	0,500	0	1	N=	27335
	between		0,497	0	1	n=	12864
	within		0,048	-0,318	1,349	T-bar=	2,125
Urban (0/1)	overall	0,371	0,483	0	1	N=	27335
	between		0,477	0	1	n=	12864
	within		0,049	-0,463	1,204	T-bar=	2,125
Rural (0/1)	overall	0,108	0,311	0	1	N=	27335
	between		0,305	0	1	n=	12864
	within		0,028	-0,725	0,942	T-bar=	2,125
Start-up (0/1)	overall	0,039	0,193	0	1	N=	27335
	between		0,206	0	1	n=	12864
	within		0,111	-0,628	0,872	T-bar=	2,125
Established (0/1)	overall	0,961	0,193	0	1	N=	27335
	between		0,206	0	1	n=	12864
	within		0,111	0,128	1,628	T-bar=	2,125
Cooperation (0/1)	overall	3,223	4,037	0	9	N=	20423
	between		3,893	0	9	n=	11062
	within		2,078	-2,777	10,723	T-bar=	1,846
Int.Competition (0/1)	overall	0,569	0,495	0	1	N=	27335
	between		0,473	0	1	n=	12864
	within		0,201	-0,264	1,403	T-bar=	2,125
Int.Cooperation (0/1)	overall	0,046	0,209	0	1	N=	27335
	between		0,142	0	1	n=	12864
	within		0,142	-0,788	0,879	T-bar=	2,125
Lack education (0/1)	overall	0,190	0,393	0	1	N=	27335
	between		0,398	0	1	n=	12864
	within		0,150	-0,643	1,024	T-bar=	2,125
Size10 (0/1)	overall	0,774	0,418	0	1	N=	27335
	between		0,365	0	1	n=	12864
	within		0,113	-0,059	1,608	T-bar=	2,125
size100 (0/1)	overall	0,204	0,403	0	1	N=	27335
	between		0,350	0	1	n=	12864
	within		0,126	-0,630	1,037	T-bar=	2,125
	overall	0,025	0,155	0	1	N=	27335

size1000 (0/1)	between		0,121	0	1	n=	12864
	within		0,053	-0,809	0,858	T-bar=	2,125
Employeed	overall	153,744	663,570	0	29175	N=	27335
	between		506,487	0	23661	n=	12864
	within		177,891	-9908	19264	T-bar=	2,125
	overall	10,977	2,157	0	18,629	N=	27335
LogTurnover	between		1,990	0,000	18,375	n=	12864
	within		0,649	-2,388	18,786	T-bar=	2,125
	overall	493929,2	3037450,000	0	123	N=	27335
Turnover	between		2064977,000	0	97100	n=	12864
	within		714858,200	-36	36700	T-bar=	2,125
	overall	3,760	1,309	0	10,281	N=	27335
Log Empoloyeed	between		1,192	0	10,072	n=	12864
	within		0,209	-1,243	7,940	T-bar=	2,125
	overall	0,426	0,495	0	1	N=	27335
Innovationoutput	between		0,428	0	1	n=	12864
	within		0,295	-0,407	1,260	T-bar=	2,125
	overall	0,192	0,394	0	1	N=	27335
Serviceoutput	between		0,333	0	1	n=	12864
	within		0,248	-0,641	1,025	T-bar=	2,125
	overall	0,262	0,440	0	1	N=	27335
Goodoutput	between		0,370	0	1	n=	12864
	within		0,245	-0,571	1,096	T-bar=	2,125
	overall	0,199	0,399	0	1	N=	27335
Processoutput	between		0,322	0	1	n=	12864
	within		0,270	-0,634	1,032	T-bar=	2,125

Appendix Table 5 – Correlation Matrix

	gp	human	startup	estab	co	intcomp	intco	lackedu	size10	size100	size1000	anst	lnoms	oms	lnanst	inno
gp	1,00															
Human capital	0,07	1,00														
startup	-0,02	0,04	1,00													
estblished	0,02	-0,04	-1,00	1,00												
co	0,08	0,10	-0,01	0,01	1,00											
intcomp	0,12	0,03	-0,05	0,05	0,12	1,00										
intco	0,07	0,15	0,01	-0,01	0,35	0,13	1,00									
lackedu	-0,22	-0,36	0,03	-0,03	-0,14	-0,11	-0,10	1,00								
size10	-0,28	-0,01	0,04	-0,04	-0,20	-0,09	-0,19	0,25	1,00							
size100	0,25	0,00	-0,04	0,04	0,16	0,08	0,12	-0,23	-0,89	1,00						
size1000	0,10	0,01	-0,01	0,01	0,12	0,03	0,17	-0,08	-0,32	-0,13	1,00					
anst	0,11	0,03	-0,02	0,02	0,12	0,05	0,21	-0,09	-0,34	0,07	0,64	1,00				
lnoms	0,33	-0,07	-0,15	0,15	0,16	0,17	0,17	-0,22	-0,59	0,46	0,35	0,36	1,00			
oms	0,08	0,04	-0,03	0,03	0,10	0,04	0,18	-0,07	-0,24	0,05	0,45	0,77	0,35	1,00		
lnanst	0,35	-0,01	-0,07	0,07	0,22	0,13	0,23	-0,31	-0,85	0,64	0,52	0,54	0,71	0,38	1,00	
inno	0,05	0,10	0,02	-0,02	0,09	0,03	0,06	-0,08	-0,06	0,04	0,04	0,04	0,04	0,03	0,06	1,00
rtot	0,03	0,06	-0,01	0,01	0,06	0,03	0,14	-0,03	-0,10	0,00	0,23	0,65	0,15	0,72	0,19	0,02

Appendix Table 6 – VIF Values

Variable	VIF	1/VIF
Log employ	5.62	0,18
Innovation output	5.16	0,19
Number of employees	3.34	0,30
size10	3.27	0,31
Coopertation	2.73	0,37
Turnover	2.27	0,44
Good output	2.12	0,47
Log turover	2.05	0,49
Size1000	1.88	0,53
Process output	1.55	0,64
Service output	1.52	0,66
Lack education	1.38	0,72
Human Capital	1.28	0,78
International competition	1.16	0,86
Internation cooperation	1.13	0,89
Start up	1.03	0,98
Mean VIF 2,34		

Appendix Table 7 – All Estimates from step 1 in the CDM-model

Year: 2004-2014		R&D intencity		R&D investment (RE Probit)	
		Coefficients	Stn.Err	Coefficients	Stn.Err
	size10	0,4387***	0,0865	-0,4025***	0,0311
	size1000	-0,1262	0,1250	0,2755***	0,0882
	startup	0,4607***	0,1193	0,0046	0,0533
	Human capital	0,0193***	0,0019	0,0092***	0,0007
	gp	0,1342**	0,0691	0,1209***	0,0258
	intcomp	0,5184***	0,1190	0,6498***	0,0235
	intco	0,9702***	0,2026	1,9239***	0,0869
	Cooperation	0,3256***	0,0367	-	-
GICS	2	0,8328***	0,2164	0,2845***	0,0451
	3	-0,5092**	0,2368	-0,5710***	0,0488
	4	0,2665	0,2474	0,0469	0,0535
	6	1,5027***	0,2729	-0,2305***	0,0686
	7	0,4079**	0,1944	0,1749**	0,0617
	8	0,6010**	0,2363	-0,1603***	0,0507
	9	2,0385***	0,5545	-0,3518***	0,1067
	11	-0,0835**	0,2459	-0,4569***	0,0640
Labour Market	1403	-1,4835**	0,6155	0,2427**	0,1151
	1404	-0,5339	0,4988	0,2739***	0,0816
	1405	-0,5631	0,4050	-0,0839	0,1054
	1406	0,0343	0,3621	0,0919	0,0740
	1407	-2,3127**	0,9681	0,3746***	0,0923
	1408	-0,6618	0,6532	0,0876	0,1586
	1409	-1,3954**	0,6912	0,4651***	0,1812
	1410	-1,5087***	0,5414	0,1025	0,1357
	1411	2,1073**	0,9543	0,3315***	0,1196
	1412	0,7305	0,6133	0,1660	0,1193
	1413	-0,1638	0,6047	-0,1592	0,1913
	1414	-0,3357	0,4334	0,4385**	0,1895
	1415	-0,3379	0,5447	0,1001	0,1823
	1416	-1,2457	0,8820	-0,3307	0,2058
	1417	-1,0862**	0,5301	0,3184	0,1481
	1418	1,4787	1,7714	0,2987	0,2046
	1419	-0,4798*	0,2817	0,0864**	0,0439

	1421	-0,1312	0,5779	0,2758**	0,1069
	1423	-0,0033	0,2341	0,1465	0,3740
	1425	-1,9204***	0,5029	-0,0049	0,0418
	1426	1,4335*	0,7785	0,3018	0,2547
	1427	-0,4062	0,2845	0,2360**	0,1011
	1428	-1,4629***	0,4346	0,1933	0,0818
	1429	-1,0716*	0,6357	0,3257**	0,1257
	1431	-1,4845**	0,5983	0,1940	0,0954
	1436	0,3318	0,5618	0,0222	0,2261
	1437	-0,3833	0,3394	0,0230	0,2887
	1438	-0,2829	0,6866	0,0716	0,0945
	1441	3,3990*	1,7894	0,2475	0,3303
	1442	-0,9156	0,5745	0,2165	0,3664
	1445	-2,2864***	0,6822	-0,0008	0,2360
	1446	-0,7834	0,6206	-0,2892	0,5326
	1448	-0,2927	0,5232	0,0759	0,0832
	1449	-0,3182	0,4259	0,0756	0,0864
	1450	-0,9485	0,6178	0,2449	0,2697
	1453	1,4351	1,1457	0,3280	0,3782
	1456	0,1669	0,5074	0,1780	0,3267
	1461	-0,6586	0,5300	-0,1151	0,2216
	1462	2,2618*	1,1939	-0,0041	0,1085
	1464	-0,3026	1,7869	0,1340	0,2151
	1471	-0,5411	0,5084	0,2619	0,2140
GICS	1512	0,1833	1,3205		
2	1401	0,2628*			
2	1403	0,5948	0,6717		
2	1404	0,1390	0,5659		
2	1405	-0,1091	0,4943		
2	1406	-1,1726***	0,3605		
2	1407	1,2467	0,9894		
2	1408	-0,3273	0,7200		
2	1409	0,3851	0,6550		
2	1410	1,2594*	0,6481		
2	1411	-2,0194**	1,0111		
2	1412	-1,9944***	0,7313		
2	1413	-2,0860**	0,9162		
2	1417	1,0010	0,6961		
2	1419	-0,0202	0,3159		
2	1421	-0,9554	0,6397		
2	1423	-0,2586	0,2469		
2	1425	0,6548	0,5540		
2	1426	-1,7885**	0,8353		
2	1428	0,4475	0,5141		
2	1431	0,4607	0,5918		
2	1436	-1,3157**	0,6507		
2	1437	-0,4047	0,3636		
2	1441	-3,1661	1,8935		
2	1442	0,8091	0,7342		
2	1446	-0,3315	0,6943		
2	1448	-0,3327	0,8471		
2	1450	0,6368	0,6912		
2	1453	-1,5497	1,2440		
2	1456	-1,1100*	0,6155		
2	1461	0,7279	0,6230		
2	1462	-2,4136**	1,2239		
2	1464	-0,8284	1,8168		
2	1471	-0,1033	0,5971		
2	1512	-0,1405	1,3996		
3	1401	-0,2916***			
3	1404	0,2853	0,6899		
3	1406	-2,1300***	0,6540		
3	1407	2,0686*	1,1753		

3	1419	-0,1103	0,3472
3	1423	-0,2675	0,3242
3	1426	-2,5467***	0,8644
3	1428	1,1015	0,7264
3	1431	0,9672	0,7363
3	1436	-0,6107	0,6717
3	1471	0,2567	0,6942
4	1401	0,1083	
4	1403	0,2649	1,0186
4	1404	0,3726	0,7149
4	1406	-0,9799	0,5281
4	1407	1,8330	0,9666
4	1411	-1,4148	1,0634
4	1419	0,3082	0,3662
4	1421	0,1890	0,7436
4	1423	-0,6275*	0,3592
4	1426	-2,1012***	0,8047
4	1436	-0,6985	0,6629
4	1442	0,4985	0,7106
4	1446	0,0130	0,7635
4	1453	-2,1195	1,3571
4	1464	-0,4628	1,8652
4	1471	0,6258	0,7162
6	1401	0,6489***	
6	1406	-0,5290	1,0370
6	1419	0,3915	0,3453
6	1423	-0,1142	0,3684
6	1426	-3,5831***	1,1388
7	1401	-0,0347	
7	1404	1,4517	0,7967
7	1419	1,3341***	0,4074
7	1423	-0,4516	0,4631
7	1425	2,7320***	0,9435
8	1401	-0,1924**	
8	1404	0,3738	0,5795
8	1406	0,5463	0,6025
8	1419	0,5185	0,3408
8	1423	0,1619	0,2715
8	1426	-1,0859	0,8804
8	1437	0,3111	0,5370
8	1442	0,0163	0,7754
8	1450	0,4249	0,7725
8	1471	-0,0981	0,6357
11	1401	-0,7014***	
11	1406	1,8625**	0,8781
11	1423	0,3516	0,3375
11	1446	1,2375	0,8834
11	1471	1,2331	0,7980

Appendix Table 8 – Estimation step 2 CDM-model

Innovation output (RE Probit)

Year: 2004-2014	Coef,	Std,Err,
lagpred	0,0744	0,0657
lagsize	-0,3744***	0,0536
laggp	0,0312	0,0464
humankapital	0,0086***	0,0019
lagco	-0,0859***	0,0228
lagconrd	0,3719***	0,0437
intcomp	0,2933***	0,0477
intco	1,1630***	0,0972
1401	-0,0033*	0,0465
1403	-0,0380	0,1848
1404	0,3472**	0,1274
1405	0,1152	0,1981
1406	0,0197	0,1106
1407	0,0931	0,1323
1408	-0,1223	0,2366
1409	-0,0383	0,2640
1410	-0,3823	0,2150
1411	0,3489	0,2178
1412	-0,2384	0,2130
1413	-0,0018	0,3222
1414	-0,1471	0,3048
1415	-0,3124	0,2891
1416	-0,6386	0,7361
1417	0,1142	0,2658
1419	-0,0196	0,0660
1421	0,2770	0,1845
1423	-0,0572	0,0625
1425	0,2125*	0,1786
1426	0,0293	0,1238
1427	0,1117**	0,2276
1428	-0,2755	0,1569
1429	0,7030	0,4447
1431	-0,1393	0,1622
1436	-0,1310**	0,1321
1437	-0,1031*	0,1416
1438	0,0982	0,4859
1441	0,0485	0,5302
1442	0,0668	0,1766
1445	-0,4926**	0,5554
1446	0,2724	0,2262
1448	0,2585	0,3015
1449	0,1492	0,3196
1450	0,0015	0,2015
1453	-0,2718**	0,2276
1456	0,2216	0,2197
1461	0,0403	0,1926
1462	0,7311*	0,5576
1464	0,3087	0,2166
1471	0,1786	0,1484
1512	-0,0034	0,5117
gics		

2	0,1941***	0,0843
3	0,0114	0,0912
4	-0,0773	0,0950
5	0,2077	0,2161
6	-0,1435	0,1269
7	0,0869	0,1086
8	0,0874	0,0956
9	0,5008	0,2914
11	-0,3549	0,1299

Appendix Table 9 –
All Estimates from step 3 in the CDM-model

Productivity (RE probit)		
Year: 2004-2014	Coefficient	Std.Err
PI2	0,0186**	0,0075
size10	0,2354***	0,0177
size1000	-0,0508***	0,0059
gp	-0,0509***	0,0067
intco	-0,0479***	0,0107
intcomp	-0,0266*	0,0106
lackedu	-0,0928***	0,0104
established	-0,0031	0,0073
stockholm	0,0113	0,0527
Eskilstuna	-0,0103	0,0647
Linkoping	0,0134	0,0577
Norrkoping	0,0279	0,0613
Jonkoping	0,0065	0,0517
Varnamo	-0,0337	0,0507
Vetlanda	-0,0157	0,0595
almhult	0,0117	0,0518
Vaxjo	-0,0483	0,0555
Ljungby	0,0161	0,0588
Kalmar	0,0360	0,0934
Oskarshamn	-0,0365	0,0577
Vastervik	-0,0188	0,0628
Vimmerby	-0,1232	0,0807
Gotland	0,0219	0,0686
Karlskrona	0,0390	0,0622
KarlshamnOlof~m	-0,0523	0,0668
MalmoLund	-0,0122	0,0525
Halmstad	-0,0098	0,0581
Goteborg	0,0358	0,0544
Stromstad	-0,0292	0,0651
TrollhattanVa~g	0,0595	0,0921
Boras	-0,0053	0,0529
LidkopingGotene	0,0635	0,0809
Skovde	-0,0149	0,0527
Torsby	0,0420	0,0674
arjang	0,0012	0,0661
Karlstad	0,0029	0,0568

Filipstad	0,0048	0,0541
Hagfors	0,1959	0,0776
Hallefors	0,0253	0,0985
orebro	-0,0153	0,0616
Vasteras	0,0010	0,0513
Fagersta	-0,1098	0,0666
Vansbro	0,1594	0,0682
MalungSalen	0,0349	0,0613
Mora	-0,0510	0,0615
la1442	0,0052	0,0526
AvestaHedemora	-0,0694	0,0655
Ludvika	0,0701	0,0574
Ljusdal	0,0267	0,0684
Gavle	0,0343	0,0514
Soderhamn	-0,0136	0,0754
BollnasOvanaker	0,0993	0,0748
Hudiksvall	0,0109	0,0563
Sundsvall	-0,0067	0,0488
Kramfors	0,0462	0,0626
Solleftea	0,0291	0,0727
ornskoldsvik	0,0057	0,0593
Stromsund	-0,0297	0,0946
Harjedalen	0,0540	0,0828
ostersund	-0,0168	0,0657
Storuman	0,0840	0,0723
asele	0,3165	0,1387
Umea	0,0246	0,0537
Lycksele	0,0255	0,0607
Arvidsjaur	-0,0243	0,0567
overkalix	-0,0908	0,1056
Gallivare	0,0000	(omitted)
Lulea	-0,0278	0,0645
Haparanda	0,0000	(omitted)
Kiruna	0,0247	0,0601